CODEN: JASMAN 4966

THE JOURNAL of the Acoustical Society of America

Vol. 103, No. 4 April 1998

SOUNDINGS SECTION

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

ERRATA

CUMULATIVE AUTHOR INDEX

NOTES CONCERNING ARTICLE ABSTRACTS

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

SOUNDINGS

This front section of the *Journal* includes acoustical news, views, reviews, and general tutorial or selected research artcles chosen for wide acoustical interest and written for broad acoustical readership.

ACOUSTICAL NEWS—USA

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

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

John C. Burgess succeeds Jeffrey L. Krolik as Associate Editor of the *Journal*

John C. Burgess

Jeffrey L. Krolik

In December 1997 Professor Emeritus John C. Burgess of the Mechanical Engineering Department of the University of Hawaii became an Associate Editor of the *Journal* for papers in Acoustic Signal Processing (PACS 43.60). He succeeded Professor Jeffrey L. Krolik of the Electrical Engineering Department at Duke University. The Editor-in-Chief expresses his appreciation to Professor Krolik who completed a three-year term of service to the *Journal*, and to his institution for support.

John C. Burgess received his B.S. degree in Engineering from Brown University in 1944, an M.S. in Mechanical Engineering from Stanford University in 1949, and his Ph.D. in Engineering Mechanics in 1955 from Stanford University.

From 1944–46 John Burgess served as Lt. (jg) in submarine service in the U.S. Navy. From 1946–53 he was an instructor, teaching assistant, or research assistant at Brown University, Stanford University, and University of Puerto Rico. He was a mechanical engineer for Stanford Research Institute from 1953–61, where he developed a high-power modulated air stream loudspeaker. He was manager of the Engineering Sciences Branch, United Technology Center in Sunnyvale, CA from 1961–66, with major responsibilities for aerodynamic, thermodynamic, and structural analysis of the Titan IIIC solid rocket.

Since 1966 Dr. Burgess has been a Professor of Mechanical Engineering at the University of Hawaii. He has served as department head; introduced courses in acoustics, vibrations, noise control, and digital spectral analysis; and also taught courses in computer graphics, methods, design, and automatic control. During sabbatical and other leaves from Hawaii he worked for five time periods in the Acoustics and Signal Processing Group at the National Research Council of Canada, and was also a resident visitor at Bell Laboratories, doing extensive research in computer applications to acoustics.

Professor Burgess is a recipient of Distinguished Service Awards from both the Acoustical Society of America and the Acoustical Society of Japan, and served as Chairman for all three joint meetings of the two societies in 1978, 1988, and 1996. He gave a two-day short course on ''Basic Digital Signal Processing in Acoustics'' at the 1995 Washington, DC meeting of the Society.

The Editor-in-Chief and his colleagues welcome Professor Burgess to membership on the Editorial Board.

DANIEL W. MARTIN
Editor-in-Chief

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.

1998

5–8 April	NOISE-CON 98, Ypsilanti, MI [Noise Control Foundation, P.O. Box 2469, Arlington Branch, Poughkeepsie, NY 12603; Tel.: 914-462-4006; Fax: 914-463-0201; E-mail: noisecon98@aol.com; WWW: users.aol.com/noisecon98/nc98_cfp.html].
13–15 May	28th Annual Symposium, Ultrasonic Industry Assn., King of Prussia, PA [Ultrasonic Industry Assn., Inc., P.O. Box 1420, Cherry Hill, NJ 08034-0054; Tel: 609-424-8998; Fax: 609-424-9248].
4–7 June	7th Symposium on Cochlear Implants in Children, Iowa City, IA [Center for Conferences and Institutes, The University of Iowa, 249 Iowa Memorial Union, Iowa City, IA 52242-1317; Tel.: 800-551-9029; Fax: 319-335-3533]. Deadline for receipt of abstracts: 1 March
20–26 June	135th meeting of the Acoustical Society of America/16th International Congress on Acoustics, Seattle, WA [ASA, 500 Sunnyside Blvd., Woodbury, NY 11797, Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org, WWW: http://asa.aip.org].
26 Jun–1 Jul	International Symposium on Musical Acoustics, ISMA 98, Leavenworth, WA [Maurits Hudig, Catgut Acoustical Society, 112 Essex Ave., Montclair, NJ 07042, Fax: 201-744-9197; E-mail: catgutas@msn.com, WWW: www.boystown.org/isma98].
7–12 July	Vienna and the Clarinet, Ohio State Univ., Columbus, OH [Keith Koons, Music Dept., Univ. of Central Florida, P.O. Box 161354, Orlando, FL 32816-1354, Tel.: 407-823-5116; E-mail: kkons@pegasus.cc.ucf.edu].
9–14 Aug.	International Acoustic Emission Conference, Hawaii [Karyn S. Downs, Lockheed Martin Astronautics, P.O. Box 179, M.S. DC3005, Denver, CO 80201; Tel.: 303-977-1769; Fax: 303-971-7698; E-mail: karyn.s.downs@lmco.com].
13–17 Sept.	American Academy of Otolaryngology—Head and Neck Surgery, San Francisco, CA [American Academy of Otolaryngology—Head and Neck Surgery, One Prince St., Alexandria, VA 22314 Tel.: 703-836-4444; Fax: 703-683-5100].
18–19 Sept.	6th Annual Conference on Management of the Tinnitus Patient, Iowa City, IA [Richard Tyler, Univ. of Iowa, Dept. of Otolaryngology—Head & Neck Surgery, 200 Hawkins Dr., C21GH, Iowa City, IA 52242; Tel: 319-356-2471; Fax: 319-353-6739; E-mail: rich- tyler@uiowa.edu].
12–16 Oct.	136th meeting of the Acoustical Society of America, Norfolk, VA [ASA, 500 Sunnyside Blvd., Woodbury, NY 11797, Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org, WWW: http://asa.aip.org].

 0001-4966/98/103(4)/1689/4/$10.00

1999

15–19 March	Joint meeting: 137th meeting of the Acoustical Society of America/Forum Acusticum [Acoustical Society of America, 500 Sunnyside Blvd., Woodbury, NY 11797, Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].
27–30 June	ASME Mechanics and Materials Conference, Blacksburg, VA [Mrs. Norma Guynn, Dept. of Engineering Science and Mechanics, Virginia Tech, Blacksburg, VA 24061-0219; Fax: 540-231-4574; E-mail: nguynn@vt.edu; WWW: http://www.esm.vt.edu/mmconf/]. Deadline for receipt of abstracts: 15 January 1999

Revisions to Membership List

New Associates

Abramov, Oleg V., Lab. of Ultrasonics, Inst. of General and Inorganic Chemistry, Leninсky Prospect 31, Moscow 117907, Russia

Abramov, Vladimir, Inst. für Werkstoffwissenschaft, Max-Planck Inst. für Metallforschung, Seest 92, Stuttgart 70174, Germany

Afshar, Marcus M., Information Resources, ReSound Corp., 220 Saginaw Drive, Redwood City, CA 94063

Alakel, Michael N., Advanced Product Development, Textron Systems Div., 201 Lowell Street, Wilmington, MA 01887

Amble, Peer, Ogden Environmental and Energy Services, Inc., 1 East Anapurau Street, Santa Barbara, CA 93101

Antopnanzas-Barroso, Norma S., 3556 Maplewood Avenue, Los Angeles, CA 90066

Aoki, Mariko, Nippon Telegraph & Telephone Corp., Speech and Acoustics Lab., Human Interface Labs., #420-C 1-1 Hikarinooka, Yokosuka, Kanagawa, 239-0847 Japan

Arthur, John A., John Arthur Design Group, Inc., 1954 NE 149th Street, North Miami, FL 33181

Aubauer, Roland, Auf Dem Kraml 5, D-55120 Mainz, Germany

Babineau, Francis J., Acoustics Lab., R-05, Johns Manville Technical Center, 10100 West Ote Avenue, Littleton, CO 80127

Bachman, Gregory M., Dodge Regupol, Inc., 715 Fountain Avenue, Lancaster, PA 17601

Barker, Terrance G., Maxwell Technologies, Geophysics, 8888 Balboa Avenue, San Diego, CA 92123

Barlow, Jay P., Southwest Fisheries Science Center, NOAA National Marine Fisheries Service, 8604 La Jolla Shores Drive, La Jolla, CA 92037

Baumgartner, Wolf D., Ebenberg 2, Siegenfeld A-2500, Austria

Bird, John S., 2445 Begbie Terrace, Port Coquitlam, BC V3C 5K8, Canada

Bland, Roger W., Physics and Astronomy, San Francisco State Univ., 1600 Holloway Avenue, San Francisco, CA 94132

Blolloch, Paul A., SDRC, Advanced Test & Analysis, 11995 El Camino Real, #200, San Diego, CA 92130

Boughman, Janette W., Psychology, Univ. of Maryland, College Park, MD 20742

Brussieux, Marc, GESMA, High Frequency Acoustics, BP 42, Brest-Naval 29240, France

Burrus, Cecilia R., SPAWAR Systems Center San Diego, Code D711, 53560 Hull Street, San Diego, CA 92152-5001

Byrnes, Gregory, 3 Totty Court, Florham Park, NJ 07932

Chen, Li-Mei, No. 2-3 Lane 29, Chung-Hsiao Road, Kuei-jen, Hsiang, Tainan City 711, Taiwan

Chenausky, Karen V., Speech Technology and Applied Research, 4 Militia Drive, Lexington, MA 02173

Clark, David S., SSC-SD, D711, 53560 Hull Street, San Diego, CA 92152

Cornuelle, Bruce D., Physical Oceanography Res. Div. 0230, Scripps Inst. of Oceanography, UCSD, 9500 Gilman Drive, La Jolla, CA 92093-0230

Cutler, E. Anne, Max Planck Inst. for Psycholinguistics, Wundtlaan 1, Nijmegen 6525 XD, The Netherlands

Dinnissen, Paul H., Kanalweg 1, Schwadernau CH-2556, Switzerland

Downs-Pruitt, Jesica C., ATR Advanced Telecommunications, Research Inst., Human Information Processing, 2-2 Hikaridi, Seika-cho, Soraku-gun, Kyoto, 619-02 Japan

Ekaterinaris, John A., Near, Inc., 526 Clyde Avenue, Mountain View, CA 94043-2212

Endo, Takuma, Center for Integrated Research in Science and Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-01 Japan

Ferrandez, Jose M., El Cristobal Sanz 21, Elche, Alicante 03201, Spain

Galembo, Alexander, c/o Olga Troinitskaya, 47 Furgtaoltskaya Street, Apt. 17, St. Petersburg 191123, Russia

Gerstein, Edmund R., Psychology, Florida Atlantic Univ., 777 Glades Road, Boca Raton, FL 33431

Giordano, Nicholas J., Physics Dept., Purdue University, Physics Building, West Lafayette, IN 47907

Glaese, Roger M., Active Systems, CSA Engineering, 2850 West Bayshore Road, Palo Alto, CA 94303

Goodall, Ashley J., Artec Consultants, Inc., 114 West 26th Street, New York, NY 10001-6812

Greenlee, Doug M., 7084 South Kendall Court, Littleton, SC 80123

Grellet, Marcos, Otolaryngology & Otorrinholanngologia, Fac. de Medicina di Ribeirao Prito USP, Ave. dos Bomdurantz 3900, Ribeirao Prito, Sao Paulo 14-049-900, Brazil

Hager, Lee D., James, Anderson & Associates, Inc., 2123 University Park Drive, Suite 130, Okemos, MI 48864

Hallam, John C. T., Dept. of Artificial Intelligence, Univ. of Edinburgh, 5 Forrest Hill, Edinburgh EH1 2QL, Scotland

Hansmann, Douglas R., EKOS Corporation, 22122 20th Avenue, SE, Suite 148, Bothell, WA 98021-4433

Helimaki, Heikki J., Engineering Office Heikki Helimaki Oy, Mantytie 30, Haarajoki FIN-04480, Finland

Hiroaki, Nomura, System Engineering, Toua Univ., 2-1 Ichinomiya, Gakuen-cho, Shimonoseki-shi, 751 Japan

Hoffman, Howard J., 9 Cherbourg Court, Potomac, MD 20854-3101

Hollman, Kyle W., 10255 Dover, Apt. 632, Westminster, CO 80021

Holmes, Frederika, 165 Southgate Road, London N1 3LE, U.K.

Holst-Jensen, Ole, Odegaard & danneskiold-Samsoe, Jylland, Beder Landevej 29, Beder 8330, Denmark

Homa, David E., Acentech, Incorporated, 33 Moulton Street, Cambridge, MA 02138

Hoover, III, Arthur M., Technical Service Group, Inc., 17015 Cloverland Court, Baton Rouge, LA 70809

Jarinko, David A., 239 Market Street, Charlestown, MD 19214

Kainz, Wolfgang, Schneiderallee 154, Langenzersdorf, Niederosterreich, A-2103 Austria

Kaminsky, Clifford S., 4906 Cole Boulevard, Ypsilanti, MI 48197

Kang, Seong-Woo, Core Technology Research Center, Samsung Electronics Co., Ltd., 416 Maetan, 3 Dong, Paldal-gu, Suwon, Kyungki-Do 442-742, Korea

Karasalo, Ilkka T., FOA, 64, Enkopingsvagen 126, S-17290 Stockholm, Sweden

Khalilullov, Chamil, 93 Coachlight Square, Montrose, NY 10548

Kim, Eunju, 428 Hawthorne Street, #114, Glendale, CA 91204

Klingborg, Kerry A., Reed Instrument, United Musical Instruments, USA, Inc., 1000 Industrial Parkway, Elkhart, IN 46516

Kook, Hyungseok, 1012 Hawkins Graduate House, West Lafayette, IN 47906

Krasilnikov, Vladimir A., Physics, Moscow State Univ., Vorobevy gory, MSU, Moscow 119899, Russia

Lambert, David E., Engineering, Test and Evaluation Dept., Naval Undersea Warfare Center, 1176 Howell Street, Newport, RI 02841

Larom, David L., 2417 Broadway, San Diego, CA 92102

Lazauski, Colin J., Code 74, Naval Undersea Warfare Center, Division Newport, 6 Howell Street, Newport, RI 02841

Lindsey, Stephen G., Cerami & Associates, 518 Fifth Avenue, New York, NY 10023

Link, Steven T., Kvernstoen Kehl & Associates, 420 North Fifth Street, Suite 1055, Minneapolis, MN 55401

Lizzi, Frederic L., Riverside Research Inst., 330 West 42nd Street, New York, NY 10036

Mack, Richard A., RAM Technologies, Inc., 33 West Higgins Road, Suite 5020, South Barrington, IL 60010

Madan, Ashvani K., Ethicon Endo-Surgery, R&D, 4545 Creek Road, Cincinnati, OH 45242

Malek, John W., Ann Arbor Audio, 1345 Rickett Road, Brighton, MI 48116

Marr, Kenneth W., FBI, Engineering Research Facility, EST-1, Bldg. 27958A, Quantico, VA 22135

Marshall, Steven E., 6633 118th Avenue, SW, Bellevue, WA 98006

Mason, Timothy J., NEX, Coventry Univ., Priory Street, Coventry CV1 5FB, U.K.

McGraw, Thomas J., McKay Conant Brook, Inc., 5655 Lindero Canyon Road, Suite 325, West Lake Village, CA 91362

McQueen, James M., Max-Planck Inst. for Psycholinguistics, PB 310, NL 6500 AH Nijmegen, The Netherlands
Mendoza, Jeffrey M., 917 Seabreeze Court, Chesapeake, VA 23320
Miller, Eric J., Architectural Testing, Inc., 130 Derry Court, York, PA 17402
Miller, Roger L., Otolaryngology, Head & Neck Surgery, Duke Univ. Medical School, Box 3550, Durham, NC 27710
Miranda, Joseph A., Code 8232, B1302/2, NUWCDIVNPT, Naval Undersea Warfare Center, Newport, RI 02841
Moller, Henrik E., Akukon Ltd., Kornetintie 4A, Helsinki 00380, Finland
Moore, Sue E., SAIC, Maritime Services Div., 3990 Old Town Avenue, #105A, San Diego, CA 92110
Motta, Mauricio S., Rue Modestino Kanto 245, Rio de Janeiro RJ, CEP 21740-280, Brazil
Mousset, Eric C., IAM, Univ. of Bern, Neubrueckstrasse 10, Bern CH-3012, Switzerland
Moyal, Ami, Natural Speech Communication, 33 Lazarou Street, Rishen Lezion 75654, Israel
Muyakshin, Sergej I., Alfred Wegener Inst. for, Polar & Marine Research, Postfach 120161, Bremerhaven D-27515, Germany
Nakasone, Hirotaka, FBI, Dept. of Justice, Engineering Research Facility, Quantico, VA 22135
Neltnor, Kenneth D., Dullane Corporation, 2900 Dukane Drive, St. Charles, IL 60174
Nordman, Jonathan D., 653 Mineral Springs Road, Owatonna, MN 55060
Nozawa, Takeshi, Dept. of English, Akashi College of Technology, 679-3 Nishioja Uozumi, Akashi, Hyogo, 674 Japan
Olalla, Pedro, Acoustic Lab. R&D Spain, Electrolux Compressors, Antonio Forrellad 2, Sant Wuirze Del Valles, Barcelona 08192, Spain
Olson, Joseph R., 7309 26th Avenue, NW, Seattle, WA 98117
Ortega, Jose C., 11746 Avenida Del Sol, Northridge, CA 91326
Patel, Aniruddh D., The Neurosciences Inst., 10640 John Jay Hopkins Drive, San Diego, CA 92121
Peremans, Herbert G. P. V., Ooilamstraat 1, Gent 9000, Belgium
Petrovska, Dijana, EPFL, Swiss Federal Inst. of Tech., Electricity, CIRC, Lausanne 1015, Switzerland
Posenecker, John A., Advanced Environmental Concepts, 14934 Webb Chapel Road, Suite 32, Farmers Branch, TX 75234
Purgue, Alejandro P., Physiological Sciences, UCLA, 621 Circle Drive, South, Los Angeles, CA 90095-1527
Rajcan, Emil, Technical Univ. of Physics and Applied Mechanics, Masarykova 24, Zvolen SK-96053, Slovakia
Reed, James D., 6734 Abrego Road, Apt. #9D, Isla Vista, CA 93117
Romano, Anthony J., Physical Acoustics Branch, Code 7132, Naval Research Lab./SFA, Washington, DC 20375-5000
Rosa, Marcelo, Dept. de Engenaria Eletrica, Escola de Engenharia de Sao Carlos—USP, Avenida Dr. Carlos Botelho 1465, C.P. 359, Curitiba, Parana 13560-250, Brazil
Sakai, Hiroyuki, Div. of Global Development Science, Graduate School of Science & Tech., Kobe University, Rokkodai, Nada, Kobe, Hyogo, 657 Japan
Saliba, Rindala, System Test Tools/Consumer Products, Philips Consumers Communications, 330 South Ranolphville Road, Piscataway, NJ 08854
Salomaa, Martti M., Dept. of Eng. Physics & Mathematics, Helsinki Univ. of Tech., P.O. Box 2200, Espoo FIN-02015 HUT, Finland
Schusterman, Ronald J., 1629 Mariposa Avenue, Palo Alto, CA 94306
Shalis, Edward, 2734 Hampshire, Ann Arbor, MI 48104
Shallcross, William D., 69 West View Road, Middletown, RI 02842
Sharma, Anu, Speech and Hearing Science, Arizona State Univ., 200 East Curry Road, Suite 146, Tempe, AZ 85287-1908
Shipley, Mitchell N., Physics, Naval Postgraduate School, Code Ph, Monterey, CA 93943
Shreiber, Isaac, Inst. for Industrial Mathematics, Hahistaarut 22, Beer-Sheva 84213, Israel
Simao, Sheila M., R. Cinco de Julho, 1011, Seropedica RJ 23835-130, Brazil
Smith, Dean J., 33 Crossings Court, Portsmouth, RI 02871
Somin, Martin E., Advanced Development Center (MS1-83), GEC—Marconi Hazeltine, 450 East Pulaski Road, Greenlawn, NY 11740
Steiger, Dianne L., 15151 Cherbourg Avenue, Irvine, CA 92604
Sullivan, Brenda M., Fluid Mechanics & Acoustics Div., NASA Langley Research Center, M/S 463, Hampton, VA 23681
Supin, Alexander Ya., Inst. of Eology and Evolution, Russian Academy of Science, Leninsky Props. 33, Moscow 117071, Russia
Szymanski, Michael D., VA Medical Center, NCHCS, Neurology (127), 150 Muir Road, Martinez, CA 94553
Tanner, Simon, Sonar Dept., DERA, Winfrith Technology Centre, Dorchester, Dorset DT2 8XJ, U.K.
Van Donselaar, Wilma A., Beyensstr. 54, 6521 Ed Nymegen, The Netherlands
Vetter, Douglas E., 2072 Diamond Street, San Diego, CA 92109
Vlahopoulos, Nickolas, Naval Architecture and Marine Eng., Univ. of Michigan, 2600 Draper Road, 214 NA&ME Bldg., Ann Arbor, MI 48109-2145
Vold, Havard I., Vold Solutions, Inc., 1716 Madison Road, Cincinnati, OH 45206
Vuono, Jeffrey P., Naval Undersea Warfare Center, Engineering Test & Evaluation, Code 74, Bldg. 103, Newport, RI 02841
Walton, Dennis C., Advanced Engineering, Abbott Labs., P.O. Box 152020, Irving, TX 75015-2020
Wei, Ruey-Chang, Inst. of Undersea Technology, National Sun Yat-Sen Univ., 70 Lian-Hae Road, Kaohsiung 804, Taiwan
White, Peta J., Centre for Advanced Studies in Music Education, Roehampton Inst. London, Downshire House, Roehampton Lane, London SW15 4HT, U.K.
Xu, Yangguang, P.O. Box 525229, Flushing, NY 11352
Yamada, Ichiro, Kobayasi Inst. of Physical Research, 3-20-41 Higashi-Motomachi, Kokubunji, Tokyo, 185 Japan
Yang, Jun, Inst. of Acoustics, Academia Sinica, Beijing, 17 Zhongguancun Street, Beijing 100080, P.R. China
Yates, Graeme K., Dept. of Physiology, Univ. of Western Australia, Nedlands, WA 6907, Australia
Yeung, Bing Kwan, 22nd Floor, Loyong Court Commercial Building, 220 Lockhart Road, Wan Chai, Hong Kong
Yi, Wesley, Seacom Div., Dukane Corp., 2900 Dukane Drive, St. Charles, IL 60174
Yuan, Gang, Ocean Research Dept., Japan Marine Science & Tech. Ctr., 2-15 Natsushima-cho, Yokosuka, 237 Japan
Zacharov, Nicholas V., Speech and Audio Systems, Nokia Research Center, P.O. Box 100, Sinitaival 6, Tampere 33721, Finland
Zampini, Mary L., Spanish & Portuguese, Univ. of Arizona, P.O. Box 210067, Tucson, AZ 85721

New Students

Abu-Al-Makarem, Ali S., 186 State Street, Bowling Green, OH 43402
Aiken, Steve J., 520-251 Platts Lane, London, ON N6H 4P4, Canada
Al-Alawi, Shaker H., Electronics and Computer Science, Univ. of Southampton, Highfield, Southampton SO17 1BJ, England
Allen, Jeffrey S., Psychology, Northeastern Univ., 125 Nightingale Hall, Boston, MA 02115
Backus, Bradford L., Ashdown House, Room 510B, 305 Memorial Drive, Cambridge, MA 02139-4303
Beckman, Sean M., 218 East Voris Street, Akron, OH 44311
Blonigen, Florian J., Physics Dept., Washington State Univ., Pullman, WA 99164-2814
Bofsford, Tim G., 1920 Birgo Terrace, #4, Madison, WI 53705
Burk, Matthew H., 2705 University Avenue, Apt. 21, Madison, WI 53705
Campos, Erik A., Jose Luis Coo #0759, Santiago RM, Chile
Cazzanti, Luca G., 1609 Madison Street, Madison, WI 53711
Chacon, Victor A., Moneda 2858 C-4, Santiago RM, Chile
Christian, Sean M., 914 Sudberry Street, Apt. 1, Staunton, VA 24401
Clement, Gregory T., Physics, Univ. of Rhode Island, 108 East Hall, Kingston, RI 02881
Dieken, David T., 1317 East Dayton Street, Madison, WI 53703
DiGiovanni, Jeffrey J., Communication Disorders, Univ. of Minnesota, 115 Shelvin Hall, 164 Pillsbury Drive, SE, Minneapolis, MN 55455
Dugan, Jennifer, 323 East Johnson Street, Madison, WI 53703
Dungan, Michael R., 2224 Stone Road, Ann Arbor, MI 48105
Ferguson, Sarah H., 2357 Worthington Lane, Bloomington, IN 47401
Fox Kruckenberg, Dirk M., Las Hualtatias 8850 Vitacura, Santiago, Chile
Frensley, Cliff, Physical Acoustics Group, Applied Physics Lab., 1013 NE 40th Street, NE, Seattle, WA 98105
Frota, Joao, 51 Rue des Arcs St. Cyprien, Toulouse F-31300, France
Gaberman, Alexander M., 43 Burt Latham Road, #A-13, Willington, CT 06279
Gehr, Sara E., Psychology Dept., Washington Univ., Campus Box 1125, One Brookings Drive, St. Louis, MO 63130

Gentner, Timothy Q., Psychology, Johns Hopkins Univ., 3400 North Charles Street, Baltimore, MD 21218
Gick, Bryan W., Haskins Laboratories, 270 Crown Street, New Haven, CT 06510
Golden, Matthew V., 801 West Aaron Drive, Apt. A5, State College, PA 16803
Gover, Bradford N., Dept. of Physics, Univ. of Waterloo, 200 University Avenue, West, Waterloo, ON N2L 3G1, Canada
Graves, Simon J., 3944 Third Avenue, San Diego, CA 92103
Han, Fei, 1077 Ray W. Herricks Labs., Purdue University, West Lafayette, IN 47906
Harrison, Daniel, 1320 Spring Street, Apt. C, Madison, WI 53715
Heinrich, Gregor, NuBbaumallee 25, Dardstadt 64297, Germany
Henderson, Paul D., 15817 Thompson Road, Charlotte, NC 28227
Henry, James K., Mechanical Engineering, Duke University, Box 90302, Durham, NC 27708-0302
Hoffmann, Cristian M., Santa Anselma 0445 La Cisterna, Santiago RM, Chile
Hoglund, Evelyn M., 3338 Timber Oak Drive, Columbus, OH 43204-4100
Hou, An, 546 Ethel Street, Atlanta, GA 30318
Iglehart, Frank, Clarke School for the Deaf, Hudgins Audiology Div., 47 Round Hill Road, Northampton, MA 01060
Ille, Jean-Francois E., 30370 GA Tech Station, Atlanta, GA 30332
Ireland, Laralee G., 1930 North 360 East, Provo, UT 84604
Jeng, Jing-Yi, 927A Eagle Heights Road, Madison, WI 53705
Jesus, Luis M. T. De, Electronics and Computer Science, Univ. of Southamptom, ISIS Research Group, Southampton SO17 1BJ, U.K.
Kang, Jungu, Ocean Engineering, Davidson Lab., Stevens Inst. of Technology, 711 Hudson Street, Hoboken, NJ 07030
Kenehan, Garrett J., 3260 Rockview Place, Apt. G, San Luis Obispo, CA 93401
Kestell, Colin D., Mechanical Engineering, Univ. of Adelaide, North Terrace, Adelaide, SA 5005, Australia
Klein, Michael D., P.O. Box 70254, Glyfada 16610, Greece
Kunes, Kay C., 721 Mustang Court, Fairfield, CA 94533
Lane, Steven A., Mechanical Engineering, Duke University, Box 90302, Durham, NC 27708-0302
Lazar, Zsolt I., Physics Inst., Univ. of Bergen, Allegaten 55, Bergen N-5007, Norway
Lenz, Peter W., 2577 North Oakland Avenue, Milwaukee, WI 53211
Lin, Dong, Physics and Astronomy Dept., Brigham Young Univ., Provo, UT 84602
Lopez, Javier F., La Canada 7355-E La Reina, Santiago RM, Chile
Luka, Steven B., 328 Morse Drive, Northlake, IL 60164
Max, Ludo, School of Speech Pathology & Audiology, Kent State University, P.O. Box 5190, Kent, OH 44242-0001
McArthur, K. Megan, 2791 Ocean Front Walk, San Diego, CA 92109
Moreton, A. Elliott, Linguistics Dept., Univ. of Massachusetts/Amherst, South College Building, Amherst, MA 01060
Omuro, Kaori, 320 Kolob Circle, P.O. Box 456, Springville, UT 84663
Overholt, Daniel J., 1752 East 8th Street, Chico, CA 95928
Poulsen, Jens K., Finsensgade 48, 2. TV, Aarhus N-8200, Denmark
Procel, Anahi M., Avenida Marathon 2582 Macul, Santiago, Chile
Ragozzine, Frank R., 9226-A Regents Road, La Jolla, CA 92037
Rouse, Jerry W., 604-2J Charleston Road, Raleigh, NC 27606
Sach, Andrew J., Psychology Dept., Univ. of York, Heslington, York YO1 5DD, England
Sato, Shin-ichi, 4-1-8-202 Shinoharanake-machi, Nada, Kobe, Hyogo, 657 Japan
Schwartz, Patricia N., 237-01 88th Avenue, Bellerose, NY 11426
Scott, Christopher D., 1657 Lynn Boulevard, Aurora, IL 60505
Seep, Benjamin C., 1004 Westrun Drive, Ballwin, MO 63021
Seymour, Patricia J., 4621 Lloyd Street, Kansas City, KS 66103
Simmons, Timothy G., NCPA, Nonlinear Acoustics, Coliseum Drive, Oxford, MS 38677
Smorodinsky, Iris, P.O. Box 206176, New Haven, CT 06520-6176
Spohnholtz, Todd W., 538 Dawn Circle, Grayslake, IL 60030-1217
Stucky, Paul A., 2617 Knapp Street, Ames, IA 50014
Svensson, Wolfgang R., Centre for Computer System Architecture, Halmstad Univ., Box 823, Halmstad S-30118, Sweden
Turgeon, Martine, Psychology, McGills Auditory Lab., 1205 Dr. Penfield Avenue, Montreal, QC H3A 1B1, Canada
Vasquez, Edna E., P.O. Box 5609, College Station, Managuez, Puerto Rico 00681
Vicente-Borruel, Luis, Ingenieria Electronics y Comunicaciones, Centro Politecnico Superior, Maria de Luna 3, Zaragoza E-50015, Spain
Viechnicki, John T., Applied Marine Physics, Rosenstiel School of Marine and Atmospheric Science, 4600 Rickenbacker Causeway, Miami, FL 33149
Wakeland, Ray S., 908 West Beaver Avenue, Apt. B, State College, PA 16801
Wang, Yue, Linguistics, Cornell University, Morrill Hall, Ithaca, NY 14853
Warren, Joseph D., Applied Ocean Physics and Engineering, Woods Hole Oceanographic Inst., Woods Hole, MA 02543
Whitley, III, Milton A., 331466 Georgia Tech Station, Atlanta, GA 30332
Wilson, Annette A., 729 Brooks, Ann Arbor, MI 48103
Zanartu, Matias, Dept. of Acoustics, Univ. Tecnologia Vicente Perez Rosales, Brown Norte 290—Nunda, Santiago, Chile
Zawaydeh, Bushra A., Linguistics Dept., Indiana Univ., Memorial Hall 322, Bloomington, IN 47406

Members Elected Fellows

X. Boutillon, A. L. Dancer, C. J. Darwin, R. J. Dooling, J. J. Eggermont, K. G. Foote, C. F. Greenlaw, T. G. Leighton, J. L. Reisse, E. W. Rubel, R. J. Schusterman, J. P. Seiler, K. Soetanto, A. Q. Summerfield, H. Sussman, K. L. Williams

Associates Elected Members

H. A. B. Alwi, H. A. Arts, C. Bean, J. Beilin, K. C. Benjamin, B. G. Berg, D. A. Berry, R. E. Causse, C. A. Champlin, G. D. Chermak, E. H. A. Colomb, P. Dalsgaard, B. B. Debus, C. X. Deng, M. P. Feeney, K. Fujiwara, A. G. H. Gabrielsson, J-P. Gagne, J. Garcia-Bonito, M. A. Gratton, S. A. Hambric, J. A. Harrison, M. S. Hedrick, M. C. Herr, J. Huang, V. F. Humphrey, T. Jupiter, O. Kirkeby, L. Kishon-Rabin, N. Krause, J. P. Kwolkoski, S. Launer, Z. Lazreq, J. Lee, J. J. Leonard, T. K. Lewis, R. Lim, J. J. Lutolf, K. I. McAnally, R. Meddis, C. D. Morgan, I. Morioka, I. Nakamura, R. M. Nielsen, K. Omori, H. Onaga, F. Orduna-Bustamante, C. A. Sammeth, A. Sarabasa, J. S. Sinclair, J. O. Smith III, K. B. Snell, H. Susaki, X. Tang, H. M. Tavossi, L. K. Thibodeau, Y. Tomoki, Y. C. Tong, K. Ueda, P. Van Dijk, R. J. M. Van Hoesel, R. Veillette, A. J. Vermiglio, S. Wang, H. Y. Woo

Students to Associates

J. A. Ekaterinaris, R. Hagiwara

Reinstated

R. F. Carlson, R. D. Kilmer, J. R. Prohs—*Members*
J. A. Henry, F.-L. Lu, J. W. Mullennix, A. Sauter, Jr., E. K. Scheer, T. Song—*Associates*
T. D. Sparks—*Student*

Resigned

H. G. Akst, P. J. Benson, I. R. Bicknell, R. P. Collier, R. M. Earsy, M. C. Lindt, B. E. Morse, C. M. Sehgal, R. Thalmann, B. E. Vircks—*Members*
H. A. Beveridge, P. Borner, G. C. J. Hofmans, D. M. Hogue, M. Kumagai, P. N. Liang, G. C. Marean, W. F. McGee, L. A. Mole, M. F. Morrison, C. M. Ort, H. H. Rose, Jr.—*Associates*
J. P. Gather, D. W. Kuhns, B. D. Scott, L. Zheng—*Students*

Deceased

C. L. Bartberger, M. A. Pedersen, I. Rudnick, A. Weissler—*Fellows*

R. W. Capaul, R. F. Fish, W. R. Goehner, J. Kritz, S. Mueller, R. J. Niederjohn, C. C. O'Brien, A. Preisman, H. B. Shaper, R. F. Sumons, W. J. Stojanowski—*Members*

Fellows	813
Members	2770
Associates	2770
Students	900
	7253

ACOUSTICAL NEWS—INTERNATIONAL

Walter G. Mayer
Physics Department, Georgetown University, Washington, DC 20057

Institute of Acoustics (UK) honors scientists

The Institute of Acoustics awards three medals in recognition of outstanding work in acoustics. During the November 1997 meeting of the IOA, the 1996 Tyndall Medal was awarded to Dr. Simon N. Chandler-Wilde for his outstanding contributions to the development and implementation of numerical methods for the modeling of sound propagation. The 1997 Rayleigh Medal was awarded to Professor Bjørnø of Denmark's Technical University for his distinguished academic and professional contributions to the discipline of acoustics. As is customary, the recipient of the Rayleigh Medal delivers the Rayleigh Lecture at the meeting. Professor Bjørnø spoke about ambient noise in the oceans. The inaugural R W B Stephens Medal was awarded to Professor Robert C. Chivers for his outstanding contribution to research and education in the field of acoustics.

Papers published in JASJ(E)

A listing of Invited Papers and Regular Papers appearing in the latest issue of the English language version of the *Journal of the Acoustical Society of Japan*, JASJ(E), was published for the first time in the January 1998 issue of the Journal. This listing is continued below.

The January issue of JASJ(E), Vol. 19, No. 1 (1998) contains the following papers:

T. Dutono, N. Ikeda, and A. Watanabe ''Effects of compound parameters on speaker-independent word recognition''
M. Hansaka and N. Mifune ''Study on the vibration damping properties of the damping material applying rubber compounding magnetic powder (DRM)''
K. Saijyou ''Data extrapolation method for near-field acoustic holography''
H. Takada and M. Ohba ''Postural stability of human body gazing at moving pictures''
J. Satonobu, K. Nakamura, and S. Ueha ''Torque accumulation for hybrid transducer ultrasonic motors using a coaxial driveshaft connection mechanism''
H. Wu and S. Ohtsuki ''An Eulerian segment-matching method for blood-flow velocity estimation using pulsed ultrasound RF echoes''
A. Yamada ''Comparative evaluation of inverse scattering quantitative CT algorithms for sound velocity image reconstruction''

International Meetings Calendar

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

April 1998

5–8 **1st Iberoamerican Congress of Acoustics**, Florianópolis-SC-Brazil. *2/98*

27–30 **Waves in Two-phase Flows (EUROMECH Colloquium)**, Istanbul. *12/97*

May 1998

10–14 **6th Meeting of the European Society of Sonochemistry**, Rostock-Warnemünde. *10/97*

18–22 **7th Spring School on Acoustooptics and Applications**, Gdańsk. *8/97*

25–27 **Noise and Planning 98**, Naples. *2/97*

June 1998

8–10 **EAA/EEAA Symposium ''Transport Noise and Vibration,''** Tallinn. *10/96*

9–12 **8th International Conference on Hand-Arm Vibration**, Umea. *6/97*

11–12 ***Swiss Acoustical Society Spring Meeting**, Lausanne, Switzerland. (Swiss Acoustical Society, Postfach, 8600 Dübendorf, Switzerland)

20–28 **Joint Meeting of the 16th International Congress on Acoustics and 135th Meeting of the Acoustical Society of America**, Seattle. *6/97*

July 1998

1–12 **NATO ASI ''Computational Hearing,''** Il Ciocco (Tuscany). *12/97*

September 1998

7–9 **Nordic Acoustical Meeting 98**, Stockholm. *10/97*

14–16 **Biot Conference on Poromechanics**, Louvain-la-Neuve. *10/97*

14–16 **ACUSTICA 98**, Lisbon. *10/97*

14–18 **35th International Conference on Ultrasonics and Acoustic Emission**, Chateau of Třešt'. *10/97*

21–25 ***4th European Conference on Underwater Acoustics**, Rome, Italy. (Secretariat ECUA 98, Istituto di Acustica—CNR, Via del Fosso del Cavaliere, 00133 Roma, Italy; e-mail: ecua98@idac.rm.cnr.it)

October 1998

4–7 ***euro-noise 98**, Munich, Germany. (CSM Congress @ Seminar Management, Industriestr. 35, 82194 Göbenzell, Germany; Fax: +49 8142 5 47 35; e-mail: csm_congress@compuserve.com)

31–3 ***AES International Conference ''Audio, Acoustics and Small Spaces,''** Copenhagen, Denmark. (Acoustical Society of Denmark, Institute of Acoustics Technology, Building 352—DTU, 2800 Lungby, Denmark; Fax: +45 45880577; e-mail: atc.das@dat.dtu.dk)

November 1998

12–15 ***Institute of Acoustics (UK) Autumn Conference: Speech and Hearing**, Windermere, UK. (Institute of Acoustics, Agriculture House, 5 Holywell Hill, St. Albans, Herts AL1 1EU, UK; Fax: +44 1727 850 533; e-mail: acoustics@clus1.ulcc.ac.uk)

16–18 **Inter-Noise 98**, Christchurch. *4/96*

20 **Recreational Noise**, Queenstown. *10/97*

23–27 **ICBEN 98: Biological Effects of Noise**, Sydney. *12/96*

30–4 **5th International Conference on Spoken Language Processing**, Sydney. *6/97*

December 1998

15–16 ***Sonar Signal Processing**, Loughborough, UK. (Institute of Acoustics, Agriculture House, 5 Holywell Hill, St. Albans, Herts AL1 1EU, UK; Fax: +44 1727 850 533; e-mail: acoustics@clus1.ulcc.ac.uk)

March 1999

15–19 **Joint Meeting of DAGA, EAA Forum Acusticum, and 137th Meeting of the Acoustical Society of America**, Berlin. *6/97*

June 1999

28–30 **1st International Congress of the East European Acoustical Association**, St. Petersburg. *10/97*

 0001-4966/98/103(4)/1694/2/$10.00

July 1999

4–9 **10th British Academic Conference in Otolaryngology**, London. *10/97*

5–8 **6th International Congress on Sound and Vibrations**, Copenhagen. *2/98*

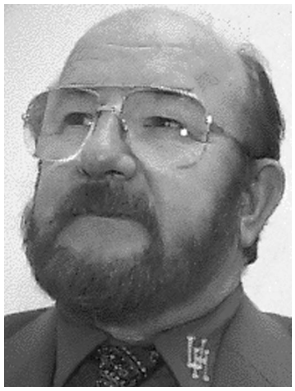
September 1999

1–4 **15th International Symposium on Nonlinear Acoustics (ISNA-15)**, Göttingen. *10/97*

OBITUARIES

This section of the Journal publishes obituaries concerning the death of Fellows of the Society and other acousticians eminent in the world of acoustics. When notified, the Editor-in-Chief solicits a summary of the person's life and contributions from an ASA member thoroughly familiar with the details, if possible. If a promised obituary is never received, a brief obituary notice may be published later.

Bill D. Cook • 1935–1997



Bill D. Cook, a Fellow of the Acoustical Society of America, died at his home in Houston, Texas on 19 April 1997. He was born in Oklahoma and graduated from Oklahoma State University in Physics in 1957. He went on to Michigan State University where he received the M.S. degree in Physics in 1959 and a Ph.D. also in Physics in 1962. After serving in the U.S. Army as a First Lieutenant from 1963 to 1965, he went to the University of Houston where he became an Assistant Professor in 1968. He attained the rank of Professor in both the Departments of Mechanical and Electrical Engineering.

While at Oklahoma State University Bill worked with G. B. Thurston and L. E. Hargrove on a problem in nonlinear acoustics. Bill's Ph.D. work was in the area of acousto-optics and several of his earlier papers were co-authored with researchers at Michigan State including E. A. Hiedemann, who was his Ph.D. advisor, M. A. Breazeale, and W. R. Klein. While working as a post doctoral fellow at Michigan State's Ultrasonics Labs on problems of light/ultrasound interactions, he and his co-workers defined an important acousto-optic parameter which is known today as the Cook–Klein parameter. At the University of Houston he continued his interest in acousto-optics but branched out into other areas such as liquid crystals, Walsh functions, holography, acoustic microscopy, room acoustics, and noise control. He and his fifteen students reported their work in over one hundred publications, including twenty-seven refereed journal papers. As late as December 1996, a picture taken by Bill and one of his students was used in the San Antonio IEEE Meeting as an illustration of how the Poynting vector and the wave vector are not always aligned in the propagation of acoustical waves in crystals. Bill's incisive scientific insight was in constant demand and was always unselfishly given.

Active in campus life, Bill served as President of the Faculty Senate and Associate Chairman and Director of Undergraduate Advising for the Department of Mechanical Engineering. He was also a member of the University Athletics Advisory Committee.

In the Acoustical Society Bill served on the Physical Acoustics Technical Committee, and was its chairman from 1971 to 1973. He served as chairman of the Houston Chapter, and was active on the Regional Chapters Committee. He served as Technical Program Committee Chairman for the 122nd meeting in Houston in 1991.

Bill is survived by his wife, Donna Scott; his daughter, Athena Adams; son Garret Cook; and former wife and mother of his children, Dolores Spencer. A celebration of Bill's life held at the University of Houston Athletics Alumni Center drew a large attendance of university administrators, faculty colleagues, students, alumni, athletes, and members of the Unitarian-Universalist congregation to which he and Donna belonged.

With Bill's passing the world has lost a fine acoustician and we have lost a friend who enjoyed life himself and made it fun for others, who was always willing to take the time to listen to a problem and to try to help. The Bill D. Cook Scholarship Endowment has been established at the Cullen College of Engineering of the University of Houston. Scholarships will support students in Mechanical Engineering who may be interested in studying acoustics. If you wish information about the endowment, please contact Vita Como at either (713) 743-4200 or vcomo@uh.edu.

ROBERT D. FINCH

Charles L. Bartberger • 1914–1997

Charles L. Bartberger, a Fellow of the Acoustical Society of America, and a major contributor to the world of modeling underwater sound propagation, died 21 July 1997 after a bout with cancer. He was born in Pittsburgh, Pennsylvania. He received a B.S. in mathematics from Allegheny College in Meadville, PA, then an M.A. in physics from Harvard University in 1936. He continued graduate study and taught at Carnegie Institute of Technology in Pittsburgh from 1936 to 1940, and then taught physics at the Altoona Undergraduate Center of the Pennsylvania State University between 1940 and 1941.

Charlie entered federal civil service at the Naval Aircraft Factory in Philadelphia in 1941, where he was introduced to the rigors of developing and testing aircraft instruments and systems. In 1952 he transferred to the new naval facility in Warminster, Pennsylvania. At Warminster he began developing computer models of underwater sound propagation, and soon discovered that large computing power was necessary. His need influenced the Naval Air Development Center (NADC) to become one of the first facilities to obtain truly large scale computational capability. NADC became a leader in large scale computational models, especially for aircraft research and development. However Charlie's speciality was underwater acoustics. He created NADC's original propagation loss prediction program, which developed into a series of models for understanding the complicated behavior of underwater sound.

During over 30 years at NADC Charlie provided acoustic modeling expertise in scientific exchanges among Navy laboratories, served on numerous Navy committees on underwater acoustics, participated in Navy symposia, and taught an in-house course at NADC on the principles of underwater sound. His ''Lecture Notes on Underwater Acoustics'' became part of the core curriculum material when Penn State introduced its underwater acoustics program. Charles Bartberger's prediction models played a major role in all Sonobuoy and Sonar research and development projects at NADC. These models have also been used by industry in system studies and new proposals to the Navy. For his contributions to underwater acoustic modeling, he was elected a Fellow of the Acoustical Society of America in 1981.

He officially retired from NADC in 1981 with 40 years of Navy Civil Service, but continued his work as a rehired annuitant, then offered his services without compensation. In 1988 Charlie was elected an NADC Fellow. In 1990 Charlie joined the Navmar Applied Sciences Corporation in Warminster, still developing the underwater acoustic models that he readily admitted he considered a ''hobby.''

Charlie was a pioneer in recycling, preserving stacks of fan-fold paper for multiple use. Many of his original mathematical derivations can be found on the reverse side of these large format papers. Charlie's humor was ever present. He penned and recited many limericks, and parodies of Gilbert and Sullivan tunes which poked fun at bureaucracy. Charlie Bartberger was a most beloved individual. He and his work will never be forgotten.

ARTHUR HORBACH
Navmar Applied Sciences Corporation

 0001-4966/98/103(4)/1696/1/$10.00

REVIEWS OF ACOUSTICAL PATENTS

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

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception Incorporated, Box 39536, Los Angeles, California 90039*
D. LLOYD RICE, *11222 Flatiron Drive, Lafayette, Colorado 80026*
CARL J. ROSENBERG, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ERIC E. UNGAR, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *University of Rochester Medical Center, 601 Elmwood Avenue, Rochester, New York 14642*

5,590,090

43.35.Zc ULTRASONIC DETECTOR USING VERTICAL CAVITY SURFACE EMITTING LASERS

Anil R. Duggal and Christopher P. Yakymyshyn, assignors to General Electric Company
31 December 1996 (Class 367/7); filed 31 March 1995

''An ultrasound/vibration detector array includes an electrically pumped vertical cavity surface emitting laser (VCSEL) array. The cavity length of each laser or pixel of the array is modulated by the acoustic field at the point where the acoustic field contacts the pixels. The resulting laser output is frequency modulated by the acoustic field. This modulation is converted to amplitude modulation at the detector head and then either detected with a charge-coupled-device (CCD) array with the information being electrically communicated to the signal processing assembly or sent directly by optical fiber to the signal processing assembly for processing. This ultrasound/vibration detector array provides high frequency bandwidth detection, fine spatial resolution, and minimal electric cabling.''—DWM

5,687,905

43.35.Zc ULTRASOUND-MODULATED TWO-FLUID ATOMIZATION

Shirley Cheng Tsai, Irvine, CA
18 November 1997 (Class 239/4); filed 5 September 1995

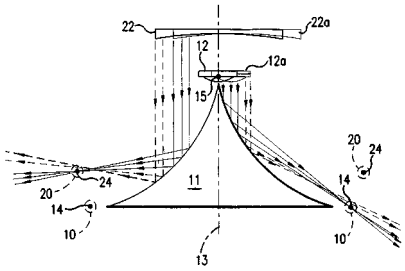
This patent describes ''a dramatic enhancement of the two-fluid atomization art through the discovery of a method of causing resonance between capillary waves in the ultrasound range in a flowing liquid stream and the waves created at the surface of that stream of liquid by an impinging gas stream. ... The surface of a stream of liquid issuing from the outlet or nozzle of an ultrasonic atomizer is impinged upon by a stream of gas. That impinging stream of gas then develops, at the surface of the liquid stream already sustaining its own wave motion, a flow of gas substantially parallel to the flow of the liquid stream that moves faster than that surface of the liquid stream. The flow of the gas at the surface of the liquid stream moves sufficiently faster than the surface of the liquid stream to generate waves at the surface of the liquid stream. The wavelength of the waves generated by the impinging gas on the surface of the liquid stream are modulated by velocity control of the impinging gas stream and resonate with the liquid stream waves. The resonance results in an atomization wherein the droplets are smaller and the droplet size distribution is reduced ... .''—DWM

5,616,892

43.38.Ja VIRTUAL IMAGING MULTIPLE TRANSDUCER SYSTEM

Michael W. Ferralli, assignor to Technology Licensing Company
1 April 1997 (Class 181/155); filed 16 January 1996

Three earlier patents to the same inventor are based on transducer–reflector combinations capable of producing virtual sources of acoustic or electromagnetic radiation. In three dimensions, paraelliptic reflector **11** is shaped something like a pointed hat. Energy from first transducer **12** is



concentrated into virtual ring **10**, whereas energy from parabolic transducer **22** is concentrated into virtual ring **24**. Because the two rings are close together, interference between the two sources is minimized.—GLA

5,661,271

43.38.Ja ACOUSTIC SPEAKER ENCLOSURE HAVING A STACKED CONSTRUCTION

Charles E. Moser, San Jose, CA
26 August 1997 (Class 181/199); filed 5 June 1995

A hollowed-out butcher block is used as a loudspeaker enclosure. Sheets of damping material are interposed between the laminations to ''avoid transmission of vibrations between structural layers.''—GLA

5,677,960

43.38.Lc ON-VEHICLE SOUND CONTROL APPARATUS

Atsushi Unno *et al.*, assignors to Victor Company of Japan
14 October 1997 (Class 381/86); filed in Japan 11 May 1995

Noise in most motor vehicles correlates well with engine speed. If a vehicle is fitted with an electronic tachometer, the same sensing signal can be used to adjust automatically the level of the sound system. If such a sensor is lacking, the patent describes a clever alternative; the frequency of the alternator's output will do just as well. For reasons not explained, the

 0001-4966/98/103(4)/1697/8/$10.00

desired signal is extracted from the dc charging circuit rather than directly from the ac alternator.—GLA

5,664,019

43.38.Si SYSTEMS FOR FEEDBACK CANCELLATION IN AN AUDIO INTERFACE GARMENT

Weija Wang and James H. Boyden, assignors to Interval Research Corporation
2 September 1997 (Class 381/71.1); filed 8 February 1995

In a separate patent application Wang and Boyden propose to replace conventional handsets, headsets, and loudspeaking telephones with a kind of coat-of-many-transducers. Such a garment-based interface gives the user full freedom of his head and hands. The present patent describes how banks of adaptive filters can be connected to suppress acoustic feedback for such a device.—GLA

5,657,391

43.38.Vk SOUND IMAGE ENHANCEMENT APPARATUS

Katsunori Jyosako, assignor to Sharp Kabushiki Kaisha
12 August 1997 (Class 381/1); filed in Japan 24 August 1994

Phase shifters, filters, and delays are combined in a circuit that seems complicated but is easier to implement than a FIR filter. ''This arrangement enables virtual speakers to be located at the back of a listener...Moreover, it is possible to reproduce reverberation sounds from the front, back and sides.''—GLA

5,666,425

43.38.Vk PLURAL-CHANNEL SOUND PROCESSING

Alastair Sibbald *et al.*, assignors to Central Research Laboratories Limited
9 September 1997 (Class 381/26); filed in the United Kingdom 18 March 1993

Audio signals from an artificial head plus several monophonic microphones are processed, equalized, and combined to make a two-channel stereophonic recording that can be played back through earphones or loudspeakers.—GLA

5,677,489

43.40.At DISTRIBUTED STRUCTURAL CHARACTERISTIC DETECTION SYSTEM USING A UNIDIRECTIONAL ACOUSTIC WAVEGUIDE

William Bert Spillman, Jr., assignor to Simmonds Precision Products, Incorporated
14 October 1997 (Class 73/598); originally filed 19 July 1993

The patent describes ''apparatus for acoustically analyzing a structure having distributed characteristics comprising a first acoustic waveguide for transmitting an acoustic signal through a portion of the structure, a second acoustic waveguide for receiving a reflected acoustic signature of the structure and transmitting the acoustic signature to a sensor; the sensor producing an output that corresponds to the acoustic signature. The apparatus and methods are particularly adaptable to detecting distributed structural characteristics.''—DWM

5,665,917

43.40.Vn METHOD FOR CONSTRUCTING SUPERSONIC SHOCK-WAVE VIBRATOR DEVICES FOR APPLYING VIBRATORY FORCE OR MEASURING PURPOSES OR TESTING PURPOSES BY USING CAVITATING SPACE

Stephen Bruce Berman, Austin, TX
9 September 1997 (Class 73/662); filed in Japan 14 November 1995

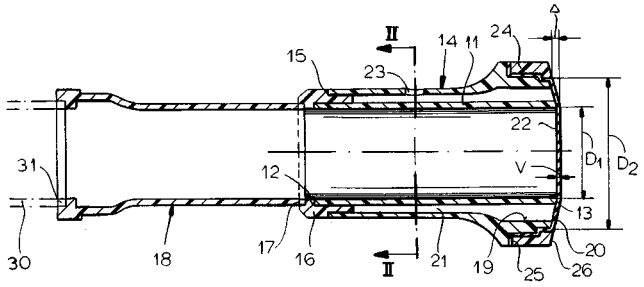
The concept of this patent consists simply of using piezoelectric drivers to create cavitation in a fluid within a container and permitting the resulting shock waves to impinge on test items. The patent also refers to temporal and spatial control devices, as well as to recording, display, and analysis devices, but it does not describe these.—EEU

5,662,064

43.50.Ed HIGH ACOUSTIC OUTPUT HORN

Laszlo Gyorgy, Pomaz, Hungary
2 September 1997 (Class 116/142 FP); filed in Hungary 30 October 1991

This is a small, lightweight horn intended to be used by sports fans to generate sound levels ''in excess of 100 dB'' with relatively little exertion. Air blown into hole **23** displaces diaphragm **22** and sound emanates



from end **12** or extension **18**. The novelty lies in the geometry of the two coaxial tubes in relation to the stretched diaphragm.—GLA

5,681,072

43.50.Gf SOUND ABSORBER FOR MOTOR VEHICLES

Klaus Stricker, assignor to Georg Naher GmbH
28 October 1997 (Class 296/39.3); filed in Germany 11 April 1995

This patent is for a double-wall sound absorber to go around an automobile motor (an engine cowling), with a perforated inner wall toward the noise source and an unperforated rear wall. Together these walls make a sound absorber, combining plate and perforated aspects, able to absorb both high and lower frequencies.—CJR

5,570,425

43.50.Ki TRANSDUCER DAISY CHAIN

Seth D. Goodman *et al.*, assignors to Digisonix, Incorporated
29 October 1996 (Class 381/71); filed 7 November 1994

An active noise attenuation system may employ numerous sensors and output transducers, all connected to a central processing unit. The patent describes ways in which these different kinds of terminals can be daisy-chained along a single communications cable.—GLA

5,684,278

43.55.Ev ACOUSTICAL CERAMIC PANEL AND METHOD

Robert Deane Yasukawa *et al.*, assignors to Lockheed Missiles & Space Company
4 November 1997 (Class 181/286); filed 18 November 1994

The patent is for a rigid acoustic insulator panel composed of a rigid matrix formed by randomly oriented, fused silica fibers. The matrix has a three-dimensionally continuous network of intercommunicating voids. The fibers can be arranged to have greater flow resistance characteristics through the panel. The patent also describes a method of preparing the panel.—CJR

5,604,810

43.55.Lb SOUND FIELD CONTROL SYSTEM FOR A MULTI-SPEAKER SYSTEM

Hirofumi Yanagawa, assignor to Pioneer Electronics Corporation
18 February 1997 (Class 381/63); filed in Japan 16 March 1993

''The sound field control system renders it possible to provide a sound effect, in a home listening room, similar to that of a reflected sound environment such as a concert hall. More particularly, delayed audio signals corresponding to the reflected sound of the direct sound (are) added to the original audio signal representing an original direct sound and the composite signal is reproduced by a multi-speaker system comprising four or more loudspeakers.''—GLA

5,652,798

43.55.Lb REVERBERANT CHARACTERISTIC SIGNAL GENERATION APPARATUS

Koichiro Mizushima, assignor to Matsushita Electric Industrial Company
29 July 1997 (Class 381/63); filed in Japan 7 June 1995

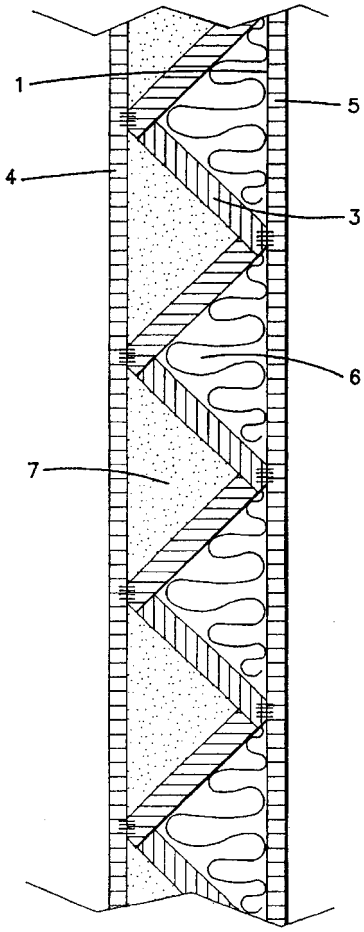
In order to measure (and later recreate) the reverberant sound characteristics of a particular concert hall, one can place a dummy head in the preferred seat and make a binaural recording of impulse response. The patent teaches that, knowing the dimensions of the hall and the geometry of the test setup, the signal from a single omnidirectional microphone can be processed to yield the equivalent binaural recording. This sounds needlessly complicated, but as a method of generating and analyzing pulse trains to yield the desired results, this is an improvement.—GLA

5,685,124

43.55.Rg WALL, CEILING OR ROOF ELEMENTS WITH HEAT INSULATION PROPERTIES ON ONE SIDE AND SOUND INSULATION PROPERTIES ON THE OTHER

Adolf Jandl, Jr., Austria
11 November 1997 (Class 52/783.11); filed in Germany 21 April 1994

This construction element has two outer skins separated by a zig-zag folded or bent plate interlayer, which separates the two outer skins. Half of the voids facing the inside are stuffed with insulation, said to provide for sound insulation properties on one side.—CJR



5,657,419

43.72.Ar METHOD FOR PROCESSING SPEECH SIGNAL IN SPEECH PROCESSING SYSTEM

Hah-Young Yoo *et al.*, assignors to Electronics and Telecommunications Research Institute
12 August 1997 (Class 395/2.32); filed in Republic of Korea 20 December 1993

This method is intended to reduce the computational load of code-excited (CELP) vocoders by preanalyzing the voice pitch and using that information to speed up the excitation codebook search. An autocorrelation-based pitch analysis allows more efficient resynthesis by narrowing the range of later pitch calculations within the codebook search loop.—DLR

5,652,843

43.72.Gy VOICE SIGNAL CODING SYSTEM

Joji Kane and Akira Nohara, assignors to Matsushita Electric Industrial Company
29 July 1997 (Class 395/217); filed in Japan 27 May 1990

This system is a controller intended for use with an unspecified speech coding system. Its purpose is to detect periods of voicing activity and to enable operation of the coding system only when voicing is detected. A variation allows the coder to be enabled also during consonant noise periods, while still rejecting nonspeech.—DLR

5,657,418

43.72.Gy PROVISION OF SPEECH CODER GAIN INFORMATION USING MULTIPLE CODING MODES

Ira Alan Gerson and Mark Antoni Jasiuk, assignors to Motorola, Incorporated
12 August 1997 (Class 395/2.16); filed 5 September 1991

This multi-mode speech coder sends a two-bit mode code identifying one of four different coding schemes used for the present frame. The modes include subframe analysis for voiceless periods and several flavors of excitation codebooks such as vector sum excited LP and pitch-related excitations. The speech coding mode is selected based primarily on a voicing periodicity measure.—DLR

5,657,420

43.72.Gy VARIABLE RATE VOCODER

Paul E. Jacobs *et al.*, assignors to Qualcomm, Incorporated
12 August 1997 (Class 395/2.32); filed 11 June 1991

This speech vocoder continually tracks the input speech activity level and adjusts the vocoding frame rate accordingly. Either a lack of input speech activity or a reduced rate of speaking at the input can result in a reduced coding rate. The coding rates drop by powers of 2 down to $\frac{1}{64}$ of the full rate.—DLR

5,659,659

43.72.Gy SPEECH COMPRESSOR USING TRELLIS ENCODING AND LINEAR PREDICTION

Victor D. Kolesnik *et al.*, assignors to Alaris, Incorporated; GT Technology, Incorporated
19 August 1997 (Class 395/228); filed 26 July 1993

This code-excited LP vocoder incorporates a trellis algorithm to encode sequences of subframes to speed up and simplify the codebook search. A trellis is a transition graph which encodes probabilites of sequences of transitions between subframe types. By basing the codebook search on the trellis sequences, the best excitation can be found much more rapidly than by pursuing an unguided search of the excitation vectors.—DLR

5,659,661

43.72.Gy SPEECH DECODER

Kazunori Ozawa, assignor to NEC Corporation
19 August 1997 (Class 395/2.37); filed in Japan 10 December 1993

The patent discloses an improved method for reducing noise in a code-excited LP vocoder. In the receiver, after the speech signal has been resynthesized, it is passed through a postsynthesis filter having coefficients that are set by a spectrum parameter transmitted along with pitch, amplitude, and excitation codes. The postfilter performs an auditory masking operation, reducing any components which are not consistent with speech perception characteristics.—DLR

5,652,828

43.72.Ja AUTOMATED VOICE SYNTHESIS EMPLOYING ENHANCED PROSODIC TREATMENT OF TEXT, SPELLING OF TEXT AND RATE OF ANNUNCIATION

Kim Ernest Alexander Silverman, assignor to NYNEX Science & Technology, Incorporated
29 July 1997 (Class 395/2.69); filed 19 March 1993

Within the limited context of telephone directory entries, this text-to-speech synthesizer system provides more natural sounding speech output by applying prosodic rules covering stress and intonation patterns and speech rate changes. Special cases are handled, such as stress reduction in anaphoric references and a reasonable stress assignment for long, complex names.—DLR

5,657,380

43.72.Ja INTERACTIVE DOOR ANSWERING AND MESSAGING DEVICE WITH SPEECH SYNTHESIS

Todd F. Mozer, assignor to Sensory Circuits, Incorporated
12 August 1997 (Class 379/88); filed 27 September 1995

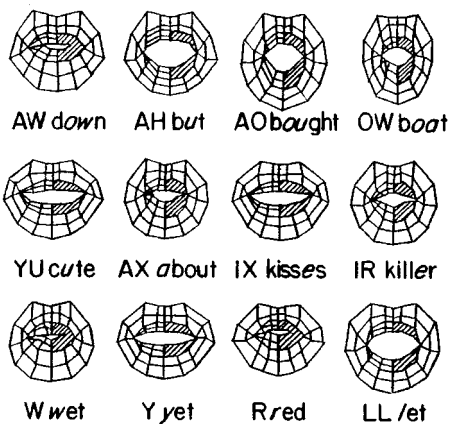
The patent shows the application of speech recognition and playback in a doorbell answering system. More elaborate than the average phone answering machine, this device uses recognition to form guesses about the new arrival's intents and desires. While a dialogue with the visitor is in progress and being recorded, the resident of the home may simultaneously and covertly converse with the system to learn more about the visitor.—DLR

5,657,426

43.72.Ja METHOD AND APPARATUS FOR PRODUCING AUDIO-VISUAL SYNTHETIC SPEECH

Keith Waters and Thomas M. Levergood, assignors to Digital Equipment Corporation
12 August 1997 (Class 395/2.85); filed 10 June 1994

The patent describes a system for synchronizing a visual display of a moving face and lips with audio playback of the corresponding speech signal. The basic arrangement is text-to-speech presentation, meaning that a



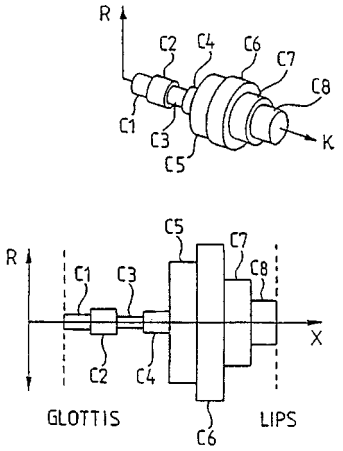
phonetic string is available without further processing. Provision is made for nonlinear motion of selected mouth parts, depending on specific phoneme durations.—DLR

5,659,658

43.72.Ja METHOD FOR CONVERTING SPEECH USING LOSSLESS TUBE MODELS OF VOCAL TRACTS

Marko Vaenskae, assignor to Nokia Telecommunications OY
19 August 1997 (Class 395/2.7); filed in Finland 12 February 1993

This uses speaker-specific characteristics of the vocal tract tube shape to synthesize voices representing different speakers. Speech sounds are converted to tube shape reflection coefficients using well-known LP-based conversions. Speaker-specific values are extracted by comparing the tube



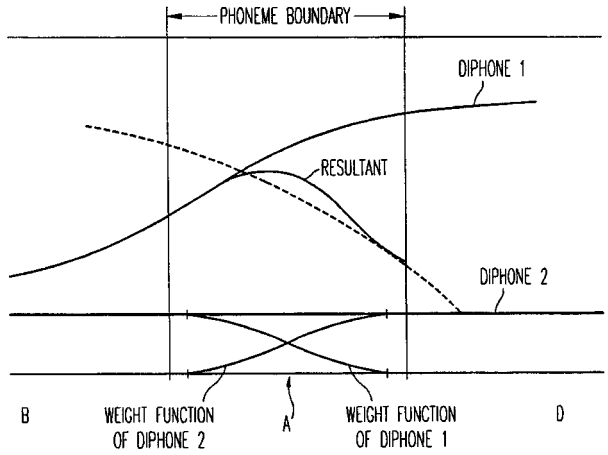
shapes with those of similar sounds averaged across speakers. These speaker differences are then reapplied to the tube shapes generated by a synthesizer system.—DLR

5,659,664

43.72.Ja SPEECH SYNTHESIS WITH WEIGHTED PARAMETERS AT PHONEME BOUNDARIES

Jaan Kaja, assignor to Televerket
19 August 1997 (Class 395/2.74); filed in Sweden 17 March 1992

This is a method of speech synthesis by recombination of multi-phoneme-sequence formant parameter sets. Transitions are generated by taking a weighted average of the parameter shapes of stored polyphones con-



taining the same phoneme sequence as that surrounding each transition period. The weights provide a blending interpolation from the averaged formant patterns across the transition period.—DLR

5,657,422

43.72.Kb VOICE ACTIVITY DETECTION DRIVEN NOISE REMEDIATOR

Thomas John Janiszewski and Michael Charles Recchione, assignors to Lucent Technologies, Incorporated
12 August 1997 (Class 395/2.37); filed 28 January 1994

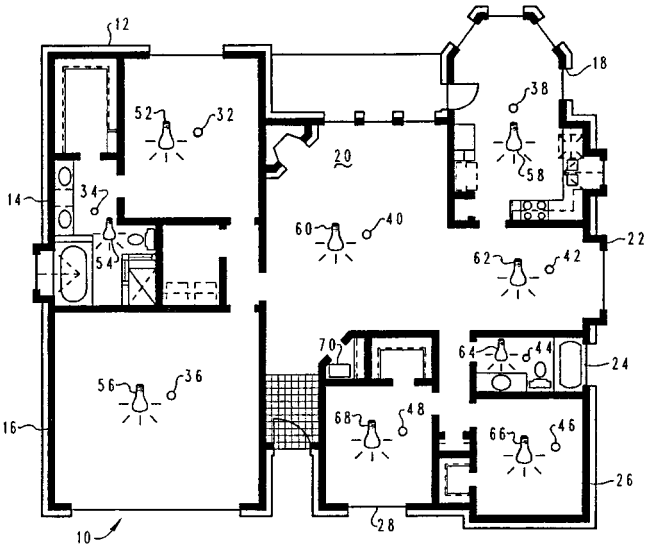
This system improves the fidelity of speech transmitted by a speech coding system, such as a cellular telephone, when operated in a noisy environment. The incoming signal is high-pass filtered at all times, depending on the detected background noise level. During times of speech pauses, a low-pass filtered noise signal is substituted for the speech signal.—DLR

5,657,425

43.72.Kb LOCATION DEPENDENT VERBAL COMMAND EXECUTION IN A COMPUTER BASED CONTROL SYSTEM

William J. Johnson, assignor to International Business Machines Corporation
12 August 1997 (Class 395/2.84); filed 15 November 1993

This is a multiple-input speech recognition system intended for use such as in a home with several rooms. Each room has a microphone, with each microphone signal being independently sampled. Each signal may have independent associated syntax and/or vocabulary. There is a bit of hand waving regarding multiple simultaneous inputs. Apparently, each signal is



independently recognized and the signal level is noted. If two or more locations produce similar outputs and the microphone levels suggest that some may be echoes of another, then the echoes are ignored. There is no mention of any form of interprocessing among signals.—DLR

5,659,597

43.72.Kb SPEECH RECOGNITION SYSTEM FOR ELECTRONIC SWITCHES IN A NON-WIRELINE COMMUNICATIONS NETWORK

Bernard F. Baries *et al.*, assignors to Voice Control Systems, Incorporated
19 August 1997 (Class 455/563); filed 13 April 1992

This system would utilize speech recognition capabilities in various ways in a cellular phone system. Recognition functions may be distributed to a base unit and/or to user hand-held units, with allowance for the resulting speaker dependence/independence. Voice commands may be used to control phone-related operations or external systems, independent of the phone.—DLR

5,659,662

43.72.Pf UNSUPERVISED SPEAKER CLUSTERING FOR AUTOMATIC SPEAKER INDEXING OF RECORDED AUDIO DATA

Lynn D. Wilcox and Donald G. Kimber, assignors to Xerox Corporation
19 August 1997 (Class 395/2.54); filed 12 April 1994

The patent discloses a method of distinguishing speakers from a recording of multiple, unidentified speakers. Speaker clustering is based on a distance measure between sets of 20 cepstral coefficients. An initial large group of clusters is defined by arbitrary segments of the speech data, each a

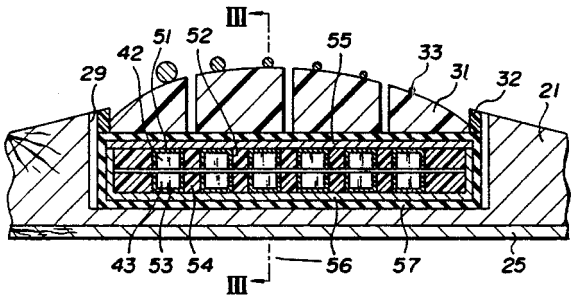
few seconds in length. Cluster merging consists of reassigning the time points of speaker change by Viterbi decoding in the graph of cluster distances. The merging process is repeated until the number of clusters (merged signal segments) equals the original number of speakers.—DLR

5,410,101

43.75.Gh PICKUP UNIT FOR ELECTRIC STRING INSTRUMENT

Hiroshi Sakurai, assignor to Yamaha Corporation
25 April 1995 (Class 84/731); filed in Japan 22 May 1992

This guitar pickup unit has dual layers **42**,**43** of piezoelectric elements clamped between a string holder **31** and the bridge **21**. The layer closer to the string holder is intended to respond chiefly to string vibrations, and the



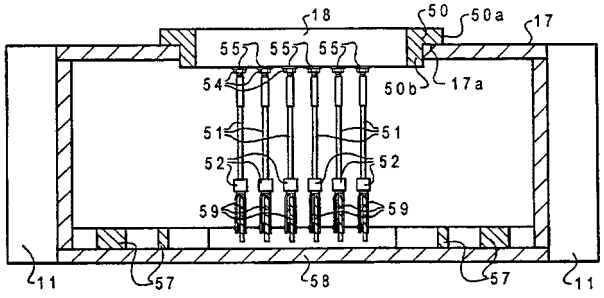
lower layer is intended to pick up guitar body vibrations. Signal outputs from the two layers can be blended together to produce a wide range of tonal timbre.—DWM

5,689,074

43.75.Gh MUSICAL INSTRUMENT

Stanley Penridge, Austin, TX
18 November 1997 (Class 84/294); filed 7 March 1996

This acoustic guitar has a special double X-shaped bracing **57** inside the bottom **58** of the guitar, and a removable insert assembly mounted on ring **50** that fits into the circular sound hole **18** in the top of the guitar. Under each of the six strings is a pair of rods **51** secured at the top to ring **50** and



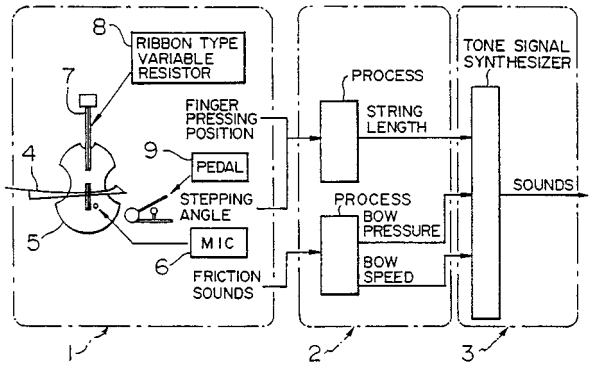
joined together at the bottom by a wooden reed, the end **59** of which is shown. The intent is to tune the vibratable members under each string to appropriate resonance frequencies that will modify the tonal characteristics of the instrument.—DWM

5,396,025

43.75.Tv TONE CONTROLLER IN ELECTRONIC INSTRUMENT ADAPTED FOR STRINGS TONE

Motoichi Tamura, assignor to Yamaha Corporation
7 March 1995 (Class 84/736); filed in Japan 11 December 1991

This electronic musical instrument synthesizes bowed string tone signal controlled by bowing and fingering on a substitute violin-shaped device **5** having a fingerboard **7** on which a ribbon-type variable resistor **8** is lo-



cated. The position of the player's finger on the resistor ribbon gives electrical control of the tonal frequency in accordance with the simulated string length. Friction sounds generated by the motion of bow **4** control tonal amplitude to correspond to the amount of bow pressure used.—DWM

5,627,334

43.75.Tv APPARATUS FOR AND METHOD OF GENERATING MUSICAL TONES

Sayoko Hirano and Yutaka Washiyama, assignors to Kawai Musical Instrument Manufacturing Company
6 May 1997 (Class 84/623); filed in Japan 27 September 1993

In contrast to many other digital electronic musical instruments in which individual tones are generated directly from stored waveforms, this system uses digitally stored ''formant'' waveforms, then modifies components to vary the spectrum across the musical scale to obtain the spectral variation required for a coherent formant corresponding to the individual tone color desired.—DWM

5,637,821

43.75.Tv STORING AND INTERPOLATING MEANS FOR A MUSICAL SOUND GENERATING DEVICE

Gen Izumisawa and Yutaka Washiyama, assignors to Kabushiki Kaisha Kawai Gakki
10 June 1997 (Class 84/604); filed in Japan 30 March 1990

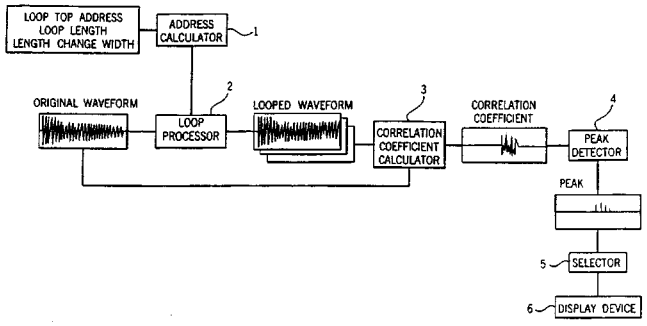
The musical sound generating device of this patent reads data repeatedly between a loop-top address and a loop-end address, in common with patent 5,672,836 reviewed below. However, this system also uses an interpolating circuit which ''allows minimization of the circuit scale and realization of a cheap musical sound generating device.''—DWM

5,672,836

43.75.Tv TONE WAVEFORM PRODUCTION METHOD FOR AN ELECTRONIC MUSICAL INSTRUMENT AND A TONE WAVEFORM PRODUCTION APPARATUS

Toshiya Yoshida, assignor to Kabushiki Kaisha Kawai Gakki
30 September 1997 (Class 84/607); filed in Japan 23 May 1995

This patent discloses apparatus and method ''for forming a repeated waveform that is employed for an electronic musical instrument, which includes the steps of: setting, as basic data, a loop top address, an initial value of a loop length, and a length change range; changing the loop length



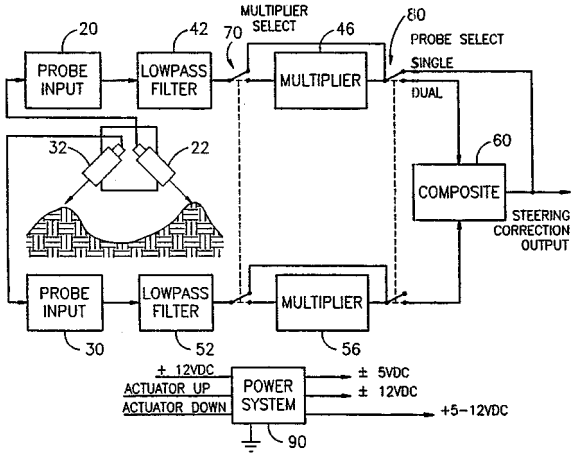
within the length change width; forming repeated waveforms that are obtained after cross-fading is performed in consonance with existing conditions; acquiring correlation coefficients of the repeated waveforms, which are obtained by cross-fading, at corresponding portions of an original waveform; and selecting a predetermined number of waveforms beginning at a waveform that is most correlative to determine a repeated waveform... .''—DWM

5,410,479

43.80.Pe ULTRASONIC FURROW OR CROP ROW FOLLOWING SENSOR

William B. Coker, Beatrice, NE
25 April 1995 (Class 364/424.02); filed 17 August 1992

For automatic steering of an agricultural vehicle along a desired track between adjacent crop rows this ultrasonic sensing system uses transducers



22, **32** wth appropriate bridge circuitry to provide a steering correction output when the two signals mismatch.—DWM

5,671,746

43.80.Vj ELEVATION STEERABLE ULTRASOUND TRANSDUCER ARRAY

William Robert Dreschel and Bradley R. Nelson, assignors to Acuson Corporation
30 September 1997 (Class 128/61.01); filed 29 July 1996

The ultrasonic energy in the imaged planes is focused in elevation by the curvature of the array. Each plane is positioned at an angle that is determined by the known curvature of the array and the known dimensions of the elements. The imaged plane is selected from a number of available planes by electronic switching.—RCW

5,671,747

43.80.Vj ULTRASOUND PROBE HAVING INTERCHANGEABLE ACCESSORIES

Brian G. Connor, assignor to Hewlett-Packard Company
30 September 1997 (Class 128/662.06); filed 24 January 1996

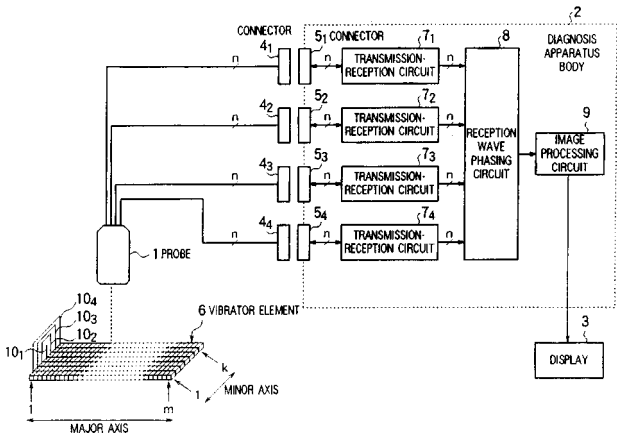
The accessories of this probe that is intended for intraoperative use adapt the shape of the probe to access body structures in the confined space of an intraoperative environment. Since the accessories are separable from the probe, they are easily sterilized and may be discarded after use. The accessories also may have low manufacturing costs relative to the cost of multiple dedicated probes, so that the overall cost of an ultrasonic imaging system may be reduced.—RCW

5,673,698

43.80.Vj MULTICHANNEL ULTRASONIC DIAGNOSIS APPARATUS

Kazutaka Okada and Akira Sasaki, assignors to Hitachi Medical Corporation
7 October 1997 (Class 128/661.01); filed in Japan 21 April 1994

In this apparatus, a two-dimensional array of transducer elements is divided into two or more groups. Elements in each group are wired to



corresponding connectors. Elements belonging to different groups may be used simultaneously.—RCW

5,673,699

43.80.Vj METHOD AND APPARATUS FOR ABERRATION CORRECTION IN THE PRESENCE OF A DISTRIBUTED ABERRATOR

Gregg E. Trahey *et al.*, assignors to Duke University
7 October 1997 (Class 128/661.01); filed 31 May 1996

A transmit aperture is used to send a focused beam from one position, and is then translated to another position from which a second beam is sent with a focus at the same point as the initial beam. An element that receives an echo from the focus of the first beam is paired with an element that receives an echo from the focus of the second beam and that has a separation equal and opposite to the spatial translation of the transmit aperture.The signals on the elements are compared to estimate phase aberrations. The

aberration estimates are then used to compensate or adjust transmit or receive signals.—RCW

5,673,700

43.80.Vj ULTRASONIC DIAGNOSIS APPARATUS AND IMAGE DISPLAYING SYSTEM

Nobuo Yamazaki and Akihiro Sano, assignors to Kabushiki Kaisha Toshiba
7 October 1997 (Class 128/661.09); filed in Japan 2 October 1992

This patent, which has resulted from a division from US Patent 5,622,174, using 95 figures, describes a system that determines the motion of structures such as heart walls from the Doppler shift of echos, and then displays the movement in color.—RCW

5,676,147

43.80.Vj ULTRASONIC RECEIVE BEAMFORMER WITH PHASED SUBARRAYS

Joseph G. Petrofsky *et al.*, assignors to Acuson Corporation
14 October 1997 (Class 128/661.01); filed 16 August 1996

Signals in this beamformer are applied to subarray processors that include at least one phase shifter and a summer. Each processor supplies a summed subarray signal to a respective beamformer processor. The phase angles for any one of the subarray processors sum essentially to zero and are independent of the time delay of the respective beamformer processor. The phase angles are updated at a slower rate than the focusing update rate of the beamformer processors. The resolution of the time delay in the beamformer is essentially as fine as the time resolution of the phase angles of the phase shifters.—RCW

The functionally and physiologically plastic adult auditory system

Catherine V. Palmer[a)]
University of Pittsburgh, Pittsburgh, Pennsylvania 15260

Charles T. Nelson
Pittsburgh Veterans Administration Medical Center, Pittsburgh, Pennsylvania 15206

George A. Lindley IV
University of Pittsburgh, Pittsburgh, Pennsylvania 15260 and Pittsburgh Veterans Administration Medical Center, Pittsburgh, Pennsylvania 15206

(Received 23 July 1997; revised 11 November 1997; accepted for publication 31 December 1997)

The potential for functional and underlying physiological change in the adult auditory system is examined through review and evaluation of several sets of literature including auditory deprivation and recovery, auditory learning after hearing aid fitting, auditory abilities of normal listeners exposed to masking noise, and neural plasticity in the sensory and motor systems of animals. This tutorial review is meant for the reader who may be interested in auditory learning and who would like to have a summary and evaluation of the various findings to date. The focus of the review is the effect that various findings of auditory learning may have on hearing aid fitting and selection. © *1998 Acoustical Society of America.* [S0001-4966(98)01604-X]

PACS numbers: 43.10.Ln, 43.66.Ts, 43.66.Ba, 43.64.Bt [DWM]

INTRODUCTION

Over the past several years, a variety of authors have reviewed the areas of auditory deprivation (Neuman, 1996), acclimatization (Turner *et al.*, 1996), and auditory plasticity (Dybala, 1997; Moore, 1993) separately. Arlinger *et al.* (1996, p. 87S) define the auditory deprivation effect as the ''systematic decrease over time in auditory performance associated with the reduced availability of acoustic information.'' Auditory acclimatization is defined as ''a systematic change in auditory performance with time, linked to a change in the acoustic information available to the listener. It involves improvement in performance that cannot be attributed purely to task, procedural, or training effects'' (Arlinger *et al.*, 1996, p. 87S). Plasticity implies a physiologic basis for change in auditory function.

For the purpose of this tutorial, auditory learning is defined as a functional change in auditory ability for the better (acclimatization) or for the worse (deprivation). Acclimatization and deprivation can be characterized as components of auditory learning that are going in two different directions. Physiological plasticity of the auditory system is examined as the possible underlying mechanism for auditory learning that is measured through functional abilities. It is the purpose of this work to review critically the research areas of acclimatization, deprivation, and physiological plasticity in order to quantify and qualify what we presently know about auditory learning and how that knowledge may impact decisions related to the remediation of hearing loss.

There are four sets of literature that may shed light on the qualification and quantification of auditory learning. One can examine the unaided ears of monaurally aided individuals over time and determine if a decrement in performance is noted and examine if any recovery takes place after a hearing aid is supplied to the nonstimulated ear. Performance by longtime hearing-impaired individuals can be compared with performance by normal-hearing listeners to try to determine if the experience of the hearing-impaired listener with a diminished signal somehow gives him/her an advantage in a reduced auditory environment. One can look at animals deprived of specific frequency regions or taught novel auditory processing tasks and actually document indirectly what happens at various levels of the auditory system, thereby identifying a possible mechanism for explaining functional changes that have been reported in humans. Conversely, one can fit a hearing aid(s) on an individual and measure performance during the weeks/months/years postfitting to determine if any change occurs. Research has been done in all four of these areas and so far has created more questions than answers. Each area will be briefly detailed below with careful attention to the methodological problems that make it difficult to interpret the data presented thus far.

I. AUDITORY DEPRIVATION AND RECOVERY

If one hypothesizes that auditory learning occurs after hearing aid fitting, the implication is that some sort of decrement has taken place. In the past decade, several investigators have reported a late-onset auditory deprivation or nonstimulation effect in unaided ears of monaurally aided hearing-impaired individuals (Boothroyd, 1993; Burkey and Arkis, 1993; Gatehouse, 1989; Gelfand, 1994, 1995; Gelfand and Silman, 1993; Gelfand *et al.*, 1987; Hattori, 1993; Hood, 1990; Hurley, 1991, 1993; Poole and Jerger, 1994; Robinson and Gatehouse, 1995; Silman *et al.*, 1984; Silman *et al.*, 1993; Silverman, 1989; Silverman and Emmer, 1993; Silverman and Silman, 1990; Stubblefield and Nye, 1989). Table I

[a)]Electronic mail: cvp@vms.cis.pitt.edu

 0001-4966/98/103(4)/1705/17/$10.00

TABLE I. Summary of investigations related to auditory deprivation and recovery.

Author Pro/Retro	*N*	Subject Age (yrs)	Hearing level Type of loss	Type of stimulation/ nonstimulation	Hearing aid description Use (hours/day)	Period of deprivation	Type of assessment[a] Behavioral	Elect.	Morph.	Level of assessment	Studied Recovery?	Findings/Comments
Donaldson (1890) P	1	9	deaf unknown	blind and deaf from birth, no reported amplification		birth to 9 years			X	central		Atrophy of cortical areas found during autopsy.
Silman *et al.* (1984) R	44	59.5 (12.2) 57.95 (10.1)	>25 dB HL PTA symmetrical sensorineural	monaural hearing aid users binaural hearing aid users		4–5 years	W-22			peripheral		39 of monaurally aided had reduced ability over time in the unaided ear.
Gelfand *et al.* (1987) R	86	21–86	>35 dB HL PTA symmetrical sensorineural	48 monaural 19 binaural 19 unaided	8 h	4–17 years	W-22			peripheral		Decrease in performance only found for unaided of monaural subjects. Had control group of unaided individuals.
Gatehouse (1989) P	24	59.3	50.7 dB HL PTA sloping, symmetrical sensorineural	monaural hearing aid users	gain=45 dB SSPL90=130 8.6 h	4.8 years	FAAF			peripheral		Aided ear better with higher presentation level, unaided ear better at lower presentation level.
Silverman (1989) R	2	54 82	>25 dB HL PTA symmetrical sensorineural	monaural hearing aid users	8 h	12 years 8 years	W-22			peripheral		Decrement in unaided ear performance. Subject #2 had an interaural difference at 4000 Hz initially.
Stubblefield *et al.* (1989) R	60 24	40–90	unknown symmetrical sensorineural	monaural hearing aid users non hearing aid users		3–6 years	W-22			peripheral		Ear difference in perfomance in monaural group, but not in non hearing aid using group.
Silverman *et al.* (1990) R	2	30 34	>25 dB HL PTA symmetrical sensorineural	monaural hearing aid users who pursued binaural amplification because of specific localization complaints	8 h	7.6 years 2.6 years	W-22			peripheral	Yes	Unaided ear performance decreased significantly within 6.5 and 2 years of monaural fitting, binaural fitting showed a partial reversal within 8 months–2.5 years.
Hood (1990) R	12 47	unknown	unknown, but groups were matched sensorineural from Meniere's	bilateral hearing loss monaural hearing loss		unknown	MRC-PB			peripheral		The unilaterally impaired ears performed worse than the matched ear from the bilaterally impaired subjects.
Hurley (1991) P	80 98	34–72 mean=52	>29 dB HL PTA symmetrical sensorineural	binaural hearing aid users monaural hearing aid users	8 h	4–6 years decrease appeared within 2 years	NU-6 SSI DDT			peripheral central		30% of monaurally aided individuals showed a decrement in performance on the word recognition and SSI test, no significant findings for the DDT were reported.
Silman *et al.* (1993) R	1	64	PTA 75, 78 sensorineural	monaural to binaural user	8 h	2 years	W-22			peripheral	Yes	Decrease in word recognition of the unaided ear, then increase in this score postbinaural fitting.
Dalzell *et al.* (1992) R	95	43–87	moderate sensorineural	monaural and binaural users		4–15.1 years	W-22 or NU-6			peripheral		Replicated Silman *et al.* (1984) and found no greater decline in word recognition in the unaided (versus aided) ears of monaurals.
Boothroyd (1993) R	1	26	Initially moderate to severe symmetrical sensorineural	predominantly monaural user from age 4–24, lost function and decided to try binaural		24 years monaural, several months binaural	AB Isophonic			peripheral	Yes	80% word recognition in the aided ear and 40% in the unaided ear after 24 years of monaural use. After several months of binaural use, the unaided ear was at 75%.
Burkey *et al.* (1993) R	20	57.4 (16.3)	41–51 dB HL PTA represents change over time symmetrical sensorineural	monaural hearing aid users to binaural		unknown retested after 1 year of binaural use	W-22			peripheral	Yes	Although the authors report significant decrease in the performance of the unaided ear and then significant improvement after 1 year of amplification,the word recognition scores changed by only 7%, which is most likely not significant (Thornton and Raffin, 1978) Reported that the more severe the hrg. loss, the more decrement. Reported no effect for length of monaural use, but did not report the lengths of monaural use.

TABLE I. (*Continued.*)

Author Pro/Retro	*N*	Subject Age (yrs)	Hearing level Type of loss	Type of stimulation/ nonstimulation	Hearing aid description Use (hours/day)	Period of deprivation	Type of assessment[a] Behavioral	Elect.	Morph.	Level of assessment	Studied Recovery?	Findings/Comments
Dieroff (1993) R	46 46	<50	unknown but groups matched symmetrical conductive	binaural hearing aid users monaural hearing aid users		10–12 years of hearing aid use	word rec (German)			peripheral		Compared aided ear of one group to unaided ear of other group, and found a decrement in the unaided ear.
Gelfand *et al.* (1993) R	10 10	5.1–7.5	41–48 dB HL SRT symmetrical sensorineural	monaurally aided binaurally aided		5–7 years	word rec			peripheral		5/10 monaurally aided children showed a decrement in performance in the unaided ear.
Hall (1993) P	8 8	28–54 mean=37.5 21–42 mean=31	59 dB HL PTA post surgery=22 <10 dB HL PTA otosclerosis conductive normal hearing	monaurally impaired individuals normal hearing controls		unknown	MLD			retrocochlear	Yes	MLD recovered 1 year postsurgery (restoration of binaural input). No change over time in controls.
Hattori (1993) R	17 18	4.8 4.9	moderate-severe to profound sensorineural	nonalternating monaural amplification alternating or binaural amplification		13–15 years	NST (Japanese)			peripheral		The nonalternating group shows a decrement between ears over time.
Hurley (1993) P	9	56–63	36.3–55 symmetrical sensorineural	monaurally aided and then followed until a significant decrement was found, then aided binaurally		decrement found in # yrs: 2 year=4 3 years=1 4 years=2 5 years=1 6 years=1	NU-6			peripheral	Yes	After significant decrement was measured, five subjects agreed to binaural amplification and a partial reversal was found within 1–2 years. The other four subjects would not tolerate the second hearing aid.
Jerger *et al.* (1993) P	4	71 66 80 81	varied, mild to severe sensorineural	monaural hearing aid viral encephalitis: left ear damaged left-sided CVA: bilat. hearing loss gradual binaural hearing loss, no aids		10 years ? ?	W-22	middle latency		peripheral central		Found ''binaural interference,'' patients did not do well with binaural input.
Silman *et al.* (1993) P	19 28 19	23–84 40–80 28–79	>25 dB HL PTA >25 dB HL PTA <25 dB HL PTA symmetrical sensorineural control group	monaurally aided (users' choice) binuaurally aided (users' choice) normal hearing, not aided	4 h	1st test 6–12 weeks following hearing aid fitting, re-test one year after fitting	W-22 SPIN NST at 40 dB SL			peripheral		Significant difference attributed to increase in aided ear performance and decrease in unaided ear performance.
Silverman *et al.* (1993) R	8 6 2	26–77	pure tone thresholds listed for each subject asymmetrical sensorineural	never aided aided monaurally previously aided monaurally at first ''test'' session		2–13 years	W-22			peripheral	Yes	Significant decrease in performance of unaided ears, no difference over time in aided ears.
Hurley and Hurley (1994) P	15 15	children 24–32 mos	transient normal conductive	transient otitis media with effusion	N/A	transient		ABR		peripheral retrocochlear		Children with patent PEtubes and controls with no history, OME group had depressed Wave V amplitude and greater latencies for I, III, V at 75 dB HL and V for all intensities.
Hurley (1994) R	15	36–85	moderate sensorineural	monaurally aided	hearing aid was fit within one year	6 years average	PB max			peripheral		Decrease in PB max in the unaided ear.
Palmer (1994) R	13	50–60	42 dB HL PTA sensorineural	monaural, binaural	>8 h	4–11 years	NU-6 SSW MLD	ABR		peripheral retrocochlear central		Some abnormal findings found on MLD, ABR, frequency discrimination, and SSW for the unaided ear of the monaural subjects.

TABLE I. (*Continued.*)

Author Pro/Retro	*N*	Subject Age (yrs)	Hearing level Type of loss	Type of stimulation/ nonstimulation	Hearing aid description Use (hours/day)	Period of deprivation	Type of assessment[a] Behavioral	Elect.	Morph.	Level of assessment	Studied Recovery?	Findings/Comments
Poole and Jerger (1994) R	?	?	? sensorineural	monaurally aided		?	PB max			peripheral		Only the unaided ear showed a significant decline between initial and final tests.
Robinson and Gatehouse (1995) P	5 9	18 to 35 elderly	normal hearing 58 dB HL at 3 kHz sensorineural	longstanding monaural hearing aid users	20 dB gain at 3 kHz linear, probe mic measure	2 years, 7 months (average)	intensity discrim for complex stimuli			peripheral		Further explored Gatehouse (1992). Saw change only at 3 kHz, no change at 250 Hz (where they had good hearing). At higher levels, aided subjects had better discrimination. At lower levels, saw the opposite. Speculated about plasticity for intensity.
Gelfand (1994, 1995) P	6	31–64	>35 dB HL PTA sensorineural	monaural to binaural user	>8 h	6.2–15.1 years	W-22 or NU-6			peripheral	Yes	Fit binaurally when deprivation was measured (at least 4 years binaural fitting). Deprivation was defined as a decrease in speech recognition score. Sorted findings into three groups: Developed deprivation w/in 2 yrs recovered in 2 years, significant but incomplete recovery. Several years to develop, did not recover.
Anderson (1985) P	1	22	? unilateral conductive	left monaural ear canal blockage for 15 years (cotton plug)	N/A	15 years	SSW W-22 in noise			peripheral retrocochlear central	Yes	Hearing normal upon removal. Initially performed significantly worse (left ear) than normal controls on W-22 in noise, SSW. Improvement bilaterally over time, especially in left ear. Tested at 8 days, 5 months, and 14 months.
Truy *et al.* (1995)	3	19 23 36	prelingually deafened, profound	no hearing aids	N/A	lifetime			Positron Emission Tomography	central		RA in the primary auditory cortex could be increased using electrical cochlear stimulation despite longstanding deafness.

[a]PTA Pure tone average
MLD Masking level difference
ABR Auditory brainstem response
SSW Staggered spondaic words
MRC-PB Medical Research Center—Phonetically Balanced words
NU-6 Northwestern University word list #6
SSI Synthetic Sentence Identification task
Retro (R) Retrospective
DDT Dichotic Digits Test
AB Isophonic Consonant-vowel-consonant words
NST Nonsense syllable test
SPIN Speech in Noise Test
RA Reticular Activation
PB-MAX Maximum score for phonetically balanced words
W-22 Central Institute for the Deaf word list #22
Pro (P) prospective

provides a summary of investigations related to adult late-onset auditory deprivation and recovery. Gatehouse (1989), while confirming Silman *et al.*'s (1984) findings of late-onset auditory deprivation, found a reverse effect when lower presentation levels were used to obtain word-recognition scores (below 75 dB SPL). Robinson and Gatehouse (1995) found that longstanding monaural hearing aid users revealed a difference in intensity discrimination ability at low and high inputs at 3000 Hz (where they received hearing aid gain) while performance was similar across intensity levels at 250 Hz where their hearing was essentially within normal limits. The lower presentation levels were meant to mimic the exposure the unaided ear had experienced while more intense inputs mimic the experience with the hearing aid. The authors hypothesized that the findings reflected auditory learning by the process of acclimatization (adaptation) to stimulation at certain intensities.

Dalzell *et al.* (1992) replicated Silman *et al.* (1984) by retrospectively investigating monaurally aided users who had been wearing their hearing aids for at least 4 years. It is worth noting that this is a truly retrospective investigation in that it used data from patient files for both the ''pre'' and ''post'' tests that were analyzed. Dalzell *et al.* (1992) found no greater decline in word recognition in the unaided ears of monaural hearing aid users.

Four of the studies investigated recovery after asymmetrical stimulation including two (Silverman and Silman, 1990), one (Boothroyd, 1993), nine (Hurley, 1993), and six (Gelfand, 1994, 1995) subjects. Partial recovery in word recognition ability after fitting the nonstimulated ear of a monaural hearing aid user was reported in all four studies, two of which were retrospective in nature. The three individuals described by Silverman and Silman (1990) and Boothroyd (1993) pursued a second hearing aid after a change in hearing status. These individuals had a strong motivation to succeed with the new hearing aid configuration and partial recovery of word recognition scores was measured in all three subjects over an extended period of time. Of Hurley's (1993) nine subjects, five tolerated binaural amplification following a decrement in performance in their unaided ears and partial recovery of word recognition scores was reported after extended use. Four of the subjects tried and rejected binaural amplification. Gelfand's (1994, 1995) six subjects were fit binaurally as soon as deprivation was measured. Amount of recovery varied greatly between individuals, and some individuals refused to try the binaural configuration because they complained that the new input was getting in the way of the other ear. This latter reaction has been termed ''binaural interference'' by Jerger *et al.* (1993). Because one cannot force individuals to try amplification, it is hard to determine how many individuals would experience some recovery of function with amplification.

A. Evaluation of auditory deprivation and recovery investigations

Unfortunately, many of the investigations of late-onset auditory deprivation are plagued with methodological problems. Specific methodological parameters can be found in Table I. These problems include: (a) a lack of data on degree of hearing loss (little deprivation would be expected until the hearing loss exceeds the level at which the weaker speech cues in normally elevated speech become inaudible), (b) a lack of data regarding hearing aid use, and (c) outcome measures that are dependent on mid- to low-frequency sounds which the individuals have probably been hearing all along as opposed to outcome measures that tap the newly restored high-frequency hearing.

The recovery data are difficult to interpret because of a lack of knowledge regarding the degree of hearing loss, the amount of audibility that was restored by the hearing aid, and the amount of binaural use. Another methodological concern is that all of the reports investigating recovery after deprivation tested subjects under earphones. If the auditory learning has taken place through the use of a hearing aid that has distinct frequency shaping and signal processing (e.g., compression throughout the dynamic range or through some part of the dynamic range), it may not make sense to perform outcome measures through an earphone that does not mimic the signal processing of the hearing aid. When one reviews all of the findings, it appears that both auditory deprivation and recovery occur in some individuals. A robust effect would most likely be found in a well-defined population (controlling for the amount of hearing loss) who was evaluated through the use of an outcome measure that would highlight a decrement in performance in the frequency range that is no longer audible to the subjects.

II. HEARING-IMPAIRED LISTENERS COMPARED WITH MASKED NORMALS AND ARTICULATION INDEX PREDICTIONS

Performance by longtime hearing-impaired individuals can be compared with performance by normal-hearing listeners to try to determine if the experience of the hearing-impaired listener with a diminished signal somehow gives them an advantage in a reduced auditory environment. This can be done by simulating hearing losses in a group of individuals with normal hearing and comparing their performance to the group with actual hearing loss. Alternatively, Articulation Index (AI) predictions can be used to determine if individuals with hearing loss are performing similarly to normal-hearing individuals tested at an identical AI value through the use of masking. Performances that are better than predicted may indicate that the individual is learning to adapt to the hearing loss.

Investigators use masking to match normally hearing subjects with hearing-impaired subjects in terms of threshold in order to investigate the difficulties experienced by hearing-impaired listeners and the contribution of audibility. A byproduct of this type of investigation is the ability to evaluate the ''abilities'' of hearing-impaired people as compared with masked normals. In at least two investigations that report individual data, several hearing-impaired subjects performed better than their matched normals (Humes *et al.*, 1987; Zurek and Delhourne, 1987). This may indicate that individuals with longstanding hearing loss adapt to making use of limited input. Other studies have found that the subjects with hearing loss and their noise-masked counterparts perform similarly on speech recognition tasks (Humes and

Roberts, 1990; Dubno and Schaefer, 1992) or that subjects with actual hearing loss perform significantly worse (Needleman and Crandell, 1995). Part of the reason for these inconsistencies may relate to the use of different types of stimuli (nonsense syllables, sentences), the use of different presentation levels and signal-to-noise (S/N) ratios, and the use of different test metrics (i.e.,% correct scores, S/N ratio at which 50% performance criterion is met).

Fabry and Van Tasell (1986) compared abilities with masked thresholds and impaired thresholds within the same subjects in order to examine differences between ears that are not explained by audibility. Six subjects with a unilateral hearing loss were given a matched, simulated hearing loss in the ear with normal hearing through the use of masking. Although individual differences varied, three of the subjects performed significantly worse in the impaired versus the noise-masked ear. The authors speculate that the results may be explained by a peripheral deprivation effect, although this may be just as readily explained by amount of hearing loss and subsequent site of lesion (e.g., more inner hair cell loss in some subjects as compared with others).

Studies that have used the AI in an attempt to predict the performance of an individual with hearing loss have produced mixed results. Generally, AI predictions are close to the actual values obtained when subjects with mild to moderate degrees of hearing loss are investigated. As the degree of hearing loss worsens, however, the degree of difference between predicted and observed scores tends to increase (Dubno *et al.*, 1988; Pavlovic, 1984). Ludvigsen (1987) and Pavlovic *et al.* (1986) found that with the incorporation of correction factors meant to account for suprathreshold deficits, AI predictions improved. A deprivation effect overlaying reduced audibility is one of many possible explanations for this suprathreshold deficit.

A. Evaluation of investigations including hearing-impaired listeners and masked normally hearing listeners

Once again, it appears that amount of hearing loss may explain variations in findings. Mild-to-moderately impaired listeners perform similarly to noise-masked normals when the signal is audible. One would expect mild-to-moderately impaired listeners to receive acoustic input across a wide frequency range throughout a normal day, thereby making it unlikely that any particular frequency region would be deprived of all acoustic input. On the other hand, more severely hearing-impaired individuals appear to continue to show a decrement in performance even at high signal presentation levels. Although it may be tempting to explain this through deprivation (they have enough hearing loss in specific frequency regions to make day-to-day acoustic signals inaudible), the physiological cause of the increased hearing loss (e.g., inner hair cell damage) actually may explain performance. Only an investigation designed to look at recovery of functional ability over time would separate these two possible explanations.

III. ADULT SENSORY AND MOTOR SYSTEM PLASTICITY ASSOCIATED WITH PERIPHERAL PATHOLOGY

A great deal of attention has been focused recently on sensory and motor system plasticity in adult mammals (Snow and Wilson, 1991; Kaas, 1991, 1995). Mature systems that were once thought of as hardwired and immutable are now viewed as malleable and amenable to modification. Mammalian central nervous systems are characterized by an orderly representation of neurons which correspond to a particular physical stimulus. For example, neurons in the dorsal portion of the inferior colliculus respond to lower frequencies and neurons in the ventral portion respond to progressively higher frequencies (Willott, 1996a). This orderly tonotopic relationship permits the plotting of neural maps in various regions of the central nervous system. Neural plasticity refers to the capability of these neuronal maps to reorganize following damage at the peripheral level.

Experimentation in adult mammalian neuroplasticity established its roots at the subcortical level through the pioneer work of Liu and Chambers (1958), Wall and Egger (1971), Dostrovsky *et al.* (1976), Devor and Wall (1978), Pollin and Albe-Fessard (1979), and Devor and Wall (1981). A plethora of experiments subsequently evolved investigating the reorganization of cortical somatosensory maps following peripheral nerve damage or deafferentation of animal appendages such as hindpaws or digits (Rasmusseon, 1982; Merzenich *et al.*, 1983a, b; Kelahan and Doetsch, 1984; Wall and Cusick, 1984; Cusick *et al.*, 1990). In addition, several studies emerged that demonstrated tonotopic refinement in the primary visual cortex secondary to unilateral or bilateral retinal lesions (Kaas *et al.*, 1990; Heinen and Shavenski, 1991; Chino *et al.*, 1992; Gilbert and Wiesel, 1992).

In recent years, it has become evident that neural plasticity also occurs in the central auditory system of adult mammals (Salvi *et al.*, 1996). Clearly, the cortical and subcortical refinement of tonotopic maps may have important functional, behavioral, and perceptual ramifications for mammalian adults. Sensorineural pathology may prompt physiologic changes at the central level which alter perceptions of sensory input derived from the peripheral level. The current neuroplasticity knowledge base has been established predominately from animal investigations that employed invasive electrophysiologic recording techniques. Recently, noninvasive procedures that measure auditory evoked potentials (mismatch negativity) and auditory magnetic fields have been used to quantify neural plasticity in human subjects (Kraus *et al.*, 1995; Vasama and Makela, 1995).

Neuroscientists have investigated the adaptive properties of the central auditory system in adult animals by creating peripheral lesions mechanically (Robertson and Irvine, 1989; Rajan *et al.*, 1993), ototoxically (Schwaber *et al.*, 1993; Popelar *et al.*, 1994), or via exposure to intense noise (Kaltenbach *et al.*, 1992). Other researchers have studied changes in the central frequency representations of inbred mice that are genetically programmed to develop progressive presbycusic-like sensorineural hearing loss similar to that observed in humans (Willott, 1984; Willott *et al.*, 1991; Willott *et al.*,1993). Central plasticity also has been quantified as a

secondary consequence of sudden idiopathic sensorineural hearing loss in adult humans (Vasama and Makala, 1995). Recanzone *et al.* (1993) utilized an innovative training approach to elicit and measure neuroplasticity in the auditory cortex of adult owl monkeys. In what follows, this literature will be reviewed as a function of the techniques used to examine neural plasticity. Table II provides a summary of investigations related to central auditory neural plasticity.

A. Mechanically induced lesions

Early evidence supporting reallocation of frequency representation in the primary auditory cortex (A1) surfaced in a study conducted by Robertson and Irvine (1989). These investigators mechanically produced an acute lesion in a restricted region of the organ of Corti in adult guinea pigs. Cortical mapping techniques were used to establish tonotopic frequency organization in the contralateral auditory cortex 35 to 81 days after cochlear damage occurred. Results revealed that the area of the A1 map deprived of its normal frequency input became responsive to acoustic frequencies adjacent to the frequency region damaged by the cochlear lesion. Rajan *et al.* (1993) found a similar effect in adult cats.

B. Ototoxically induced lesions

Schwaber *et al.* (1993) created a high-frequency cochlear hearing loss in adult macque monkeys using ototoxic drugs (kanamycin and furosemide). Tonotopic mapping of the primary auditory cortex was completed 80 to 90 days after the ototoxic drugs were administered. Recordings showed that A1 regions formerly responsive to tones above 10 kHz were currently responsive to tones of 10 kHz or less. Thus, depriving the primary auditory cortex of its normal source of activation for high-frequency sounds resulted in an enlarged A1 representation from the area of the cochlea adjacent to the damaged area (lesion edge or boundary frequencies).

Popelar *et al.* (1994) injected the contralateral cochlea of adult guinea pigs with sisomicin, an ototoxic agent known to cause profound and rapid destruction of the organ of Corti (Cazals *et al.,* 1983). Ipsilateral and contralateral evoked responses were recorded from the inferior colliculus and the auditory cortex in response to high-frequency stimulation of the intact ear. Ipsilateral and contralateral differences (ICDs) were then evaluated. Data revealed a change in thresholds and amplitudes of ipsilaterally evoked responses following hair cell destruction in the contralateral cochlea. Changes were more rapid and profound in the auditory cortex than in the inferior colliculus. Higher (elevated) thresholds and lower amplitudes were observed in the inferior colliculus and ICDs were smaller than those recorded in the auditory cortex. These findings provided further support for the contention that neural plasticity occurs following unilateral cochlear hearing loss at the cortical and to a lesser extent the subcortical levels.

C. Noise-induced lesions

Kaltenbach *et al.* (1992) conducted an experiment to ascertain the effect of intense noise exposure on the topographic organization of the dorsal cochlear nucleus (DCN) in a group of adult hamsters. The hamsters were exposed to a 10-kHz tone at 120–126 dB SPL from 30 min to 6 h causing cochlear lesions that were characterized by loss of stereocilia in a restricted portion of the organ of Corti. Tonotopicity of the dorsal cochlear nucleus (30–76 days postexposure) revealed a shift in the frequency representations in the region of DCN that normally corresponded to the damaged area of the cochlea. These results were consistent with previous investigations that showed immediate broadening of neuronal tuning curves in the inferior colliculus of C57 mice (Willott and Lu, 1982) and the auditory cortex of adult cats (Calford *et al.*, 1993) following noise-induced threshold shifts.

D. Presbycusic-induced lesions

The C57BL/67 (C57) mice exhibit a genetically determined effect on the cochlea (progressive outer hair cell and spiral ganglion cell degeneration) that serves as a convenient model for the study of central auditory system plasticity secondary to sensorineural peripheral pathology (Willott, 1996a, b, c). The C57 mice progressively develop high-frequency sensorineural hearing loss beginning at 2–3 months of age (young adulthood) with elevated thresholds extending into the middle and lower frequencies by 1 year of age (Willott, 1996a). The predictable pathologic nature and genetic similarity of the C57 auditory system obviates the need for inducing hearing loss mechanically, ototoxically, or via noise exposure (Willott, 1996a). The C57 mice, therefore, are ideal for the study of central plasticity associated with peripheral presbycusis.

Willott (1984) initially demonstrated and later reaffirmed (Willott, 1986; Willott *et al.*, 1988) that as C57 mice age and lose their ability to hear high-frequency stimuli, neurons in the high-frequency tonotopic region of the inferior colliculus became responsive to the middle and lower frequencies that remain audible. Willott *et al.* (1993) showed a similar modification in the primary auditory cortex of C57 mice. Again, as high-frequency sensitivity progressively diminished at the peripheral level, high-frequency representation at the cortical level shifted downward to the middle frequencies. Virtually the entire primary auditory cortex became devoted to frequencies in the 10–13-kHz range for which peripheral hearing sensitivity remained relatively intact. High frequencies that were lost to sensorineural peripheral pathology were no longer represented physiologically at the A1 level.

E. Plasticity induced by sudden sensorineural hearing loss

Vasama and Makela (1995) recorded auditory evoked magnetic fields from eight adult human subjects (35–48 years of age) 2 to 5 years after the onset of unilateral sudden idiopathic sensorineural hearing loss. Data analysis revealed response modifications suggestive of auditory pathway reorganization in five out of eight subjects. Although the time course and rate of reorganization cannot be defined from these data, Vasama and Makela were able to provide evi-

TABLE II. Summary of investigations related to central auditory system plasticity.

Author/date	Adult species	Techniques used to examine neural plasticity	Central auditory system (CAS) structure examined	Findings
Robertson and Irvine (1989)	guinea pigs	mechanical	Primary auditory cortex (A1)	Enlarged A1 cochleotopic representation of adjacent frequencies into the frequency region damaged by the cochlear lesion.
Rajan *et al.* (1993)	cats	mechanical	Primary auditory cortex (A1)	Enlarged A1 cochleotopic representation of adjacent frequencies into the frequency region damaged by the cochlear lesion.
Schwaber *et al.* (1993)	macque monkeys	ototoxic	Primary auditory cortex (A1)	Enlarged A1 cochleotopic representation of adjacent frequencies into the frequency region damaged by the cochlear lesion.
Popelar *et al.* (1994)	guinea pigs	ototoxic	Inferior colliculus (IC) Primary auditory cortex (A1)	Ipsilateral and contralateral evoked potential differences (ICDs) were more rapid and profound in the primary auditory cortex as compared with the inferior colliculus.
Kaltenbach *et al.* (1992)	hamsters	noise	Dorsal cochlear nucleus (DCN)	Refinement of frequency representations in the region of the DCN that normally corresponded to the damaged area of the cochlea.
Willott (1984)	C57 mice	presbyacusis	Inferior colliculus (IC)	As presbycusic-induced high-frequency sensorineural hearing loss develops in C57 mice, neurons in the high-frequency tonotopic region of the IC became responsive to the middle and lower frequencies that remained audible.
Willott (1986)	C57 mice	presbyacusis	Inferior colliculus (IC)	As presbycusic-induced high-frequency sensorineural hearing loss develops in C57 mice, neurons in the high-frequency tonotopic region of the IC became responsive to the middle and lower frequencies that remained audible.
Willott *et al.* (1988)	C57 mice	presbyacusis	Inferior colliculus (IC)	As presbycusic-induced high-frequency sensorineural hearing loss develops in C57 mice, neurons in the high-frequency tonotopic region of the IC became responsive to the middle and lower frequencies that remained audible.
Willott *et al.* (1993)	C57 mice	presbyacusis	Primary auditory cortex (A1)	As presbycusic-induced high-frequency sensorineural hearing loss develops in C57 mice, neurons in the high frequency region of A1 became responsive to the middle and lower frequencies that remained audible.
Willott *et al.* (1991)	C57 mice	presbyacusis	Ventral cochlear nucleus (VCN) Dorsal cochlear nucleus (DCN) Inferior colliculus (IC)	Minimal shifting of the frequency representation from the higher to middle frequencies occurred in the ventral cochlear nucleus after cochlear pathology developed in middle-aged mice. The extent of neural plasticity differs in the VCN, IC, and A1 of aging C57 mice.
Vasama and Makela (1995)	humans	idiopathic	Auditory cortex	Cortical response modifications suggestive of reorganization in the central auditory system subsequent to sudden unilateral sensorineural hearing loss.
Recanzone *et al.* (1993)	owl monkeys	training	Primary auditory cortex (A1)	Trained monkeys showed significant improvement in perceptual frequency discrimination.
Kraus *et al.* (1995)	humans	training	Auditory thalmocortical pathways	Behavioral training in the discrimination of synthetic phonemes resulted in a significant change in the duration and magnitude of the mismatch negativity (MMN) response implying that auditory training may influence neurophysiological changes in the central auditory system.

dence of auditory neuroplasticity in adult humans using a noninvasive technique (magnetoencephalography).

F. Training-induced plasticity

Another set of investigators employed training in order to evaluate cortical and/or subcortical modification in the neural pathways of the central auditory system. In the somatosensory domain, Recanzone *et al.* (1992a, b) analyzed the cortical maps of adult owl monkeys behaviorally trained to complete a tactile frequency-discrimination task. Cortical representations of trained hands were compared with the opposite, untrained hand and also to the hands of control monkeys that performed an auditory frequency-discrimination task while vibratory stimuli were being delivered to the skin (passive stimulation). Results indicated that cortical representations of the trained hands were more complex in topographic detail than the representations of unstimulated or passively stimulated control hands. In other words, training-induced cortical changes occurred that paralleled improvements in behavioral performance.

Recanzone *et al.* (1993) postulated that the topographic representation of the cochlea in the primary auditory cortex may be altered as a consequence of training. These investigators trained three adult owl monkeys in an auditory frequency discrimination task for several weeks and the tonotopic representation of the primary auditory cortex was electrophysiologically mapped after significant improvements in performance were measured. Neuronal response properties were compared with two groups of control monkeys, those receiving no training and those that received auditory stimulation while concurrently engaged in an unrelated tactile discrimination task. The trained monkeys showed a significant improvement in perceptual frequency discrimination that paralleled an enlargement of cortical representations for the trained frequencies.

Kraus *et al.* (1995) measured central plasticity associated with speech discrimination training in normal hearing adult humans. Mismatch negativity (a passively elicited event-related potential that reflects discrimination of small acoustic contrasts) was recorded in response to two slightly different speech stimuli (synthetic phonemes). Behavioral training in the discrimination of these synthetic phonemes resulted in a significant change in the duration and magnitude of the mismatch negativity response. This finding in conjunction with the adult owl monkey data of Recanzone *et al.* (1993) lends credence to the notion that intense listening training can alter the neurophysiologic responses of the central auditory system. Both of these investigations used normal hearing subjects. The impact of listening training on hearing-impaired subjects has not yet been determined.

G. Potential perceptual consequences of central pathway plasticity

Harrison *et al.* (1991) proposed a key question beyond the basic science of neuroplasticity, namely, what are the perceptual consequences associated with cortical and/or subcortical reorganization following long-term sensorineural hearing loss? The answer to this question has potential implications for auditory rehabilitation in humans. Several investigators have collected data that highlight the potential perceptual consequences of reorganization.

Willott *et al.* (1994) investigated behavioral changes in C57 (presbycusis), DBA (accelerated presbycusis), and CBA (control) mice using a prepulse inhibition startle modification paradigm. The paradigm was designed so that prepulse tone pips (S1) were presented prior to the presentation of a broadband noise startle stimulus (S2) in order to determine if the tone pip inhibited or reduced the amplitude of the acoustic startle response. High-frequency tone pips (4 to 24 kHz) were used to parallel the characteristic high-frequency loss that inevitably develops in C57 and at an accelerated rate in DBA mice. The CBA mice exhibit minimal age-related hearing loss and therefore served as controls. Results indicated that as high-frequency sensitivity progressively diminished in C57 mice from 1 to 12 months, the efficacy of low-frequency tone pips became significantly enhanced. The DBA mice exhibited similar changes to those observed in C57 mice but the changes occurred at an accelerated rate. The control mice (CBA) exhibited no age-related changes in startle modification. This enhanced behavioral saliency of low and middle frequencies in C57 and DBA mice is consistent with previous physiological studies that demonstrated a downward frequency shift of representation in the auditory cortex and inferior colliculus of C57 mice (Willott, 1984; Willott, 1986; Willott *et al.*, 1988, 1993). Willott (1996b) discussed the potential negative consequences of frequency map reorganization on the neural coding of sounds. If neurons in high-frequency tonotopic regions become responsive to adjacent middle frequencies, altered perceptions may occur. As Willott (1996c) indicated, this may contribute to perceptual difficulties beyond the loss of auditory sensitivity.

Rajan *et al.* (1993) demonstrated that rearrangement of central frequency representations may have a direct impact on binaural auditory processing. These investigators mechanically produced unilateral cochlear lesions in adult cats. After a brief recovery period, tonotopic maps were recorded in the primary auditory cortex both ipsilateral and contralateral to the lesioned cochlea. Data analysis revealed an enlarged representation of the lesion edge frequencies in the primary auditory cortex contralateral to the damaged cochlea. However, tonotopic reorganization did not occur in the primary auditory cortex ipsilateral to the lesioned cochlea. Therefore, a mismatch was created between the contralateral and ipsilateral cortical maps. The perceptual consequences of such a mismatch may be significant for humans with unilateral sensorineural hearing loss and/or for humans exposed to long-term monaural amplification, thereby producing a nonstimulated ear. The results of this study may also reflect the deprivation and acclimatization effects that are often associated with long-term monaural hearing aid use. On the other hand, based on the body of evidence for acclimatization, one might expect the brain to compensate and/or even use the mismatched information to advantage after a period of adjustment.

None of the animal data or even the recent human data relate directly to trying to restimulate an area of the cortical map that has not received stimulation over some period of

time. In the lesion work, it is not possible to restore the peripheral loss and therefore recovery cannot be examined. In the novel task work greater representation is seen, but this is a representation of frequencies that have continually been represented in the cortical map. Restimulation implies that the cortical map will reallocate space currently occupied by boundary frequencies to newly audible sound if reallocation is indeed a mechanism for auditory learning in humans.

Indirectly, evidence related to restimulation has emerged from the work of Snyder *et al.* (1990) who documented changes in central frequency representations related to chronic intracochlear stimulation. These investigators demonstrated an expansion of the frequency map in the inferior colliculus of neonatally deafened adult cats that corresponded to electrical stimulation of a restricted portion of the basilar membrane. This study demonstrated that stimulation of central auditory system neurons can induce plasticity in a system that has been acoustically impoverished from birth. Further experimentation would be needed to ascertain if a similar effect occurs in adventitiously impaired adults.

H. Evaluation of adult sensory and motor system plasticity investigations

The entire body of literature in this area supports the phenomenon of plasticity in the adult sensory system. The findings also appear to provide a mechanism for adult human auditory learning. At this juncture, it is not clear whether hearing aid amplification enhances or hinders auditory perceptions associated with tonotopic reorganization for a given individual. If the acoustic input from a hearing aid reactivates neurons that have become responsive to boundary or lesion-edge frequencies, neural coding and perceptual confusion may evolve. If perceptual problems develop as a result of ineffectual neural coding, hearing aid use or the specific frequency response of a hearing aid may be contraindicated. Conversely, if cortical and subcortical topographic maps currently occupied by boundary frequencies reallocate space for those frequencies that have their audibility restored, auditory learning may enhance communicative function. In addition, Willott (1996c) hypothesized that it may be better for amplification to focus on making more neurons responsive to the frequency region where sound has been audible and therefore one would expect neural representation rather than attempting to reallocate space for the frequency region that has not had input over time due to peripheral impairment. These are issues that clearly need to be addressed in order to delineate further the impact of central auditory system plasticity on hearing aid benefit and hearing aid evaluation, selection, and fitting protocols.

IV. AUDITORY LEARNING POSTHEARING AID FITTING

Watson and Knudsen (1940) were the first to publish data related to word recognition improving as a function of posthearing aid fitting time. In a study designed to evaluate selective amplification, the investigators noted that one individual (out of 17) showed improvement over time. Since that time, a variety of investigators have reported improvement over time in a variety of speech related tasks in groups of subjects with newly fitted hearing aids (Arkis and Burkey, 1994; Cox and Alexander, 1992; Cox *et al.*, 1996; Gatehouse, 1992, 1993; Horwitz, 1995; Robinson and Gatehouse, 1996; Walden *et al.*, 1981; Yund and Buckles, 1995). Three investigations designed for other purposes have reported no change over time for newly fitted hearing aid wearers (Bentler *et al.*, 1993a, b; Humes *et al.*, 1995; Taylor, 1992, 1993). These studies were investigating different hearing aid processing schemes and the ability to predict outcome measures from prefit measures, respectively. In addition, Saunders and Cienkowski (1997), in a prospective study designed to look at acclimatization with hearing aids, did not find a substantial acclimatization effect over 3 mon of hearing aid use.

A revealing way to examine these data is to evaluate the investigations as a function of experimental control variables that can be identified *a priori* as important. Each of these variables could be advanced as a reason for conflicting reports or large intersubject variability, making it hard to determine if acclimatization truly occurs and just how robust (and therefore clinically meaningful) it is in any given population. The experimental control variables are addressed below. Table III provides a summary of investigations related to acclimatization.

A. Audibility in day-to-day listening and in the testing situation

When assessing return of function posthearing aid fitting, audibility is a critical factor. One would not expect a change in performance if new sounds have not become audible during practice and/or testing. The investigations to date have all fit the new or refitted hearing aid users with some type of prescriptive formula [most have used the revised National Acoustics Laboratory (NAL-R) fitting strategy Byrne and Dillon, 1986]. This type of hearing aid fitting does not necessarily produce audibility for the quieter, high-frequency speech cues. The investigations to date have not attempted to account for audibility during practice (day-to-day listening). With the use of NAL-R target fitting, it is difficult to determine what is and is not audible for a person in day-to-day listening. Two studies have controlled gain during test presentation (Cox and Alexander, 1992; Horwitz, 1995). Cox *et al.* (1996) monitored presentation level with a probe microphone in the ear canal during testing. This in and of itself does not guarantee audibility unless a target is defined that is known to be audible to the listener.

Gatehouse (1992) reported improvement in a sentence task over 18 weeks postfitting and measured the insertion gain of the hearing aid for each test session. These data reveal that the insertion gain was actually increasing over time which would imply that the signal was becoming more audible over time. In this case, it would be impossible to separate auditory learning from increased audibility. Although it may be tempting to suggest that auditory learning can be accounted for by audibility alone, Horwitz (1995) controlled for this element by providing the test signal at two volume control levels for her experimental group and found the same improvement in performance with time at both levels. Gate-

TABLE III. Summary of investigations related to acclimatization.

Study Retro (R) Pro (P)	*N*	Subjects Age (yrs)	Hearing level Type of loss	Type of stimulation	Hearing aid description Hearing aid use hrs/day	Followed postfitting	Type of assessment[a]	Findings/comments
Watson and Knudsen (1940) P	1	?	SRT=63 dB HL sensorineural	uniform amplified signal supplied and selective amplification	18 dB gain headphones	3 months	articulation lists CVCs	Word recognition continued to improve up to four months for both types of amplification.
Walden *et al.* (1981) P	35	19–68	mild-to-moderate sensorineural	investigated learning/ training in new hearing aid wearers		2 weeks and 7 hours training	W-22	All subjects received traditional 2 week AR program. In addition, ten received additional 7 h of auditory training and ten received additional 7 h of visual training. Most improvement noted with auditory group, less with visual and traditional AR groups.
Malinoff and Weinstein (1989) P	25	55–90	moderate sensorineural	new hearing aid users		3 weeks, 3 months, 1 year	HHIE	HHIE scores improved at 3 weeks, poorer at 3 months (not as poor as original), stabilized by 1 year.
Gatehouse (1992) P	4	?	mild-moderately severe symmetrical sensorineural	provided a monaural hearing aid for each subject, some testing done with frequency shaped headphones	insertion gain for maximum intelligibility is reported by frequency	12 weeks of hearing aid use	FAAF psychoacoustic	Over 6–12 weeks of hearing aid use, the fitted ear scores increased, the nonamplified ear did not do any better with the amplified ear's hearing aid response on it.
Mulrow (1990) P	192	mean=72	mild-moderately severe sensorineural (3% mixed)	quality of life assessment for new hearing aid wearers	Most >4 h per day	baseline, 4, 8, 12 months	scales	At 4 months postfitting scale scores improved. Score were sustained at 8 and 12 months. The cognitive measures reverted to baseline at 12 months. (HHIE, GDS, SPMSQ, QDS).
Cox and Alexander (1992) P	17	52–81	mild-moderate-severe, sloping sensorineural	new and experienced hearing aid users, all received new hearing aids	prescriptive fitting, real ear measures 8 h (4 for new)	soon after fitting, 10 weeks	CST PHAB	Similar time-related changes over 10 weeks of hearing aid use. No improvement in noise or reverb with visual cues, improvement in low noise and noise w/o visual cues, improvement in PHAB score.
Gatehouse (1993) P	36	46–81 mean=64	31, 33, and 43 at 0.5, 1, 2 kHz symmetrical sensorineural	monaural users new users	insertion gain, then prescribed NAL based insertion gain for new hearing aid, more high frequency gain	12–15 months using first aid, then 8–16 weeks with replacement aid	FAAF Sentence test	The benefit from the new prescription was not measured until 8–16 weeks of practice.
Bentler *et al.* (1993)[a,b] P	65	21–84	moderately severe flat and sloping sensorineural	investigated learning in new and experienced users	prescriptive insertion gain (NAL) 20%<4 h 80% 4–16 h	investigated learning, followed for 12 months	SPIN, NST quiet and noise HPI, quality judgments	Individuals chose to be monaural or binaural users. No improvement on any objective measures. The quiet section of the HPI (subjective) showed significant change.
Taylor (1992, 1993) P	58	65–81	>30 dB HFPTA sensorineural	new hearing aid users, 37 monaural, 21 binaural	hrg aid performed w/in manuf specs. at each test	prefitting, 3 weeks, 3 months, 6 months, 1 year	HHIE NU-6 quiet NU-6 noise	Handicap reduced at 3 weeks (compared with pre-score). Increased at 3 months and then stabilized. No change in speech scores (initial test was 3 weeks postfitting).
Arkis and Burkey (1994) R	105	adults	moderate sensorineural	looked at pre-fitting word recognition scores versus a few months post-fitting			WRS	5% increase for the aided ears after several months of hearing aid use.
Horwitz (1995) P	26	adults	mild-to-moderate sloping sensorineural	13 new hearing aid users, 13 longstanding monaural hearing aid users		18 weeks for new users	NST APHAB	Benefit (objective) increased for new users, not for longstanding users (control group). Lack of control over hearing aid fitting, and volume control in day-to-day use. Tested in two volume control conditions, no impact on new users. Longstanding users did better w/adjusted volume control, may mean it was more audible. No subjective benefit increase.
Humes *et al.* (1995) P	20	63–78	mild-to-moderate sloping sensorineural	10 new hearing aid users (not used within 2 years) 10 experienced hearing aid users monaural and binaural	NAL-type formula datalogger of objective use and subjective use	0, 7, 15, 30, 60, 90, 180 days	NST quiet noise 100-item HINT HAPI, HHIE	Measured over a 24-week period. No increase for anything in any group. Testing monaurally although individuals were fit binaurally. HINT data are invalid for test sessions beyond 60 days.

TABLE III. (*Continued.*)

Study Retro (R) Pro (P)	*N*	Subjects Age (yrs)	Hearing level Type of loss	Type of stimulation	Hearing aid description Hearing aid use hrs/day	Followed postfitting	Type of Assessment[a]	Findings/comments
Yund and Buckles (1995) P	15	48–79	moderate sensorineural	laboratory study, subjects exposed to multichannel compression in lab no trial-by-trial feedback	multichannel compression only in lab	within lab use	NST SNR=−5 to 15	14 of 15 individual subjects showed significant improvement across experiments. Consonant-confusion matrices showed improvement in discrimination occurred for both manner and place information, consistent with improved use of high frequency information.
Robinson and Gatehouse (1996) P	5	38–83	moderate sensorineural	new monaural users	insertion gain NAL, linear, peak clipping ?? use time	prefitting, 4, 12, 18 weeks	discrim of intensity for 250 and 3000 Hz	After 15–18 weeks, aided ear had better intensity discrimination than the unaided ear at 3000 Hz for 95 dB signal, no difference at 250 Hz (where hearing had been normal).
Cox *et al.* (1996) P	22	elderly	? sensorineural	1st time users, control group of experienced users (small group)	level held constant in listener's ear canal during test		CST SPAC analyze by speech features	After 12 weeks, 4% improvement in CST (significant) with no increase in the control group. Acclimatization should be mainly in high-frequency speech, but the SPAC data did not support this.
Saunders and Cienkowski (1997) P	24 24	mean=69.1 mean=68.9	mild to moderate sensorineural	new users experienced users (mean=12.7 years)	Linear, PC, WDRC, Input compression. Fitting schemes varied	0, 30, 60, and 90 days postfitting	Spondees, HINT (objective and subjective)	No significant changes over the test sessions. Low-to-mid frequency emphasis test material was used.
Neuman (1997) P	?	adults	mild to moderate sensorineural	new users	NAL—R	2 months (as of this particular report)	Subjective ratings Spondee threshold CVC words in quiet and noise NST in noise	Preliminary data indicate no change in performance after 2 months of hearing aid use. It is unlikely that hearing aids fit to the NAL-R are providing an audible signal in the high frequencies where one would expect a change in performance if there was to be one.

[a] Retro (R) Retrospective
NAL National Acoustics Laboratory
WDRC Wide dynamic range compression
PC Peak clipping
CVC Consonant–vowel–consonant
W-22 Central Institute for the Deaf, Word list 22
HHIE Hearing Handicap Inventory for the Elderly
FAAF Four Alternative Auditory Feature test
CST Connected Speech Test
PHAB Profile of Hearing Aid Benefit
SPIN Speech in Noise Test
NST Nonsense syllable test
NAL-R National Acoustics Laboratory—Revised
Pro (P) Prospective
NU-6 Northwestern University word list #6
WRS Word recognition score
HPI Hearing Performance Inventory
APHAB Abbreviated Profile of Hearing Aid Benefit
HINT Hearing in Noise Test
HAPI Hearing Aid Performance Inventory
HFPTA High-frequency pure tone average
SPAC Speech Pattern Contrast Test
GDS Geriatric Depression Scale
SPMSQ Short portable mental status questionnaire
QDS Quantified Denver Scale of Communication Function
AR Auditory Rehabilitation

house (1993) purposely refit hearing aid users to provide more high-frequency amplification and reported that the benefit from the newly audible signal was not evident until 8–16 weeks postfitting.

B. Amount of day-to-day exposure at various levels of audibility

Investigations related to auditory deprivation have documented self-reported hours of use per day. These reports vary from less than 4 h per day to 16 h per day. If experience is needed to use new signals, this variability may explain some of the individual variability found in the investigations. Brooks (1979, 1981) objectively measured use-time in body worn and behind-the-ear hearing aids and found that subjects often overestimated use-time. Humes *et al.* (1995) reported data supporting the notion that self-reported hours of use and objective measures are not necessarily in agreement.

C. Impact of various signal processing schemes

In an experiment designed to assess various compression processing schemes, Yund and Buckles (1995) reported that 14 of their 15 subjects showed a significant improvement in performance with high-frequency stimuli. These subjects were only exposed to the multichannel compression in a laboratory setting (2 h of exposure for each visit) for repeated practice and test sessions over a 1-year period. These data suggest that a time course for learning may be associated with particular signal processing separate from or in conjunction with returning audibility. Palmer (1995) reported that longstanding linear/peak clipping hearing aid wearers judged sound quality differently from matched nonhearing aid wearers. These subjects were more tolerant of distortion than matched hearing-impaired nonhearing aid users. Again, this indicates that learning related to distinct signal processing may take place over time. Gatehouse (1992) is the only investigator to employ the same hearing aid circuitry for all subjects ($N=4$). One possibility for the intersubject variability observed in various investigations may be accounted for by subjects' previous experience with signal processing, and/or the fact that the individuals in the study are all being exposed to different signal processing and some signal processing may lend itself to faster adaptation than other types of signal processing.

D. Relationship of test material to real-world practice

The small magnitude of improvement that has been reported may be explained by the lack of relationship between day-to-day listening and the outcome measures. The studies to date have all relied on day-to-day listening as the practice source, which means that the practice material is mostly connected discourse. Mid-frequency response is important in understanding connected discourse and the task may be an open or closed set depending on context available. The majority of outcome measures have been nonspeech, high-frequency weighted tests [e.g., Nonsense Syllable Test (Kalikow *et al.*, 1979)] using a closed set.

Gatehouse (1993) used sentence material and found improvement over time in the aided ears. Humes *et al.* (1995) employed a test most representative of day-to-day communication [Hearing In Noise Test (Soli and Nilsson, 1994)] but reported that the data could not be interpreted past the 60-day test because of procedural problems. It is not unreasonable to suggest that increase in performance could come after 60 days. Saunders and Cienkowski also used the HINT and obtained data at 0, 30, 60, and 90 days postfitting and reported slight but nonsignificant changes in performance. Data past 3 months could not be collected because subjects changed hearing aid configuration every 3 months in this investigation. Cox and Alexander (1992) used quiet, noise, and reverberation conditions with visual cues in order to more closely represent the real-world practice condition. Improvement for their subjects was found only in the low noise and noise with visual cues conditions. The authors suggest that these conditions were most similar to at home listening.

E. Age range of subjects

All of the investigations have either used an extensive age range (21–84 years) (Arkis and Burkey, 1994; Bentler *et al.*, 1993a, b; Cox and Alexander, 1992; Gatehouse, 1992, 1993; Horwitz, 1995; Robinson and Gatehouse, 1996; Saunders and Cienkowski, 1997; Taylor, 1992, 1993; Yund and Buckles, 1995) or have used an exclusively elderly population (Cox *et al.*, 1996; Humes *et al.*, 1995). Considering the animal data suggesting possible mechanisms for auditory learning (at the level of the inferior colliculus and primary auditory cortex) it is unclear how an aging auditory system might influence plasticity. Neural activity and synaptic strengths of an elderly listener may decrease the likelihood of measuring auditory learning. Although the area is controversial, several investigators have reported decreased central auditory processing abilities in aging adults while controlling for audibility (Humes and Christopherson, 1991; Humes *et al.*, 1994). These investigators clearly demonstrated that the majority of difficulty experienced by the aging listeners can be accounted for by lack of audibility, but that there is a small component that cannot be explained by lack of audibility. A younger group of adults may increase the likelihood of measuring a robust auditory learning effect.

F. Degree of hearing loss

Three of the investigations that reported a lack of improvement over time in newly aided individuals (Bentler *et al.*, 1993a, b; Humes *et al.*, 1995; Taylor, 1992, 1993) used subjects with mild high-frequency hearing loss. It is reasonable to suggest that high-frequency signals were always audible to some extent for these individuals and that one would not expect ''recovery.'' All listeners are exposed to their own voices, which contain high-frequency speech cues. A person's own voice regularly reaches the ear at a level of 75–80 dB SPL as points in the discussion are emphasized. This corresponds to an audiometric hearing level of 65 dB in which the quieter speech cues are at approximately 40–45 dB HL. Only when hearing loss exceeds 45 dB HL, therefore, would any loss of processing be expected.

The investigations reporting various amounts of improvement over time have used subjects with a fairly wide range of hearing abilities.

G. Accounting for task/procedural learning via a control group

Task and procedural learning can only be separated from stimulus learning by comparing a control group to the experimental group(s). Three of the investigations have employed control groups (Cox *et al.*, 1996; Horwitz, 1995; Saunders and Cienkowski, 1997). In the Cox *et al.* (1996) and Horwitz (1995) studies, neither control group revealed significant changes in performance over time whereas the experimental groups showed at least a small level of improvement. Saunders and Cienkowski (1997) did not demonstrate change in the experimental or control groups.

H. Relationship of day-to-day or practice listening configuration with testing configuration

If auditory learning can be characterized as learning to perceive a particular stimulus, then one would expect learning to be ear specific and specific to frequencies and input levels newly provided to the listener. Several investigations of auditory learning (acclimatization) appear to support this hypothesis (Cox and Alexander, 1992; Gatehouse, 1992; Robinson and Gatehouse, 1995). Gatehouse (1992) found frequency-specific improvement in performance of monaurally aided hearing aid users while Robinson and Gatehouse (1995) measured level-dependent changes in longstanding monaurally aided and unaided ears consistent with the levels to which the ears would have been exposed. Cox and Alexander (1992) found that a group of subjects showed no improvement in the noise or reverberation component of the task, but demonstrated improvement in low noise and noise with visual cues. The authors proposed that these two conditions were most similar to the everyday listening experiences of this group. Three of the studies that found no improvement in performance over time (Bentler *et al.*, 1993a, b; Humes *et al.*, 1995; Taylor, 1992, 1993) employed testing configurations that differed from the day-to-day practice configurations. Humes *et al.* (1995) tested in the monaural condition while the subjects spent their days listening binaurally and Bentler *et al.* (1993a, b) and Taylor (1992, 1993) tested all of the subjects binaurally although there were monaural and binaural users included in the experiment.

I. Evaluation of investigations related to auditory learning post hearing aid fitting

The greatest deterrent to interpreting the work reported in auditory learning is the lack of control over original hearing loss (was it enough to make certain sounds inaudible on a day-to-day basis?), the audibility supplied by the new hearing aid (did it return audibility consistently?), the amount of hearing aid use at various volume control settings postfitting (how often were audible signals stimulating the ear or how much time during the day could ''learning'' be taking place?), the independence of the outcome measure and expected day-to-day practice, the lack of control groups to sort out procedural and task learning, and the potential impact of differing signal processing schemes on the extent and rate of learning. It is possible that these factors could account for a great deal of the variability associated with the findings reported thus far. Since there are experimental methods and testing tools that will allow us to control these factors, a more complete description of auditory learning in a specifically identified group is possible. Turner *et al.* (1996) recommend that further investigation of auditory learning be limited to groups where one would reasonably expect acclimatization if it were to exist. It would appear that one should expect auditory learning to occur in adult humans who have enough hearing loss to make day-to-day sounds in a particular frequency range inaudible, but who do not have so much hearing loss that it is not possible to supply audible stimulation through a hearing aid. One would expect that the amount of time the hearing aid is worn day-to-day to impact either the final amount or rate of auditory learning. Currently the data from the acclimatization literature indicate that learning continues up until and beyond the 18th week after fitting. The deprivation literature documents learning (relearning) for up to 2 years (Hurley, 1993). Of the investigations that have reported auditory learning over time, none have been designed to allow the investigators to continue measurements until an asymptote was reached. This limitation has made it difficult to try to examine the rate of learning as well as the extent of auditory learning.

V. CONCLUSION

The potential for auditory learning has been examined through review of data related to auditory deprivation and recovery, hearing-impaired listeners compared with masked normals, adult sensory plasticity associated with peripheral pathology, and auditory learning post-hearing-aid fitting. Although data in each area support the notion of auditory learning, aspects of experimental design and control leave many unanswered questions. The extent, rate, population, and possible physiologic basis is unclear at this time. This offers a broad research area that demands the expertise of investigators from both basic and applied areas related to auditory physiology, psychoacoustics, engineering, and clinical evaluation, selection, fitting, and validation of amplification.

ACKNOWLEDGMENTS

The authors would like to thank two anonymous reviewers who provided detailed reviews of this work and assisted in shaping it into the current tutorial. In addition, George Lindley acknowledges the generous support of the Veterans Administration through a doctoral fellowship.

Anderson, K. L. (**1985**). ''A case study of central processing following long-standing unilateral conductive hearing loss,'' J. Aud. Res. **25**, 201–213.

Arkis, P., and Burkey, J. (**1994**). ''What WRS's say about client performance, adjustment to hearing aids. Word recognition scores: Do they support adaptation?,'' Hearing Instrum. **45**(1), 24–25.

Arlinger, S., Gatehouse, S., Bentler, R. A., Byrne, D., Cox, R. M., Dirks, D. D., Humes, L., Neuman, A., Ponton, C., Robinson, K., Silman, S., Sum-

merfield, 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**(3), 87S–90S.
Bentler, R., Niebuhr, D. Getta, J., and Anderson, C. (**1993a**). ''Longitudinal study of hearing aid effectiveness. I: Objective Measures,'' J. Speech Hear. Res. **36**, 808–819.
Bentler, R., Niebuhr, D., Getta, J., and Anderson, C. (**1993b**). ''Longitudinal study of hearing aid effectiveness. II: Subjective Measures,'' J. Speech Hear. Res. **36**, 820–831.
Boothroyd, A. (**1993**). ''Recovery of Speech Perception Performance after prolonged auditory deprivation: Case Study,'' J. Am. Acad. Audiol. **4**, 331–337.
Brooks, D. (**1979**). '' Counseling and its effects on hearing aid use,'' Scand. Audiol. **8**, 101–107.
Brooks, D. (**1981**). ''Use of post-aural aids by national health service patients,'' Br. J. Audiol. **15**, 79–86.
Burkey, J. M., and Arkis, P. N. (**1993**). ''Word recognition changes after monaural, binaural amplification,'' Hearing Instrum. **44**, 8–9.
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 Hearing **7**(4), 257–265.
Calford, M. B., Rajan, R., and Irvine, D. R. F. (**1993**). ''Rapid changes in the frequency tuning of neurons in cat auditory cortex resulting from pure-tone-induced temporary threshold shift,'' Neuroscience (Oxford) **55**, 953–964.
Cazals, Y., Aran, J.-M., Erre, J.-P., Guillhaume, A., and Arousseaou, C. (**1983**). ''Vestibular acoustic reception in the guinea pig: A saccular function,'' Acta Oto-Laryngol. **95**, 211–217.
Chino, Y. M., Kaas, J. H., Smith III, E. L., Langston, A. L., and Cheng, H. (**1992**). ''Rapid reorganization of cortical maps in adult cats following restricted deafferentation in retina,'' Vision Res. **32**, 789–796.
Cox, R. M., and Alexander, G. C. (**1992**). ''Maturation of hearing aid benefit: Objective and subjective measurements,'' Ear Hear. **13**(3), 131–141.
Cox, R., Alexander, G., Taylor, I., and Gray, G. (**1996**). ''Benefit acclimatization in elderly hearing aid users,'' J. Am. Acad. Audiol. **7**, 28–441.
Cusick, C. G., Wall, J. T., Whiting, J. H., and Wiley, R. G. (**1990**). ''Temporal progression of cortical reorganization following nerve injury,'' Brain Res. **537**, 355–358.
Dalzell, L. E., Merle, K. S., and Dalzell, S. M. (**1992**). ''Speech recognition decline for monaural hearing aid users: Aided versus unaided ears,'' Paper presented at the Meeting of the American Speech-Language-Hearing Association, San Antonio, TX.
Devor, M., and Wall, P. D. (**1978**). ''Reorganization of spinal cord sensory map after peripheral nerve injury,'' Nature **275**, 75–76.
Devor, M., and Wall, P. D. (**1981**). ''Effect of peripheral nerve injury on receptive fields of cells in the cat spinal cord,'' J. Comp. Neurol. **199**, 277–291.
Dieroff, H. G. (**1993**). ''Late-onset auditory inactivity (deprivation) in persons with bilateral essentially symmetric and conductive hearing impairment,'' J. Am. Acad. Audiol. **4**, 347–350.
Donaldson, H. H. (**1890**). ''Anatomical observations on the brain and several sense organs of the blind deaf-mute, Laura Dewey Bridgman,'' Am. J. Psychol. **3**, 293–342.
Dostrovsky, J. O., Millar, J., and Wall, P. D. (**1976**). ''The immediate shift of the afferent drive of dorsal column nucleus cells following deafferentation: A comparison of acute and chronic deafferentation in gracile nucleus and spinal cord,'' Exp. Neurol. **52**, 480–495.
Dybala, P. (**1997**). ''Effects of peripheral hearing loss on tonotopic organization of the auditory cortex,'' Hearing J. **50**, 49–52.
Dubno, J. R., and Schaefer, A. B. (**1992**). ''Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners,'' J. Acoust. Soc. Am. **91**, 2110–2121.
Dubno, J. R., Dirks, D. D., and Schaefer, A. B. (**1988**). ''Stop-consonant recognition for normal hearing listeners and listeners with high-frequency hearing loss. II: Articulation Index predictions,'' J. Acoust. Soc. Am. **85**, 355–364.
Fabry, D. A., and van Tasell, D. J. (**1986**). ''Masked and filtered simulation of hearing loss: Effects on consonant recognition,'' J. Speech Hear. Res. **29**, 170–178.
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 on the selection of frequency responses for hearing aids,'' J. Am. Acad. Audiol. **4**(5), 296–306.
Gelfand, S. A. (**1994**). ''Auditory deprivation recovery and no recovery with binaural hearing aids: Case reports,'' paper presented at the Convention of the American Academy of Audiology, Richmond, VA.
Gelfand, S. A. (**1995**). ''Long-term recovery and no recovery from the auditory deprivation effect with binaural amplification: Six cases,'' J. Acad. Audiol. **6**, 141–149.
Gelfand, S. A., and Silman, S. (**1993**). ''Apparent auditory deprivation in children: Implications of monaural versus binaural amplification,'' J. Am. Acad. Audiol. **4**, 313–318.
Gelfand, S. A., Silman, S., and Ross, L. (**1987**). ''Long-term effects of monaural, binaural, and no amplification in subjects with bilateral hearing loss,'' Scand. Audiol. **16**, 201–207.
Gilbert, C. D., and Wiesel, T. N. (1992). ''Receptive field dynamics in adult primary visual cortex,'' Nature **356**, 150–152.
Hall, J., and Grose, J. (**1993**). ''Short-term and long-term effects on the masking level difference following middle ear surgery,'' J. Am. Acad. Audiol. **4**, 307–312.
Harrison, R. V., Nagasawa, A., Smith, D. W., Stanton, S., and Mount, R. J. (**1991**). ''Reorganization of auditory cortex after neonatal high frequency cochlear hearing loss,'' Hearing Res. **54**, 11–19.
Hattori, H. (**1993**). ''Ear dominance for nonsense-syllable recognition ability in sensorineural hearing-impaired children: Monaural versus Binaural amplification,'' J. Am. Acad. Audiol. **4**, 319–330.
Heinen, S. J., and Skavenski, A. A. (**1991**). ''Recovery of visual responses in foveal *V*1 neurons following bilateral foveal lesion in adult monkey,'' Exp. Brain Res. **83**, 670–674.
Hood, J. D. (**1990**). ''Problems in central binaural integration in hearing loss cases,'' Hearing Instrum. **41**(4), 6–11, 56.
Horwitz, A. (**1995**). ''The time course of hearing aid benefit,'' unpublished dissertation, Syracuse University.
Humes, L. and Christopherson, L. (**1991**). ''Speech identification difficulties of hearing-impaired elderly persons: The contributions of auditory processing deficits,'' J. Speech Hear. Res. **34**, 686–693.
Humes, L. E., and Roberts, L. (**1990**). ''Speech recognition difficulties of the hearing-impaired elderly: The contributions of audibility,'' J. Speech Hear. Res. **33**, 726–735.
Humes, L., Dirks, D., Bell, T., and Kincaid, G. (**1987**). ''Recognition of nonsense syllables by hearing-impaired listeners and by noise-masked normal hearers,'' J. Acoust. Soc. Am. **81**, 765–773.
Humes, L., Halling, D., Schmit, K., Coughlin, M., Wilson, D., and Kinden, S. (**1995**). ''Longitudinal study of hearing-aid use, satisfaction and benefit in a group of elderly hearing-aid users,'' presented at the Hearing Aid Research and Development Conference, National Institutes of Health, Bethesda, MD.
Humes, L., Watson, C., Christensen, L., Cokely, C., Halling, D., and Lee, L. (**1994**). ''Factors associated with individual differences in clinical measures of speech recognition among the elderly,'' J. Speech Hear. Res. **37**, 465–474.
Hurley, R. (**1991**). ''Hearing aid use and auditory deprivation: A prospective study,'' paper presented at the meeting of the American Academy of Audiology. Denver, CO.
Hurley, R. (**1993**). ''Monaural hearing aid effects: Case presentations,'' J. Am. Acad. Audiol. **4**(5), 285–294.
Hurley, R. (**1994**). ''Auditory deprivation,'' paper presented at the Convention of the American Academy of Audiology, Richmond, VA.
Hurley, A., and Hurley, R. (**1994**). ''The effect of protracted infant middle ear diseases disease on auditory function in children,'' paper presented at the Convention of the American Academy of Audiology, Richmond, VA.
Jerger, J., Silman, S., Lew, H. L., and Chimel, R. (**1993**). ''Case studies in binaural interference: Converging evidence from behavioral and electrophysiologic measures,'' J. Am. Acad. Audiol. **4**, 122–131.
Kaas, J. H. (**1991**). ''Plasticity of sensory and motor maps in adult mammals,'' Ann. Rev. Neurosci. **14**, 137–167.
Kaas, J. H. (**1995**). ''The reorganization of sensory and motor maps in adult mammals,'' in *The Cognitive Neurosciences*, edited by M. S. Gazzaniga (MIT, Cambridge), pp. 51–71.
Kaas, J. H., Krubitzer, L. A., Chino, Y. M., Langston, A. L., Polley, E. H., and Blair, N. (**1990**). ''Reorganization of retinotopic cortical maps in adult mammals after lesions of the retina,'' Science **248**, 229–231.
Kalikow, D., Stevens, K., and Elliot, L. (**1979**). ''Development of a test of speech intelligibility in noise using sentence materials with controlled

word predictability,'' J. Acoust. Soc. Am. **61**, 1337–1351.
Kaltenbach, J. A., Czaja, J. M., and Kaplan, C. R. (**1992**). ''Changes in the tonotopic map of the dorsal cochlear nucleus following induction of cochlear lesions by exposure to intense sound,'' Hearing Res. **59**, 213–223.
Kelahan, A. M., and Doetsch, G. S. (**1984**). ''Time-dependent changes in the functional organization of somatosensory cerebral cortex following digit amputation in adult racoons,'' Somatosens. Res. **2**, 49–81.
Kraus, N., McGee, T., Carrell, T., King, C., Tremblay, K., and Nicol, T. (**1995**). ''Central auditory system plasticity associated with speech discrimination training,'' J. Cogn. Neurosci. **7**, 25–32.
Liu, C. N., and Chambers, W. W. (**1958**). ''Intraspinal sprouting of dorsal root axons,'' Arch. Neurol. Psychiatr. **79**, 46–61.
Ludvigsen, C. (**1987**). ''Prediction of speech intelligibility for normal-hearing and cochlearly hearing-impaired listeners,'' J. Acoust. Soc. Am. **82**, 1162–1171.
Malinoff, R., and Weinstein, B. (**1989**). ''Changes in self-assessment of hearing handicap over the first year of hearing aid use by older adults,'' J. Acad. Rehab. Audiol. **22**, 54–60.
Merzenich, M. M., Kaas, J. H., Wall, J. T., Nelson, R. J., Sur, M., and Felleman, D. J. (**1983a**). ''Topographic reorganization of sonatosensory cortical areas 3 Band 1 in adult monkeys following restricted deafferentation,'' Neuroscience (Oxford) **8**, 33–55.
Merzenich, M. M., Kaas, J. H., Wall, J. T., Sur, M., Nelson, R. J., and Felleman, D. J. (**1983b**). ''Progression of change following median nerve section in the cortical representation of the hand in areas 3*b* and 1 in adult owl and squirrel monkeys,'' Neuroscience (Oxford) **10**, 639–665.
Moore, D. R. (**1993**). ''Plasticity of binaural hearing and some possible mechanisms following late-onset deprivation,'' J. Am. Acad. Audiol. **4** 227–283.
Mulrow, C., Aguilar, C., Endicott, J., Tuley, M., Velez, R., Charlip, W., Rhodes, M., Hill, J., and DeNino, L. (**1990**). ''Quality-of-life changes and hearing impairment, a randomized trial,'' Ann. Internal Med. **113**, 188–194.
Needleman, A. R., and Crandell, C. C. (**1995**). ''Speech recognition in noise by hearing-impaired and noise-masked normal-hearing listeners,'' J. Am. Acad. Audiol. **6**, 414–424.
Neuman, A. C. (**1996**). ''Late-onset auditory deprivation: A review of past research and an assessment of future research needs,'' Ear Hearing **17**(3), 3S–13S.
Neuman, A. C. (**1997**). ''Acclimatization to Hearing Aids,'' Poster presented at the Second Biennial Hearing Aid Research and Development Conference. The National Institute on Deafness and Other Communication Disorders and The Department of Veterans Affairs.
Palmer, C. (**1994**). ''Variables to consider when interpreting the impact of monaural amplification,'' J. Am. Acad. Audiol. **5**, 286–290.
Palmer, C. (**1994**). ''The application of hearing aid sound quality judgments,'' International Hearing Aid Conference, Department of Speech Pathology and Audiology and The Department of Otolaryngology-Head and Neck Surgery, The University of Iowa, Iowa City, IA (June, 1995).
Pavlovic, C. V. (**1984**). ''Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment,'' J. Acoust. Soc. Am. **75**, 1253–1258.
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.
Pollin, B., and Albe-Fessard, P. (**1979**). ''Organization of somatic thalamus in monkeys with and without section of dorsal spinal track,'' Brain Res. **173**, 431–449.
Poole, C., and Jerger, J. (**1994**). ''Auditory deprivation in monaural hearing aid users,'' Annual Convention of the American Academy of Audiology (April 1994), Richmond, VA.
Popelar, J., Erre, J.-P., Aran, J.-M., and Cazals, Y. (**1994**). ''Plastic changes in ipsi-contralateral differences of auditory cortex and inferior colliculus evoked potentials after injury to one ear in the adult guinea pig,'' Hearing Res. **72**, 125–134.
Rajan, R., Irvine, D. R. F., Wise, L. Z., and Heil, P. (**1993**). ''Effects of unilateral partial cochlear hearing lesions in adult cats on the representation of lesioned and unlesioned cochleas on primary auditory cortex,'' J. Comp. Neurol. **338**, 17–49.
Rasmusson, D. D. (**1982**). ''Reorganization of racoon somatosensory cortex following removal of the fifth digit,'' J. Comp. Neurol. **205**, 313–326.
Recanzone, G. H., Merzenich, M. M., and Jenkins, W. M. (**1992a**). ''Frequency discrimination training engaging a restricted skin surface results in an emergence of a cutaneousresponse zone in cortical area 3*a*,'' J. Neurophysiol. **67**, 1057–1070.
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**, 1.
Recanzone, G. H., Merzenich, M. M., Jenkins, W. M., Grajski, K. A., and Dinse, H. R. (**1992**). ''Topographic reorganization of the hand representation in cortical area 3*b* of owl monkeys trained in a frequency-discrimination task,'' J. Neurophysiol. **67**, 1031–1056.
Robertson, D., and Irvine, D. R. F. (**1989**). ''Plasticity of frequency organization in auditory cortex of guinea pigs with partial unilateral deafness,'' J. Comp. Neurol. **282**, 456–471.
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**, 1–4.
Salvi, R. J., Henderson, D., Fiorino, F., and Colletti, V. (**1996**). *Auditory System Plasticity and Regeneration* (Thieme Medical, New York).
Saunders, G. H., and Cienkowski, K. M. (**1997**). ''Acclimatization to hearing aids,'' Ear Hearing **18**(2), 129–139.
Schwaber, M. K., Garraghty, P. E., and Kaas, J. H. (**1993**). ''Neuroplasticity of the adult primate auditory cortex following cochlear hearing loss,'' Am. J. Otol. **14**(3), 252–258.
Silman, S., Gelfand, S. A., and Silverman, C. A. (**1984**). ''Late-onset auditory deprivation: Effects of monaural versus binaural hearing aids,'' J. Acoust. Soc. Am. **76**, 1357–1362.
Silman, S., Silverman, C. A., Emmer, M. B., and Gelfand, S. A. (**1993**). ''Effects of prolonged lack of amplification on speech-recognition performance: Preliminary findings,'' J. Rehab. Res. **30**(3), 326–332.
Silverman, C. A. (**1989**). ''Auditory Deprivation,'' Hearing Instrum. **40**(9), 26–32.
Silverman, C. A., and Emmer, M. B. (**1993**). ''Auditory deprivation and recovery in adults with asymmetric sensorineural hearing impairment,'' J. Am. Acad. Audiol. **4**, 338–346.
Silverman, C. A., and Silman, S. (**1990**). ''Apparent auditory deprivation from monaural amplification and recovery with binaural amplification: Two case studies,'' J. Am. Acad. Audiol. **1**, 175–180.
Snow, P. J., and Wilson, P. (**1991**). ''Plasticity in the somatosensory system of developing and mature mammals,'' *in Progress in Sensory Physiology, Vol. II* (Springer-Verlag, New York).
Snyder, R. L., Rebscher, S. J., Cao, K., Leake, P. A., and Kelly, K. (**1990**). ''Chronic intracochlear electrical stimulation in the neonatally deafened cat I: Expansion of central representation,'' Hearing Res. **50**, 7–34.
Soli, S., and Nilsson, M. (**1994**). ''Assessment of communication handicap with the HINT,'' Hear. Instrum. **12**(1), 15–16.
Stubblefield, J., and Nye, C. (**1989**). ''Aided and unaided time-related differences in word discrimination,'' Hearing Instrum. **40**(9), 38–43, 78.
Taylor, K. (**1992**). ''Self-perceived and audiometric evaluations of hearing aid benefit in the elderly,'' poster presented at the annual meeting of the American Speech-Language-Hearing Association, San Antonio, TX.
Taylor, K. (**1993**). ''Self-perceived and audiometric evaluations of hearing aid benefit in the elderly,'' Ear Hearing **14**, 390–394.
Thornton, A., and Raffin, M. (**1978**). ''Speech discrimination scores modeled as a binomial variable,'' J. Speech Hear. Res. **21**, 507–518.
Turner, C., Humes, L., Bentler, R., and Cox, R. (**1996**). ''A review of past research on changes in hearing aid benefit over time,'' Ear Hear. **17**, 14S–28S.
Truy, E., Deiber, M., Cinotti, L., Mauguiere, F., Froment, J., and Morgan, A. (**1995**). ''Auditory cortex activity changes in long-term sensorineural deprivation during crude electrical stimulation: Evaluation by positron emission tomography,'' Hearing Res. **86**, 34–42.
Vasama, J., and Makela, J. P. (**1995**). ''Auditory pathway plasticity in adult humans after unilateral idiopathic sudden sensorineural hearing loss,'' Hearing Res. **87**, 132–140.
Walden, B., Erdman, S., Montgomery, A., Schwartz, D., and Prosek, R. (**1981**). ''Some effects of training on speech recognition by hearing-impaired adults,'' J. Speech Hear. Res. **24**, 207–216.
Wall, J. T., and Cusick, C. G. (**1984**). ''Cutaneous responsiveness in primary somatosensory (S-1) hindpaw cortex before and after partial hindpaw deafferentation in adult rats,'' J. Neurosci. **4**, 1499–1515.
Wall, P. D., and Egger, M. D. (**1971**). ''Formation of new connections in adult rat brains after partial deafferentation,'' Nature **232**, 542–545.

Watson, N. A., and Knudsen, K. O. (**1940**). ''Selective amplification in hearing aids,'' J. Acoust. Soc. Am. **11**, 406–419.
Willot, J. (**1984**). ''Changes in frequency representation in the auditory system of mice with age-related hearing impairment,'' Brain Res. **309**, 159–162.
Willott, J. F. (**1986**). ''Effects of aging, hearing loss, and anatomical location on thresholds of inferior colliculus neurons in C57BL/6 and CBA mice,'' J. Neurophysiol. **56**, 391–408.
Willott, J. F. (**1996a**). ''Auditory system plasticity in the adult C57BL/6J mouse,'' in *Auditory System Plasticity and Regeneration* edited by R. J. Salvi, D. Henderson, F. Fiorino, and V. Colletto (Thieme Medical, New York), pp. 297–316.
Willott, J. F. (**1996b**). ''Anatomic and physiologic aging: A behavioral neuroscience perspective,'' J. Am. Acad. Audiol. **7**, 41–151.
Willott, J. F. (**1996c**). ''Physiological plasticity in the auditory system and its possible relevance to hearing aid use, deprivation effects, and acclimatization,'' Ear Hearing **17** (Suppl.) 66s–77s.
Willott, J. F., and Lu, S. (**1982**). ''Noise-induced hearing loss can alter neural coding and increase excitability in the central nervous system,'' Science **216**, 1331–1332.
Willott, J. F., Aitkin, L. M., and McFadden, S. L. (**1993**). ''Plasticity of auditory cortex associated with sensorineural hearing loss in adult C57BL/6J mice,'' J. Comp. Neurol. **329**, 402–411.
Willott, J. F., Carlson, S., and Chen, H. (**1994**). ''Prepulse inhibition of the startle response in mice: Relationship to hearing loss and auditory system plasticity,'' Behav. Neurosci. **108** 703–713.
Willott, J. F., Parham, K., and Hunter, K. P. (**1988**). ''Response properties of inferior colliculus neurons in middle-aged C57BL/6J mice with presbycusis,'' Hearing Res. **37**, 15–28.
Willott, J. F., Parham, K., and Hunter, K. P. (**1991**). ''Comparison of the auditory sensitivity of neurons in the cochlear nucleus and inferior colliculus of young and aging C57BL/6J and CBA/J mice,'' Hearing Res. **53**, 78–94.
Yund, W., and Buckles, K. (**1995**). ''Discrimination of multichannel-compressed speech in noise: Long-term learning in hearing-impaired subjects,'' Ear Hear. **16**, 417–427.
Zurek, R., and Delhourne, L. (**1987**). ''Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment,'' J. Acoust. Soc. Am. **82**, 1548–1559.

Temporal backward planar projection of acoustic transients

G. T. Clement, R. Liu, and S. V. Letcher
Department of Physics, University of Rhode Island, Kingston, Rhode Island 02881

P. R. Stepanishen
Department of Ocean Engineering, University of Rhode Island, Narragansett, Rhode Island 02882

(Received 30 September 1997; accepted for publication 22 December 1997)

Backward projection schemes use data in front of a transmitter to reconstruct a field at closer points. Existing techniques have concentrated propagating radial and temporal information along constant planar cross sections in front of a planar source. This approach requires a careful consideration of evanescent waves, as the transfer function used to backproject in space causes evanescent wave solutions to increase exponentially with the projected distance. Erroneous signals may result from exponentially increasing noise, experimental error or roundoff error. A method is presented that is designed to work with imaging methods that record three-dimensional spatial data at constant times. Several widely used optical methods are of this type. Our algorithm projects the field backward in time via linear wave theory. The approach is similar to previously reported methods but is designed to work with time-constant data and avoids problems associated with evanescent waves. © *1998 Acoustical Society of America.* [S0001-4966(98)01504-5]

PACS numbers: 43.20.Bi [ANN]

INTRODUCTION

Linear planar propagation methods may be used for full field characterization in space-time using limited data. These numerical methods generally operate in the wave vector frequency domain using Fourier transforms[1] or Hankel transforms.[2,3] A wave vector, time-domain approach[4] has also been described. These methods use field information along a planar cross-section at a constant distance z_0 from a baffled planar source. The signal may be forward projected away from the source or backprojected toward the source. Backward projections of experimentally measured harmonic pressure amplitudes,[5,6] $P(\mathbf{r},z_0)$ have been reported. Recently, experimental time dependent transient fields, $p(\mathbf{r},z_0,t)$,[7,8] were forward projected.

Backward projection of time-dependent transients, however, is more complicated. Fleischer and Axelrad[9] have addressed problems associated with exponential growth of evanescent wave solutions. Williams and Maynard[5] used data very close to the source to avoid complications. Filtering methods[6] and approximation techniques[10] have also been used to partially amend this problem.

An alternative to projecting the field backward in space is to record spatial field information and time reverse the field. To provide distinction from the recent work of Fink *et al.*[11,12] we refer to this technique as temporal backward projection. Fink's work concentrates on the use of acoustic time-reversal mirrors. These mirrors physically reproject the conjugate of their received signals, focusing the field back on the original source. Through an iterative process a strong target in a multitarget media may be singled out. Prada *et al.*[13] provide an elegant analysis of this process. In contrast, the temporal backward projection described here assumes full knowledge of the field in a medium and is designed to reproduce acoustic signal information at arbitrary points at earlier times. A single measurement at a given time allows the full spatial field to be reconstructed at any earlier time.

This temporal approach is well suited for imaging systems which record field data at constant times. Examples of these systems include tomographically reconstructed optical pulsed schlieren, interferometry, and holographic techniques. The transfer function employed in this temporal projection method is mathematically similar to existing techniques, however it exhibits advantages over spatial backward projections. The transfer function does not display the exponential growth associated with spatial projections. In addition, the time constant imaging method is capable of recording complete field information, while spatial projection methods record temporal and radial data over a constant plane and ultimately must truncate the field.

We present a practical algorithm for temporal forward/backward projection. Axisymmetric sources are assumed, allowing Hankel transform techniques to be used. The underlying theory is discussed in Sec. II. An impulse response approach[14] is used to generate numerical solutions to the field resulting from a Gaussian pulse input into a baffled circular radiator. Data obtained at a specific time, t_0 is projected backward to a new time t and compared with direct numerical calculations. To demonstrate the range of the method, we select a starting acoustic pulse far from the source. The back-projection is shown to reconstruct major field features at earlier times. Results and procedures are outlined and discussed in Sec. III.

I. THEORY

A general pressure field in Cartesian coordinates may be written in terms of its three dimensional Fourier transform

$$p(\mathbf{r},t)=\frac{1}{(2\pi)^{3/2}}\int\int\int_{-\infty}^{\infty} P(\mathbf{K},t)e^{i\mathbf{K}\cdot r}\,d^3\mathbf{K}, \tag{1}$$

 0001-4966/98/103(4)/1723/4/$10.00

where the wavenumber $K^2=k_x^2+k_y^2+k_z^2$. It is assumed that this function satisfies the linear wave equation

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

with sound velocity c. Substitution of the right hand side of Eq. (1) into Eq. (2) yields a Helmholtz equation,

$$\frac{d^2}{dt^2}P(\mathbf{K},t)+c^2K^2P(\mathbf{K},t)=0. \tag{3}$$

The solutions to this second-order equation describe the advanced and retarded propagation, respectively. The advanced solution is used for the projection problem:

$$P(\mathbf{K},t)=P(\mathbf{K},t_0)e^{-ic(t-t_0)K}, \tag{4}$$

where the exponential term is the transfer function associated with the projection. The field $P(K,t_0)$ is obtained by direct measurement of the field at t_0. It is noted that the exponent of the transfer function is pure imaginary, thus eliminating the exponentially increasing signals associated with evanescent waves.

The pressure fields considered in this paper are axisymmetric, thus a cylindrical coordinate system is more natural for describing the projections. The radiator is assumed to be a circular disk in the $y-z$ plane and centered about the z-axis. A change of variables is made so that $k_x=k_\rho\cos\phi$, $k_y=k_\rho\sin\phi$, $K=\sqrt{k_\rho^2+k_z^2}$, and the field is independent of the polar angle ϕ. The pressure field in wave-vector space is now represented by

$$P(k_\rho,k_z,t)=P(k_\rho,k_z,t_0)e^{-ic(t-t_0)\sqrt{k_\rho^2+k_z^2}}. \tag{5}$$

To reconstruct the field at the time t, the zeroth-order Hankel transform pair is introduced:

$$P(k_\rho,k_z,t)=\int_0^\infty p(\rho,k_z,t)J_0(k_\rho\rho)\rho\ d\rho, \tag{6}$$

$$p(\rho,k_z,t)=\int_0^\infty P(k_\rho,k_z,t)J_0(k_\rho\rho)k_\rho\ dk_\rho. \tag{7}$$

The zeroth-order Bessel function is given by $J_0(k_\rho\rho)$. Using the conjugate transform of Eq. (1) and Eq. (6), the field may be expressed as

$$p(\rho,z,t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\int_0^\infty P(k_\rho,k_z,t_0)\times e^{-ic(t-t_0)\sqrt{k_\rho^2+k_z^2}}J_0(k_\rho\rho)e^{ik_zz}k_\rho\ dk_\rho\ dk_z. \tag{8}$$

The pressure $P(k_\rho,k_z,t_0)$ is obtained by direct measurement of the field at t_0.

II. PROCEDURES AND RESULTS

The impulse response method is used to simulate acoustic fields from an underwater ultrasonic source. A Gaussian input signal, full width at half-maximum of 1 μs, is considered. The radiator is assumed to be a 2.8-cm-diam baffled circular disk.

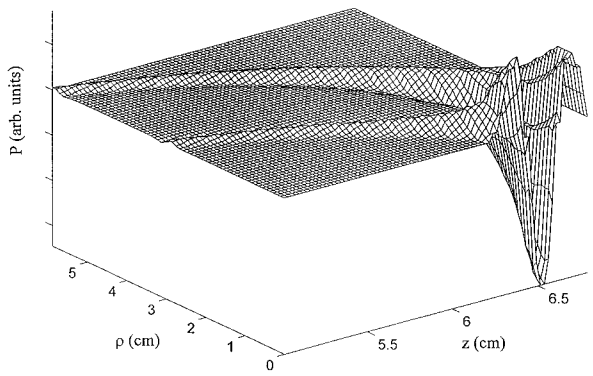


FIG. 1. Acoustic signal at 30 μs.

A simulated field is calculated for the time $t_0=30$ μs after the start of the launching of the Gaussian pulse. This field is then propagated backward in time using a matrix based algorithm to calculate a discrete approximation to Eq. (8). The field is represented by a 90×150 matrix describing the ρ and z dimensions, respectively. Specifically, fields reconstructed at 15, 9, and 3 μs are presented. The projected fields are compared with direct impulse response calculations over a series of time slices. The starting pressure field at 30 μs is shown in Fig. 1 as a function of its spatial coordinates, ρ and z. The field is measured in the region $0\leqslant\rho\leqslant6$ cm and $5\leqslant z\leqslant6.8$ cm, and is assumed zero elsewhere. This field displays the characteristic shape of a Gaussian signal far from the source. The impulse response field calculation at 3 μs is shown in Fig. 2(a). This signal is compared with the reconstructed back-projection from 30 μs to 3 μs, in Fig. 3(a). It is noted that the backprojection carries additional small oscillations along both the radial and propagation directions. The projection algorithm is, of course, discrete and band limited and ultimately cannot completely reconstruct the surface describing the field. The signal at 9 μs is shown in Fig. 2(b) and its backprojection from 30 μs in Fig. 3(b). Direct comparison reveals the backprojection's ability to reconstruct the half-ring shaped waveform centered about the transducer edge. At 15 μs the field now exhibits two local minima on-axis; the peak at 2.4 μs being from the main signal and the second a result of the acoustic ''edge wave.'' This feature, shown in Fig. 2(c), is reproduced in Fig. 3(c), the projection from 30 μs to 15 μs. This projection, however, reproduces a slightly broader edge wave than the direct numerical calculation predicts. Once again, this seems to result from the band limitations set by the numerical algorithm.

Each projection took under one minute to compute using a 133-MHz Pentium processor. The spatial sampling is 16 cm^{-1} in the radial direction and 80 cm^{-1} along z for each of the fields in Fig. 3. The spatial-frequency sampling used in the algorithm is 21 cm over $0\leqslant k_z\leqslant60$ cm^{-1} and 50 cm over $0\leqslant k_\rho\leqslant20$ cm^{-1}.

III. SUMMARY

The temporal backward projection algorithm is shown to accurately reconstruct major features of the pressure field

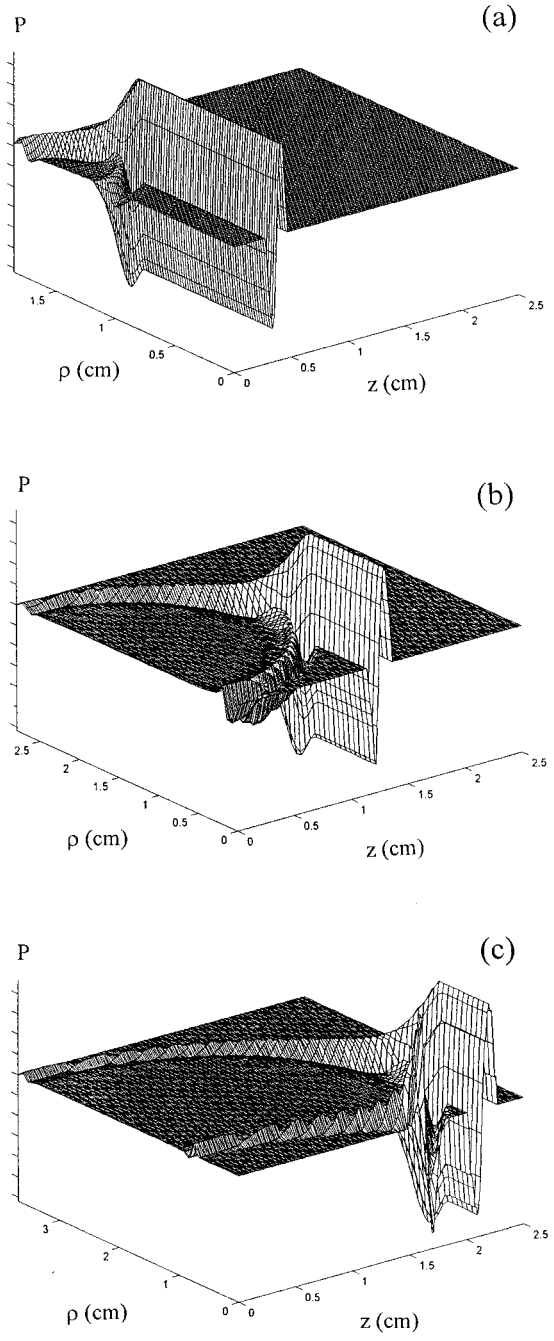


FIG. 2. Impulse response calculation of pressure field **P** in arbitrary units at (a) $t=3$ μs, (b) $t=9$ μs, and (c) $t=15$ μs.

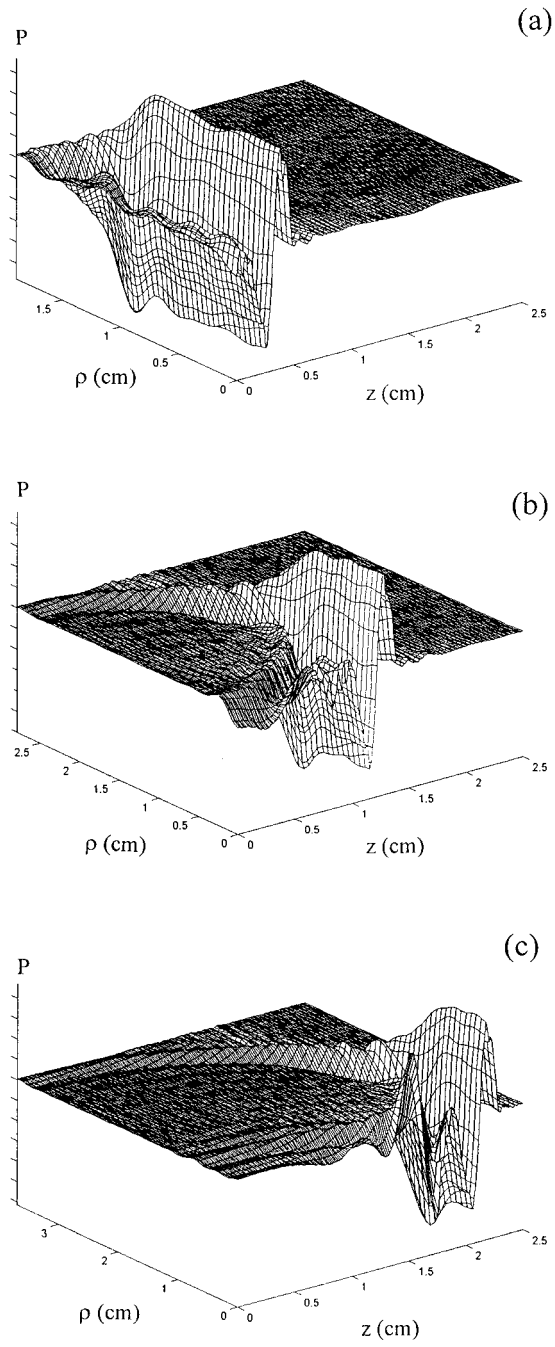


FIG. 3. Backward projection of pressure field from $t_0=30$ μs to (a) $t=3$ μs, (b) $t=9$ μs, and (c) $t=15$ μs.

resulting from an axisymmetric Gaussian input signal. Failure of the Hankel transform based algorithm to reconstruct sharp features in the field near the source seems to result from band limitations along both the radial and propagation axes. In practice, fields measured experimentally tend to be smooth and in this respect may prove easier to reconstruct than synthetic data. There are advantages to the reported method over previously reported spatial projection techniques. Field imaging systems, such as the modulated schlieren apparatus, which record spatial field information at a specific time, may more readily use the temporal method for projecting fields. Further the transfer function implemented in the temporal backward projection method does not exhibit the complications associated with evanescent waves. Finally, since the field as a specific time is spatially finite for pulsed signals, an entire field may be imaged where previous methods require truncation. Further study will concentrate on the backward projection of experimental transient signals. In addition, a numerical investigation will be conducted concerning the effects of noise on the algorithm.

[1] P. R. Stepanishen and K. C. Benjamin, ''Forward and backward projection of acoustic fields using FFT methods,'' J. Acoust. Soc. Am. **71**, 803–812 (1982).
[2] A. F. Medeiros and P. R. Stepanishen, ''The forward and projection of

acoustic fields from axisymmetric ultrasonic radiators using impulse response and Hankel transform techniques,'' J. Acoust. Soc. Am. **75**, 1732–1740 (1984).
[3] P. T. Christopher and K. J. Parker, ''New approaches to the linear propagation of acoustic fields,'' J. Acoust. Soc. Am. **90**, 507–521 (1991).
[4] M. Forbes, S. Letcher, and P. Stepanishen, ''A wave vector, time-domain method of forward projecting time-dependent pressure fields,'' J. Acoust. Soc. Am. **90**, 2782–2793 (1991).
[5] E. G. Williams and J. D. Maynard, ''Holographic imaging without the wavelength resolution limit,'' Phys. Rev. Lett. **45**, 554–557 (1980).
[6] R. Reibold and F. Holzer, ''Complete mapping of ultrasonic fields from optically measured data in a single cross-section,'' Acustica **58**, 11–16 (1985).
[7] G. T. Clement, R. Liu, S. V. Letcher, and P. R. Stepanishen, ''Forward projection of transient signals obtained from a fiber-optic pressure sensor,'' submitted to J. Acoust. Soc. Am.
[8] G. T. Clement, R. Liu, S. V. Letcher, and P. R. Stepanishen, ''Forward projection of transient signals obtained from a fiber-optic pressure sensor,'' J. Acoust. Soc. Am. **101**, 3033(A) (1997).
[9] H. Fleischer and V. Axelrad, ''Restoring and acoustic source from pressure data using Weiner filtering,'' Acustica **60**, 172–175 (1986).
[10] D. Liu and R. Waag, ''Propagation and backpropagation for ultrasonic wavefront design,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 1–13 (1997).
[11] A. Derode, P. Roux, and M. Fink, ''Robust Acoustic Time Reversal with High-Order Multiple Scattering,'' Phys. Rev. Lett. **75**, 4206–4209 (1995).
[12] M. Fink, ''Time Reversed Acoustics,'' Phys. Today **50**, 34–40 (1997); M. Fink, ''Time-reversed ultrasound,'' J. Acoust. Soc. Am. **101**, 3088(A) (1997).
[13] 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).
[14] P. R. Stepanishen, ''Acoustic transients from planar axisymmetric vibrators using the impulse response approach,'' J. Acoust. Soc. Am. **70**, 1176–1181 (1981).

Bounds and approximations for elastodynamic wave speeds in tetragonal media

Q. H. Zuo
Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

K. D. Hjelmstad
Department of Civil Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

(Received 25 July 1997; revised 18 December 1997; accepted 20 December 1997)

This paper presents an analytical study of elastodynamic waves propagating along an arbitrary direction in anisotropic materials with tetragonal symmetry. Upper and lower bounds are developed on the wave speeds through an additive decomposition of the acoustic tensor into an associated hexagonal counterpart and a rank-one modification, the eigenproperties of which can be determined analytically. The bounds are obtained by applying the minimax property of eigenvalues. Linear approximations of the wave speeds are obtained from a first-order expansion of the eigenvalues of the acoustic tensor, with respect to a tetragonality index, about the value of that index for which tetragonal symmetry degenerates to hexagonal symmetry. A numerical example shows that the linear approximations agree remarkably well with the exact values of the wave speeds. © *1998 Acoustical Society of America.* [S0001-4966(98)00704-8]

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

INTRODUCTION

In an anisotropic material the elastodynamic wave speeds depend on the propagation direction, and in general the polarization directions are neither parallel nor perpendicular to the propagation direction. Unless the propagation direction coincides with one of the ''privileged'' directions in the material, the wave speeds and the polarization directions often have to be calculated numerically (Fedorov, 1968; Musgrave, 1970; Pao and Kaul, 1974; Auld 1990). The so-called acoustic axes in a crystal refer to those directions along which a pure-longitudinal or a pure-shear wave can propagate. Along an acoustic axis the characteristic equation for the wave speeds can be factored into linear and quadratic parts, and hence can be solved analytically. Much research has been done in identifying the acoustic axes for crystals with various symmetry classes, and to examine the acoustic behavior along those axes. In particular, Fedorov (1965), Truesdell (1966), Kolodner (1966), and Stippes (1966) have proved that in a material satisfying the positive longitudinal elasticity condition there exist at least three distinct directions along which pure-longitudinal waves can propagate. Al'shits and Lothe (1979) have conducted a thorough study of the existence of the acoustic axes and the related degeneracy problem in triclinic crystals (crystals with no symmetry).

Several researchers have proposed approximations and established bounds for wave speeds in anisotropic media. For waves propagating along an arbitrary propagation direction in a tetragonal material, Fedorov (1968) presented a technique for finding a hexagonal material to approximately characterize the acoustic behavior of the tetragonal material. The elastic constants of the hexagonal material are chosen to minimize the difference between the wave speeds of the approximate hexagonal and the original tetragonal materials over all propagation directions. More recently, Shuvalov and Every (1997), in studying the distortion of the acoustic slowness surface near a conical degeneracy, provided approximations of the eigenvalues of the acoustic tensor for the propagation directions that are close to the acoustic axes. Ditri (1997) proved bounds on the changes in wave speeds due to changes in the elastic moduli of an anisotropic material. The key result is that the sum of the square of the changes in the eigenvalues of an acoustic tensor is bounded from above by the square of the Euclidean norm of the changes in the acoustic tensor.

For materials with cubic symmetry, Zuo and Hjelmstad (1997) developed upper and lower bounds on the waves speeds, and provided a first-order approximation for the speeds, in terms of a parameter that characterizes the degree of anisotropy of the material and the wave propagation direction. For waves propagating along one of the crystal faces, their first-order approximation compared extremely well with the exact results, even for copper, the most anisotropic material among the commonly used metals with cubic symmetry. The success of the approximation for cubic materials motivates the development presented here for materials with tetragonal symmetry.

For specific values of the material constants and a given wave propagation direction, one can numerically perform the spectral decomposition of the symmetric three by three acoustic tensor. However, the simple analytical expressions we present here lend deeper insights into the effects of the material constants and the propagation direction on the wave speeds. The acoustic tensor for a tetragonal crystal depends on eight independent parameters (six material constants and two propagation angles). Hence the numerical parameter studies required to gain the same level of insight would be quite tedious to compute and difficult to represent. We develop a simple approximate expression for the wave speeds in terms of a tetragonality index and the propagation direction. We study the sensitivity of the approximated wave

speeds to the change of the propagation direction and show that the directions with extreme sensitivity are those special directions in a tetragonal crystal along which pure waves can propagate. Such intriguing features cannot be revealed by numerically computing the eigenvalues of an acoustic tensor.

I. PRELIMINARIES

Consider plane waves propagating in an anisotropic elastic solid along direction $\mathbf{n}$ ($\mathbf{n}\cdot\mathbf{n}=1$). The displacement vector $\mathbf{U}$ can be expressed as

$$\mathbf{U}=\mathrm{Re}\{\mathbf{u}e^{i\xi(\mathbf{n}\cdot\mathbf{x}-ct)}\}, \tag{1}$$

where $\mathbf{u}$ is the polarization direction, $\mathbf{x}$ the position vector, t time, ξ the wave number, and c the wave speed. The wave speeds and the polarization directions are the solutions of the following eigenvalue problem (the Christoffel equation, Auld, 1990):

$$\mathbf{A(n)u}=\varrho c^2\mathbf{u}, \tag{2}$$

where ϱ is the density of the material. The acoustic tensor $\mathbf{A}(\mathbf{n})$ is defined by

$$\mathbf{A(n)}\equiv\mathbf{n}\cdot\mathbf{En}, \tag{3}$$

where $\mathbf{E}$ is the fourth-order elasticity tensor of the material which maps strain $\mathbf{e}$ to stress $\boldsymbol{\sigma}$ as

$$\boldsymbol{\sigma}=\mathbf{Ee}. \tag{4}$$

Restrictions on E

If the strain energy density function exists for the elastic material, then $\mathbf{E}$ must possess both minor and major symmetries, namely, $E_{ijkl}=E_{jikl}=E_{ijlk}=E_{klij}$. For a given propagation direction $\mathbf{n}$, it follows from Eq. (2) that a wave can propagate with a real velocity if and only if the corresponding eigenvalue of the acoustic tensor is positive. According to Truesdell (1966), Hadamard first showed that along any propagation direction, all three waves have real velocities (and hence can propagate) for materials satisfying the strong ellipticity condition, namely,

$$\mathbf{m}\cdot\mathbf{A(n)m}=(\mathbf{m}\otimes\mathbf{n}):\mathbf{E}(\mathbf{m}\otimes\mathbf{n})>0 \tag{5}$$

for any unit vector $\mathbf{m}$. The strong ellipticity condition turns out to be intimately related to stability of equilibrium in an elasto-plastic material. Neilsen and Schreyer (1993) suggested that the loss of strong ellipticity be the criterion for predicting the initiation of a discontinuous bifurcation, which causes deformation to localize and ultimately leads to material failure.

Truesdell (1966), by using the fixed-point theorem, proved that in a material satisfying the positive longitudinal elasticity condition there exists at least one direction along which longitudinal waves can propagate. Under the same assumption, but using the Poincaré theorem in topology, Kolodner (1966) proved that there exist at least three such directions, and hence improved Truesdell's results. The positive longitudinal elasticity condition assumed in the two proofs is

$$\mathbf{n}\cdot\mathbf{A(n)n}=(\mathbf{n}\otimes\mathbf{n}):\mathbf{E}(\mathbf{n}\otimes\mathbf{n})>0. \tag{6}$$

Obviously, the strong ellipticity condition (5) implies the positive longitudinal elasticity condition, but not vice versa. In this paper, we restrict the elasticity tensors to be positive definite, which requires, for any second-order tensor $\mathbf{M}$,

$$\mathbf{M}\cdot\mathbf{EM}>0 \tag{7}$$

Since $\mathbf{m}\otimes\mathbf{n}$ is a special second-order tensor (it is rank-one), the condition of positive definiteness is even stronger than the condition of strong ellipticity.

II. ANISOTROPIC ELASTICITY TENSOR

The constitutive equation (4) is often rewritten as

$$\{\sigma\}=[C]\{e\}, \tag{8}$$

where $\{\sigma\}^T=\{\sigma_{11},\sigma_{22},\sigma_{33},\sqrt{2}\sigma_{23},\sqrt{2}\sigma_{31},\sqrt{2}\sigma_{12}\}$ and $\{e\}^T=\{e_{11},e_{22},e_{33},\sqrt{2}e_{23},\sqrt{2}e_{31},\sqrt{2}e_{12}\}$ are the vector version of the second-order stress and strain tensors, respectively, and $[C]$ is the stiffness matrix (e.g., Auld, 1990). The factor $\sqrt{2}$ is chosen to preserve the 2-norm of stress and strain. In the materials engineering and acoustics literature (e.g., Hertzberg, 1987), a material is usually described by its stiffness matrix $[C]$. However, to formulate the acoustic tensor defined in (3), we need the fourth-order elasticity tensor $\mathbf{E}$ of the material. The fourth-order elasticity tensor of a material with general anisotropy can be represented as

$$\mathbf{E}=\sum_{i=1}^{6}\sum_{j=1}^{6}D_{ij}\mathbf{N}_i\otimes\mathbf{N}_j, \tag{9}$$

where the six base tensors are chosen to be

$$\begin{aligned}
&\mathbf{N}_1=\frac{1}{\sqrt{3}}\,\mathbf{i},\quad \mathbf{N}_2=\frac{1}{\sqrt{6}}\,(\mathbf{e}_1\otimes\mathbf{e}_1+\mathbf{e}_2\otimes\mathbf{e}_2-2\mathbf{e}_3\otimes\mathbf{e}_3),\\
&\mathbf{N}_3=\frac{1}{\sqrt{2}}\,(\mathbf{e}_1\otimes\mathbf{e}_1-\mathbf{e}_2\otimes\mathbf{e}_2),\\
&\mathbf{N}_4=\frac{1}{\sqrt{2}}\,(\mathbf{e}_2\otimes\mathbf{e}_3+\mathbf{e}_3\otimes\mathbf{e}_2),\quad \mathbf{N}_5=\frac{1}{\sqrt{2}}\,(\mathbf{e}_1\otimes\mathbf{e}_3+\mathbf{e}_3\otimes\mathbf{e}_1),\\
&\mathbf{N}_6=\frac{1}{\sqrt{2}}\,(\mathbf{e}_1\otimes\mathbf{e}_2+\mathbf{e}_2\otimes\mathbf{e}_1),
\end{aligned} \tag{10}$$

where $\mathbf{i}$ is the second-order identity tensor. The number of independent components in Eq. (9), $D_{ij}=D_{ji}$, ranges from 21 for a triclinic crystal to 2 for an isotropic material. For a triclinic crystal, the coefficients D_{ij}, in terms of the components of the stiffness matrix $[C]$, are found to be

$$\begin{aligned}
D_{11}&=\tfrac{1}{3}(C_{11}+C_{22}+C_{33}+2(C_{12}+C_{13}+C_{23})),\\
D_{22}&=\tfrac{1}{6}(C_{11}+C_{22}+4C_{33}+2(C_{12}-2C_{13}-2C_{23})),\\
D_{33}&=\tfrac{1}{2}(C_{11}+C_{22}-2C_{12}),\\
D_{12}&=\frac{\sqrt{2}}{6}\,(C_{11}+C_{22}-2C_{33}+(2C_{12}-C_{13}-C_{23})),
\end{aligned} \tag{11}$$

$$D_{13}=\frac{1}{\sqrt{6}}(C_{11}-C_{22}+C_{13}-C_{23}),$$

$$D_{23}=\frac{\sqrt{3}}{6}(C_{11}-C_{22}+2(C_{23}-C_{13})),$$

and for $j=4,5,6$,

$$D_{1j}=\frac{1}{\sqrt{3}}(C_{1j}+C_{2j}+C_{3j}),\quad D_{2j}=\frac{1}{\sqrt{6}}(C_{1j}+C_{2j}-2C_{3j}),$$
$$D_{3j}=\frac{1}{\sqrt{2}}(C_{1j}-C_{2j}) \tag{12}$$

and $D_{ij}=C_{ij}$, for $i=4,5,6$, and $j=4,5,6$, corresponding to the shear moduli of the material.

Tetragonal symmetry

For an anisotropic material with tetragonal symmetry, if an orthonormal basis (Cartesian coordinate system) is chosen along symmetry axes of the material, the stiffness matrix is

$$[C]=\begin{bmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & 0 \\ C_{12} & C_{11} & C_{13} & 0 & 0 & 0 \\ C_{13} & C_{13} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & 2C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & 2C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & 2C_{66} \end{bmatrix}, \tag{13}$$

in which the six independent material parameters are C_{11}, C_{12}, C_{13}, C_{33}, C_{44}, and C_{66}. The rotational symmetry about $\mathbf{e}_3$ is reflected in Eq. (13). The general representation (9) reduces to

$$\begin{aligned}\mathbf{E}=&D_{11}\mathbf{N}_1\otimes\mathbf{N}_1+D_{12}(\mathbf{N}_1\otimes\mathbf{N}_2+\mathbf{N}_2\otimes\mathbf{N}_1)+D_{22}\mathbf{N}_2\otimes\mathbf{N}_2\\&+(C_{11}-C_{12})\mathbf{N}_3\otimes\mathbf{N}_3+2C_{66}\mathbf{N}_6\otimes\mathbf{N}_6\\&+2C_{44}(\mathbf{N}_4\otimes\mathbf{N}_4+\mathbf{N}_5\otimes\mathbf{N}_5),\end{aligned} \tag{14}$$

where the three modified constants are

$$\begin{aligned}D_{11}&=\tfrac{1}{3}(2C_{11}+C_{33}+2C_{12}+4C_{13}),\\D_{22}&=\tfrac{1}{3}(C_{11}+2C_{33}+C_{12}-4C_{13}),\\D_{12}&=\frac{\sqrt{2}}{3}(C_{11}-C_{33}+C_{12}-C_{13}).\end{aligned} \tag{15}$$

Positive definiteness of **E**. The elasticity tensor of a tetragonal material is positive definite, as defined in Eq. (7), if and only if its eigenvalues are positive. The eigenvalues are (e.g., Schreyer and Zuo, 1995)

$$2\lambda^*_{1,2}=C_{33}+C_{11}+C_{12}\pm\sqrt{(C_{33}-C_{11}-C_{12})^2+8(C_{13})^2}, \tag{16}$$

$$\lambda^*_3=C_{11}-C_{12},\quad \lambda^*_4=\lambda^*_5=2C_{44},\quad \lambda^*_6=2C_{66}. \tag{17}$$

Hence, the requirement for **E** to be positive definite is $C_{11}>C_{12}$, $C_{44}>0$, $C_{66}>0$, and

$$C_{33}(C_{11}+C_{12})>2C_{13}^2. \tag{18}$$

The positive definiteness of the elasticity tensor of an elastic material guarantees that all waves in the material propagate with real velocities.

III. ACOUSTIC TENSOR OF A TETRAGONAL MATERIAL

Substitution of the expression for the elasticity tensor in Eq. (9) into (3) yields the expression for the acoustic tensor of a generally anisotropic material

$$\mathbf{A}(\mathbf{n})=\sum_{i=1}^{6}\sum_{j=1}^{6}D_{ij}\mathbf{n}_i\otimes\mathbf{n}_j, \tag{19}$$

where $\mathbf{n}_i=\mathbf{N}_i\mathbf{n}$ are the deflections of the propagation direction by the base tensors,

$$\begin{aligned}\mathbf{n}_1&=\frac{1}{\sqrt{3}}\mathbf{n},\quad \mathbf{n}_2=\frac{1}{\sqrt{6}}(n_1\mathbf{e}_1+n_2\mathbf{e}_2-2n_3\mathbf{e}_3),\\\mathbf{n}_3&=\frac{1}{\sqrt{2}}(n_1\mathbf{e}_1-n_2\mathbf{e}_2),\end{aligned} \tag{20}$$

$$\begin{aligned}\mathbf{n}_4&=\frac{1}{\sqrt{2}}(n_3\mathbf{e}_2+n_2\mathbf{e}_3),\quad \mathbf{n}_5=\frac{1}{\sqrt{2}}(n\mathbf{e}_1+n_1\mathbf{e}_3),\\\mathbf{n}_6&=\frac{1}{\sqrt{2}}(n_2\mathbf{e}_1+n_1\mathbf{e}_2).\end{aligned} \tag{21}$$

In Eqs. (20)–(21), $n_i\equiv\mathbf{n}\cdot\mathbf{e}_i$ $(i=1,2,3)$ are the components of the propagation direction in the chosen basis. For a tetragonal material the acoustic tensor given in Eq. (19) reduces to

$$\begin{aligned}\mathbf{A}(\mathbf{n})=&D_{11}\mathbf{n}_1\otimes\mathbf{n}_1+D_{12}(\mathbf{n}_1\otimes\mathbf{n}_2+\mathbf{n}_2\otimes\mathbf{n}_1)+D_{22}\mathbf{n}_2\otimes\mathbf{n}_2\\&+2C_{44}(\mathbf{n}_4\otimes\mathbf{n}_4+\mathbf{n}_5\otimes\mathbf{n}_5)+(C_{11}-C_{12})\mathbf{n}_3\otimes\mathbf{n}_3\\&+2C_{66}\mathbf{n}_6\otimes\mathbf{n}_6,\end{aligned} \tag{22}$$

where D_{11}, D_{22}, and D_{12} have been defined in Eq. (15).

IV. BOUNDS FOR TETRAGONAL MATERIALS

For an elastodynamic wave propagating along an arbitrary direction in a tetragonal material a cubic equation must be solved numerically to find the waves speeds. In what follows we will first provide upper and lower bounds, then develop the first-order approximation for the wave speeds, in terms of the tetragonality index γ and the propagation direction.

A. Decomposition of *A*

That the eigenvalues and eigenvectors of an acoustic tensor of a hexagonal material can be solved analytically motivates the decomposition of the acoustic tensor of a tetragonal material into a hexagonal counterpart and an anisotropic modification. Following Fedorov (Fedorov, 1968), we write

$$\mathbf{A}=\mathbf{A}^H+\mathbf{A}^a, \tag{23}$$

where

$$\mathbf{A}^H(\mathbf{n}) \equiv D_{11}\mathbf{n}_1\otimes\mathbf{n}_1 + D_{12}(\mathbf{n}_1\otimes\mathbf{n}_2 + \mathbf{n}_2\otimes\mathbf{n}_1) + D_{22}\mathbf{n}_2\otimes\mathbf{n}_2 + (C_{11}-C_{12})(\mathbf{n}_3\otimes\mathbf{n}_3 + \mathbf{n}_6\otimes\mathbf{n}_6) + 2C_{44}(\mathbf{n}_4\otimes\mathbf{n}_4 + \mathbf{n}_5\otimes\mathbf{n}_6) \tag{24}$$

and

$$\mathbf{A}^a \equiv (C_{11}-C_{12})(\gamma-1)\mathbf{n}_6\otimes\mathbf{n}_6, \tag{25}$$

where we have introduced the tetragonality index $\gamma \equiv 2C_{66}/(C_{11}-C_{12})$. Tetragonal symmetry degenerates to hexagonal symmetry when $\gamma=1$. Hence, $\gamma-1$ measures the deviation of a tetragonal material from hexagonal symmetry.

Let us refer to $\mathbf{A}^H$ as the hexagonal counterpart of $\mathbf{A}$. As such, $\mathbf{A}^a$ can be viewed as the modification to $\mathbf{A}^H$ due to the deviation of the tetragonal material from hexagonal symmetry. The modification tensor $\mathbf{A}^a$ vanishes in the hexagonal limit. Fedorov (1968) presented a technique for determining equivalent elastic constants of a hexagonal material to best approximate the acoustic behavior of the crystal. The elastic constants of the hexagonal material are chosen to minimize the difference between the wave speeds of the approximate hexagonal and the original tetragonal materials over all propagation directions.

B. Hexagonal symmetry (transverse isotropy)

In materials with hexagonal symmetry, there always exists at least one pure-shear wave, along an arbitrary propagation direction, where the polarization direction is orthogonal to both the propagation direction and the axis of rotational symmetry of the material. Consequently, the cubic equations for wave speeds factors into a linear and a quadratic algebraic equations and hence can be solved analytically (Auld, 1990). The eigenvalue corresponding to the pure-shear wave with the polarization direction of $\mathbf{u}_3 = n_2\mathbf{e}_1 - n_1\mathbf{e}_2$ is

$$\lambda_3^H = C_{44}n_3^2 + \tfrac{1}{2}(C_{11}-C_{12})(1-n_3^2) = C_{44}\cos^2\theta + \tfrac{1}{2}(C_{11}-C_{12})\sin^2\theta, \tag{26}$$

where the superscript H is used to indicate hexagonal symmetry, and θ measures the angle between the propagation direction and the $\mathbf{e}_3$ axis, i.e., $\cos\theta \equiv n_3$. The eigenvalues of the so-called ''quasi-longitudinal'' and ''quasi-shear'' waves are

$$\lambda_{1,2}^H = \tfrac{1}{2}(C_{44} + C_{11}\sin^2\theta + C_{33}\cos^2\theta \pm \Delta), \tag{27}$$

where

$$\Delta^2 \equiv \{(C_{11}-C_{44})\sin^2\theta + (C_{44}-C_{33})\cos^2\theta\}^2 + (C_{13}+C_{44})^2\sin^2 2\theta. \tag{28}$$

The polarization directions for the two quasi-waves are $\mathbf{u}_i = \cos\theta(n_1\mathbf{e}_1 + n_2\mathbf{e}_2) + m_i\mathbf{e}_3$, $i=1,2$ where

$$m_{1,2} = \frac{C_{33}\cos^2\theta - C_{11}\sin^2\theta - C_{44}\cos 2\theta \pm \Delta}{2(C_{13}+C_{44})}. \tag{29}$$

C. Bounds

The eigenvalues λ_i of $\mathbf{A}$ are closely related to those of $\mathbf{A}^H$ given in Eqs. (26) and (27). From the minimax principle of eigenvalue problem (Wilkinson, 1965), we have

$$\mu_3 \leqslant \lambda_i - \lambda_i^H \leqslant \mu_1, \quad i=1,2,3, \tag{30}$$

where μ_1 and μ_3 are the largest and smallest eigenvalues of $\mathbf{A}^a$, respectively. Hence, the differences between the eigenvalues of the acoustic tensor $\mathbf{A}$ and those of its hexagonal counterpart $\mathbf{A}^H$ are bounded by the smallest eigenvalue of the anisotropic modification $\mathbf{A}^a$ from below and the largest from above. The eigenvalues of the rank-one tensor $\mathbf{A}^a$ are

$$\langle\mu_i\rangle = \tfrac{1}{2}(C_{11}-C_{12})(\gamma-1)\sin^2\theta\langle 1,0,0\rangle, \tag{31}$$

where $\mathbf{n}_6\cdot\mathbf{n}_6 = (n_1^2+n_2^2)/2 = \sin^2\theta/2$ has been noted. Recall that $\varrho c_i^2 = \lambda_i$. Thus for a given propagation direction the differences between the squared wave speeds in a tetragonal material and those in its hexagonal counterpart are within $\frac{1}{2}(C_{11}-C_{12})|\gamma-1|\sin^2\theta/\varrho$.

Equation (30) implies upper and lower bounds on λ_i of

$$\lambda_i^H + \mu_3 \leqslant \lambda_i \leqslant \lambda_i^H + \mu_1. \tag{32}$$

Substitution of Eq. (31) into (32) yields, for a tetragonal material with $\gamma \geqslant 1$ $(C_{11} \leqslant C_{12}+2C_{66})$,

$$\lambda_i^H \leqslant \lambda_i \leqslant \lambda_i^H + \tfrac{1}{2}(C_{11}-C_{12})(\gamma-1)\sin^2\theta, \tag{33}$$

and for a tetragonal material with $\gamma \leqslant 1$,

$$\lambda_i^H - \tfrac{1}{2}(C_{11}-C_{12})(1-\gamma)\sin^2\theta \leqslant \lambda_i \leqslant \lambda_i^H. \tag{34}$$

Since the eigenvalues of the acoustic tensor are related to the waves speeds, we have derived upper and lower bounds on the wave speeds in terms of the elastic constants of the tetragonal material and the propagation direction. An immediate corollary result from inequalities (33) and (34) is that waves in a tetragonal material with $\gamma \geqslant 1$ $(\gamma \leqslant 1)$ cannot be slower (faster) than those of its hexagonal counterpart.

The bounds on the quasi-shear wave speeds are particularly simple. Substitution of Eq. (26) into (33) and (34) yields the bounds on the speed of the quasi-shear wave in a tetragonal material as

$$\begin{aligned} &C_{44}\cos^2\theta + \tfrac{1}{2}(C_{11}-C_{12})\sin^2\theta \\ &\quad \leqslant \varrho c_s^2 \leqslant C_{44}\cos^2\theta + C_{66}\sin^2\theta, \\ &\qquad \text{for } C_{11} \leqslant C_{12}+2C_{66}, \\ &C_{44}\cos^2\theta + C_{66}\sin^2\theta \\ &\quad \leqslant \varrho c_s^2 \leqslant C_{44}\cos^2\theta + \tfrac{1}{2}(C_{11}-C_{12})\sin^2\theta \\ &\qquad \text{for } C_{11} \geqslant C_{12}+2C_{66}. \end{aligned} \tag{35}$$

For a given material, the bounds given in Eqs. (33) and (34) are loosest when the propagation direction is in the $\mathbf{e}_1 - \mathbf{e}_2$ plane $(\theta=\pi/2)$, and become tighter as the propagation direction moves towards the $\mathbf{e}_3$ axis. Indeed, when the direction is along $\mathbf{e}_3$ $(\theta=0)$, $\mu_1=\mu_3=0$, i.e., the lower and upper bounds on the eigenvalues coincide and yield the exact results $\lambda_i = \lambda_i^H$. Hence, for the plane waves propagating along $\mathbf{e}_3$ in a tetragonal material, the propagation speeds are the same as their hexagonal counterparts, a well-known result.

Since $0 \leqslant \sin^2\theta \leqslant 1$, looser, but more general, bounds than the ones given in Eqs. (33) and (34) can be established for all propagation directions

$$\begin{aligned} &\lambda_i^H \leqslant \lambda_i \leqslant \lambda_i^H + (C_{66} - \tfrac{1}{2}(C_{11} - C_{12})), \\ &\qquad \text{for } C_{11} \leqslant C_{12} + 2C_{66}, \\ &\lambda_i^H - (\tfrac{1}{2}(C_{11} - C_{12}) - C_{66}) \leqslant \lambda_i \leqslant \lambda^H, \\ &\qquad \text{for } C_{11} \geqslant C_{12} + 2C_{66}. \end{aligned} \tag{36}$$

Therefore, the eigenvalues of the acoustic tensor of a tetragonal material can differ by no more than $|\frac{1}{2}(C_{11} - C_{12}) - C_{66}|$ from their hexagonal counterparts. In particular, corresponding to the quasi-shear wave, Eq. (36) implies that

$$\begin{aligned} &\text{Min}\{C_{44}, C_{66}, \tfrac{1}{2}(C_{11} - C_{12})\} \leqslant \varrho c_s^2 \\ &\qquad \leqslant \text{Max}\{C_{44}, C_{66}, \tfrac{1}{2}(C_{11} - C_{12})\}, \end{aligned} \tag{37}$$

a result to be expected based on tetragonal symmetry of the material.

V. APPROXIMATIONS OF THE WAVE SPEEDS

It follows from Eq. (23) that the acoustic tensor **A** of a tetragonal material depends analytically on the tetragonality index γ. Hence its eigenvalues are also analytic functions of γ. For values of γ close to one, then, by Rellich's theorem on Taylor expansions of eigenvalues of perturbed linear operators (Kato, 1984), the first-order approximations of the eigenvalues of **A**, in terms of the γ, are

$$\lambda_i(\gamma) \approx \overline{\lambda}_i(\gamma) = \lambda_i(1) + (\gamma - 1)\lambda_i'(1), \quad i = 1,2,3, \tag{38}$$

where $\lambda_i(1)$ and $\lambda_i'(1)$ are the eigenvalues and their derivatives, respectively, of **A** evaluated at $\gamma = 1$. At $\gamma = 1$ the acoustic tensor degenerates to that of a hexagonal material. Thus $\lambda_i(1) = \lambda_i^H$, given in Eqs. (26) and (27).

The derivatives of the eigenvalues are (Fox and Kapoor, 1968)

$$\lambda_i'(1) = \frac{\mathbf{u}_i \cdot \mathbf{A}' \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i}, \quad i = 1,2,3, \tag{39}$$

where $\mathbf{u}_i$ is an eigenvector of $\mathbf{A}^H$. The tensor $\mathbf{A}'$ is the derivative of **A** with respect to the parameter γ, which can be found by differentiating Eq. (23)

$$\mathbf{A}' = (C_{11} - C_{12})\mathbf{n}_6 \otimes \mathbf{n}_6. \tag{40}$$

Since $\mathbf{A}'$ is a positive semi-definite tensor with the nonzero eigenvalue equal to $\frac{1}{2}(C_{11} - C_{12})\sin^2\theta$, application of Rayleigh's principle to Eq. (39) results

$$0 \leqslant \lambda_i'(1) \leqslant \tfrac{1}{2}(C_{11} - C_{12})\sin^2\theta. \tag{41}$$

The significance of Eq. (41) is that it gives the limits on how fast the eigenvalues of acoustic tensor can change with the deviation of a tetragonal material from hexagonal symmetry. An immediate implication of Eq. (41) is that $\lambda_i'(1) = 0$ for $\theta = 0$, consequently, Eq. (38) predicts $\lambda_i(\gamma) = \lambda_i(1)$, as expected since $\mathbf{A} = \mathbf{A}^H$. For all values of θ, it follows from Eqs. (38) and (41) that

$$0 \leqslant |\overline{\lambda}_i(\gamma) - \lambda_i(1)| \leqslant \tfrac{1}{2}(C_{11} - C_{12})|\gamma - 1|\sin^2\theta. \tag{42}$$

Therefore, the linear approximations in (38) fall within the upper and lower bounds on the eigenvalues given in Eqs. (33) and (34).

Substituting the expressions for λ_i^H and Eq. (39) into (38) yields the linear approximations of the elastic wave speeds along an arbitrary direction in a tetragonal material. For the quasi-shear wave,

$$\begin{aligned} \varrho c_s^2 = \lambda_3(\gamma) \approx{}& C_{44}\cos^2\theta + \tfrac{1}{2}(C_{11} - C_{12})\sin^2\theta \\ &+ \tfrac{1}{2}(C_{11} - C_{12})(\gamma - 1)\frac{(n_2^2 - n_1^2)^2}{\sin^2\theta}, \end{aligned} \tag{43}$$

and for the other quasi-shear and the quasi-longitudinal waves,

$$\begin{aligned} \varrho c_1^2 = \lambda_1(\gamma) \approx{}& \tfrac{1}{2}(C_{44} + C_{11}\sin^2\theta + C_{33}\cos^2\theta + \Delta) \\ &+ 2(C_{11} - C_{12})(\gamma - 1)\frac{\cos^2\theta}{\cos^2\theta \sin^2\theta + m_1^2} n_1^2 n_2^2, \end{aligned} \tag{44}$$

$$\begin{aligned} \varrho c_2^2 = \lambda_2(\gamma) \approx{}& \tfrac{1}{2}(C_{44} + C_{11}\sin^2\theta + C_{33}\cos^2\theta - \Delta) \\ &+ 2(C_{11} - C_{12})(\gamma - 1)\frac{\cos^2\theta}{\cos^2\theta \sin^2\theta + m_2^2} n_1^2 n_2^2, \end{aligned} \tag{45}$$

where Δ and m_1, m_2 have been defined in Eqs. (28) and (29).

VI. NUMERICAL EXAMPLE AND CONCLUSIONS

As an example, consider a tin crystal with elastic constants of $C_{11} = 8.39$, $C_{33} = 9.67$, $C_{12} = 4.87$, $C_{13} = 2.81$, $C_{44} = 1.75$, $C_{66} = 0.74$ (Sutcliffe, 1992). The unit of the constants is 10^{11} dyn/cm^2. With the anisotropic index $\gamma = 2C_{66}/(C_{11}$ $\gamma = 2C_{66}/(C_{11} - C_{12}) = 0.42 < 1$, the second equation of Eq. (35) applies, and gives the bounds on the speed of the quasi shear wave as

$$\sqrt{\cos^2\theta + 0.42\sin^2\theta} \leqslant \frac{c_s}{c_0} \leqslant 1, \tag{46}$$

where $c_0 = \sqrt{C_{44}/\varrho}$, the speed of a pure-shear wave along $\mathbf{e}_3$ axis. The particularly simple expression of the upper bound is a result of the coincidence that $2C_{44} \approx C_{11} - C_{12}$ for this material. The inequality (46) becomes an equality when $\theta = 0$ and yields the exact result. On the other extreme, when $\theta = \pi/2$, the lower bound reaches the minimum of $C_{66}/C_{44} = 0.65$. The corresponding weaker bounds independent of the propagation direction are

$$0.65 \leqslant \frac{c_s}{c_0} \leqslant 1. \tag{47}$$

The lower bound corresponds to the speed of a pure-shear wave when it propagates along the $\mathbf{e}_1$ (or $\mathbf{e}_2$) axis. Since both the upper and lower bounds are attainable, the proposed

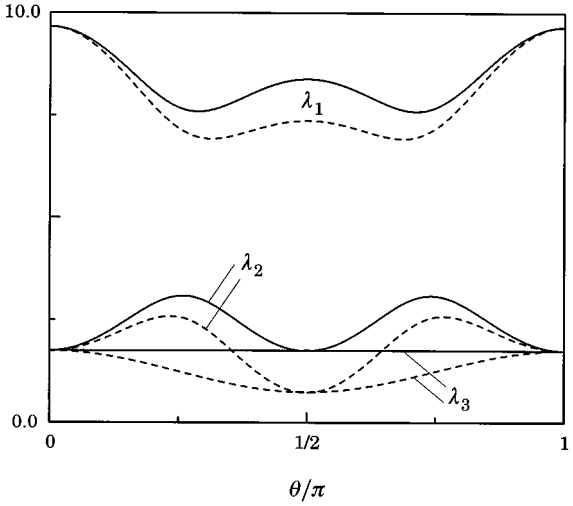


FIG. 1. Bounds on the eigenvalues as functions of the propagation direction.

bounds are actually the tightest bounds on the quasi-shear wave speed.

Figure 1 plots the bounds on the three eigenvalues given by Eq. (34) for all propagation directions. The range of the bounds changes with the direction; it starts at zero when $\theta = 0$ and reaches the maximum at $\theta = \pi/2$. The curves are symmetric about $\theta = \pi/2$, reflecting the symmetry of the material. It is interesting to observe that the bounds depend on the propagation direction only through θ, the angle it makes with the $\mathbf{e}_3$ axis.

The angles between the propagation direction and the $\mathbf{e}_1$ and $\mathbf{e}_2$ axes influence the linear approximations of the eigenvalues. Let α denote the angle between the propagation direction and the $\mathbf{e}_1$ axis. Then for a given value of θ the eigenvalues of the acoustic tensor are functions of the angle α only. Figure 2 gives a comparison of such functions for the approximations and the exact values from numerically solving the eigenvalue problem. In Fig. 2, $\theta = \pi/4$ has been chosen, consequently, the angle α ranges from $\pi/4$ to $3\pi/4$. It can be seen that the first-order linear approximations compare very well with the exact values obtained by numerically solving the eigenvalue problem. Figure 2 also shows the lower and upper bounds on the eigenvalues, which do not change with α. Observe that both the approximations and the exact solutions fall inside the bounds.

We have examined the differences between the approximated and the numerical eigenvalues for all values of θ and α. The maximum difference is found to be in λ_3 at $\theta = 70°$ and $\alpha = 148°$, where the approximated and the exact values are 1.39 and 1.34, respectively.

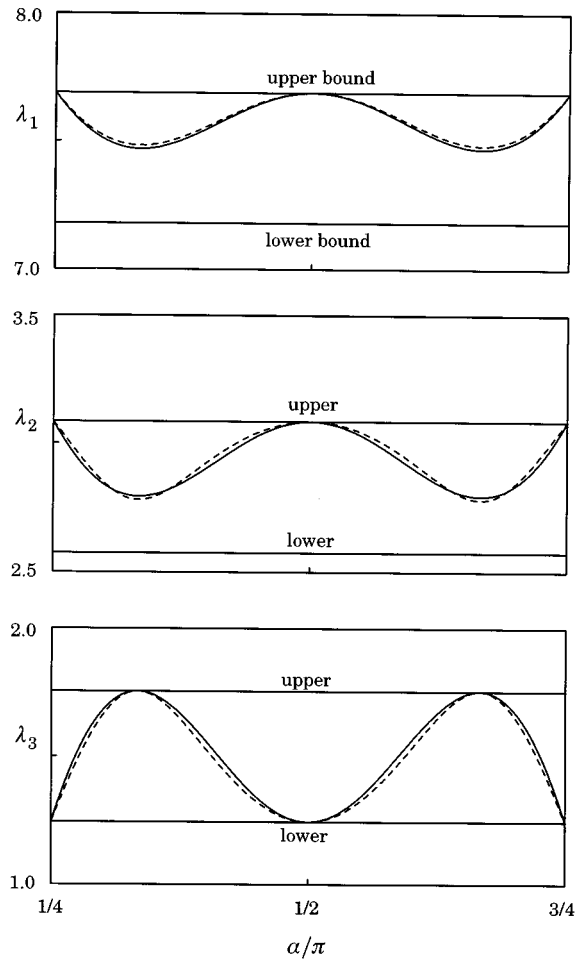


FIG. 2. Eigenvalues as functions of the propagation angle α for $\theta = \pi/4$.

ACKNOWLEDGMENTS

This paper was prepared from a study conducted in the Center of Excellence for Airport Pavement Research which is funded in part by the Federal Aviation Administration under Research Grant Number 95-C-001. Technical discussions with F. L. Addessio, P. J. Maudlin of the Los Alamos National Laboratory and H. L. Schreyer of the University of New Mexico were very helpful and hence greatly appreciated.

Al'shits, V. I., and Lothe, J. (**1979**). ''Elastic waves in triclinic crystals I, II, III,'' Sov. Phys. Crystallogr. **24**, 387–392, 393–398, 644–648.

Auld, B. A. (**1990**). *Acoustic Fields and Waves in Solids* (Krieger, Florida), 2nd ed., Vol. 1, Chap. 7, pp. 191–264.

Ditri, J. J. (**1997**). ''Bounds on phase velocity changes of plane waves in elastic media due to changes in elastic moduli,'' Int. J. Solids Struct. **34**, 2105–2114.

Fedorov, F. I. (**1968**). *Theory of Elastic Waves in Crystals* (in Russian, Nauka, Moscow, 1965; translated from Russian to English by J. E. S. Bradley (Plenum, New York).

Fox, R. L., and Kapoor, M. P. (**1968**). ''Rates of change of eigenvalues and eigenvectors,'' AIAA J. **6**, 2426–2429.

Hertzberg, R. W. (**1987**). *Deformation and Fracture Mechanics of Engineering Materials* (Wiley, New York), 3rd ed., Chap. 1, pp. 3–47.

Kato, T. (**1984**). *Perturbation Theory for Linear Operators* (Springer-Verlag, New York).

Kolodner, I. I. (**1966**). ''Existence of longitudinal waves in anisotropic media,'' J. Acoust. Soc. Am. **40**, 730–731.

Musgrave, M. J. P. (**1970**). *Crystal Acoustics* (Holden-Day, California).

Neilsen, M. K., and Schreyer, H. L. (**1993**). ''Bifurcations in elastic-plastic materials,'' Int. J. Solids Struct. **30**, 521–544.

Pao, Y.-H., and Kaul, R. K. (**1974**). ''Waves and vibrations in isotropic and anisotropic plates,'' in *R. D. Mindlin and Applied Mechanics*, edited by G. Herrmann (Pergamon, New York), pp. 149–195.

Schreyer, H. L., and Zuo, Q. H. (**1995**). ''Anisotropic yield surfaces based on elastic projection operators,'' J. Appl. Mech. **62**, 780–785.

Stippes, M. (**1965**). ''Steady state waves in anisotropic media,'' Annual Meeting of Society of Engineering Sciences.

Shuvalov, A. L., and Every, A. G. (**1997**). ''Shape of the acoustic slowness surface of anisotropic solids near points of conical degeneracy,'' J. Acoust. Soc. Am. **101**, 2381–2383.

Sutcliffe, S. (**1992**). ''Spectral Decomposition of the Elasticity Tensor,'' J. Appl. Mech. **59**, 762–773.

Truesdell, C. (**1966**). ''Existence of longitudinal waves,'' J. Acoust. Soc. Am. **40**, 729–730.

Wilkinson, J. H. (**1965**). *The Algebraic Eigenvalue Problem* (Clarendon, Oxford).

Zuo, Q. H., and Hjelmstad, K. D. (**1997**). ''Bounds and approximations for elastic wave speeds in cubic crystals,'' J. Acoust. Soc. Am. **101**, 3415–3420.

Zuo, Q. H., and Schreyer, H. L. (**1995**). ''A note on pure-longitudinal and pure-shear waves in cubic crystals,'' J. Acoust. Soc. Am. **98**, 580–583.

Moving frame technique for planar acoustic holography

Hyu-Sang Kwon and Yang-Hann Kim[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 24 October 1996; accepted for publication 22 December 1997)

Acoustic holography is one of the best methods to visualize sound fields. The quality of the visualized sound is primarily determined by the size of the hologram, its microphone spacing, and the number of microphones. This paper describes a way to virtually increase the hologram size and the spatial resolution of the holograph. For a stationary sound field, the method continuously sweeps the sound field by a line array of microphones. For moving sound sources, radiating sound is measured by using a line array of microphones fixed in space. In both cases, the measured signals have Doppler effects. The theoretical formulation has been systematically addressed by employing a moving coordinate which has relative motion between the measurement coordinate and the hologram coordinate. Simulations and experiments support the proposed theory. The drawback is that the method is only applicable to discrete frequencies. © *1998 Acoustical Society of America.* [S0001-4966(98)01004-2]

PACS numbers: 43.20.Fn, 43.35.Sx [ANN]

INTRODUCTION

The beauty of the acoustic holograph method[1] is that one can construct an entire sound field by only measuring sound pressure on a hologram plane. This enables one to see every detail of the sound field of interest. These include not only the pressure distribution but also all other acoustic variables such as velocity, kinetic/potential energy, and intensity distribution over any surface of interest. However, limitations also exist. These are mainly due to the finite size of the hologram plane and microphone spacing; for details, see Refs. 1 and 2. The former can be regarded as a two dimensional window effect on the hologram, and the latter is associated with spatial sampling. These limitations cannot be completely avoided.

In order to get a better holograph, one simply needs a larger hologram size and smaller microphone spacing. For stationary sound fields, one can increase the size of the hologram as well as the microphone spacing by simply scanning an array microphone over the hologram plane. An efficient method of this kind was introduced by Hald.[3] It uses a reference microphone or microphones, depending on the number of coherent sources, and a line array of microphone which scans over the hologram plane. This method works only for a non-moving sound source which is emitting a stationary sound field. The sound field must be stationary for each measurement set. The scanning in this case cannot be performed continuously. This is done at discrete positions which one could call ''step-by-step'' scanning.

This paper addresses a scanning method which is similar to that of the original acoustic holograph[1] and the step-by-step scanning technique[3,4] but generalizes them so that one can also use it for the visualization of a sound field induced by moving noise sources. For a stationary sound field, the method continuously scans over a desired measurement plane. The line array microphone travels over the measurement plane at a constant speed. For moving sources, the method measures sound pressure by a line array which is fixed in space. In both of these cases there is relative velocity between the array and sources. This velocity will introduce a Doppler shift to the measured sound pressure signals.

This paper addresses a procedure to handle this well known effect on acoustic holograms. The moving frame technique is suggested for this purpose. This essentially takes into consideration three different coordinate systems; the absolute coordinate, the hologram coordinate, and the microphone coordinate.

I. THEORETICAL BACKGROUND OF MOVING FRAME TECHNIQUE

A general configuration of hologram measurement by the moving frame technique is depicted in Fig. 1. Three co-

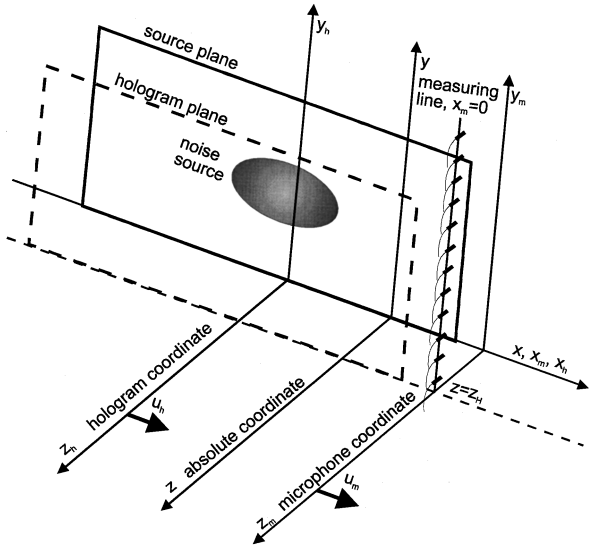


FIG. 1. Coordinate systems (absolute coordinate, microphone coordinate, and hologram coordinate) and their relative motions which are determined by the locations of source and microphone array.

[a)]Electronic mail: yhkim@sorak.kaist.ac.kr

 0001-4966/98/103(4)/1734/8/$10.00

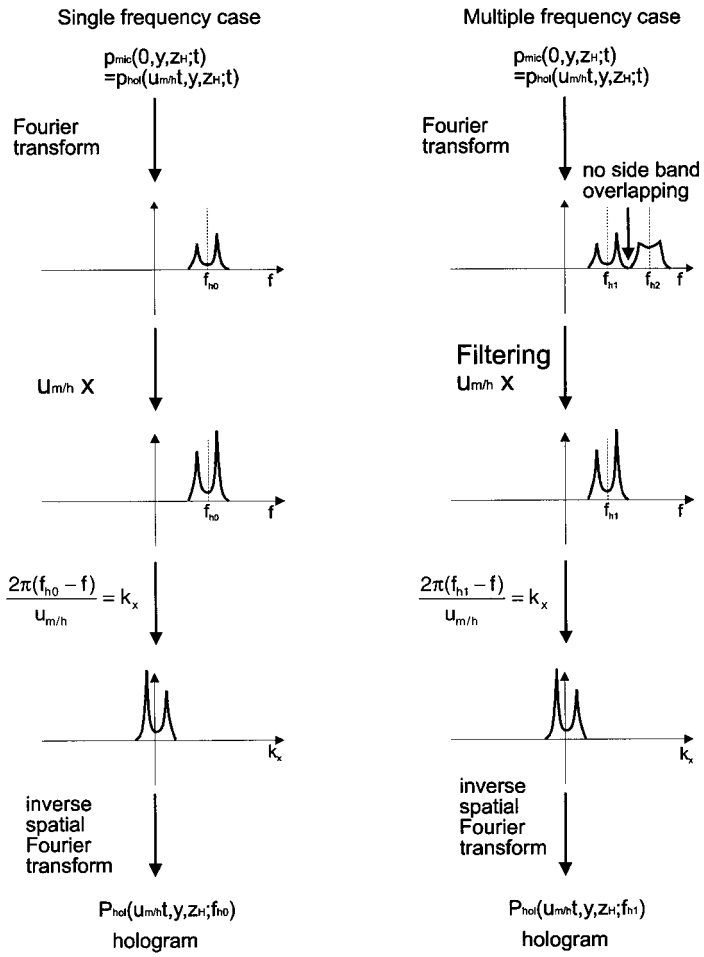


FIG. 2. Procedure to calculate the hologram from the measured pressure signal in moving frame technique.

ordinates are employed in this measurement system. One is the ''absolute coordinate'' which is fixed in space, the other one is the ''microphone coordinate'' which is allowed to move in space, and the last one is the ''hologram coordinate.'' The hologram coordinate must move with the same speed as the sound sources.

We denote these coordinate systems as

(x,y,z): absolute coordinate system,
(x_m,y_m,z_m): microphone coordinate system,
(x_h,y_h,z_h): hologram coordinate system.

Sound pressure on each coordinate system can be written as $p(x,y,z;t)$, $p_{\mathrm{mic}}(x_m,y_m,z_m;t)$, and $p_{\mathrm{hol}}(x_h,y_h,z_h;t)$. The subscript mic and hol express microphone and hologram coordinate system respectively. We assume that all coordinates are in parallel and that their origins are at the same location at $t=0$. Then, at arbitrary time t,

$$y=y_m=y_h, \tag{1a}$$

$$z=z_m=z_h, \tag{1b}$$

$$x=u_m t+x_m, \tag{1c}$$

$$x=u_h t+x_h. \tag{1d}$$

The relative velocity between the microphone coordinate and the hologram coordinate; $u_{m/h}$, is

$$u_{m/h}=u_m-u_h. \tag{2}$$

Then, Eqs. (1c) and (1d) give the relation between x_h and x_m. That is

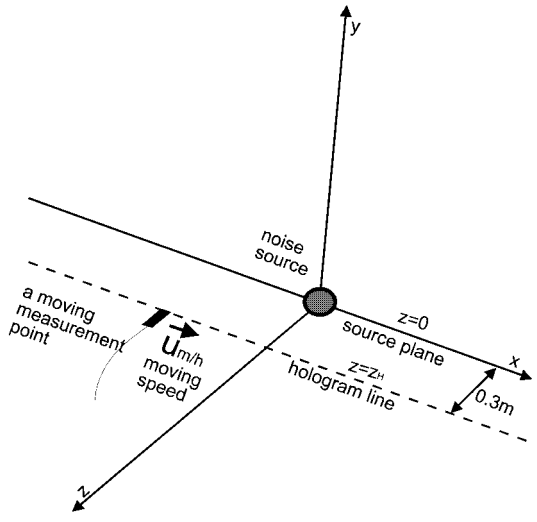


FIG. 3. Simulation configuration; microphone moves in the x direction, 0.3 m away from the source plane.

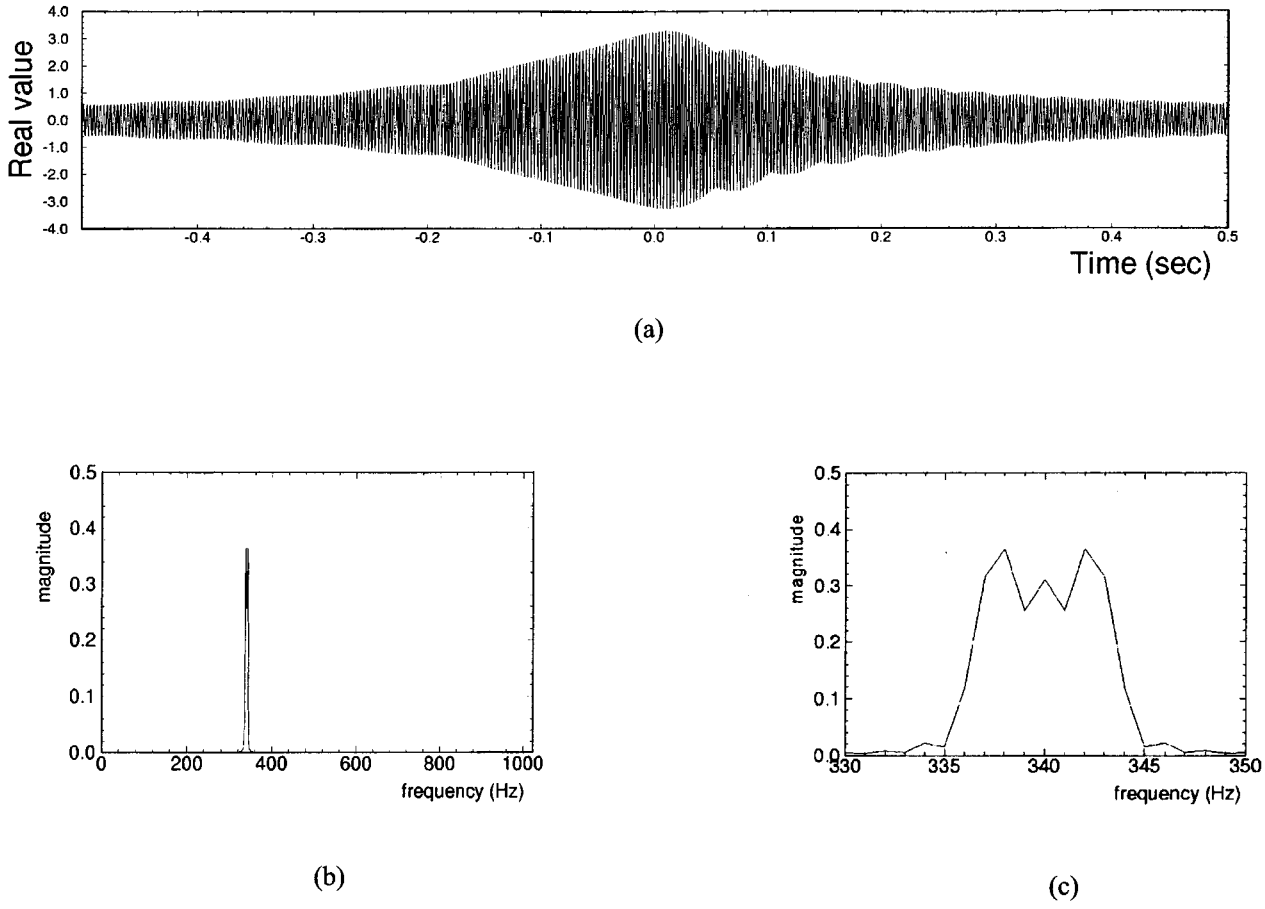


FIG. 4. Time signal and its spectrum obtained by FFT in a computer simulation; a monopole of 340 Hz located at (0 m, 0 m, 0 m) was used. (a) Time signal of a moving microphone. (b) Magnitude of the Fourier transformed signal. (c) Details of Fourier transformed signal around 340 Hz.

$$x_h = x_m + u_{m/h} t. \tag{3}$$

If one assumes that the microphone array is rigidly attached to the microphone coordinate as shown in Fig. 1, then the pressures measured by the microphone and hologram coordinate will be

$$p_{\mathrm{mic}}(x_m, y_m, z_H; t) = p_{\mathrm{hol}}(x_h, y_h, z_H; t) = p_{\mathrm{hol}}(x_m + u_{m/h} t, y_h, z_H; t). \tag{4}$$

Note that one can always make x_m zero in most cases, therefore Eq. (4) can be simplified to

$$p_{\mathrm{mic}}(0, y_m, z_H; t) = p_{\mathrm{hol}}(u_{m/h} t, y_h, z_H; t). \tag{5}$$

This equation essentially allows us to construct acoustic pressure on the hologram plane.

A hologram expresses a spatial distribution of pressure with respect to frequency; that is, the temporal Fourier transform of $p_{\mathrm{hol}}(x_h, y_h, z_H; t)$. This cannot be obtained by taking the temporal Fourier transform of Eq. (5). Equation (5) provides the relationship between measured pressure signals with regard to the microphone and hologram coordinates. It is noteworthy that the first argument of pressure on the hologram is time dependent. In other words, the measured signal p_{mic} is Doppler shifted. This effect can be easily dealt with if sound pressure propagates in one direction only. However, it will not be straightforward if the pressure field of interest is more than two dimensional. For example, directivity pattern of sources could be dipole or quadrupole even if these have the same temporal frequency.

Progress comes from the realization that the wave number spectrum could express this effect in a more explicit form. This is simply because the wave number transform does not explicitly have anything to do with time; it has to do with space. Mathematical derivations associated with this idea are presented in the following.

First one could write,

$$p_{\mathrm{hol}}(u_{m/h} t, y_h, z_H; t) = F_T^{-1}\{P_{\mathrm{hol}}(u_{m/h} t, y_h, z_H; f_h)\} = \int_{-\infty}^{\infty} P_{\mathrm{hol}}(u_{m/h} t, y_h, z_H; f_h) e^{-j2\pi f_h t}\, df_h. \tag{6}$$

Then the temporal Fourier transform of Eq. (6), in other words, the hologram is

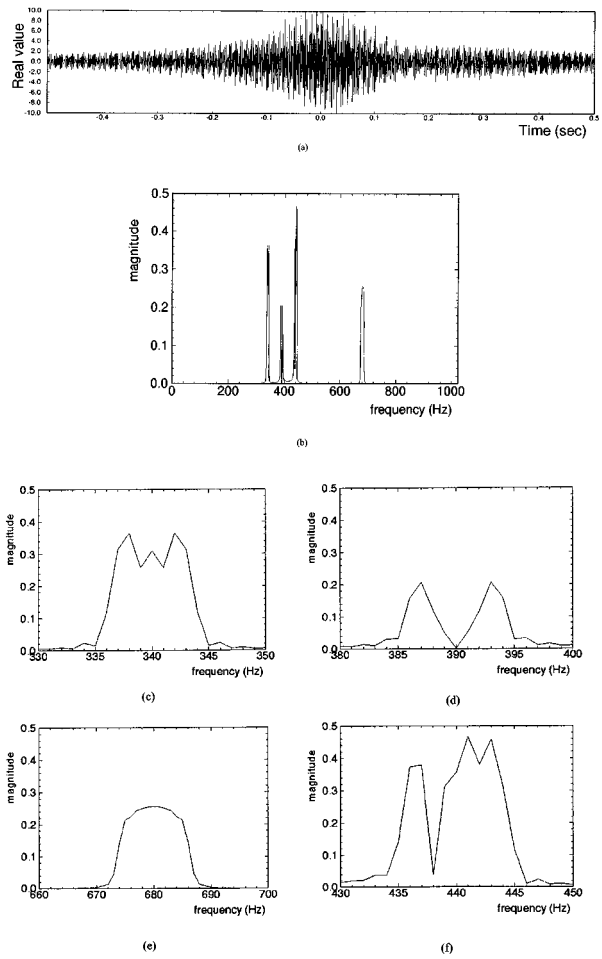


FIG. 5. Time signal and its spectrum obtained by FFT in a computer simulation; a monopole of 340 Hz located at (0 m, 0 m, 0 m), an x-polarized dipole of 390 Hz located at (0 m, 0 m, 0 m), a z-polarized dipole of 680 Hz located at (0 m, 0 m, 0 m), and two monopoles of 440 Hz located at (0 m, 0 m, 0 m) and (0.5 m, 0.5 m, 0 m) were used. (a) Time signal of a moving microphone, (b) magnitude of the Fourier transformed signal. (c) Details of Fourier transformed signal around 340 Hz. (d) Details of Fourier transformed signal around 390 Hz. (e) Details of Fourier transformed signal around 680 Hz. (f) Details of Fourier transformed signal around 440 Hz.

$$F_T\{p_{\mathrm{hol}}(u_{m/h}t,y_h,z_H;t)\}$$
$$=\int_{-\infty}^{\infty} p_{\mathrm{hol}}(u_{m/h}t,y_h,z_H;t)e^{j2\pi ft}\,dt$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_{\mathrm{hol}}(u_{m/h}t,y_h,z_H;f_h)e^{-j2\pi f_h t}\,df_h e^{j2\pi ft}\,dt. \tag{7}$$

Note that P_{hol} in Eq. (7) expresses the spatial distribution of sound pressure of frequency, f_h. This can be also obtained from the wave number spectrum, that is,

$$P_{\mathrm{hol}}(x_h=u_{m/h}t,y_h,z_H;f_h)$$
$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{P}_{\mathrm{hol}}(k_x,y_h,z_H;f_h)e^{jk_x x_h}\,dk_x. \tag{8}$$

Equations (7) and (8) lead one to

$$F_T\{p_{\mathrm{hol}}(u_{m/h}t,y_h,z_H;t)\}$$
$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\hat{P}_{\mathrm{hol}}(k_x,y_h,z_H;f_h)e^{jk_x u_{m/h}t}$$
$$\times e^{-j2\pi f_h t}e^{j2\pi ft}\,dt\,dk_x\,df_h$$
$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\hat{P}_{\mathrm{hol}}(k_x,y_h,z_H;f_h)\int_{-\infty}^{\infty}e^{jk_x u_{m/h}t}$$
$$e^{-j2\pi f_h t}e^{j2\pi ft}\,dt\,dk_x\,df_h$$
$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\hat{P}_{\mathrm{hol}}(k_x,y_h,z_H;f_h)$$
$$\times\delta\left(\frac{k_x u_{m/h}}{2\pi}-(f_h-f)\right)dk_x\,df_h, \tag{9}$$

where $\delta(\cdot)$ is the delta function.

Equation (9) can be simplified by using the property of the delta function, that is

$$F_T\{p_{\mathrm{hol}}(u_{m/h}t,y_h,z_H;t)\}$$
$$=\frac{1}{u_{m/h}}\int_{-\infty}^{\infty}\hat{P}_{\mathrm{hol}}\left(\frac{2\pi(f_h-f)}{u_{m/h}},y_h,z_H;f_h\right)df_h. \tag{10}$$

This equation forms the basis of the proposed ''moving frame technique'' method. The left-hand side of Eq. (10) is the temporal Fourier transform of sound pressure with respect to the microphone coordinate, or measured signals [Eq. (5)]. The integrand of the right-hand side of the equation is the Doppler shifted wave number spectrum in the x direction. The inverse Fourier transform of the wave number spectrum is what we want. This is the sound pressure distribution of a selected frequency on a plane of interest. Therefore, one can get a desired hologram at $z=z_H$ by using Eq. (10).

However, one must note that the equations are valid only for discrete frequency. Therefore, this method has very severe drawbacks compared with other conventional method.[1,3,4] This is what the method has to pay for greatly reducing the number of microphones. This method uses a line array, for example, N sensors, instead of $N\times N$ sensors as the conventional method does. Next, the issues associated with the discrete frequency will be addressed.

If one has a sound field generated by a discrete frequency; single frequency (f_{h0}), then Eq. (10) can be simply written as

$$F_T\{p_{\mathrm{hol}}(u_{m/h}t,y_h,z_H;t)\}/\Delta f_h$$
$$=\frac{1}{u_{m/h}}\hat{P}_{\mathrm{hol}}\left(\frac{2\pi(f_{h0}-f)}{u_{m/h}},y_h,z_H;f_{h0}\right). \tag{11}$$

Figure 2 illustrates the way in which Eq. (11) could be implemented in practice. If one has multiple, discrete frequency components, then Eq. (10) reduces to

$$F_T\{p_{\mathrm{hol}}(u_{m/h}t,y_h,z_H;t)\}/\Delta f_h$$
$$=\frac{1}{u_{m/h}}\sum_{i=1}^{N}\hat{P}_{\mathrm{hol}}\left(\frac{2\pi(f_{hi}-f)}{u_{m/h}},y_h,z_H;f_{hi}\right). \tag{12}$$

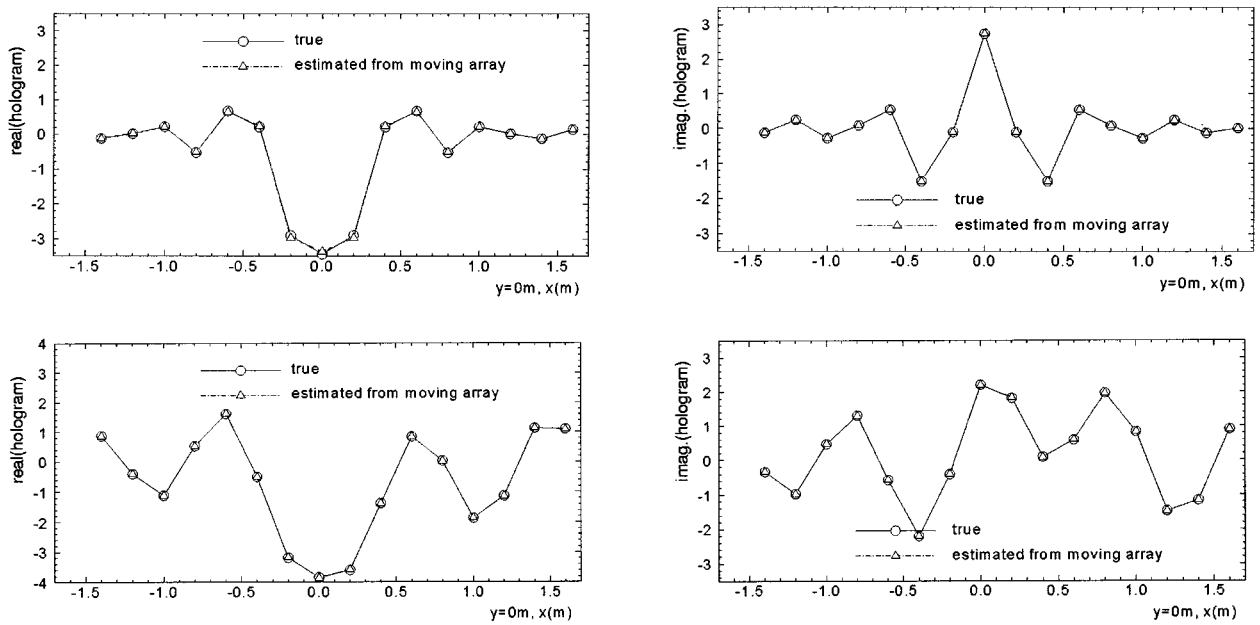


FIG. 6. Comparisons between the holograms which are estimated from the moving microphone signal by using moving frame technique and the true holograms. (a) Comparison of hologram at 680 Hz [a z-polarized dipole, at (0 m, 0 m, 0 m)]. (b) Comparison of hologram at 440 Hz [two monopoles, at (0 m, 0 m, 0 m) and (0.5 m, 0.5 m, 0.5 m)].

As illustrated in Fig. 2, it is possible to have a frequency band which is common to the discrete frequencies. In other words, side band overlapping can occur (Fig. 2). The way to avoid this undesirable, side band overlapping follows.

Let's assume that we have N frequencies whose magnitude distribution is

$$f_{h1}<f_{h2}<\cdots<f_{hi-1}<f_{hi}<f_{hi+1}<\cdots<f_{hN}.$$

Corresponding wave numbers will be

$$k_i=\frac{2\pi f_{hi}}{c},$$

where c is the speed of sound propagation.

It is well known that the wave number in the x direction, k_x, has to be less than two times the free space wave number to avoid spatial aliasing, that is

$$-2k_i\leqslant k_x\leqslant 2k_i \tag{13}$$

or,

$$-\frac{4\pi f_{hi}}{c}\leqslant k_x\leqslant\frac{4\pi f_{hi}}{c}$$

and the relation between k_x and frequencies for the moving frame system is

$$\frac{2\pi(f_{hi}-f\,)}{u_{mlh}}=k_x. \tag{14}$$

Therefore Eq. (13) can be rewritten as

$$-\frac{4\pi f_{hi}}{c}\leqslant\frac{2\pi(f_{hi}-f\,)}{u_{mlh}}\leqslant\frac{4\pi f_{hi}}{c}. \tag{15}$$

Rearranging this equation with respect to f gives

$$(1-2M)f_{hi}\leqslant f\leqslant(1+2M)f_{hi}, \tag{16}$$

where $M(=u_{m/h}/c)$ is a Mach number.

For the adjacent discrete frequency components, one can write Eq. (16) as

$$(1-2M)f_{hi-1}\leqslant f\leqslant(1+2M)f_{hi-1}$$

and (17)

$$(1-2M)f_{hi+1}\leqslant f\leqslant(1+2M)f_{hi+1}.$$

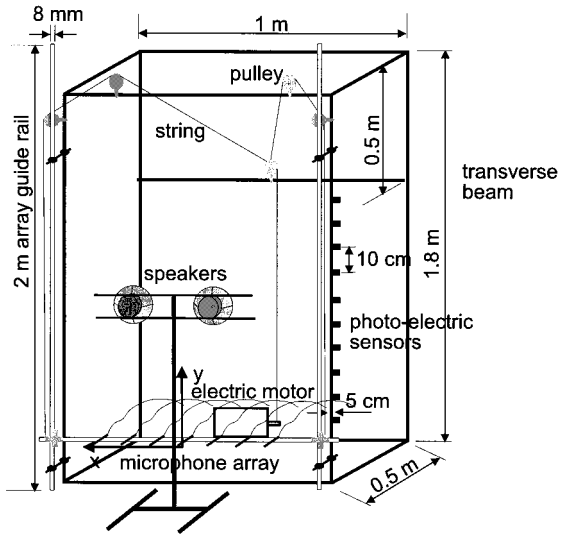


FIG. 7. Experimental setup to measure the hologram by using a microphone array which moves along vertical direction; The microphone array moves along the array guide rail with commanded speed which is controlled by electric motor. A string and pulley system provides the desired, vertical movement of the array. Ten photo-electric sensors are used to measure the velocity of the array.

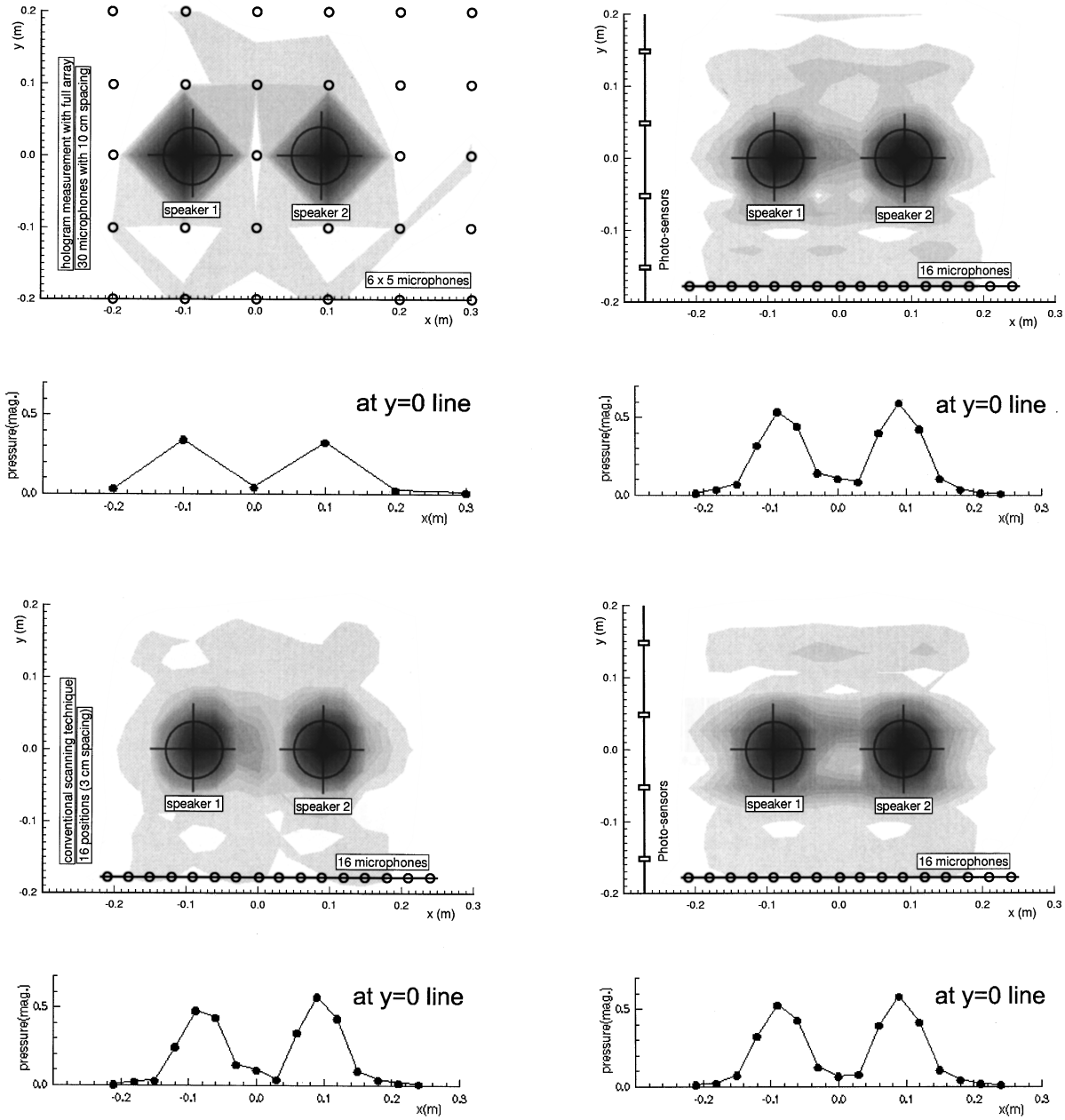


FIG. 8. Predicted results of moving array experiments by using 2 identical speakers which radiate a pure tone of 500 Hz. Comparison with conventional hologram measurement techniques; spatial distribution of pressure magnitude on source plane, including that of $y=0$ line. (a) Simultaneous measurement method, 6×5 points (spacing spacing 10 cm (x), 10 cm (y)). (b) Step-by-step scanning method, 16×16 points (sampling spacing 3 cm (x), 3 cm (y)). (c) Moving frame technique (velocity=0.204 m/s), 16×16 points (sampling spacing 3 cm (x), 2.55 cm (y)). (d) Moving frame technique (velocity=0.448 m/s), 16×16 points (sampling spacing 3 cm (x), 2.80 cm (y)).

Equations (16) and (17) essentially express how the discrete frequency component spreads out in the frequency domain due to the Doppler shift. It is now clear that the upper boundary of frequency spreading by f_{hi} must not be overlapped by the lower boundary of f_{hi+1}, and the lower boundary by f_{hi} is not also allowed to be overlapped by that due to f_{hi-1}. In other words

$$(1-2M)f_{hi+1}>(1+2M)f_{hi}$$

and (18)

$$(1-2M)f_{hi}>(1+2M)f_{hi-1}$$

or,

$$f_{hi+1}>\frac{1+2M}{1-2M}f_{hi} \quad \text{and} \quad f_{hi-1}<\frac{1-2M}{1+2M}f_{hi}.$$

One should also note that Eq. (18) determines a Mach number to be less than 0.5. This restriction comes from the condition to avoid spatial aliasing (Eq. (13)). For example, for a noise source moving with the speed of $M=0.1$, which is about 123 km/h, Eq. (18) determines

$$f_{hi+1}>1.5f_{hi} \quad \text{and} \quad f_{hi-1}<0.67f_{hi} \tag{19}$$

which is an acceptable requirement in practice.

We now summarize this section and briefly address unsolved problems which will affect the performance of the proposed method when it is applied to practical problems.

We have found the relationship between sound pressure on the hologram coordinate and the microphone coordinate [Eq. (10)]. This enables one to either sweep a microphone array over a stationary sound field or to allow sound sources to move with constant speed by keeping the array stationary in space. The desired acoustic pressure on the hologram can be obtained in both cases. A condition to avoid undesirable errors which are associated with side band overlapping due to the Doppler effect has been obtained in Eq. (18). When one has a discrete spectrum which violates the condition, or a discrete spectrum with continuous background spectrum, the proposed method will be plagued by errors. These errors may be controlled by the relative magnitude between the discrete frequency component's and that of side band's. Further theoretical development is required to apply the method to more general cases.

II. SIMULATIONS AND EXPERIMENTS

Numerical simulations have been performed to demonstrate what we have proposed in the previous section. Monopole and dipole sources have been used for the simulations. We assume that the sources are located on the source plane which is defined as $z=0$, at the fixed coordinate, (x,y,z) (see the details in Fig. 3). As illustrated in the same figure, a microphone was assumed to be located 0.3 m away from the source plane and be moving along the x axis with Mach number, 0.01. A hologram was constructed based on the pressure measured by the moving microphone for 1 s: The sampling rate was 1/2048 s, providing 2048 data points.

In the first instance, a monopole is located at the (0 m, 0 m, 0 m) position and radiates the sound of 340 Hz. Figure 4(a) shows the variation of pressure magnitude and frequency due to the Doppler effect. The Fourier transform of this signal with respect to time is shown in Fig. 4(b). As explained in the previous section, we have the signal distributed over a band of frequency, [Fig. 4(b) and (c)], whose center frequency is 340 Hz. If one examines the frequency representation of the signal [Fig. 4(c)], then one may realize that it has a similar shape as the wave number spectrum of the monopole sound source, shifted by 340 Hz. This is an expression of Eq. (10), which says that the temporal Fourier transform of the measured signal will be same as its wave number spectrum but shifted by f_h, due to the Doppler effect.

To have general confidence on the proposed method, a more complicated case was also examined (Fig. 5). Four types of sources having different frequencies construct the pressure signal;

a monopole of 340 Hz located at (0 m, 0 m, 0 m),
a x-polarized dipole of 390 Hz located at (0 m, 0 m, 0 m),
a z-polarized dipole of 680 Hz located at (0 m, 0 m, 0 m), and
two monopoles of 440 Hz located at (0 m, 0 m, 0 m), and (0.5 m, 0.5 m, 0 m).

Figure 5(a) and (b) shows the time signal, depicted by the moving microphone, and its temporal Fourier transform. As one can see, the Doppler shifted spectrum are so well separated that frequency band overlappings are not possible. It is also interesting to note that the details of each Doppler shifted spectrum [Fig. 5(c) to (f)] mimic corresponding wave number spectra.

Based on the moving frame technique, the holograms were constructed and compared with true holograms. As one would expect, a perfect match is obtained. Some of the typical results are shown in Fig. 6.

The experiment, which is illustrated in Fig. 7, was conducted to compare the proposed moving microphone method with other existing methods: the conventional holograph method and the step-by-step scanning method [i.e., Spatial Transformation of Sound Fields (STSF)]. The microphone array (Fig. 7) moves along the array guide rail (Fig. 7) with speed which is controlled by an electric motor (Fig. 7) A string, pulley system (Fig. 7) provides the desired, vertical movement of the array. Ten photo electric sensors, each separated by 5 cm, are used to measure the velocity of the array. Two identical speakers which have no outer casing (Fig. 7) were positioned. Then the speakers were driven by 500 Hz pure tone. Typical experimental results are summarized in Fig. 8. Four different types of holograms were obtained for comparison. One is the hologram, 6×5 points measurement, in the x and y directions respectively [Fig. 8(a)]. The others were the holograms created by using the step-by-step scanning method [Fig. 8(b)], and two cases were obtained by the moving frame method [Fig. 8(c) and (d)]. Compared with Fig. 8(a) the other methods give better resolution. This is simply because the methods increase the effective number of microphones, therefore increase spatial resolution. The difference between Fig. 8(b) and (c) and (d) is that the former measures sound fields by conventional scanning techniques, but the latter two use the moving array method. Comparing the magnitude of pressure along $y=0$ lines clearly shows that the moving frame method recovers the desired hologram from the Doppler shifted wave number spectrum [Eq. (10)].

III. SUMMARY AND CONCLUDING REMARKS

The moving frame technique to measure the hologram was introduced. This method considers the relative movement between the source/hologram and the measuring microphone array. Theoretical formulation demonstrates how the idea of utilizing the wave number spectrum constructs the desired hologram. Numerical simulations and experiments demonstrate that the proposed method is valid and has great potential in various applications.

ACKNOWLEDGMENT

This work has been partially supported by the grant of KOSEF (Korea Science and Engineering Foundation).

[1] J. D. Maynard, E. G. Williams, and Y. Lee, ''Near field acoustic holography (NAH): 1. Theory of generalized holography and the development of NAH,'' J. Acoust. Soc. Am. **78**, 1395–1413 (1985).
[2] W. A. Veronesi and J. D. Maynard, ''Near field acoustic holography (NAH): 2. Holographic reconstruction algorithms and computer implementation,'' J. Acoust. Soc. Am. **81**, 1307–1322 (1987).
[3] J. Hald, ''Development of STSF with emphasis on the influence of bandwidth,'' Noise-Con 88 Proceedings, 529–536 (1988).
[4] T. Loyau, J.-C. Pascal, and P. Gaillard, ''Broadband acoustic holography reconstruction from acoustic intensity measurements I: Principle of the method,'' J. Acoust. Soc. Am. **84**, 1744–1750 (1988).

Acoustic bullets/transient Bessel beams: Near to far field transition via an impulse response approach

Peter R. Stepanishen
Department of Ocean Engineering, University of Rhode Island, Kingston, Rhode Island 02891

(Received 4 October 1997; accepted for publication 20 November 1997)

Acoustic transient Bessel beams (TBB) are a new class of waves which maintain their peak amplitude and general shape as they propagate in free space from an infinite planar aperture. The on axis and far field space-time properties of acoustic TBB waves generated by a finite spatial aperture are investigated using impulse response methods. The on-axis transition from the near field to the far field is investigated and several analogies to the acoustic field generated from a circular piston source are observed. In particular, the importance of an edge generated wave in determining the transition for the TBB field is clearly noted. Space-time properties of the far field are also investigated as a function of polar angle using impulse response methods. Numerical results are presented to illustrate the general characteristics of the on-axis and far field impulse response for a finite aperture. © *1998 Acoustical Society of America.* [S0001-4966(98)02204-8]

PACS numbers: 43.20.Fn [ANN]

INTRODUCTION

Acoustic bullets are defined as waves which maintain their general shape and amplitude as they propagate in space and are localized at any time to a compact region of support in space. Moses and Prosser[1] presented a similar definition in addressing the general inverse initial value problem for electromagnetic and acoustic fields. The generation of acoustic bullets and the space-time properties of the wave fields are clearly important as a result of the vast number of potential applications for such wave fields.

Within the last decade, several localized wave (LW) solutions of the homogeneous wave equation which exhibit the properties of bullets have been discovered. In 1983, Brittingham discovered the first localized wave (LW) solution to the homogeneous Maxwell's equation which he termed the focus wave mode (FWM).[2] Two years later Ziolkowski[3] constructed new families of localized wave (LW) solutions which are based on the FWM solution.[3–8] Shaarawi[9] recently addressed the generation of such localized waves from a planar dynamic Gaussian aperture. Sedky[10] subsequently addressed the diffraction length of localized wave fields launched from dynamic apertures.

In a recent paper,[11] Stepanishen and Sun presented a general development for a new class of acoustic bullets, designated as acoustic transient Bessel beams (TBB). This class of localized wave solutions includes the X waves, developed by Lu and Greenleaf[12,13] as a special case. The general theory of acoustic TBB fields was presented for both baseband and modulated carrier wave fields using time-domain, Fourier and Hilbert transform techniques. The TBB solution maintains its temporal and spatial shape as the acoustic bullet propagates in free space from an infinite planar aperture. Although the exact TBB solution in all space can obviously not be generated from a finite planar aperture, the general characteristics of acoustic bullets can be realized in a limited space-time region via the use of an appropriate space-time source distribution.

Consider now a finite planar circular aperture as shown in Fig. 1 with a specified axisymmetric space-time source distribution. For the case of a piston source with a spatially uniform harmonic source distribution it is well known that there is a collimated near field which extends from the source to the Rayleigh distance which provides a measure of the near to far field transition range. Beyond this distance the field must exhibit an inverse range dependence due to the spatial divergence of the field. In contrast to the piston source, an acoustic bullet with a smaller support region than the circular aperture can also be launched from the finite aperture[11] as illustrated in Fig. 1. It is clear however there must also be an associated near and far field region for this field. In the near field region the field must exhibit properties similar to those observed for the infinite aperture case, and in the far field region the field must exhibit an inverse range dependence.

The present paper investigates the space-time properties of the near to far field transition and the far field for acoustic bullets or acoustic TBB fields which are generated from finite planar apertures. In Section I, the general theory of acoustic TBB fields for signals is reviewed and the space-time planar aperture distributions which are required to generate acoustic TBB fields from planar apertures are presented. Space-time surface integral, impulse response and associated Fourier transform representations for the fields are presented. General properties of the on-axis near field and the transition to the far field for TBB fields which are generated using finite apertures are then developed in Section II using an impulse response approach and compared to the analogous results for a circular piston source. The far field for acoustic TBB fields which are generated from finite planar apertures is then investigated in Section III using Fourier transform and impulse response methods. General characteristics of the far field impulse responses for acoustic TBB fields which are generated from a finite circular aperture are presented.

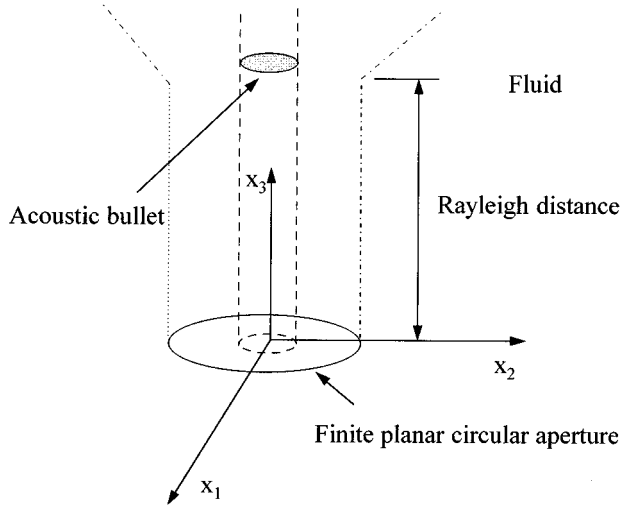


FIG. 1. A finite planar circular aperture.

I. TRANSIENT BESSEL BEAMS—GENERAL THEORY

The basic problem of interest here is to determine the near and far field space-time properties of acoustic bullets or TBB fields for signals which are generated by axisymmetric space-time source distributions over the circular planar aperture shown in Fig. 2. The acoustic wave equation for the axisymmetric pressure fields of interest $p(\rho,z,\tau)$ can be expressed in the normalized cylindrical coordinate system illustrated in Fig. 2 as

$$\left[\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial}{\partial\rho}\right)+\frac{\partial^2}{\partial z^2}-\frac{\partial^2}{\partial\tau^2}\right]p(\rho,z,\tau)=0, \tag{1}$$

where (ρ,z) are the normalized spatial coordinates and the normalized time is $\tau=c_0t/L$, where t denotes the unscaled time variable, c_0 is the constant sound speed of the media, and L is a characteristic length of interest.

For practical reasons of generating the acoustic bullets or TBB fields the planar space-time source distributions of interest are considered to be normal velocity $v_z(\rho,0,\tau)$ or acceleration distributions $a_z(\rho,0,\tau)$ over the surface $z=0$. A standard Green's function solution to the initial Neumann boundary value problem then leads to the following well known retarded time solution[14] for the pressure field:

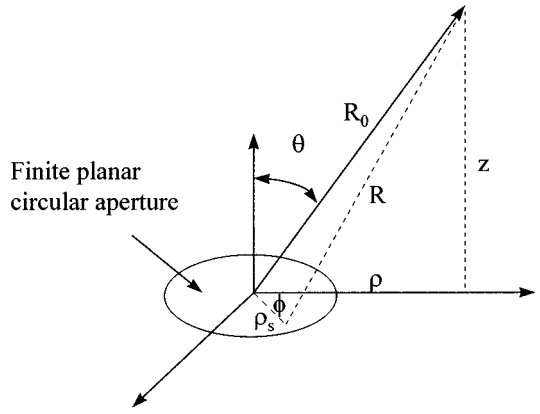


FIG. 2. Geometry of interest.

$$p(\rho,z,\tau)=\int_0^\infty \rho_s\,d\rho_s\int_0^{2\pi}d\phi\,\frac{\delta(\tau-R)}{2\pi R}\otimes a_z(\rho,0,\tau_s), \tag{2}$$

where $\delta(\ )$ denotes the Dirac delta function, $\otimes$ denotes the convolution operation and $R=\sqrt{\rho^2+\rho_s^2-2\rho\rho_s\cos\phi+z^2}$. The well known Rayleigh space-time surface integral[15] representation for the field then follows after noting the convolution property of the Dirac delta function:[16]

$$p(\rho,z,\tau)=\int_0^\infty \rho_s\,d\rho_s\int_0^{2\pi}d\phi\,\frac{1}{2\pi R}\{a_z(\rho,0,\tau)|_{\tau_s=\tau-R}\}, \tag{3}$$

where the surface acceleration remains to be determined.

In order to determine the source distributions of interest which are required to generate acoustic bullets it is first noted that the following solution to the free space scalar wave equation in Eq. (1) can be readily obtained via separation of variables

$$p_{k1,k2}(\rho,z,\tau)=J_0(k_1\rho)e^{i(k_3\tau-k_2z)}, \tag{4}$$

where $k_1^2+k_2^2=k_3^2$. A new family of solutions to the free space scalar wave equation can then be constructed from the basis function in Eq. (4) by first setting $k_1=\Omega\sin\zeta$ and $k_2=\Omega\cos\zeta$, multiplying by an arbitrary weight function $W(\Omega)$ and then integrating over the parameter Ω to obtain the following TBB solution to the free space scalar wave equation,

$$p(\rho,z,\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}W(\Omega)J_0(\Omega\sin\zeta\rho)e^{i\Omega(\tau-z\cos\zeta)}\,d\Omega, \tag{5}$$

where $0<\zeta<\pi/2$.

The normalized axial acceleration of a fluid particle for the TBB field can now be readily obtained from the usual linearized momentum equation, i.e.,

$$a_z(\rho,z,\tau)=-\frac{\partial p(\rho,z,\tau)}{\partial z}. \tag{6}$$

It then follows from Eq. (5) that the normalized axial acceleration can be expressed as

$$a_z(\rho,z,\tau)=\frac{\cos\zeta}{2\pi}\int_{-\infty}^{\infty}i\Omega W(\Omega)J_0(\Omega\sin\zeta\rho)e^{i\Omega(\tau-z\cos\zeta)}\,d\Omega, \tag{7}$$

and the associated scalar axial velocity field can be expressed as follows:

$$v_z(\rho,z,\tau)=\frac{\cos\zeta}{2\pi}\int_{-\infty}^{\infty}W(\Omega)J_0(\Omega\sin\zeta\rho)e^{i\Omega(\tau-z\cos\zeta)}\,d\Omega. \tag{8}$$

It is thus apparent that the pressure and axial velocity for TBB fields are scaled replicas of one another at all points in the field, i.e.,

$$p(\rho,z,\tau)=\cos\zeta\,v_z(\rho,z,\tau). \tag{9}$$

Since Eq. (5) is a Fourier integral, it is now obvious from the delay property of Fourier transforms[16] that the on-axis TBB fields of interest can be expressed as follows:

$$p(0,z,\tau)=w(\tau-z\ \cos\ \zeta) \tag{10}$$

and

$$v_z(0,z,\tau)=\cos\ \zeta w(\tau-z\ \cos\ \zeta), \tag{11}$$

where $w(\tau)$ and $W(\Omega)$ are a Fourier transform pair, i.e., $w(\tau)\Leftrightarrow W(\Omega)$

$$w(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty} W(\Omega)e^{i\Omega\tau}\,d\Omega, \tag{12}$$

$$W(\Omega)=\int_{-\infty}^{\infty} w(\tau)e^{-i\Omega\tau}\,d\Omega. \tag{13}$$

The importance of the weighting function $w(\tau)$ which represents a normalized axial velocity is thus obvious.

More generally it is observed from Eq. (5) and the convolution property of Fourier transforms that $p(\rho,z,\tau)$ and $v_z(\rho,z,\tau)$ for the TBB field can be expressed as follows:

$$p(\rho,z,\tau)=w(\tau)\otimes j(\rho,z,\tau) \tag{14}$$

and

$$v_z(\rho,z,\tau)=\cos\ \zeta w(\tau)\otimes j(\rho,z,\tau), \tag{15}$$

where

$$j(\rho,z,\tau)$$
$$=\frac{1}{2\pi}\int_{-\infty}^{\infty} J_0(\Omega\rho\ \sin\ \zeta)e^{j\Omega(\tau-z\ \cos\ \zeta)}\,d\Omega$$
$$=\frac{1}{\pi[(\rho\ \sin\ \zeta)^2-(\tau-z\ \cos\ \zeta)^2]^{1/2}},$$
$$-\rho\ \sin\ \zeta+z\ \cos\ \zeta<\tau<\rho\ \sin\ \zeta+z\ \cos\ \zeta. \tag{16}$$

For the on-axis case where $\rho=0$ it is noted that

$$j(0,z,\tau)=\delta(\tau-z\ \cos\ \zeta) \tag{17}$$

and the sifting property of $\delta(\ )$ then leads to the results in Eqs. (10) and (11).

A space-time surface integral representation for the acoustic TBB fields generated by a finite aperture of interest can now be obtained by combining Eqs. (2), (6), and (14):

$$p_\sigma(\rho,z,\tau)=\int_0^\sigma \rho_s d\rho_s\int_0^{2\pi} d\theta\frac{1}{2\pi R}$$
$$\times\left\{w(\tau)\otimes\frac{\partial j(\rho_s,0,\tau_s)}{\partial z}_{\tau_s=\tau-R}\right\}, \tag{18}$$

where σ is the radius of the aperture. Since

$$-\frac{\partial j(\rho,z,\tau)}{\partial z}=\cos\ \zeta\frac{\partial j(\rho,z,\tau)}{\partial\tau} \tag{19}$$

and

$$r(\tau)\otimes\frac{ds(\tau)}{d\tau}=s(\tau)\otimes\frac{dr(\tau)}{d\tau}, \tag{20}$$

then it is also apparent that $p_\sigma(\rho,z,\tau)$ be expressed as

$$p_\sigma(\rho,z,\tau)=\cos\zeta\int_0^\sigma\rho_s\,d\rho_s\int_0^{2\pi}d\phi\frac{1}{2\pi R}$$
$$\times\left\{\frac{dw(\tau)}{d\tau}\otimes j(\rho_s,0,\tau_s)_{\tau_s=\tau-R}\right\}. \tag{21}$$

For the infinite aperture case the surface integrals in Eqs. (18) and (21) simply yield the TBB solution in Eq. (14).

It is now noted that the terms involving $w(\tau)$ can be removed outside the surface integrals, e.g.,

$$p_\sigma(\rho,z,\tau)=\cos\ \zeta\frac{dw(\tau)}{d\tau}\otimes\int_0^\sigma\rho_s\ d\rho_s\int_0^{2\pi}d\phi\frac{1}{2\pi R}$$
$$\times j(\rho_s,0,\tau_s)_{\tau_s=\tau-R}. \tag{22}$$

Some useful forms for the pressure field from a finite aperture are then introduced as follows:

$$p_\sigma(\rho,z,\tau)=\frac{dw(\tau)}{d\tau}\otimes h_\sigma(\rho,z,\tau) \tag{23}$$
$$=w(\tau)\otimes\frac{dh_\sigma(\rho,z,\tau)}{d\tau}, \tag{24}$$

where the space-time impulse response $h_\sigma(\rho,z,\tau)$ is defined as follows:

$$h_\sigma(\rho,z,\tau)$$
$$=\cos\ \zeta\int_0^\sigma\rho_s\ d\rho_s\int_0^{2\pi}d\phi\frac{1}{2\pi R}j(\rho_s,0,\tau_s)_{\tau_s=\tau-R}. \tag{25}$$

Although the pressure fields of interest from a finite aperture can be evaluated using the space-time integral solutions in Eqs. (18), (21) and (22) or the convolution integral solutions in Eqs. (23) and (24), an alternative transform domain approach is also presented here for completeness. After noting the convolution property of Fourier transforms, it is evident that the Fourier transforms of Eqs. (18), (21) and (22) lead to the following equation for the finite aperture case:

$$P_\sigma(\rho,z,\Omega)$$
$$=i\Omega W(\Omega)\cos\ \zeta\int_0^\sigma\rho_s\ d\rho_s\int_0^{2\pi}d\phi\frac{e^{-i\Omega R}}{2\pi R}J_0(\Omega\ \sin\ \zeta\rho_s), \tag{26}$$

which can also be expressed as

$$P_\sigma(\rho,z,\Omega)=i\Omega W(\Omega)H_\sigma(\rho,z,\Omega), \tag{27}$$

where

$$H_\sigma(\rho,z,\Omega)$$
$$=\cos\ \zeta\int_0^\sigma\rho_s\ d\rho_s\int_0^{2\pi}d\phi\frac{e^{-i\Omega R}}{2\pi R}J_0(\Omega\sin\zeta\rho_s). \tag{28}$$

It is noted that Eq. (26) is equivalent to the usual Rayleigh surface integral representation for a harmonic field.

In light of the preceding development it is then obvious that

$$h_\sigma(\rho,z,\tau)\Leftrightarrow H_\sigma(\rho,z,\Omega), \tag{29}$$

$$p_\sigma(\rho,z,\tau)\Leftrightarrow P_\sigma(\rho,z,\Omega). \tag{30}$$

It then follows that $p_\sigma(\rho,z,\tau)$ can be simply expressed as the following inverse Fourier transform

$$p_\sigma(\rho,z,\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty} P_\sigma(\rho,z,\Omega)e^{i\Omega\tau}\,d\Omega. \tag{31}$$

The Fourier transform representations for the acoustic TBB fields are noted to be particularly useful for evaluating the far field properties of acoustic TBB fields generated from a finite aperture.

II. TRANSIENT BESSEL BEAMS—ON-AXIS

The surface integral representations presented in the previous section are now used to investigate the properties of the on-axis for the acoustic TBB fields generated by a finite aperture of interest where the weighting function $w(\tau)$ is considered to be arbitrary. Although the entire space-time acoustic TBB pressure field can be numerically evaluated from the surface integrals in Eqs. (18), (21) and (22), the on-axis case is particularly simple. As a result of the axisymmetry of the source distribution, the on-axis pressure $p(0,z,\tau)$ for a specified $w(\tau)$ is readily evaluated from line integrals, i.e., the ϕ integration is simply performed in the surface integrals.

It now follows directly from Eq. (22) or from Eqs. (24) and (25) that $p(0,z,\tau)$ can be expressed as

$$p_\sigma(0,z,\tau)=\frac{dw(\tau)}{d\tau}\otimes h_\sigma(0,z,\tau)$$

$$=w(\tau)\otimes\frac{dh_\sigma(0,z,\tau)}{d\tau}, \tag{32}$$

where $h_\sigma(0,z,\tau)$ is expressed as the following line integral:

$$h_\sigma(0,z,\tau)=\int_0^\sigma \frac{\cos\zeta}{R} j(\rho_s,0,\tau_s)_{\tau_s=\tau-R}\rho_s\,d\rho_s. \tag{33}$$

Since $j(\rho_s,0,\tau_s)$ is a time and space limited function which is bounded by the lines $\tau_s=\pm\,\rho_s\sin\zeta$ and $\rho_s=\sigma$, care must be exercised in establishing the limits of the integration for various z and τ. In general, the path of integration for a fixed z and τ is along the hyperbolic contour defined by $\tau_s=\tau-\sqrt{\rho_s^2+z^2}$.

As an aside it is noted that an alternative representation of the on-axis pressure is simply obtained via the use of a superposition of two integrals which leads to

$$p_\sigma(0,z,\tau)=p(0,z,\tau)-\frac{dw(\tau)}{d\tau}\otimes\int_\sigma^\infty \frac{\cos\zeta}{R}$$

$$\times j(\rho_s,0,\tau_s)_{\tau_s=\tau-R}\rho_s\,d\rho_s, \tag{34}$$

where $R=\sqrt{\rho_s^2+z^2}$. The first term simply corresponds to $p_\infty(\rho,z,\tau)$, i.e., the free space solution in Eq. (14) and the latter term in Eq. (34) corresponds to an edge generated wave associated with the finite aperture. The importance of the edge generated wave, which has been previously observed in the solution for the on-axis pressure from a piston type of source, will become apparent from the following development.

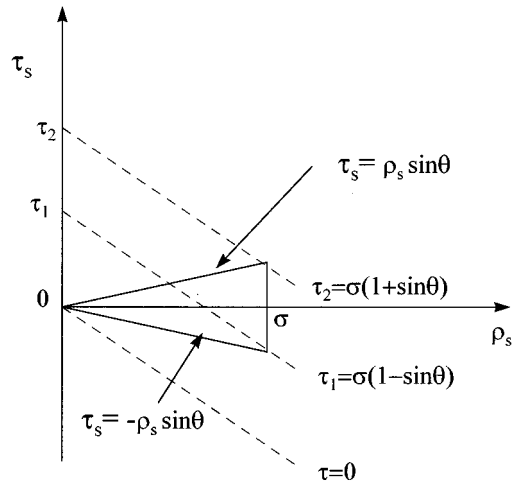


FIG. 3. Spatial region of integration at fixed times for $z=0$.

Prior to addressing the on-axis pressure for the TBB field, the on-axis pressure for the piston source is first addressed using Eq. (34) with $\zeta=0$. For this case it is apparent from Eq. (34) that

$$j(0,z,\tau)=\delta(\tau-z), \tag{35}$$

thus

$$p_\sigma(0,z,\tau)=w(\tau-z)-\frac{dw(\tau)}{d\tau}\otimes\int_\sigma^\infty\frac{\delta(\tau-R)}{R}\rho_s\,d\rho_s, \tag{36}$$

Since

$$\int_\sigma^\infty\frac{\delta(\tau-R)}{R}\rho_s\,d\rho_s=\int_{\sqrt{z^2+\sigma^2}}^\infty\delta(\tau-R)dR$$

$$=u(\tau-\sqrt{z^2+\sigma^2}), \tag{37}$$

where $u(\tau)$ is the familiar Heaviside function, and

$$\frac{du(\tau)}{d\tau}=\delta(\tau), \tag{38}$$

then

$$p_\sigma(0,z,\tau)=w(\tau-z)-w(\tau-\sqrt{z^2+\sigma^2}). \tag{39}$$

As noted earlier,[14] it is thus apparent that the on-axis pressure for the piston source can be decomposed into a plane wave component and an edge wave contribution.

As a special case of interest consider now the on-axis pressure for the TBB field where $z=0$. The contour of integration for a fixed τ in Eq. (33) corresponds to a straight line defined by $\tau_s=\tau-\rho_s$ as illustrated in Fig. 3. It is thus apparent from the figure that the limits of the ρ_s integration along a contour are a function of τ which is in the range $0\leqslant\tau\leqslant\sigma(1+\sin\zeta)$. Three different contours corresponding to $\tau=0$, $\sigma(1-\sin\zeta)$ and $\sigma(1+\sin\zeta)$ are illustrated in the figure. It is clear that the space-time impulse response of interest $h_\sigma(0,0,\tau)=0$ for $\tau<0$ and $\tau>\sigma(1+\sin\zeta)$. The time $\tau=\sigma(1-\sin\zeta)$ corresponds to the propagation time of the initial edge ($\rho_s=\sigma$) generated signal to propagate to the ori-

gin ($\rho_s=0$) and the time $\tau=\sigma(1+\sin\zeta)$ corresponds to the propagation time of the final edge generated signal to propagate to the origin.

The space-time impulse response of interest, $h_\sigma(0,0,\tau)$, can now be expressed as follows:

$$\begin{aligned} h_\sigma(0,0,\tau)&=0, \quad \tau<0\\ &=\int_{\tau/(1+\sin\zeta)}^{\tau/(1-\sin\zeta)}\cos\zeta j(\rho_s,0,\tau_s-\rho_s)d\rho_s,\\ &\quad 0\leqslant\tau\leqslant\sigma(1-\sin\zeta)\\ &=\int_{\tau/(2+\sin\zeta)}^{\sigma}\cos\zeta j(\rho_s,0,\tau_s-\rho_s)d\rho_s,\\ &\quad \sigma(1-\sin\zeta)\leqslant\tau\leqslant\sigma(1+\sin\zeta)\\ &=0, \quad \sigma(1+\sin\zeta)\leqslant\tau. \end{aligned} \tag{40}$$

After noting that[17]

$$\begin{aligned} &\int j(\rho_s,0,\tau_s-\rho_s)d\rho_s\\ &=\int\frac{d\rho_s}{\sqrt{c\rho_s^2+b\rho_s+a}}\\ &=-\frac{1}{\sqrt{-c}}\sin^{-1}\left\{\frac{2c\rho_s+b}{\sqrt{b^2-4ac}}\right\}, \quad c<0 \text{ and } b^2>4ac, \end{aligned} \tag{41}$$

where $c=-\cos^2\zeta$, $b=2\tau$ and $a=-\tau^2$, it is then easily shown that

$$\begin{aligned} h_\sigma(0,0,\tau)&=0, \quad \tau<0\\ &=u(\tau), \quad 0\leqslant\tau\leqslant\sigma(1-\sin\zeta)\\ &=\frac{1}{2}-\frac{1}{\pi}\sin^{-1}\left(\frac{\tau-\sigma\cos^2\zeta}{\tau\sin\zeta}\right),\\ &\quad \sigma(1-\sin\zeta)\leqslant\tau\leqslant\sigma(1+\sin\zeta)\\ &=0, \quad \sigma(1+\sin\zeta)\leqslant\tau. \end{aligned} \tag{42}$$

General characteristics of $h_\sigma(0,0,\tau)$ are evident from the numerical results in Fig. 4 which present $h_\sigma(0,0,\tau)$ versus a normalized time τ/σ for $\zeta=0$, 4, 8 and 16 °. It is again noted that $\zeta=0°$ corresponds to the circular piston case.

It is now readily apparent that $dh_\sigma(0,0,\tau)/d\tau$ can be expressed as

$$\frac{dh_\sigma(0,0,\tau)}{d\tau}=\delta(\tau)-e(0,0,\tau), \tag{43}$$

where $e(0,0,\tau)$ is clearly an edge generated wave defined as follows:

$$\begin{aligned} e(0,0,\tau)&=\frac{1}{\pi\sqrt{1-[(\tau-\sigma\cos^2\zeta)/\tau\sin\zeta]^2}}\ \frac{\cos^2\zeta\sigma}{\sin\zeta\tau^2},\\ &\quad \sigma(1-\sin\zeta)\leqslant\tau\leqslant\sigma(1+\sin\zeta). \end{aligned} \tag{44}$$

The importance of $dh_\sigma(0,0,\tau)/d\tau$ is observed from Eq. (32) which leads to the following expression for the pressure $p_\sigma(0,0,\tau)$:

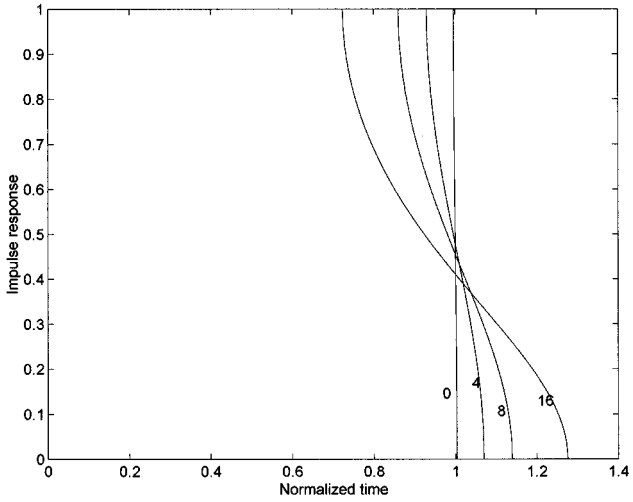


FIG. 4. Impulse response for $z=0$ and $\zeta=0$, 4, 8 and 16°.

$$p_\sigma(0,0,\tau)=w(\tau)-w(\tau)\otimes e(0,0,\tau). \tag{45}$$

The first term is the infinite aperture result and the second term is an edge generated pressure wave for the TBB field at $z=0$. For $\zeta=0$ the above equation reduces to the piston result in Eq. (39).

Consider now the on-axis space-time impulse response of interest $h_\sigma(0,z,\tau)$ for the TBB field where $\dot{z}>0$. The contour of integration for a fixed τ in Eq. (33) now corresponds to a hyperbola defined by $\tau_s=\tau-\sqrt{\rho_s^2+z^2}$ and the limits of the ρ_s integration along a contour are a function of z and τ. At a fixed z, four different contours corresponding to $\tau=z\cos\zeta$, z, $\sqrt{\sigma^2+z^2}-\sigma\sin\zeta$ and $\sqrt{\sigma^2+z^2}+\sigma\sin\zeta$ are illustrated on the (τ_s,ρ_s) plane in Fig. 5. The contour defined by $\tau=z\cos\zeta$ corresponds to a tangent point between the hyperbola and $\tau_s=-\rho_s\sin\zeta$. This important condition defines a near to far field transition as noted below and does not exist for $z>z_t$ where $z_t=\sigma/\tan\zeta$. The time $\tau=\sqrt{\sigma^2-z^2}-\sigma\sin\zeta$ corresponds to the propagation time of

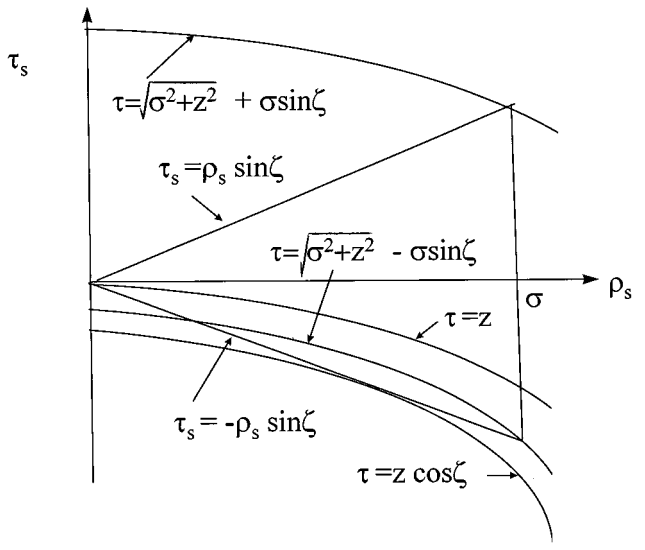


FIG. 5. Contours of integration for constant τ.

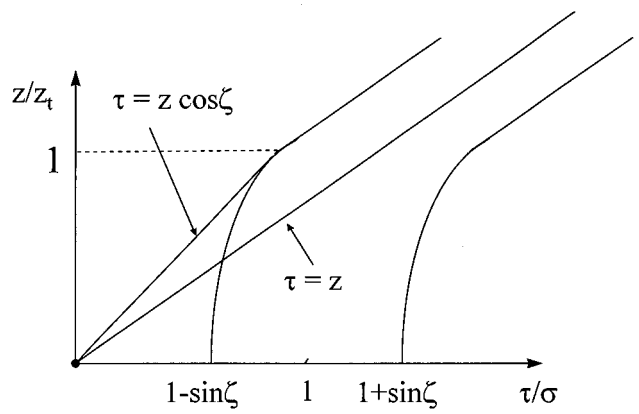


FIG. 6. Boundaries of the impulse response in the normalized τ-z plane.

the initial edge ($\rho_s=\sigma$) generated signal to propagate to the location z whereas the time $\tau=\sqrt{\sigma^2+z^2}+\sigma\sin\zeta$ corresponds to the propagation time of the final edge generated signal to propagate to z.

Some general characteristics of the space-time impulse response of interest $h_\sigma(0,z,\tau)$ can now be obtained by mapping the contours from the (τ_s,ρ_s) plane to the boundaries in the normalized (z,τ) plane which are shown in Fig. 6. Since no contour of integration exists for $\tau<z\cos\zeta$ or $\tau>\sqrt{\sigma^2+z^2}+\sigma\sin\zeta$, it is clear that $h_\sigma(0,z,\tau)=0$ for $\tau<z\cos\zeta$ and for $\tau>\sqrt{\sigma^2+z^2}+\sigma\sin\zeta$. For $z<z_t$ and $z\cos\zeta<\tau<\sqrt{\sigma^2+z^2}-\sigma\sin\zeta$ it is also noted that $h_\sigma(0,z,\tau)=u(\tau-z\cos\zeta)$ which is the TBB field for the infinite aperture case, i.e., the edge effects occurring at $\rho=\sigma$ have not influenced the on-axis field. In this space-time region the phase velocity of the wavefront along the z axis is obviously $1/\cos\zeta$ and the field exhibits the characteristics of an acoustic supersonic bullet. For $z>z_t$ it is also apparent that $h_\sigma(0,z,\tau)$ is determined by the edge contribution.

In light of the preceding development it is now obvious that $h_\sigma(0,z,\tau)$ can be expressed in the following form for $z/z_t<1$:

$$h_\sigma(0,z,\tau)=u(\tau-z\cos\zeta)-u(\tau-\sqrt{\sigma^2+z^2}+\sigma\sin\zeta)+d_\sigma(0,z,\tau), \tag{46}$$

where

$$d_\sigma(0,z,\tau)=\cos\zeta\int_{\rho_1}^{\rho_2}\frac{1}{\sqrt{\rho_s^2+z^2}}\times j(\rho_s,0,\tau-\sqrt{\rho_s^2+z^2})\rho_s\,d\rho_s \tag{47}$$

and $\rho_1(\tau)$ and $\rho_2(\tau)$ are functions of z. The transformation

$$r=\sqrt{\rho_s^2+z^2} \tag{48}$$

then leads to the following expression for $d_\sigma(0,z,\tau)$:

$$d_\sigma(0,z,\tau)=\cos\zeta\int_{r_1}^{r_2}j(\rho_s,0,\tau-r)dr \tag{49}$$

$$=\cos\zeta\int_{r_1}^{r_2}\frac{dr}{\sqrt{cr^2+br+a}}, \tag{50}$$

where $c=-\cos^2\zeta$, $b=2\tau$ and $a=-\tau^2+(z\sin\zeta)^2$.

Closed form expressions for $d_\sigma(0,z,\tau)$ can now be obtained from Eq. (49) via the use of Eq. (41). For $0<z/z_t<1$ it is easily shown that

$$d_\sigma(0,z,\tau)=-\frac{1}{\pi}\left\{\sin^{-1}\left[\frac{\tau-r_\sigma\cos^2\zeta}{\sin\zeta\sqrt{\tau^2-z^2\cos^2\zeta}}\right]-\sin^{-1}\left[\frac{\tau-r_1\cos^2\zeta}{\sin\zeta\sqrt{\tau^2-z^2\cos^2\zeta}}\right]\right\},\quad \sqrt{\sigma^2+z^2}-\sigma\sin\zeta<\tau<\sqrt{\sigma^2+z^2}+\sigma\sin\zeta, \tag{51}$$

where $r_\sigma=\sqrt{\sigma^2+z^2}$, $r_1=\sqrt{\rho_1^2+z^2}$ and

$$\rho_1=\frac{-\tau\sin\zeta+\sqrt{\tau^2-z^2\cos^2\zeta}}{\cos^2\zeta}. \tag{52}$$

For $z/z_t>1$ it is also easily shown that

$$h_\sigma(0,z,\tau)=-\frac{1}{\pi}\left\{\sin^{-1}\left[\frac{\tau-r_\sigma\cos^2\zeta}{\sin\zeta\sqrt{\tau^2-z^2\cos^2\zeta}}\right]-\sin^{-1}\left[\frac{\tau-r_1\cos^2\zeta}{\sin\zeta\sqrt{\tau^2-z^2\cos^2\zeta}}\right]\right\},\quad \sqrt{\sigma^2+z^2}-\sigma\sin\zeta<\tau<z$$

$$=-\frac{1}{\pi}\left\{\sin^{-1}\left[\frac{\tau-r_\sigma\cos^2\zeta}{\sin\zeta\sqrt{\tau^2-z^2\cos^2\zeta}}\right]-\sin^{-1}\left[\frac{\tau-r_2\cos^2\zeta}{\sin\zeta\sqrt{\tau^2-z^2\cos^2\zeta}}\right]\right\},\quad z<\tau<\sqrt{\sigma^2+z^2}+\sigma\sin\zeta, \tag{53}$$

where $r_1=\sqrt{\rho_1^2+z^2}$ with

$$\rho_1=\frac{\tau\sin\zeta-\sqrt{\tau^2-z^2\cos^2\zeta}}{\cos^2\zeta} \tag{54}$$

and $r_2=\sqrt{\rho_2^2+z^2}$ with

$$\rho_2=\frac{-\tau\sin\zeta+\sqrt{\tau^2-z^2\cos^2\zeta}}{\cos^2\zeta}. \tag{55}$$

In order to illustrate the general characteristics of $h_\sigma(0,z,\tau)$ numerical results are now presented for $h_\sigma(0,z,\tau)$ versus the normalized time $(\tau-z)/\sigma$ for $\zeta=4°$. The results in Fig. 7 correspond to $0<z/z_t<1$ whereas the results in Fig. 8 correspond to $z/z_t>1$. In the region where $0\leqslant z/z_t<1$ each impulse response exhibits an initial step discontinuity and is then constant until the effect of the edge wave arrives. Following the arrival of the edge wave, each impulse response monotonically decreases to zero as indicated in Fig. 7. In the region where $z/z_t>1$ each impulse response is a continuous function. As z/z_t increases the peak value of the impulse response decreases and the response is noted to exhibit a more symmetric response as a function of the normalized time parameter.

As a final case of interest, consider now the on-axis space-time impulse response of interest $h_\sigma(0,z,\tau)$ for the

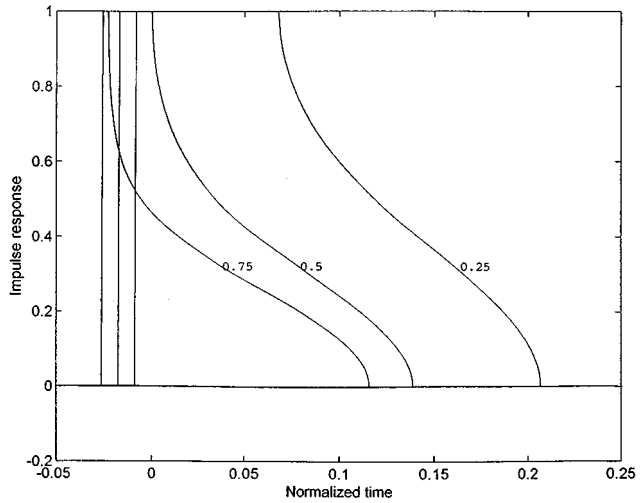


FIG. 7. Impulse response for $z/z_t = 0.25,\ 0.5$ and 0.75.

region where $z \gg z_t$. The curvature of the hyperbolic contour of integration for a fixed τ in Fig. 5 can now be neglected. The resulting contour of integration which is illustrated in Fig. 9 then leads to the following expression for $h_\sigma(0,z,\tau)$ via the use of Eq. (33):

$$h_\sigma(0,z,\tau) \approx \frac{\cos\zeta}{\pi z}\int_{(\tau-z)/\sin\zeta}^{\sigma} \frac{\rho_s d\rho_s}{\sqrt{\rho_s^2-(\tau-z)^2}},$$

$$z-\sigma\sin\zeta < \tau < z+\sigma\sin\zeta. \tag{56}$$

It then follows that the on-axis far field impulse response can be simply expressed as

$$h_\sigma(0,z,\tau) \approx \frac{\cos\zeta}{\pi z}\sqrt{\sigma^2-\left(\frac{\tau-z}{\sin\zeta}\right)^2},$$

$$z-\sigma\sin\zeta < \tau < z+\sigma\sin\zeta \tag{57}$$

which clearly indicates the inverse range dependence of the on-axis far field and the symmetry about the pulse center at $\tau = z$ as noted from Fig. 8.

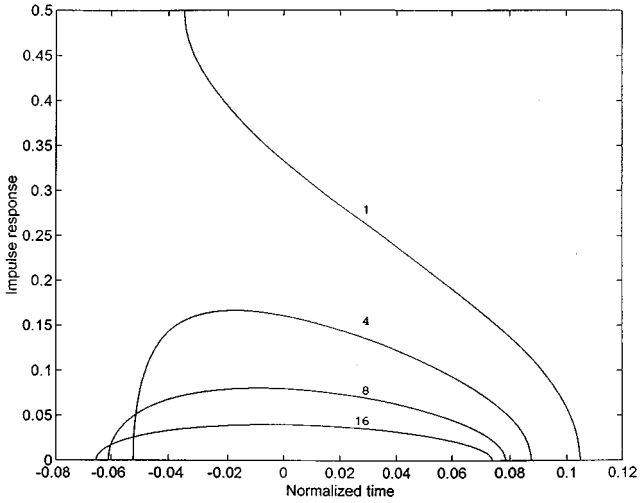


FIG. 8. Impulse response for $z/z_t = 1,\ 4,\ 8$, and 16.

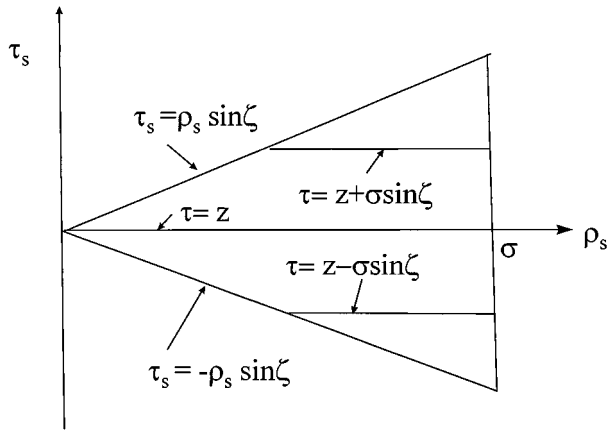


FIG. 9. Contours of integration for far field z and constant τ.

III. TRANSIENT BESSEL BEAMS—FAR-FIELD

Consider now the far field region where $z \gg z_t$ for TBB fields which are generated using a finite aperture as shown in Fig. 2. In contrast to a direct space-time approach to evaluate the far field, the Rayleigh surface integral representation for the Fourier transform of $p_\sigma(\rho,z,\tau)$ in Eq. (26) provides a convenient starting point, i.e.,

$$P_\sigma(\rho,z,\Omega) = i\Omega W(\Omega)\cos\zeta\int_0^\sigma \rho_s\, d\rho_s \int_0^{2\pi} d\phi \frac{e^{-i\Omega R}}{2\pi R}$$

$$\times J_0(\Omega\sin\zeta\rho_s). \tag{58}$$

In light of Fig. 2 the usual far field approximation for R then leads to:

$$\frac{e^{-i\Omega R}}{2\pi R} \sim \frac{e^{-i\Omega R_0}}{2\pi R_0} e^{i\Omega\rho_s \sin\theta\cos\phi}. \tag{59}$$

Since

$$J_0(y) = \frac{1}{2\pi}\int_0^{2\pi} e^{iy\cos\phi} dy \tag{60}$$

it then follows that

$$P_\sigma(\rho,z,\Omega) \to P_\sigma(\theta,\Omega), \tag{61}$$

where

$$P_\sigma(\theta,\Omega) = i\Omega W(\Omega)\cos\zeta \frac{e^{-i\Omega R_0}}{2\pi R_0}\int_0^\sigma \rho_s J_0(\Omega\sin\zeta\rho_s)$$

$$\times J_0(\Omega\sin\theta\rho_s) d\rho_s. \tag{62}$$

After noting that[17]

$$\int \rho_s J_0(\alpha\rho_s) J_0(\beta\rho_s) d\rho_s$$

$$= \rho_s \frac{\beta J_0(\alpha\rho_s) J_1(\beta\rho_s) - \alpha J_0(\beta\rho_s) J_1(\alpha\rho_s)}{\beta^2-\alpha^2}, \tag{63}$$

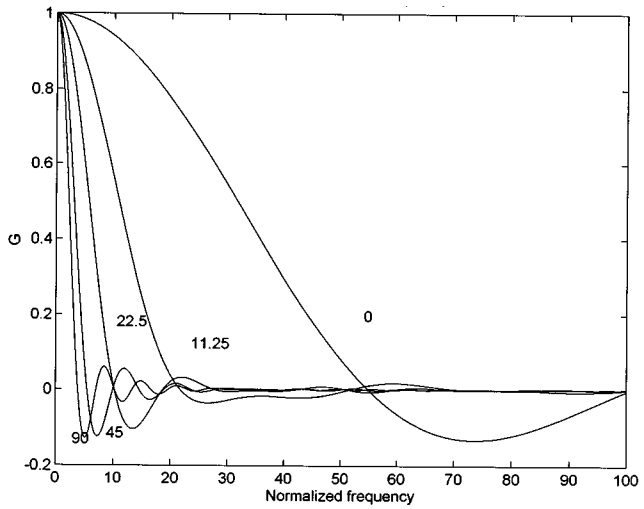


FIG. 10. $G_\sigma(\theta,\zeta,\Omega)$ as a function of normalized frequency $\sigma\Omega$ for $\zeta=4°$ and $\theta=0°, 11.25°, 22.5°, 45°$, and $90°$.

the following far field Fourier transform is simply obtained

$$P_\sigma(\theta,\Omega)=i\Omega W(\Omega)H^f_\sigma(\theta,\zeta,\Omega), \tag{64}$$

where

$$H^f_\sigma(\theta,\zeta,\Omega)=\frac{e^{-i\Omega R_0}}{2\pi R_0}\pi\sigma^2\cos\zeta G_\sigma(\theta,\zeta,\Omega) \tag{65}$$

and

$$G_\sigma(\theta,\zeta,\Omega)\equiv\frac{1}{[\sin^2\zeta-\sin^2\theta]}\times\left[\sin^2\zeta J_0(\Omega\sigma\sin\theta)\frac{2J_1(\Omega\sigma\sin\zeta)}{\Omega\sigma\sin\zeta}-\sin^2\theta J_0(\Omega\sigma\sin\zeta)\frac{2J_1(\Omega\sigma\sin\theta)}{\Omega\sigma\sin\theta}\right], \tag{66}$$

which is noted to be real. The low pass nature of the Fourier transform $G_\sigma(\theta,\zeta,\Omega)$ is apparent from the numerical results in Fig. 10 which are presented as a function of the normalized frequency $\Omega\sigma$ for $\zeta=4$. These results illustrate the general oscillatory behavior of $G_\sigma(\theta,\zeta,\Omega)$ as a function of frequency and the decreasing bandwidth as θ increases.

Several special cases of interest are now considered prior to addressing the time-dependent problem of interest. For $\zeta=0$ it is first observed from Eqs. (64) and (66) that:

$$P_\sigma(\theta,\Omega)=i\Omega W(\Omega)\frac{e^{-i\Omega R_0}}{2\pi R_0}\pi\sigma^2\frac{2J_1(\Omega\sigma\sin\theta)}{\Omega\sigma\sin\theta}, \tag{67}$$

which is the expected classical result[18] for a piston with a uniform velocity $W(\Omega)$. For all ζ and $\theta=0$, i.e., on-axis, it is also observed from Eq. (64) that:

$$P_\sigma(0,\Omega)=i\Omega W(\Omega)\frac{e^{-i\Omega R_0}}{2\pi R_0}\pi\sigma^2\cos\zeta\frac{2J_1(\Omega\sigma\sin\zeta)}{\Omega\sigma\sin\zeta}, \tag{68}$$

which of course reduces to the on-axis result for the piston when $\zeta=0$. An interesting analogy between the far field on-axis pressure for the TBB field and the off-axis pressure field for the piston case is clearly observed from the above results. Finally, for $\theta=\zeta$ it readily follows from Eqs. (64) and (66) using L'Hospitals rule that

$$P_\sigma(\zeta,\Omega)=i\Omega W(\Omega)\frac{e^{-i\Omega R_0}}{2\pi R_0}\pi\sigma^2[J_0^2(\Omega\sigma\sin\zeta)+J_1^2(\Omega\sigma\sin\zeta)]. \tag{69}$$

It now follows from Eq. (64) via an inverse Fourier transform that the far field pressure can be expressed as

$$p_\sigma(\theta,\tau)=\frac{dw(\tau)}{d\tau}\otimes h^f_\sigma(\theta,\zeta,\tau), \tag{70}$$

where the far field impulse response $h^f_\sigma(\theta,\zeta,\tau)$ is simply expressed as

$$h^f_\sigma(\theta,\zeta,\tau)=\frac{\sigma^2}{2R_0}\cos\zeta g_\sigma(\theta,\zeta,\tau-R_0). \tag{71}$$

The form of the solution for the far field pressure in Eq. (70) is noted to be consistent with the more general result in Eq. (23). In contrast to the general case relatively simple closed form expressions for $h^f_\sigma(\theta,\zeta,\tau)$ can be simply developed as shown in the following paragraphs.

In order to evaluate the on-axis far field impulse response $h^f_\sigma(0,\zeta,\tau)$ the following transform pair is first observed

$$\frac{2J_1(\Omega\sigma\sin\zeta)}{\Omega\sigma\sin\zeta}\Leftrightarrow h_1(\zeta,\tau), \tag{72}$$

where[17]

$$h_1(\zeta,\tau)=\frac{1}{\pi\sigma\sin\zeta}\left[1-\left(\frac{\tau}{\sigma\sin\zeta}\right)^2\right]^{1/2}, \qquad \left|\frac{\tau}{\sigma\sin\zeta}\right|<1$$
$$=0, \qquad \left|\frac{\tau}{\sigma\sin\zeta}\right|<1. \tag{73}$$

It then follows that the on-axis far field impulse response can be expressed as:

$$h^f_\sigma(0,\zeta,\tau)=\frac{\sigma^2\cos\zeta}{2z}h_1(\zeta,\tau-z), \tag{74}$$

which is identical to Eq. (57). As $\sigma\sin\zeta\to0$, it is simply noted that

$$h^f_\sigma(0,\zeta,\tau)\to\frac{\sigma^2\cos\zeta}{2z}\delta(\tau-z). \tag{75}$$

Similar results for $\zeta=0$ have been previously observed for the far field of a baffled piston source.[19]

The far field impulse response $h^f_\sigma(\zeta,\zeta,\tau)$ is also readily evaluated after noting the following transform pair:

$$iJ_1(\Omega\sigma\sin\zeta)\Leftrightarrow j_1(\zeta,\tau), \tag{76}$$

where

$$j_1(\zeta,\tau)=\frac{1}{\pi}\frac{\tau/\sigma\ \sin\ \zeta}{[1-(\tau/\sigma\ \sin\ \zeta)^2]^{1/2}}, \qquad \left|\frac{\tau}{\sigma\ \sin\ \zeta}\right|<1$$

$$=0, \qquad \left|\frac{\tau}{\sigma\ \sin\ \zeta}\right|<1. \tag{77}$$

Since

$$J_1^2(\Omega\sigma\ \sin\ \zeta)\Leftrightarrow -j_1(\zeta,\tau)\otimes j_1(\zeta,\tau) \tag{78}$$

and

$$J_0^2(\Omega\sigma\ \sin\ \zeta)\Leftrightarrow j(\sigma,0,\tau)\otimes j(\sigma,0,\tau), \tag{79}$$

where $j(\sigma,0,\tau)$ is known from Eq. (16), it then follows that $h_\sigma^f(\zeta,\zeta,\tau)$ can be expressed as

$$h_\sigma^f(\zeta,\zeta,\tau)=\frac{\sigma^2}{2R_0}\cos\ \zeta[j(\sigma,0,\tau-R_0)\otimes j(\sigma,0,\tau) - j_1(\zeta,\tau-R_0)\otimes j_1(\zeta,\tau)]. \tag{80}$$

The preceding results can now be used to develop a general representation for the far field impulse response $h_\sigma^f(\theta,\zeta,\tau)$ via the use of Eq. (71). In light of the special cases it is now apparent that $g_\sigma(\theta,\zeta,\tau)$ can be simply expressed in the following form:

$$g_\sigma(\theta,\zeta,\tau)=\frac{1}{[\sin^2\ \zeta-\sin^2\ \theta]}[\sin^2\ \zeta j(\sigma,0,\tau)\otimes h_1(\zeta,\tau) - \sin^2\ \theta\hat{j}(\sigma,0,\tau)\otimes h_1(\theta,\tau)], \tag{81}$$

where

$$\hat{j}(\sigma,0,\tau)=j(\sigma,0,\tau)|_{\zeta=\theta}. \tag{82}$$

Since $j(\sigma,0,\tau)$ and $h_1(\theta,\tau)$ are time limited functions, it then follows that $g_\sigma(\theta,\zeta,\tau)$ and $h_\sigma^f(\theta,\zeta,\tau)$ are thus also time limited, i.e., functions of finite time duration. It is also observed that

$$g_\sigma(\theta,\zeta,\tau)=g_\sigma(\zeta,\theta,\tau). \tag{83}$$

In order to illustrate typical behavior of $g_\sigma(\zeta,\theta,\tau)$ numerical results for $g_\sigma(\theta,\zeta,\tau)$ are presented in Fig. 11 as a function of the normalized time τ/σ for $\zeta=4$. Since $g_\sigma(\theta,\zeta,\tau)$ is an even function of τ, the results are presented only for $\tau\geqslant 0$. These results illustrate the general behavior of $g_\sigma(\zeta,\theta,\tau)$ and thus $h_\sigma^f(\theta,\zeta,\tau)$ as a function of normalized time and polar angle θ. As θ increases the time duration of the impulse response is noted to increase which is to be expected from the results in Fig. 10. Since the impulse responses are relatively simple, the convolution operation in Eq. (70) can be readily performed and the characteristics of the far field can be readily investigated for any specified $w(\tau)$.

IV. SUMMARY AND CONCLUSIONS

Acoustic transient Bessel beams (TBB) or acoustic bullets can be generated via infinite space and time-dependent planar aperture distributions and have been previously shown to exhibit unique properties. In particular, the time-dependent pressure at any on-axis point has been observed to be identical apart from a time delay and the TBB field does not spatially disperse as it propagates at a supersonic wave speed from the aperture—the normalized wave speed of the acoustic bullet is $1/\cos\ \zeta$. Although the exact TBB solution in all space can obviously not be generated from a finite planar aperture, the general characteristics of acoustic bullets can be realized in a limited space-time region via the use of an appropriate space-time planar aperture distribution.

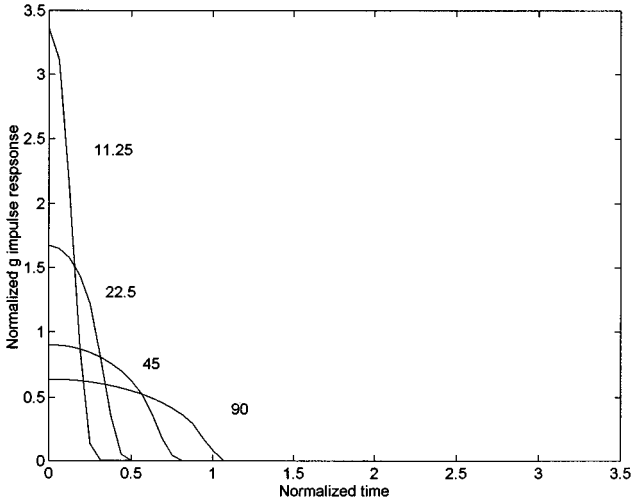


FIG. 11. Far field normalized impulse response $g_\sigma(\theta,\zeta,\tau)$ as a function of the normalized time τ/σ for $\zeta=11.25°$, 22.5°, 45°, and 90°.

An impulse response approach has been presented here to investigate the the near to far field transition for acoustic transient Bessel beams (TBB) which are generated from both space and time limited aperture distributions for arbitrary weighting functions $w(\tau)$. Relatively simple closed form expressions for the on-axis and far field impulse responses were developed. These impulse responses are surprisingly simple in form given the nature of the problem. In addition, the space-time impulse responses provide important physical insight into the space-time characteristics of acoustic (or electro-magnetic) TBB fields. In particular, the supersonic nature of the on-axis acoustic bullet is readily observed to result from the noncausal[16] time-dependent planar aperture distributions. Several interesting analogies to a previously developed impulse response approach for simple piston sources were noted.

In contrast to the generation of TBB fields via an infinite space and time-dependent planar aperture distribution, it was shown that important edge effects are present in the generation of TBB fields from finite apertures. The importance of edge effects had been previously observed for acoustic piston sources[14] and more recently for analogous electromagnetic field problems.[20] For the TBB field, these edge effects result in a well defined on-axis near to far field transition range. Beyond this transition range the TBB field begins to exhibit an inverse range dependence. For axial distances less than the transition range the TBB field exhibits the characteristics of the acoustic bullet generated from an infinite aperture over a limited time window, e.g., a supersonic wave described by $w(\tau-z\ \cos\ \zeta)$ is readily observed on-axis. In contrast to the

Rayleigh distance which is proportional to the frequency of excitation for a uniformly excited aperture with a pulsed harmonic excitation, the transition range is noted to be independent of frequency.

Although it was previously noted that TBB fields can be realized with a lateral beam dimension that is much narrower than the physical source aperture size, the near field lateral properties of the TBB fields generated from a finite aperture have not been addressed here. A more extensive investigation of the near field for TBB fields for various weighting functions is in progress using the impulse response method and will be the subject of a later paper.

[1] H. E. Moses and R. T. Prosser, ''Initial conditions, sources, and currents for prescribed time-dependent acoustic and electromagnetic fields in three dimensions, Part I: The inverse initial value problem, acoustic and electromagnetic bullets,'' expanding waves, and imploding waves,'' IEEE Trans. Antennas Propag. **AP-34**, 188–196 (1986).

[2] J. N. Brittingham, ''Focus wave modes in homogeneous Maxwell's equations: transverse electric mode,'' J. Appl. Phys. **54**, 1179–1189 (1983).

[3] R. W. Ziolkowski, ''Exact solutions of the wave equation with complex source locations,'' J. Math. Phys. **26**, 861–863 (1985).

[4] R. W. Ziolkowski, ''Localized transmission of wave energy,'' Eng. Res. pp. 16–23 (1988).

[5] R. W. Ziolkowski, I. M. Besieris, and A. M. Shaarawi, ''Localized wave representations of acoustic and electromagnetic radiation,'' Proc. IEEE **79**, 1371–1378 (1991).

[6] R. W. Ziolkowski, ''Properties of electromagnetic beams generated by ultrawide bandwidth pulse-driven arrays,'' IEEE Trans. Antennas Propag. **40**, 888–905 (1992).

[7] R. W. Ziolkowski and J. B. Judkins, ''Propagation characteristics of ultrawide-bandwidth pulsed Gaussian beams,'' J. Opt. Soc. Am. A **9**, 2021–2030 (1992).

[8] R. W. Ziolkowski, I. M. Besieris, and A. M. Shaarawi, ''Aperture realizations of exact solutions to homogeneous-wave equations,'' J. Opt. Soc. Am. A **10**, 75–87 (1993).

[9] A. M. Shaarawi, I. M. Besieris, R. W. Ziolkowski, and S. M. Sedky, ''Generation of approximate focus-wave-mode pulses from wide-band dynamic Gaussian apertures,'' J. Opt. Soc. Am. A **12**, 1954–1964 (1995).

[10] S. M. Sedky, A. M. Shaarawi, I. M. Besieris, and F. Taiel, ''Diffraction length of localized waves generated by dynamic apertures,'' J. Opt. Soc. Am. A **13**, 1719–1727 (1996).

[11] P. R. Stepanishen and J. Sun, ''Acoustic bullets: Transient Bessel beams generated by planar apertures,'' J. Acoust. Soc. Am. (accepted for publication).

[12] J. Lu and J. F. Greenleaf, ''Nondiffracting x waves-exact solutions to free-space scalar wave equation and their finite aperture realization,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 19–31 (1992).

[13] J. Lu and J. F. Greenleaf, ''Experimental verification of nondiffracting x waves,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 441–446 (1992).

[14] P. R. Stepanishen, ''Transient radiation from pistons in an infinite planar baffle,'' J. Acoust. Soc. Am. **49**, 1629–1638 (1971).

[15] J. W. S. Rayleigh, *The Theory of Sound* (Dover, New York, 1945).

[16] A. Papoulis, *Signal Analysis* (McGraw-Hill, New York, 1980).

[17] I. S. Gradshetyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, Orlando, 1980).

[18] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 1982).

[19] P. R. Stepanishen, ''Acoustic transients in the far field of a baffled circular piston using the impulse response approach,'' J. Sound Vib. **32**, 295–310 (1974).

[20] H. L. Bertoni, L. Carin, and L. B. Felsen, *Ultra-Wideband, Short-Pulse Electromagnetics* (Plenum, New York, 1993), pp. 285–292.

On coupled modes theory of two-dimensional wave motion in elastic waveguides with slowly varying parameters in curvilinear orthogonal coordinates

V. B. Galanenko
Acoustics Chair, Kiev Polytechnic Institute, av. Pobedy 37, Kiev, 252056, Ukraine

(Received 26 September 1996; revised 15 October 1997; accepted 6 November 1997)

Coupled modes theory, as applied to 2-D wave propagation in range-dependent isotropic elastic waveguides in curvilinear orthogonal coordinates, is proposed. The model consists of one or more curvilinear layers with interfaces and outer boundaries departing slowly from the coordinate surfaces. The material parameters are supposed to be slowly dependent on the longitudinal coordinate. Four-dimensional vector-valued functions with the displacements, rotation, and divergence as the components are shown to be a convenient form of wave field description to extend the coupled modes theory from liquid to elastic waveguides. For this case, the cross section operator is defined and the generalized bi-orthogonality property of its eigenfunctions is established. Using the generalized bi-orthogonality, the coupling equations are derived. Wave propagation in a tube which narrows slowly to a conic rod and in a disk of radially dependent thickness are considered as the examples. The differences between the results obtained by the coupled modes theory and those the multiple-scales theory leads to are discussed. The theory modification for dealing with a cutoff phenomenon and the general algorithm of computation of the wave field excited by a point source are considered. [S0001-4966(98)04403-8]

PACS numbers: 43.20.Mv, 43.20.Ks [JEG]

LIST OF SYMBOLS

$\mathbf{a} = \{a_1, a_2, a_3, a_4\}$ is a row vector with a_1, a_2, a_3, a_4 as the components
$\mathbf{a} = \{a_1; a_2; a_3; a_4\}$ is a column vector with a_1, a_2, a_3, a_4 as the components
$\mathbf{A} = \{\mathbf{a}_1; \mathbf{a}_2; \mathbf{a}_3; \mathbf{a}_4\}$ is a matrix with $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, $\mathbf{a}_4$ as the rows
$\mathbf{A} = \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\}$ is a matrix with $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, $\mathbf{a}_4$ as the columns

INTRODUCTION

Unremitting attention to wave propagation in elastic guiding environments such as single- or multilayered shells, rods of various shapes, etc., was heated over the past few decades by a number of important applications of ultrasound, for example, at nondestructive evaluation, material testing, vibration control, and so on. Though there have been extensive studies on wave propagation based on the shell theories (see Ref. 1 as an example) there are many elastic structures of practical interest that have not been covered by these studies because of the constraints inherent to all shell theories. A wider class of practical problems relating to wave propagation may be successfully treated by using exact equations of elastodynamics. This paper suggests one approach to solve a two-dimensional problem of wave propagation along layered isotropic structures with outer surfaces and interfaces that are slowly departing from the coordinate surfaces in the appropriate curvilinear orthogonal coordinate system. Let s, h, h' be curvilinear coordinates in which the longitudinal axis s is colinear to the wave propagation direction, while the h axis and the h' axis belong to the cross section. The elastic structure's parameters (both in terms of material and shape) as well as the parameters of the wave field are supposed to be invariant to the h' coordinate. The angle between the tangent to the boundary or interface $h = h_i(s)$ at any point of the surface and the $h = \text{const}$ coordinate line passing through this point is assumed to be small. Lame's constants $\lambda(h,s)$, $\mu(h,s)$ and the density $\rho(h,s)$ are supposed to be slowly variable functions of the longitudinal coordinate s. All the layers are rigidly bounded while the outer surfaces of the waveguide are stress-free.

Under such assumptions the structures of interest have to be treated as elastic range-dependent waveguides. Specific examples are rods and tubes of varying radius and/or thickness and elastic layers bounded by surfaces, which are close to cylindrical, spherical, or spheroidal ones, and the like. Guided waves in such the structures have found extensive application in nondestructive evaluation. There are several efficient theoretical tools for dealing with range dependence: the ray theory, the parabolic equation technique, the multiple-scales technique and the theory of coupled modes (known also as the cross section method). Both last theories lead to local normal mode representation, a convenient form of wave field description for many ultrasound applications. In the paper below we are focusing on coupled modes theory. Although the multiple-scales method and coupled modes theory often lead to close leading-order solutions, their methodological difference may be substantial for solving some specific problems. In addition, there are some cases (shown below), where the above theories lead to quite different results, even in their leading-order approximations.

 0001-4966/98/103(4)/1752/11/$10.00

The essence of distinction between these theories will be considered below in brief.

Both methods utilize (each by its own way) the large difference between wavelength and the scale on which the structure parameters change significantly. In fact, the multiple-scales theory employs (explicitly or implicitly) some preliminary assumption on the law of wave propagation along the longitudinal axis to construct one of the particular solutions as an asymptotic series.[2–6] For example, by using this assumption in an explicit form, the solution may be represented by JWKB anzats as a wave with a slowly varying phase speed (e.g. Ref. 6). The wave transverse structure, which depends slowly on the longitudinal coordinate, is represented by power expansion in the small parameter ε, which is proportional to the derivatives of the waveguide's parameters with respect to the longitudinal coordinate. If we introduce fast and slow scales and use either integer or fractional powers of ε, it means that we use the above assumption implicitly.[1] The zeroth-order approximation resulting from the multiple scales theory gives us an eigenfunction of a so-called ''companion waveguide,'' i.e., a range-independent waveguide with the parameters that are equal to the current parameters of the guiding structure under consideration. The higher-order corrections result in changing of the wave's transverse structure, while the propagation factor does not change and the wave retains its identity (there is no conversion between the different wave types). In contrast to the multiple-scales theory the coupled modes method is based on the assumption of the wave-field transverse structure. In this case, the general solution is represented by expansion into the eigenfunctions of a differential operator (a so-called cross section operator). The expansion coefficients depend on the longitudinal coordinate and can be treated as propagation factors of the respective waves. Higher-order corrections result in mutual conversion of different wave types and do not change the wave's transverse structure. The local eigenfunctions and eigenvalues are not always identical to those of the respective ''companion waveguide.'' In this case, the leading-order solution (a so-called adiabatic approximation) does not coincide with that given by the multiple-scales theory.

Since the coupled modes method is based on the completeness and orthogonality property of the eigenfunction set it can be combined with the method of ''partial regions'' and efficiently applied to all problems in which waveguide parameters jump and, consequently, in which wave conversion is an essential feature. Wave reflection at the elastic waveguide's end, wave diffraction at steps, hollows, inclusions, delaminations and the alike are particular examples of such situations.

While the multiple-scales method is widely used in studies on sound wave propagation in both liquid and elastic guiding environments,[2–6] the coupled modes method has been applied (as far as we know) only to liquid waveguides in the context of underwater sound problems.[7–9] Note that this method has also found application to studies of wave propagation in electromagnetic waveguides.[10,11] The main purpose of this paper is to extend the coupled modes method to studies of wave propagation in range-dependent elastic waveguides.

The main part of the paper is devoted to a systematic exposition of the coupled modes theory formalism as applied to elastic waveguides (Secs. I, III, IV). In the first section of this paper a cross section operator is defined, wave field expansion into eigenfunctions is obtained and the main equations of the coupled modes theory are derived. The second section contains two examples which are considered to illustrate some characteristic features of the method as applied to elastic waveguides. Wave propagation in a conic shell which is tapered to a conic rod and axially symmetric waves within a layer of radially dependent thickness are chosen as the examples. In particular, these examples demonstrate a difference between the results that can be obtained by using the coupled modes theory and the multiple-scales method. They also show that a cutoff phenomenon in the case of elastic waveguides (as regards to the P-SV waves) is related to the merging of the eigenvalues corresponding to the forward and backward waves. The theory modification for the cutoff cross sections is discussed in the third section of the paper. Since the whole mathematical formulation of the propagation problem involves the specification of the source and the border conditions, these issues are treated in the last section where computation of the wave field excited by a point source is considered.

I. COUPLED MODES EQUATIONS

To generalize the coupled modes method to elastic guiding environments it is first necessary to define the cross section (local) operator in every cross section of the waveguide. The operator's domain is a set of so-called cross section functions, i.e., of such differentiable functions of the variables s and h which comply with the boundary conditions as functions of the transverse coordinate h only. The differential expression corresponding to this operator must involve differentiation of the cross section functions only with respect to h and has to be extracted from the equation governing wave motion. It is not a trivial problem because its feasibility for elastic waveguides depends significantly on an accepted form of wave field description. For example, the conventional description in potentials does not permit + to define the set of cross section functions because, in this case, the boundary conditions contain differentiation of the potentials with respect to the longitudinal coordinate. The description using the 4-D vector-valued ''displacements-stresses'' functions (Kennet[12]) is not appropriate as well. We get a convenient description by using displacements, divergence and rotation vector components. Such a description has been successfully used for the first time (as far as we know) by Bobrovnitskii[13] to study the Lamb's waves orthogonality problem.

It is known that the 2-D wave motion breaks up into two independent families of waves: P-SV waves and SH waves. Analysis of the equation for the SH waves demonstrates the proximity of the SH-waves' mathematical nature to that inherent to waves in liquid waveguides. So, in the case of the SH waves, the coupled modes method formalism is basically

the same as that of sound waves in liquid waveguides. We focus below on the *P-SV* waves.

As far as the boundary conditions are evident from the continuity of displacements and stresses at interfaces and the vanishing of stresses at outer boundaries we express the stresses in terms of the displacements u,w (u is a longitudinal displacement, while w is a transverse one), the divergence d and the rotation component c (tangential to h' axis). Stress vector τ with the normal τ_n and tangential τ_l components belongs to the ''s,h'' coordinate surface. Starting from the known relation for the stress vector τ at an area element with $\mathbf{n}$ as its normal

$$\tau=\lambda\mathbf{n}\ \mathrm{div}\ \mathbf{u}+2\mu\frac{\partial\mathbf{u}}{\partial n}+\mu(\mathbf{n}\times\mathrm{rot}\ \mathbf{u}) \tag{1}$$

(where $\mathbf{u}$ is the displacement vector) and taking the use of expressions of rot $\mathbf{u}$ and div $\mathbf{u}$ we obtain

$$\tau_n=\frac{1}{1+\varepsilon^2}\left\{(\lambda+\varepsilon^2\chi)d+2\mu\left[\frac{1-\varepsilon^2}{H_2}\frac{\partial w}{\partial h}-\frac{2\varepsilon}{H_2}\frac{\partial u}{\partial h}-\varepsilon c+\left((1-\varepsilon^2)\frac{H_{2s}}{H_{12}}-\varepsilon^2\frac{H_{3s}}{H_{13}}\right)u+\left(\varepsilon\frac{2H_{2s}}{H_{12}}-\varepsilon^2\frac{H_{3h}}{H_{13}}\right)w\right]\right\}, \tag{2}$$

$$\tau_l=\frac{\mu}{1+\varepsilon^2}\left\{(1-\varepsilon^2)c+2\left[\frac{1-\varepsilon^2}{H_2}\frac{\partial u}{\partial h}+\frac{2\varepsilon}{H_2}\frac{\partial w}{\partial h}-\varepsilon d+\varepsilon\left(\frac{2H_{2s}}{H_{12}}+\frac{H_{3s}}{H_{13}}\right)u-\left((1-\varepsilon^2)\frac{H_{2s}}{H_{12}}-\varepsilon\frac{H_{3h}}{H_{23}}\right)w\right]\right\}, \tag{3}$$

where $H_1=H_1(s,h)$, $H_2=H_2(s,h)$, $H_3=H_3(s,h)$ are Lame's coefficients corresponding to a specific curvilinear coordinate system; the subscripts h and s refer to differentiation with respect to h or s; $H_{123}=H_1H_2H_3$; $H_{ik}=H_iH_k$; $\varepsilon=\varepsilon_i=(H_2/H_1)dh_i/ds$ (geometrically $\varepsilon_i=\tan\alpha_i$ where α_i is the angle of layer boundary departure from the coordinate surface $h=$const containing the considered point). Since expressions (2), (3) and, hence, the boundary conditions at interfaces and outer boundaries do not contain differentiation of the accepted vector-valued functions with respect to the longitudinal coordinate s, the set of all piecewise differentiable 4-D vector-valued functions $\mathbf{p}$:$\{u;c;w;d\}$ may be considered as the set of cross section functions.

Lame's equation for inhomogeneous isotropic medium:[14]

$$\chi\ \mathrm{grad\ div}\ \mathbf{u}-\mu\ \mathrm{rot\ rot}\ \mathbf{u}+\mathrm{grad}\ \lambda\cdot\mathrm{div}\ \mathbf{u}+2(\mathrm{grad}\ \mu,\mathrm{grad})\mathbf{u}+\mathrm{grad}\ \mu\times\mathrm{rot}\ \mathbf{u}+\omega^2\rho\mathbf{u}=0$$

(where $\chi=\lambda+2\mu$) being expressed in the terms of displacements, divergence and rotation component, reads:

$$\mathbf{A}_1\frac{\partial\mathbf{p}}{\partial h}+\mathbf{B}_1\mathbf{p}+\mathbf{C}_1\frac{\partial\mathbf{p}}{\partial s}=0, \tag{4}$$

where $\mathbf{C}_1=\{1/H_1,\ 0,\ 0\ 0;\ 0,\ \mu/H_1,\ 2\mu_s/{H_1}^2,\ 0;\ 0,\ 0,\ 1/H_1,\ 0;\ 2\mu_s/H_1^2,\ 0,\ 0,\ \chi/H_1\}$. Due to the matrix $\mathbf{C}_1$ nonsingularity we can transform the last equation to the form with partially separated variables:

$$\mathbf{A}\frac{\partial\mathbf{p}}{\partial h}+\mathbf{B}\mathbf{p}+\frac{\partial\mathbf{p}}{\partial s}=0, \tag{5}$$

where the matrix-valued coefficients $\mathbf{A}$ and $\mathbf{B}$ make: $\mathbf{A}=H_1/H_2\{0,0,1,0;2\mu_s/\mu H_1,0,2\mu_h/\mu H_1,\chi/\mu;-1,0,0,0;2\mu_h/\chi H_2,-\mu/\chi,-2\mu_s/\chi H_1,0\}$, $\mathbf{B}=\{(H_{23})_s/H_{23},0,(H_{13})_h/H_{23},\ -H_1;\ (H_{1h}/H_{12})(2\mu_s/\mu),\ H_{3s}/H_3+\mu_s/\mu,\ H_1k_t^2-(H_{2s}/H_{12})(2\mu_s/\mu),\ (H_1/H_2)(\lambda_h/\mu);\ -H_{1h}/H_2,\ -H_1,\ H_{2s}/H_2,\ 0;\ H_1k_l^2-(H_{23})_s/H_{123}(2\mu_s/\chi),\ -(H_1/H_2)(\mu/\chi)H_{3h}/H_3-\mu_h/\chi),-(H_{13})_h/H_{123})(2\mu_s/\chi),\chi_s/\chi\}$. In the case of multilayered structure, the vector-valued function $\mathbf{p}$ and the matrix-valued coefficients $\mathbf{A}$ and $\mathbf{B}$ are defined as piecewise functions. For each waveguide's cross section ($s=$const), the local operator $\mathbf{L}=\mathbf{A}d/dh+\mathbf{B}$ [extracted from (5)] may be accepted as the cross section operator with the set of cross section functions as the domain.

Any solution of the problem may be represented as $\mathbf{p}=\mathbf{L}^{-1}\mathbf{b}$ (where $\mathbf{b}=-\partial\mathbf{p}/\partial s$). Introducing operator $\mathbf{L}+i\gamma\mathbf{I}$ (where $\mathbf{I}$ is the identity operator) one can transform the last expression in the usual way[15] to the sum of residues of the poles of the meromorphic matrix-valued Green's function $\mathbf{G}(h,h',\gamma)$ of the operator $(\mathbf{L}+i\gamma\mathbf{I})^{-1}$ in complex γ-plane. As every residue is proportional to the respective eigenfunction of the operator $\mathbf{L}$, any solution of the wave propagation problem may be represented by a series in eigenfunctions (all poles are assumed to be simple) :

$$p=\sum_{n=1}^{\infty}[a_n^+(s)\mathbf{p}_n^+(h,s)+a_n^-(s)\mathbf{p}_n^-(h,s)], \tag{6}$$

where $\mathbf{p}_n^+$ and $\mathbf{p}_n^-$ correspond to the poles belonging, respectively, to the right ($\mathrm{Re}(\gamma)>0$) or left ($\mathrm{Re}(\gamma)<0$) half-plane. Thus the cross section eigenfunctions show completeness in the set of all possible solutions of the wave propagation problem.

The eigenfunctions may be derived as the solutions of the equation:

$$\mathbf{A}\frac{\partial\mathbf{p}_n}{\partial h}+(\mathbf{B}+i\gamma\mathbf{I})\mathbf{p}_n=0 \tag{7}$$

under the following boundary conditions:

$$(\mathbf{T}_{i+1}\mathbf{p})_{h_i+0}-(\mathbf{T}_i\mathbf{p})_{h_i-0}+\frac{2i\gamma(\mu_i-\mu_{i+1})}{H_1}(\mathbf{R}\mathbf{u})_{h_i}=0 \tag{8}$$

at interfaces between the ith and $(i+1)$th layers and

$$\left(\mathbf{T}\mathbf{p}-\frac{2i\gamma\mu}{H_1}\mathbf{R}\mathbf{u}\right)_{h=h_0,h_N}=0 \tag{9}$$

at the outer boundaries. Here $\mathbf{T}$ and $\mathbf{R}$ are the 2×4 and 2×2 matrices:

$$\mathbf{T}=\{-2\mu\chi_{11},-(1-\varepsilon^2)\mu,\ 2\mu\chi_{13},\ 2\varepsilon\mu;\ -2\mu\chi_{21},\ 2\varepsilon\mu,-2\mu\chi_{23},\chi+\varepsilon^2\lambda\}$$

$$\mathbf{R}=\{2\varepsilon,-1+\varepsilon^2;1-\varepsilon^2,2\varepsilon\}$$

where

$$\chi_{11}=(1-\varepsilon^2)H_{1h}/H_{12}+\varepsilon H_{3s}/H_{13};$$

$$\chi_{13}=-\varepsilon(2H_{1h}/H_{12}+H_{3h}/H_{23});$$

$$\chi_{21}=(H_{3s}/H_{13}-2\varepsilon H_{1h}/H_{12});$$

$$\chi_{23}=(1-\varepsilon^2)H_{1h}/H_{12}+H_{3h}/H_{23}.$$

The next question which arises is whether the eigenfunctions show orthogonality property. It turns out that the eigenfunctions $\mathbf{p}_n^+$ and $\mathbf{p}_n^-$ are bi-orthogonal to the eigenfunctions $\mathbf{q}_n^+$ and $\mathbf{q}_n^-$ of the adjoint operator under the respective definition of a scalar product denoted here as $[\mathbf{p},\mathbf{q}]$ (see Appendix A for more details on the generalized bi-orthogonality condition). Inserting (6) to (5) and using generalized bi-orthogonality property one can obtain the system of coupled equations with respect to the propagation factors $a_n^+(s)$ and $a_n^-(s)$

$$\frac{da_n^+}{ds}-i\gamma_n^+(s)a_n^+ +\frac{1}{N_n^+}\sum_{m=1}^{\infty}(C_{mn}^{++}a_m^+ +C_{mn}^{-+}a_m^-)=0,$$
$$\frac{da_n^-}{ds}-i\gamma_n^-(s)a_n^- +\frac{1}{N_n^-}\sum_{m=1}^{\infty}(C_{mn}^{+-}a_m^+ +C_{mn}^{--}a_m^-)=0, \tag{10}$$

where γ_n^+ and γ_n^- are eigenvalues belonging to the right or the left half-plane, $C_{mn}=[\partial\mathbf{p}_m/\partial s,\ \mathbf{q}_n]$ are the nonnormalized coupling coefficients (with the first superscript corresponding to $\mathbf{p}$ and the second corresponding to $\mathbf{q}$); $N_n=[\mathbf{p}_n,\ \mathbf{q}_n]$ is the generalized norm of the eigenfunction $\mathbf{p}_n$. On condition that the coefficients of the intermode coupling C_{mn} (for $m\neq n$) are sufficiently small to be negligible, we obtain equations of the adiabatic theory as follows:

$$N_n^+\left(\frac{da_n^+}{ds}-i\gamma_n^+a_n^+\right)+C_{nn}^{++}a_n^+ +C_{nn}^{-+}a_n^-=0,$$
$$N_n^-\left(\frac{da_n^-}{ds}-i\gamma_n^-a_n^-\right)+C_{nn}^{--}a_n^- +C_{nn}^{+-}a_n^+=0. \tag{11}$$

These equations ignore the intermode coupling but account for the intramode coupling, i.e., the coupling of the forward and backward waves.

If $N_n^+\gg C_{nn}^{++}$, C_{nn}^{-+} and $N_n^-\gg C_{nn}^{--}$, C_{nn}^{+-} then the adiabatic approximation (in a narrow sense) leads to the solution consisting of independent forward and backward waves:

$$a_n^+(s)=a_n^+(s_1)\exp\left(i\int_{s_1}^{s}\gamma_n^+(s)ds\right),$$
$$a_n^-(s)=a_n^-(s_1)\exp\left(i\int_{s_1}^{s}\gamma_n^-(s)ds\right). \tag{12}$$

Note that the eigenvalues and the eigenfunctions used here are not identical to those of the ''companion waveguide'' because the boundary conditions [see (8) and (9)] and the matrix-valued coefficients of Eq. (5) contain derivatives of the waveguide's parameters with respect to the longitudinal coordinate. Although these derivatives are small values of $O(\varepsilon)$ they can sometimes affect a solution substantially. In such a case, adiabatic approximation differs essentially from the zeroth-order approximation of the multiple-scales method based on the ''companion waveguide'' approximation and leads, as we suppose, to more correct results. Below we illustrate such a situation by some examples.

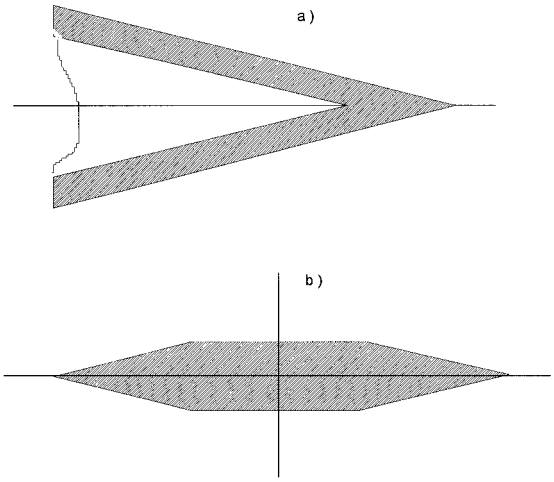


FIG. 1. The range-dependent configurations that are considered as the examples. (a) A tapering tube. (b) A disk of the variable thickness.

Variation of the adiabatic amplitudes controlled formally by imaginary parts of the eigenvalues stems substantially from the energy dissipation, cross section square variation, etc. Consequently, the imaginary part of the eigenvalue may be either positive or negative. In some instances it may be reasonable to isolate the factor corresponding to the cross section square variation. Then, introducing $\mathbf{p}'(h,s)=(\nu(s))^{1/2}\,\mathbf{p}(h,s)$ (where $\nu(s)$ is a relative cross section square) one can transform Eq. (5) to the following equation about $\mathbf{p}'(h,s)$:

$$\mathbf{A}\frac{\partial\mathbf{p}'}{\partial h}+\left(B-\frac{\nu_s}{2\nu}\mathbf{I}\right)\mathbf{p}'+\frac{\partial\mathbf{p}'}{\partial s}=0. \tag{13}$$

The solution of this equation has the form of (6) with the propagation factors $a_n'^{+}(s)$ and $a_n'^{-}(s)$ that are the solutions of equation system (10) with $\gamma_n'^{+}=\gamma_n^+-i\nu_s/2\nu$ and $\gamma_n'^{-}=\gamma_n^-+i\nu_s/2\nu$ in place of γ_n^+ and γ_n^-.

As far as the coefficients of the coupled modes equation system (10) depend on variation of local eigenvalues and eigenfunctions along the longitudinal coordinate s, the first step we have to take in order to solve Eq. (10) is to analyze the behavior of the eigenvalues and the eigenfunctions with the s-coordinate variation. The next section contains two examples of such an analysis.

II. AXIALLY SYMMETRIC WAVES IN HOMOGENEOUS TUBE, ROD AND LAYER WITH STRESS-FREE BOUNDARIES

To illustrate the eigenvalues' and eigenfunctions' evolution caused by range dependence we consider a circular tube of slowly decreasing radius (transforming to a tapering rod) and a disk of radially dependent thickness as the examples (see Fig. 1). For definiteness sake, the thickness of the tube wall is supposed to be invariant while the tube radius (as well as the disk thickness) is assumed to be varying linearly with the longitudinal coordinate. It is convenient to choose

the cylinder coordinate system with the longitudinal coordinate $s=z$ and the transverse coordinate $h=r$ for a tube and $s=r$, $h=z$ for a disk.

Having substituted $H_1=H_2=1$, $H_3=r$, $\mu_z=\mu_r=\lambda_z=\lambda_r=0$ into the matrix-valued coefficients of Eq. (10) for a tube one can obtain its general solution as follows: $\mathbf{Z}=\{-i\gamma K_0(\beta_l r),\ i\gamma K_0(-\beta_l r),\ \beta_t K_0(\beta_t r),\ -\beta_t K_0(-\beta_t r);\ 0,\ 0,\ -k_t^2 K_1(-\beta_t r);\ \beta_l K_1(\beta_l r),\ -\beta_l K_1(-\beta_l r),\ i\gamma K_1(\beta_t r),\ i\gamma K_1(-\beta_t r);\ k_l^2 K_0(\beta_l r),\ k_l^2 K_0(-\beta_l r),0,0\}$. Here $\beta_l=(\gamma^2-k_l^2)^{1/2}$, $\beta_t=(\gamma^2-k_t^2)^{1/2}$ and $K_n(\beta r)$ are the modified Bessel functions. In the case of a disk bounded by surfaces $z=z_1(r)$ and $z=z_2(r)$, it is preferable to use equation (13) (with $\partial\mathbf{p}'/\partial s$ replaced formally with $i\gamma\mathbf{p}'$) to account for variation of the cross section square with the radial distance increase. Its general solution is as follows: $\mathbf{Z}=\{-i\gamma' c_l,\ 0,\ \beta_l s_l,\ k_l^2 c_l;\ -i\gamma' s_l/\beta_l,\ 0,\ -c_l,\ k_l^2 s_l/\beta_l;\ -\beta_t s_t,\ k_t^2 c_t,\ i\gamma'' c_t,\ 0;\ -c_t,\ -k_t^2 s_t/\beta_t,\ i\gamma'' s_t/\beta_t,\ 0\}^T$ where: superscript T indicates matrix transpose; $i\gamma'=i\gamma-1/(2r)-(\varepsilon_2-\varepsilon_1)/2(z_2-z_1)$, $i\gamma''=i\gamma+1/(2r)-(\varepsilon_2-\varepsilon_1)/2(z_2-z_1)$, $\varepsilon_k=dz_k/dr$ $(k=1,2)$; $\beta^2=\gamma'\gamma''-k^2$; $c=\cosh(\beta(z+z_0))$, $s=\sinh(\beta(z-z_0))$, $z_0=(z_1+z_2)/2$ with the subscripts l and t corresponding to longitudinal or shear wave's parameters.

Let $\mathbf{S}=\{\mathbf{S}_1;\ \mathbf{S}_2\}$ be the boundary condition matrix with $\mathbf{S}_1$ and $\mathbf{S}_2$ being 2×4 matrices which correspond to a stress released boundary and may be obtained from (9). For a tapering tube: $\mathbf{S}_n=\{-2i\gamma\mu(1-\varepsilon_n^2),\ 2\varepsilon_n\mu,\ -2\mu/r_n-4i\gamma\mu\varepsilon_n,\ \chi+\varepsilon_n^2\lambda;\ -4i\gamma\varepsilon_n,\ -1+\varepsilon_n^2,\ -2\varepsilon_n/r_2+2i\gamma(1-\varepsilon_n^2),\ 2\varepsilon_n\}$ $(n=1,2)$; for a rod: $\mathbf{S}_1=\{0,r,0,0;0,0,r,0\}_{r\to0}$ and for a disk: $\mathbf{S}_k=\{-2i(\gamma''-\varepsilon_k^2\gamma'),\ 2\varepsilon_k,\ -4i\gamma'\varepsilon_k,\ (\chi+\varepsilon_k^2\lambda)/\mu;\ 4\varepsilon_k(i\gamma-(\varepsilon_2-\varepsilon_1)/(z_2-z_1)),\ -1+\varepsilon_k^2,\ 2i\gamma'(1-\varepsilon_k^2),\ 2\varepsilon_k\}_{z=z_k}$. The local eigenvalues can be found as the solutions of equation $\det(\mathbf{SZ})=0$.

Having solved this equation numerically, the local eigenvalues and eigenfunctions of the waveguides (plotted schematically in Fig. 1) were computed. A steel conic tube (with the wall thickness equal to 0.2) which gets narrow in the positive direction of the z axis and turns to a conic rod of which the conic angle $\theta\approx11.5°$ $(\varepsilon=-0.1)$ and a steel disk of the same initial thickness which then decreases linearly (with $\varepsilon=-0.1$) were chosen as the specific examples. Hereafter we use dimensionless distance, sizes and wave numbers (eigenvalues) normalized to respectively a shear wave's wavelength and a wave number. There are only two Lamb's waves that can propagate in a flat steel layer of the same thickness: symmetric and antisymmetric ones. For further presentation we prefer to use the same designations (''symmetric'' and ''antisymmetric'') as well for the waves propagating in tapering tube, rod and disk.

Due to the variation of the cross section geometry with the longitudinal coordinate (z for a tube or a rod, r for a disk), the eigenvalues (or, what is same, wave numbers) of both waves move along some trajectories in a complex γ plane and the waves change their transverse structure (described by the eigenfunctions) to comply with the local geometry. The trajectories of the eigenvalues corresponding to the ''symmetric'' and ''antisymmetric'' (forward and backward) waves in a tube, tapering to a rod, are plotted in Fig. 2. (The history of the eigenvalues corresponding to the cylindrical waves in the tapering disk is shown in Figs. 4 and 5.)

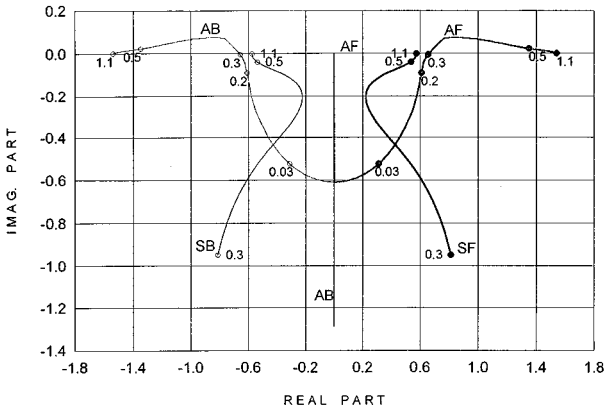


FIG. 2. The trajectories of the local eigenvalues in the complex γ plane for a tapering tube transforming to a conic rod. The labels AF, AB indicate the trajectories for the ''antisymmetric'' forward and backward waves respectively; SF, SB indicate the same for the ''symmetric'' waves. The labels on the curves indicate the normalized local outer radius.

It is interesting to compare the local eigenvalues of the tapering tube and rod with those supported by the respective ''companion'' waveguides (see Fig. 3). As long as the outer radius of the tube is sufficiently big $(10>r'>1)$, the compared eigenvalues are close to one another and to the respective wave number of a plane Lamb's wave. Due to $\varepsilon\neq0$, the local eigenvalues of a tapering tube acquire small imaginary parts (negative for the ''symmetric'' wave, first positive and then negative for the ''antisymmetric'' one), while the ''companion'' tube's eigenvalues are real. But with the further decreasing of the tube radius, these eigenvalues vary more widely. Whereas the ''symmetric'' wave's local wave number supported by the ''companion'' tube vanishes at $r_2'=0.3913$ and then appears as an imaginary value, the respective wave number for the tapering tube does not vanish. Its real part first decreases down to 0.223 at $r_2'=0.3925$ but afterward increases (see Fig. 3), the absolute value of its imaginary part grows rapidly and its trajectory in a complex

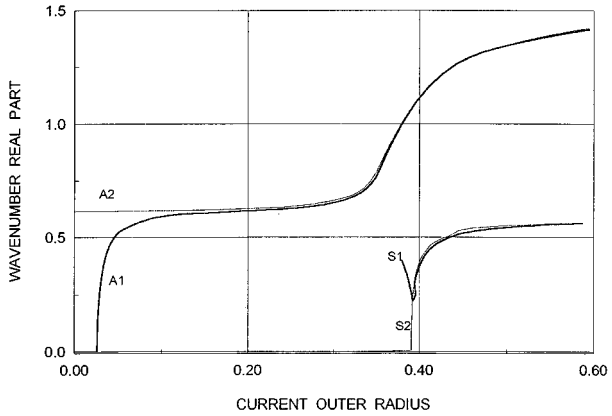


FIG. 3. The local wave number real part of the ''symmetric'' and ''antisymmetric'' wave versus the normalized current outer radius. Labels A1 and S1 indicate the curves corresponding to the ''antisymmetric'' and ''symmetric'' wave, respectively, in a tapering tube transforming to a conic rod. Labels A2 and S2 indicate the same for the respective ''companion waveguide.''

γ-plane turns back. The local wave number of the ''antisymmetric'' wave in a tapering tube gives a smooth way to that corresponding to the first mode of a conic rod at $r_2'=0.2$ (where the conic tube transforms into the conic rod). At first, its real part is close to the respective eigenvalue of the ''companion waveguide,'' but, with further diminution of the rod radius, these eigenvalues behave variously: the ''companion'' rod's eigenvalue tends to some constant value while the real part of the tapering rod's eigenvalue vanishes at $r_2'=0.025$ (see Fig. 3).

Thus the coupled mode theory and the theory based on the ''companion waveguide'' approximation give quite different predictions on the behavior of the ''symmetric'' and ''antisymmetric'' waves in a tapering tube and rod. In particular, the coupled mode theory predicts a cutoff of the compressional mode in a tapering rod and points that the local eigenvalues of the forward and backward waves merge together at the cutoff cross section (see Fig. 2).

The transverse structure of the ''symmetric'' and ''antisymmetric'' modes in a tapering tube first allows to categorize the ''symmetric'' mode as a compressional wave and the ''antisymmetric'' mode as a wave being close to the flexural (as long as the tube radius is sufficiently big). With the tube's radius decreasing, the distributions of the vertical and horizontal displacements corresponding to these modes change progressively. Above $r_2'=0.39$ the ''symmetric'' mode becomes closer to the flexural wave, while the ''antisymmetric'' mode rearranges gradually into a wave closely related to the compressional one within interval $0.2<r_2'<0.4$. At $r_2'=0.2$, where the conic tube transforms into the conic rod, the distributions of horizontal and vertical displacements and divergence over the tube cross section pass gradually into those being typical for the first mode of a rod. Distribution of the rotation varies in a different way. Over the tube's cross sections that are close to the intermediate one (which separates the tube from the rod), the rotation decreases almost linearly in the direction from the outer to the inner boundary (in the same manner as over the rod's cross section) except for the narrow region nearby the inner boundary where it increases rapidly up to some finite value. The closer the cross section is to the intermediate one, the narrower this region is. It tends to zero with $r_1'\to 0$. Thus within intermediate cross section ($r_2'=0.2$), the vector valued eigenfunctions of a narrowing tube and a rod coincide, except for point $r=0$ where the rotation shows a finite jump caused by the boundary conditions jump.

Analysis of variation of the eigenvalues and eigenfunctions corresponding to the ''symmetric'' and ''antisymmetric'' cylindrically crested waves in a tapering disk leads to closely resembled results. In particular, the symmetric (compressional) mode in a tapering disk shows a cutoff (see Fig. 4) at the radial distance $r=2{,}865$ (the respective local thickness is equal to 0.027), where the eigenvalue's real part vanishes and the eigenvalues of the forward and backward waves merge together. Note that the eigenvalue of the symmetric mode for the corresponding ''companion disk'' does not vanish. The eigenvalue of the antisymmetric wave shows a progressive increase of its real and imaginary parts towards the apex (see Fig. 5). A peculiarity of cylindrically crested

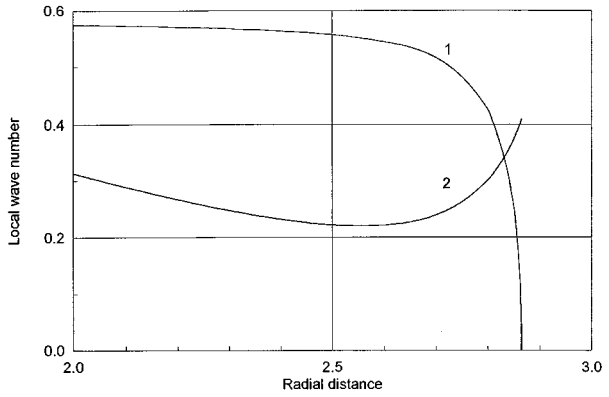


FIG. 4. The history of the local wave number of the ''symmetric'' wave in a tapering disk (1—wave number real part; 2—wave number imaginary part $\times 100$).

waves in a disk is that the intramode and intermode coupling exists even for a layer of the constant thickness. This fact is caused formally by the dependence of the matrix-valued coefficients of equations (10) and (13) on the longitudinal coordinate r. Hence the local general solution (see the above expression) and, consequently, the local eigenvalues and eigenfunctions depend on the longitudinal coordinate, even though $\varepsilon_1=\varepsilon_2=0$. In the considered case of cylindrical waves, this fact has a clear physical explanation: because the cross section curvature varies with the radial coordinate r, the motion of the elastic medium varies in its turn. It is quite equivalent to the variation of the waveguide's shape parameters.

Consider axisymmetric waves within a layer of constant thickness ($z_1=-\Delta$, $z_2=\Delta$) in more details. The equations in eigenvalues of the cylindrical modes are shown to be as follows:

$$(2\gamma'\gamma''-k_t^2)c_l\frac{s_t}{\beta_t}-4\gamma'\gamma''\beta_l s_l c_t=0,$$

$$(2\gamma'\gamma''-k_t^2)c_t\frac{s_l}{\beta_l}-4\gamma'\gamma''\beta_t s_t c_l=0.$$

Here $c=\cosh(\beta\Delta)$, $s=\sinh(\beta\Delta)$ with β being provided by the respective subscript.

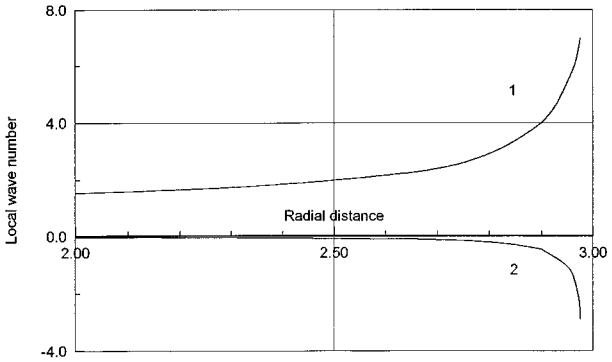


FIG. 5. The history of the local wave number of the ''antisymmetric'' wave in a tapering disk (1—wave number real part; 2—wave number imaginary part $\times 100$).

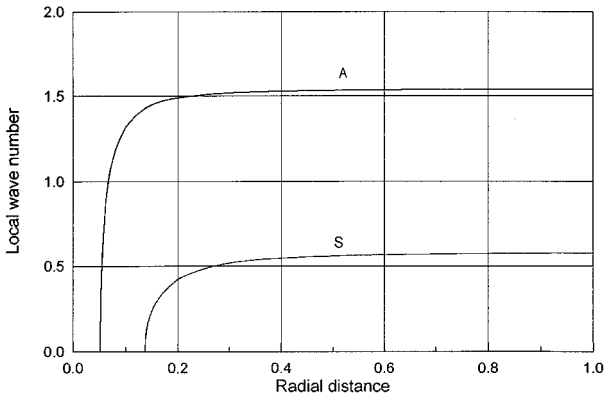


FIG. 6. The local wave numbers of the cylindrically crested waves in a flat homogeneous layer versus the radial distance (A—antisymmetric wave; S—symmetric wave).

With $r\to\infty$, these equations tend to those corresponding to the symmetric and antisymmetric Lamb's waves. Thus the elastic layer response to a point source passes into Lamb's waves only asymptotically.

Figure 6 illustrates the wave numbers' variation via the radial coordinate for a steel layer with the relative thickness equal to 0.2. There are two propagating Lamb's waves in such a layer. The wave numbers of their cylindrical counterparts differ substantially from those of Lamb's waves at small radial distances only. There are cutoff cross sections for every type of wave (at $r=r_n$), where the eigenvalues of the waves that spread towards or away from the center vanish and merge together.

In general case we must take into account that the origin is a singularity point of Eqs. (5) and (13). Straightforward usage of the coupled mode technique in the vicinity involving the origin is expected to be unjustified. As a consequence, the coupled modes method does not permit to compute the amplitudes of the waves excited by a point source. One possible way is to subdivide the layer into two regions by the cylindrical surface of a radius r_0. For the inner region ($r<r_0$), the layer of real geometry is supposed to be approximated with a layer of constant thickness. The exact solution of Eq. (5) for such a layer can be found (see Appendix B). The solution for the external region ($r>r_0$) can be represented by expansion (10). The border values $a_n^+(r_0.)$ and $a_n^-(r_0.)$ can be found using the generalized bi-orthogonality condition (see Appendix A).

Considered examples point to some special situations which can arise when using the coupled modes method. (1) If the coefficients of Eq. (5) or (13) are singular at some point, the usage of the coupled modes method is available only in combination with the method of partial regions. In this case, the exact solution within the nearest vicinity of the singularity point obtained by any other method has to be complied with the coupled mode solution given by series (6). (2) There are regions of fast variation of the local eigenfunctions and eigenvalues, where mode coupling may be substantial. In particular, intramode coupling is expected to result in wave reflection. (3) There are some cross sections where eigenvalues of the forward and backward waves merge together (the cutoff cross sections). The coupled modes theory has to be complemented by special consideration of this case because the above assumption on simple poles of the Green's matrix is not true. The next section of the paper is devoted to cutoff cross sections.

III. A CUTOFF CROSS SECTION

In this section we analyze the mth adiabatic mode passing through its cutoff cross section (the mode number m is omitted for the further notation). Green's matrix possesses a pole of the second order in a cutoff cross section ($s=s_0$) where eigenvalues of the forward and backward waves merge together. Hence, the expansion in eigenfunctions given by (6) should be replaced by the another one corresponding to a pole of the second order:

$$\mathbf{p}=\sum_{n\neq m}(a_n^+\mathbf{p}_n^+ + a_n^-\mathbf{p}_n^-)+b_1\mathbf{p}_{m1}+b_2\mathbf{p}_{m2}. \tag{14}$$

At the same time expansion (6) is available for any small punctured vicinity of the cutoff cross section. Equation (11) (governing adiabatic mode propagation) is to be solved separately for the region $s<s_0$ (on the left side of a cutoff cross section) and for $s>s_0$ (on the right). The solutions are to be complied with one another by the condition of the wave field continuity. The first question which arises is whether the general solution of the coupling equation (11) is regular or singular at $s\to s_0$. This depends on the behavior of the eigenvalues γ^+ and γ^-, the eigenfunctions $\mathbf{p}^+$ and $\mathbf{p}^-$ and the coupling coefficients C^{++}, C^{-+}, C^{+-} and C^{--} under $s\to s_0$. Let $x=s_0-s$. With $x\to 0$ from the left side of a cutoff cross section ($x>0$), the eigenvalues γ^+ and γ^- move in a complex γ-plane toward one another along the trajectories which are symmetric to the imaginary axis (see Fig. 2). Their real parts tend to zero and their imaginary parts tend to the common limit $\gamma=\gamma_0$ in a merging point at the imaginary axis. On the right side of the cutoff cross section ($x<0$), the eigenvalues move apart along the imaginary axis (this corresponds to transformation of travelling waves into nonpropagating ones). Their locations at the imaginary axis are not symmetric to the merging point. Such a behavior of the eigenvalues enable us to assume that the functions $\gamma^+(x)$ and $\gamma^-(x)$ can be expanded in powers of $t=x^{1/2}$:

$$\gamma^+-\gamma_0=\sum_{k=1}^{\infty}\nu_k t^k;\quad \gamma^--\gamma_0=\sum_{k=1}^{\infty}(-1)^k\nu_k t^k. \tag{15}$$

With $x\to 0$, the eigenfunctions $\mathbf{p}^+$ and $\mathbf{p}^-$ tend to the common limit $\mathbf{p}_0$, the adjoint eigenfunctions tend to $\mathbf{q}_0$ and their generalized norms N^+ and N^- tend to zero. When considering the behavior of the eigenfunctions in a small vicinity of a cutoff cross section one needs to keep in view that there are two kinds of their dependence on x: by way of the eigenvalues $\gamma^+(x)$, $\gamma^-(x)$ and by way of the material and shape parameters of the waveguide. The last sort of dependence introduces the values of $O(\varepsilon)$ on computing the derivatives of the eigenfunctions with respect to the longitudinal coordinate. According to the conventional definition of the adiabatic approximation, these values should be neglected. Therefore we neglect this kind of dependence on x below.

This enables us to express the eigenfunctions as power series in $(\gamma^{\pm}-\gamma_0)$ near a cutoff cross section:

$$\mathbf{p}^{\pm}=\mathbf{p}_0+\sum_{k=1}^{\infty}\frac{\mathbf{p}_0^{(k)}}{k!}(\gamma^{\pm}-\gamma_0)^k;$$
$$\mathbf{q}^{\pm}=\mathbf{q}_0+\sum_{k=1}^{\infty}\frac{\mathbf{q}_0^{(k)}}{k!}(\gamma^{\pm}-\gamma_0)^k. \tag{16}$$

The abovementioned constraint is not essential for treating a wave field near a cutoff cross section. It is used here only for more brief and clear presentation. The more exact and detailed computation which is free of this constraint gives the close result differing only by the values of $O(\varepsilon)$.

Using (15) and (16), one can ensure that the generalized norms N^+ and N^- are values of $O(x^{1/2})$ and the coupling coefficients C^{++}, C^{-+}, C^{+-}, and C^{--} are values of $O(x^{-1/2})$. Taking into account such a behavior of the equation coefficients one can expect the solution of Eq. (11) to be singular at $x\to 0$. At the same time the wave field of the adiabatic mode defined by the expression: $\mathbf{p}=a^+\mathbf{p}^+ + a^-\mathbf{p}^-$ is known not to be singular in a cutoff neighborhood. Using (15) and (16), this expression can be written in a small vicinity of a cutoff cross section as follows:

$$\mathbf{p}\approx(a^+ + a^-)\mathbf{p}_0+(a^+ - a^-)2\nu_1 x^{1/2}\mathbf{p}_0{}^{(1)}, \tag{17}$$

where the bracketed superscript indicates the order of the derivative with respect to the eigenvalue at γ^+ and $\gamma^- \to \gamma_0$. In line with this formula the sum $(a^+ + a^-)$ is expected not to be singular, while the remainder $(a^+ - a^-)$ is expected to be singular with the order of the singularity not stronger than $O(x^{-1/2})$. Moreover, one can expect that the function $(a^+ - a^-)$ shows just the same order of the singularity because expression (17) has to possess no less than two arbitrary constants at $x\to 0$.

Introducing the new variable $t=x^{1/2}$, Eq. (11) can be expressed in the following form:

$$t\,\frac{d\mathbf{y}}{dt}=\mathbf{F}\mathbf{y}, \tag{18}$$

where $y_1=a^+$, $y_2=a^-$ are the components of the vector-valued function $\mathbf{y}(t)$ and the elements of the $\mathbf{F}$ matrix are as follows:

$$f_{11}=\frac{2tC^{++}}{N_1^+}+2i\gamma^+t^2;\quad f_{12}=\frac{2tC^{-+}}{N_1^+},$$
$$f_{12}=\frac{2tC^{+-}}{N_1^-};\quad f_{22}=\frac{2tC^{--}}{N_1^-}+2i\gamma^-t^2, \tag{19}$$

where $N_1^+=N^+/t$, $N_1^-=N^-/t$.

This equation is classified as an equation with a regular (or weak) singular point.[17,18] Its two linearly independent solutions can be shown to be represented as: $\mathbf{y}_1=y_0\mathbf{y}(t)$ and $\mathbf{y}_2=g_0\mathbf{g}(t)/t$, where $\mathbf{y}(t)$ and $\mathbf{g}(t)$ are some Taylor series and y_0 and g_0 are the arbitrary coefficients. Accounting only for the first terms of the Taylor series in a small vicinity of a cutoff, we obtain

$$a^+(x)=-g_0/(2x^{1/2})+y_0+\xi g_0/2+O(x^{1/2}),$$
$$a^-(x)=-g_0/(2x^{1/2})+y_0+\xi g_0/2+O(x^{1/2}), \tag{20}$$

where $\xi=\nu_2/\nu_1-\nu_1[\mathbf{p}_0^{(2)},\mathbf{q}_{0s}]/[\mathbf{p}_0^{(1)},\mathbf{q}_0]$. We see that the sum $(a^+ + a^-)$ is really regular at $x\to 0$ and $(a^+ - a^-)$ behaves like $O(x^{-1/2})$. The abovementioned Taylor series which can be obtained by the respective recurrence procedures are not suitable for practical computation. We can reduce this computational problem to numerical solving some initial value problems. Let $\mathbf{y}(t)$ be the solution of Eq. (18) with the initial values $y_{10}=1$, $y_{20}=1$ at $t=0$ and $\mathbf{g}(t)$ be the solution of the equation

$$t\frac{d\mathbf{g}}{dt}=(\mathbf{F}+\mathbf{I})\mathbf{g} \tag{21}$$

(where $\mathbf{I}$ is the identity matrix) with the initial values $g_{10}=-1$, $g_{20}=1$. Then

$$a^+(x)=y_0y_1(x^{1/2})+g_0g_1(x^{1/2})/x^{1/2},$$
$$a^-(x)=y_0y_2(x^{1/2})+g_0g_2(x^{1/2})/x^{1/2}. \tag{22}$$

The wave field of the adiabatic mode can be computed as follows:

$$\mathbf{p}=0.5[(a^+(x)+a^-(x))(\mathbf{p}^+ + \mathbf{p}^-) + (a^+(x)-a^-(x))(\mathbf{p}^+ - \mathbf{p}^-)]. \tag{23}$$

Both terms of the sum (23) are regular at $x\to 0$.

The arbitrary coefficients which appear in the general solution given by expression (22) for $x>0$ (on the left side of a cutoff cross section: $y_0=y_{0l}$, $g_0=g_{0l}$) and for $x<0$ (on the right: $y_0=y_{0r}$, $g_0=g_{0r}$) should be matched by the continuity condition formulated in this case as follows:

$$(a^+\mathbf{p}^+ + a^-\mathbf{p}^-)_{x=\delta x}-(b_1\mathbf{p}_1+b_2\mathbf{p}_2)=O_1(\delta x),$$
$$(a^+\mathbf{p}^+ + a^-\mathbf{p}^-)_{x=-\delta x}-(b_1\mathbf{p}_1+b_2\mathbf{p}_2)=O_2(\delta x). \tag{24}$$

Here $x=-\delta x$ and $x=\delta x$ are the coordinates of some cross sections close to the cutoff (on the left and on the right side). By multiplying the first equality of (24) scalarly (in terms of the generalized scalar product) first into $\mathbf{q}^+(-\delta x)$ and then into $\mathbf{q}^-(-\delta x)$ and the next equality into $\mathbf{q}^+(\delta x)$ and $\mathbf{q}^-(\delta x)$ one can obtain two systems of algebraic equations. In vector-matrix notation:

$$\mathbf{Q}_l\mathbf{b}=\mathbf{N}_l\mathbf{a}_l+\Delta_l,\quad \mathbf{Q}_r\mathbf{b}=\mathbf{N}_r\mathbf{a}_r+\Delta_r,$$

where $\mathbf{a}_l=\{a^+(-\delta x);a^-(-\delta x)\}$, $\mathbf{a}_r=\{a^+(\delta x);a^-(\delta x)\}$, $\mathbf{b}=\{b_1;b_2\}$, $\mathbf{N}_l=\{N^+(-\delta x),0;N^-(-\delta x),0\}$, $\mathbf{N}_r=\{N^+(\delta x),0;N^-(\delta x),0\}$, $\Delta_l=O(\delta x)$, $\Delta_r=O(\delta x)$. The elements of the matrices $\mathbf{Q}_l$ and $\mathbf{Q}_r$ are the generalized scalar products of the eigenfunctions $\mathbf{p}_1$ and $\mathbf{p}_2$ by $\mathbf{q}^+(-\delta x)$ and $\mathbf{q}^-(-\delta x)$ (as for $\mathbf{Q}_l$) or by $\mathbf{q}^+(\delta x)$, $\mathbf{q}^-(\delta x)$ (as for $\mathbf{Q}_r$). Having $\mathbf{b}$ found from the first equation and substituted to the second, one can obtain

$$\mathbf{a}_r=\mathbf{M}\mathbf{a}_l+\Delta, \tag{25}$$

where $\mathbf{M}=\mathbf{N}_r{}^{-1}\mathbf{Q}_r\mathbf{Q}_l{}^{-1}\mathbf{N}_l$; $\Delta=\mathbf{N}_r{}^{-1}\mathbf{Q}_r\mathbf{Q}_l^{-1}\Delta_l-\mathbf{N}_r{}^{-1}\Delta_r$.

By expanding the adjoint eigenfunctions $\mathbf{q}^+(\pm\delta x)$ and $\mathbf{q}^-(\pm\delta x)$ into the power series in δx one can show that $\mathbf{M}=0.5\{1-i,\ 1+i;\ 1+i,\ 1-i\}$, and $\Delta=O(\delta x^{1/2})$. Then, by use of matrix $\{1,\ 1;\ 1,\ -1\}$, the vector equality (25) can be rearranged to the pair of independent scalar equalities:

$$a^+(\delta x)+a^-(\delta x)=a^+(-\delta x)+a^-(\delta x)+O(\delta x^{1/2}), \tag{26}$$

$$a^+(\delta x)-a^-(\delta x)=-i(a^+(-\delta x)+a^-(-\delta x)) + O(\delta x^{1/2}). \tag{27}$$

After substituting $a^+(\delta x)$ and $a^-(\delta x)$ from (20) into (26) and (27), multiplying (27) by $\delta x^{1/2}$ and computing the limit under $\delta x\to 0$ one can obtain that $y_{0r}=y_{0l}$, $g_{0r}=g_{0l}$.

IV. WAVES EXCITED BY A SOURCE IN A WAVEGUIDE WITH SLOWLY VARYING PARAMETERS

The complete mathematical formulation of the wave propagation problem involves the specification of a source and border conditions at the waveguide ends. Let the source be point one or distributed along the coordinate line h' in such a manner to produce a two-dimensional wave field. Vector $\mathbf{f}\delta(h-h_1,\ s-s_1)/H_{123}$ placed to the right hand side of equation (4) corresponds to such a source. We find vector $\mathbf{C}^{-1}\mathbf{f}\delta(h-h_1,\ s-s_1)/H_{123}$ at the right hand side of (5). Here $f=\{-d_1,\ f_h,-c_1,\ f_s\}$; d_1, c_1 are divergence and rotation, f_h, f_s are the components of the volume distributed force; $\delta(.)$ is Dirac's δ function. We suppose below that the border conditions at waveguide ends $s=s_l$ and $s=s_r$ correspond to the situation in which the waves leave a waveguide without reflection (this is the condition of outgoing waves). We also suppose that the interval $(s_l,\ s_r)$ contains neither singularity points nor cutoff cross sections. Under such assumptions, we consider a correspondence between the source parameters and the wave amplitudes.

Inserting (6) to (5) (with the mentioned above right hand part) and applying the generalized bi-orthogonality condition lead to equation system (10) with $m_n^{\pm}=(\mathbf{q}_n^{\pm}(h_1,s_1)\mathbf{C}^{-1}\mathbf{f}\delta(s-s_1)/H_{123})/N_n^{\pm}$ at its right hand side. Being written in matrix-vector form, this equation reads:

$$\frac{d\mathbf{a}}{ds}+(\mathbf{Q}+i\Gamma)\mathbf{a}=\mathbf{m}\cdot\delta(s-s_1), \tag{28}$$

where vector-valued function $\mathbf{a}(s)$ comprises all $a_n^+(s)$ and all $a_n^-(s)$ $(n=1,2,\ldots,\infty)$: $\mathbf{a}=\{\mathbf{a}^+;\ \mathbf{a}^-\}$; Γ is the diagonal matrix consisted of all $\gamma_n^+(s)$ and $\gamma_n^-(s)$; $\mathbf{Q}$ is the matrix of the coupling coefficients with the diagonal blocks consisted of $C_{mn}^{++}(s)/N^+$ and $C_{mn}^{--}(s)/N^-$ and with the side blocks consisted of $C_{mn}^{+-}(s)/N^-$ and $C_{mn}^{-+}(s)/N^+$; vector $\mathbf{m}=\mathbf{m}(h_1,s_1)$ is the vector with $m_n^{\pm}$ as the components. Since in practice the infinity-dimensional equation system is usually truncated to $2N$-dimensional one, we shall consider the reduced system. A solution of such an equation system complemented by border conditions at the waveguide ends can be expressed in terms of the solutions to two initial problems: $\mathbf{A}_l(s)$ and $\mathbf{A}_r(s)$. $\mathbf{A}_l(s)$ is the $(2N\times N)$ matrix-valued function complied with the border condition at the left end, $\mathbf{A}_r(s)$ is the same for the right end. Every column of the matrix $\mathbf{A}_r(s)$ is the solution to nth partial initial problem with $a_n^+(s_r)=1$ and with all the others $a_m^+(s_r)=0$ $(m\neq n)$. All the border values $a_n^-(s_r)$ are to be found by using respective border conditions. Having the values $a_n^-(s_r)$ already known one can obtain the solution $\mathbf{A}_r(s)$ [and then $\mathbf{A}_l(s)$ in an analogous way].

Let $\mathbf{A}$ be square $(2N\times 2N)$ matrix consisted of $\mathbf{A}_l(s)$ and $\mathbf{A}_r(s)$: $\mathbf{A}=\{\mathbf{A}_l,\ \mathbf{A}_r\}$. Then Green's matrix of reduced equation (28) is

$$\mathbf{M}=\begin{cases}-\mathbf{A}(s)\mathbf{D}_2\mathbf{A}^{-1}(s_1), & s<s_1,\\ \mathbf{A}(s)\mathbf{D}_1\mathbf{A}^{-1}(s_1), & s>s_1,\end{cases} \tag{29}$$

where $\mathbf{D}_1$ is the diagonal matrix with unity elements at the first N positions and with zeroes at other positions; $\mathbf{D}_2$ shows an opposite disposition of the elements. Then, we obtain the solution of Eq. (28) as follows:

$$\mathbf{a}(s,h_1,s_1)=\mathbf{M}(s,s_1)\mathbf{m}(h_1,s_1). \tag{30}$$

Thus the computational algorithm involves the following steps: (1) the numerical evaluation of the eigenvalues $\gamma_n(s_k)$, eigenfunctions $\mathbf{p}_n(h,s_k)$ and coupling coefficient matrix $\mathbf{Q}(s_k)$ in a discrete set of the cross sections $s=s_k\in[s_l,\ s_r]$ with a subsequent interpolation; (2) the computation of the border values of the vector $\mathbf{a}(s)$ by using the boundary conditions at the waveguide ends; (3) the numerical solving of two initial problems and the evaluation of the wave field excited by the source of an arbitrary location within a waveguide by use of (29) and (30).

Let the condition of ''outgoing waves'' be specified at both waveguide ends. Then, the nth column of the matrix $\mathbf{A}_r(s)$ is the solution of equation (28) complying with the following border conditions: $a_n^+(s_r)=1$, $a_m^+(s_r)=0$ $(m\neq n)$ and $a_n^-(s_r)=0$. If the adiabatic approximation (in a narrow sense) is available, expressions (29)–(30) provide:

$$\mathbf{p}=\sum_{n=1}^{N}\pm\frac{\mathbf{q}_n^{\pm}(h_1,s_1)\mathbf{C}^{-1}\mathbf{f}}{[\mathbf{q}_n^{\pm},\mathbf{p}_n^{\pm}]H_{123}(h_1,s_1)}\mathbf{p}_n^{\pm}(h,s) \times\exp\left(\pm i\int_{s_1}^{s}\gamma_n^{\pm}(s)ds\right),$$

where the upper sign ''+'' corresponds to $s>s_0$ and the lower sign ''−'' corresponds to $s<s_0$.

V. CONCLUSION

(1) Using the 4-D vector-valued functions with displacements, rotation and divergence as the components for the description of a 2-D wave field, the coupled mode method has been extended to isotropic elastic waveguides. This form of description enables us to define the set of the cross section functions and the cross section operator with this set as the domain. As a sequence, the complete expansion of the wave field in eigenfunctions of this operator can be obtained. The coupling equations, governing *P-SV* wave propagation within range-dependent waveguides, can be derived by using the generalized bi-orthogonality of the eigenfunctions.

(2) The sets of the local eigenvalues and eigenfunctions supported by a range-dependent waveguide and employed by the coupled modes theory differ from those corresponding to

the respective ''companion waveguide.'' These eigenvalues and eigenfunctions depend on derivatives of the waveguide current parameters with respect to the longitudinal coordinate and, consequently, account for not only local geometry and the current material parameters but their local variability as well. The difference between the adiabatic approach suggested by this theory and zeroth-order solution based on the ''companion waveguide'' approximation is essential near cutoff cross sections. Moreover, the predictions on existence of a cutoff suggested by the coupled modes method and those of multiple scales theory may be quite different.

(3) The local eigenvalues of the forward and backward waves merge in a cutoff cross section. The respective local expansion in eigenfunctions corresponds to a pole of the second order and differs from those respected to arbitrarily close cross sections. Thus the cutoff cross section subdivides the waveguide range into two parts in which the coupling equations have to be solved individually with further their linking. The expansion coefficients $a_n^+(s)$ and $a_n^-(s)$ are shown to be singular (like $O(x^{-1/2})$, $x\to 0$) in a vicinity of the cutoff cross section but the sum of the ''forward'' and ''backward'' modes is regular at the same time.

(4) There are several kinds of guiding environments which lead to curvilinear coordinate systems (as preferable to use) that cause the coefficients of the governing equation (5) to be singular in the origin. Coupled modes formalism is not available at the origin in this case. If the region of interest involves the origin one needs to subdivide it into two subregions. For that which involves the origin the solution has to be obtained by another method (analytically or numerically) and then to be conformed with the coupled modes solution for the external subregion by using the bi-orthogonality property of the eigenfunctions.

(5) The computation of a point source response within range-dependent waveguide is reduced down to the solving of some initial value problems for the coupling equation system. If both waveguide's ends are nonreflecting, the adiabatic point source response is represented as the sum of outgoing adiabatic modes.

APPENDIX A: GENERALIZED BI-ORTHOGONALITY

In this section we consider the generalized bi-orthogonality relations which play important role in the coupled modes theory. The derivation of the bi-orthogonality relationships related to elastic layers of constant thickness is based for the most cases on Betti–Rayleigh theorem.[19,20] Traditionally, when deriving these relationships the wave's horizontal structure is assumed to be described by the harmonic space-time factor: $\exp(-i\omega t+kx)$. As such the assumption is not true in the case being considered, we prefer to appeal to the well known technique based on the Lagrange identity.

Consider equation (7) in eigenfunction $\mathbf{p}$ with $\gamma=\gamma_1$ and the respective adjoint equation in function $\mathbf{q}$ with $\gamma=\gamma_2$, where γ_1 and γ_2 are some eigenvalues and $\mathbf{q}=\{q_1,q_2,q_3,q_4\}$ is a row vector. Multiplying the first equation into $\mathbf{q}$ and the second one into $\mathbf{p}$ and subtracting one from another we obtain, after integrating over cross-section:

$$i(\gamma_1-\gamma_2)\int_{h_0}^{h_N}\mathbf{q}'\mathbf{p}dh+(\mathbf{q}'\mathbf{A}\mathbf{p})_{h=h_N}-(\mathbf{q}'\mathbf{A}\mathbf{p})_{h=h_0}+\sum_{n=1}^{N}[(\mathbf{q}'\mathbf{A}\mathbf{p})_{h=h_n-0}-(\mathbf{q}'\mathbf{A}\mathbf{p})_{h=h_n+0}]=0, \tag{A1}$$

where h_n is the local coordinate of the nth interface, h_0 and h_N are the same of the outer boundaries.

By expressing the components $c(h_n+0)$ and $d(h_n+0)$ of the vector $\mathbf{p}(h_n+0)$ in terms of $u=u(h_n+0)=u(h_n-0)$, $w=w(h_n+0)=w(h_n-0)$, $c(h_n-0)$, $d(h_n-0)$ and by using the associated boundary conditions one can derive the following relation:

$$((\mathbf{q}'\mathbf{A}\mathbf{p})_{h=h_n+0}-(\mathbf{q}'\mathbf{A}\mathbf{p})_{h=h_n-0})=i(\gamma_1-\gamma_2)s_n,$$

where

$$s_n=\frac{2(\mu_{n+1}-\mu_n)}{H_2}\left[u(h_n)\frac{\chi_{n+1}}{\mu_{n+1}}b_{21}q_2(h_n+0)-\frac{\mu_{n+1}}{\chi_{n+1}}q_4(h_n+0)(b_{11}u(h_n))+b_{12}w(h_n)\right],$$

$$b_{11}=-2\varepsilon_n(\chi_{n+1}+\mu_{n+1}+\varepsilon_n^2(\chi_{n+1}-\mu_{n+1}))/b_{00};$$

$$b_{12}=\mu_{n+1};\quad b_{21}=-(1+\varepsilon_n^2)^2\mu_{n+1}/b_{00};$$

$$b_{00}=\mu_{n+1}((1-\varepsilon_n^2)(\chi_{n+1}+\varepsilon_n^2\lambda_{n+1})-4\varepsilon_n^2\mu_{n+1}).$$

In an analogous way: $\mathbf{q}'\mathbf{A}\mathbf{p}|_{h=h_0}=i(\gamma_1.-\gamma_2.)s_0.$, $\mathbf{q}'\mathbf{A}\mathbf{p}|_{h=h_N}=i(\gamma_1.-\gamma_2.)s_N$, where s_0 and s_N are defined by the same expressions with the respective subscripts:

$$s=\frac{2}{H_2}\left[(1+\varepsilon^2)\frac{\chi}{\mu b}uq_2-\left(2\varepsilon(1+\varepsilon^2)\frac{\mu-\chi}{\mu b}u+w\right)\frac{\mu}{\chi}q_4\right],$$

where $b=(1+\varepsilon^2)((1-\varepsilon^2)\chi/\mu+2\varepsilon^2)$.

Substituting the above obtained relations to (A1) one can ensure that the eigenfunctions $\mathbf{p}_n$ are bi-orthogonal to the adjoint eigenfunctions $\mathbf{q}_n$ on the condition that the generalized scalar product is defined as

$$[\mathbf{p},\mathbf{q}]=\int_{h_0}^{h_N}\mathbf{q}'\mathbf{p}\,dh+(s_N-s_0)-\sum_{n=1}^{N-1}s_n. \tag{A2}$$

APPENDIX B: POINT SOURCE RESPONSE WITHIN A FLAT HOMOGENEOUS LAYER

We derive now the expression for the wave field produced by a point source within a homogeneous isotropic layer of constant thickness in terms of the above defined 4-D vector-valued functions. Let, for the first time, the force be specified as vertical and uniformly distributed along the horizontal circle $r=r_0$, $z=z_0$. The wave field excited by this source is the solution of Eq. (5) with $\mathbf{f}/r\,\delta(r-r_0)\,\delta(z-z_0)$ in its right hand side (where $\mathbf{f}=\{0;F_z/\mu;0;0\}$) and with the matrix-valued coefficients: $\mathbf{A}=\{0,0,1,0;\ 0,0,0,\chi/\mu;-1,0,0,0;\ 0,\mu/\chi,0,0\}$, $\mathbf{B}=\{1/r,0,0,-1;\ 0,1/r,k_t^2,0;\ 0,-1,0,0;\ k_l^2,0,0,0\}$. The boundary conditions corresponding to the stress released boundaries of the layer are the following: $\lambda d+2\mu(\partial w/\partial z)$, $c+\partial u/\partial z$ at $z=\pm H$.

Using Hankel's transform with $J_1(\gamma r)$ as the kernel for the first two equation's rows and $J_0(\gamma r)$ for the last ones, we can obtain that kth component of the solution vector $\mathbf{p}(z,r)$ may be represented as follows:

$$p^{(k)}=\int_0^\infty t^{(k)}(\gamma,z)J_{\nu_k}(\gamma r)\,\gamma d\gamma, \tag{B1}$$

where $\nu_k=1$ for $k=1$ or 2, $\nu_k=0$ for $k=3$ or 4. The vector-valued function $\mathbf{t}(\gamma,z)$ with $t^{(k)}(\gamma,z)$ as the components is the solution of the following equation:

$$\mathbf{A}_2 t+(\mathbf{B}_2+\gamma\mathbf{I})\mathbf{t}=\mathbf{f}\delta(z-z_0). \tag{B2}$$

The matrix-valued coefficients are those: $\mathbf{A}_2=\mathbf{A}$, $\mathbf{B}_2=\{0,0,0,1;0,0,-k_t^2,0;0,-1,0,0;k_l^2,0,0,0\}$. The respective Green's matrix is a meromorphic matrix-valued function with the poles governed by the following equations:

$$\begin{aligned}(k_t^4-4\gamma^4)c_l\frac{s_t}{\beta_t}-4\gamma^2c_t\beta_l s_l=0,\\ (k_t^4-4\gamma^4)c_t\frac{s_l}{\beta_l}-4\gamma^2c_l\beta_t s_t=0,\end{aligned} \tag{B3}$$

where $c=\mathrm{ch}(\beta H)$, $s=\mathrm{sh}(\beta H)$, $\beta_t^2=\gamma^2-k_t^2$; $\beta_l^2=\gamma^2-k_l^2$.

By using analytical properties of Bessel's functions[16] we transform (B1) into the integral along the real γ axis and then to the series of residues of the poles belonging to the upper γ half-plane. The residues of the Green's matrix are proportional to the eigenfunctions $\mathbf{t}_n(z)$ [with the components $u_n(z)$, $c_n(z)$, $w_n(z)$, $d_n(z)$] and to the adjoint eigenfunctions $\mathbf{r}_n(z)$, which, in this case, may be simply expressed in terms of each other. By use of the residue theorem we obtain:

$$p^{(k)}(z,r)=-F_z\frac{\pi i}{2\chi}\sum_{n=1}^{\infty}\gamma_n\frac{w_n(z_0)t_n^{(k)}(z)}{(\mathbf{r}_n,\mathbf{t}_n)}\times\begin{Bmatrix}H_0^{(1)}(\gamma_n r_0)J_{\nu_k}(\gamma_n r)\\ J_0(\gamma_n r_0)H_{\nu_k}(\gamma_n r)\end{Bmatrix}, \tag{B4}$$

where upper row corresponds to $r<r_0$ and lower row corresponds to $r>r_0$; $(\mathbf{r}_n,\mathbf{t}_n)$ is a scalar product defined by the usual way.

If the layer is excited by a horizontal radially directed force distributed uniformly over the abovementioned circle, the components $p^{(k)}$ may be represented by the analogous expression with F_r, $H_0^{(1)}(\gamma_n r_0)$, J_0, $(\gamma_n r_0)$ and $u_n(z_0)$ in place of F_z, $H_1^{(1)}(\gamma_n r_0)$, $J_1(\gamma_n,r_0)$ and $w_n(z_0)$. If F_z is given as f/r_0, we obtain from (B4) in limit $r_0\to 0$ the components $p^{(k)}$ of the wave field caused by a point vertical force:

$$p^{(k)}=-f\frac{\pi i}{2\chi}\sum_{n=1}^{\infty}\gamma_n^2\frac{w_n(z_0)}{(\mathbf{r}_n,\mathbf{t}_n)}p_n^{(k)}(z)H_{\nu_k}^{(1)}(\gamma_n r). \tag{B5}$$

As Eq. (B3) does not coincide with the Lamb's dispersional equation, the terms of the sum (B4) [or (B5)] cannot be treated as Lamb's waves. Moreover, they are not independent waves at all, since different components of the vector-valued eigenfunction $\mathbf{t}_n(z)$ are multiplied by different propagation factors.

It results from the above equations that the wave field limited at origin may be represented as follows:

$$\mathbf{p}=\sum_{n=1}^{\infty}b_n\mathbf{J}(\gamma_n r)\mathbf{t}_n(z), \tag{B6}$$

where $\mathbf{J}(\gamma_n r)$ is the diagonal matrix with the elements: $J_1(\gamma_n r)$, $J_1(\gamma_n r)$, $J_0(\gamma_n r)$, $J_0(\gamma_n r)$.

[1] Y. P. Guo, ''Normal mode propagation on conical shells,'' J. Acoust. Soc. Am. **96**, 256–264 (1995).
[2] F. C. Karal and J. B. Keller, ''Elastic wave propagation in homogeneous and inhomogeneous media,'' J. Acoust. Soc. Am. **31**, 694–702 (1959).
[3] G. Rosenfeld and J. B. Keller, ''Wave propagation in nonuniform elastic rods,'' J. Acoust. Soc. Am. **57**, 1094–1096 (1975).
[4] H. Weinberg and R. Burridge, ''Horizontal ray theory for ocean acoustics,'' J. Acoust. Soc. Am. **55**, 63–72, 1094–1096 (1974).
[5] V. E. Nomofilov, ''Propagation of quasi-stationary Raleigh waves in a nonuniform anisotropic elastic medium,'' Steclov's Math. Inst., Math. Prob. Wave Prop. Theory **10**, 234–245 (1979).
[6] J. Tromp and F. A. Dahlen, ''Surface wave propagation in a slowly warying anisotropic waveguide,'' Geophys. J. Int. **113**, 239–249 (1993).
[7] A. D. Pierce, ''Extension of the method of normal modes to sound propagation in an almost stratified medium,'' J. Acoust. Soc. Am. **37**, 19–27 (1965).
[8] F. B. Jensen and W. A. Kuperman, ''Sound propagation in a wedge-shaped ocean with a penetrable bottom,'' J. Acoust. Soc. Am. **67**, 1564–1566 (1980).
[9] J. M. Arnold and I. B. Felsen, ''Couple mode theory of intrinsic modes in wedge,'' J. Acoust. Soc. Am. **79**, 31–40 (1986).
[10] B. Z. Katzenelenbaum, *Theory of Irregular Waveguides of Slowly Varying Parameters* (Academy of Science of the USSR, 1961) (in Russian).
[11] V. V. Shevchenko, *Continuous Transitions in Open Waveguides* (Golem, Boulder, 1971).
[12] B. L. N. Kennet, ''Seismic waves in laterally inhomogeneous media,'' Geophys. J. R. Astron. Soc. **27**, 301–325 (1972).
[13] Yu. I. Bobrovnitskii, ''Orthogonality conditions for Lamb waves'' Acust. Zh. **18**, 513–515 (1972).
[14] V. M. Eving, W. S. Jardetsky, and F. Press, *Elastic Waves in Layered Media* (New York, 1957).
[15] M. A. Neimark, *Linear Differential Operators* (Nauka, Moscow, 1969), p. 108 (in Russian).
[16] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. Stegun (National Bureau of Standards, Washington, DC, 1964), p. 183.
[17] E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (New York, 1955).
[18] E. Kamke, *Differentialgleichungen* (1943).
[19] C. L. Scandrett and C. L. Frenzen, ''Bi-orthogonality relationships involving porous media,'' J. Acoust. Soc. Am. **98**, 1199–1203 (1995).
[20] M. V. Fedoriuk, ''Orthogonality relations for solid waveguides,'' Acust. Zh. **20**, 513–515 (1974).

An alternative formulation for predicting sound radiation from a vibrating object

Sean F. Wu and Qiang Hu[a)]
Department of Mechanical Engineering, Wayne State University, Detroit, Michigan 48202

(Received 10 March 1997; accepted for publication 9 September 1997)

An alternative formulation is derived for predicting acoustic radiation from a vibrating object in an unbounded fluid medium. The radiated acoustic pressure is shown to be expressible as a surface integral of the particle velocity, which is determinable by using a nonintrusive laser velocimeter. This approach is in contrast with the Kirchhoff integral formulation which requires the knowledge of both the normal component of the surface velocity and the surface acoustic pressure prior to predicting the radiated acoustic pressure. Solutions thus obtained are unique. Moreover, the efficiency of numerical computations is high because the surface integration can be readily implemented numerically by using standard Gaussian quadratures. This alternative formulation may be desirable to analyze the acoustic and vibration responses of a lightweight, a flexible, or a structure under an adverse environment for which a nonintrusive laser measurement technique must be used. Validations of this alternative formulation are demonstrated both analytically and numerically for vibrating spheres and right circular cylinders. [S0001-4966(98)03201-9]

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

INTRODUCTION

A. Background

The Kirchhoff integral formulation is one of the most widely used methods for predicting acoustic radiation and scattering fields from an elastic structure in engineering practice.[1–8] The advantage of using this integral formulation is a reduction of the dimensionality of the problem by one. The first step of this approach is to determine the pertinent acoustic quantities on the surface. For an acoustic radiation problem, the normal component of the surface velocity (or the surface acoustic pressure for an acoustic scattering problem) is calculated or measured by an accelerometer. Next, the surface acoustic pressure (or the normal component of the surface velocity for an acoustic scattering problem) is specified by solving an integral equation. Once these quantities are known, the radiated acoustic pressure anywhere is completely determined.

Oftentimes in engineering applications, we may find it difficult to use a conventional accelerometer to measure the vibration responses of a flexible or a lightweighted structure, such as a loudspeaker membrane or a passenger vehicle fuel pump, because the weight of an accelerometer may alter the desired signal. In other cases, we may simply find it unfeasible to use an accelerometer in an adverse environment such as an engine oil pan, where the surface temperature is extremely high. Under these circumstances, we must rely on a nonintrusive measurement technique.

One approach commonly adopted in practice is to use a laser vibrometer to measure the normal component of the particle velocity, which is equal to that of the surface velocity at the interface, and then solve an integral equation for the surface acoustic pressure. The shortcomings of such an approach are well-known: (1) the surface Kirchhoff integral equation may fail to yield a unique solution whenever the frequency is close to one of the eigenfrequencies associated with the related interior boundary value problem;[9] and (2) the numerical computation may become quite involved. This is because for an arbitrary surface, we must discretize the surface into many segments with several hundreds or even more nodes. Accordingly, we must solve a large number of simultaneous integral equations for the acoustic pressures at these nodes using boundary element method (BEM). Since the central processing unit (CPU) time increases cubically with the number of discrete nodes, the computation process may be excessively time consuming.

Actually, the laser technique can be used to measure the displacement and velocity vectors of a suspended microparticle in an insonified medium. The work in this area, however, has received much less attention than that of measurements of the out-of-plane motion of a vibrating structure. Summarized below are the basic principles and applications of nonintrusive laser measurement techniques to measurements of particle displacement and velocity vectors both in fluids and in air.

B. Laser techniques

1. Laser Doppler velocimeter (LDV)

LDV has become a standard tool for nonintrusive measurements of fluid particle velocities since the pioneering work of Yeh and Cummins.[10] The basic premise in the LDV measurements is that motion of a microparticle in the fluid (either due to natural impurities or due to seeded particles) will scatter an incident light, and produce a Doppler shift in the scattered light which can be detected with appropriate electronics and signal processing.[11,12] Taylor[13,14] extended

[a)]On leave from Department of Mechanical Engineering, Zhejiang University, People's Republic of China.

 0001-4966/98/103(4)/1763/12/$10.00

the LDV used in fluid mechanics to acoustics by measuring in-air particle velocities associated with steady-state time-harmonic standing waves and travelling waves inside a tube. Dubbelday and Schau[15] used laser Doppler anemometer (LDA) for a remote detection of sound. The technique of LDA consists of measurements of the velocity of neutrally buoyant microparticles suspended in an acoustic field by analyzing the spectral content of Doppler-shifted laser light scattered by the microparticles. Vignola *et al.*[16] utilized LDV to measure the acoustic particle velocity in fluids. In particular, Vignola *et al.*[17] investigated the measurements of acoustic particle displacements using different LDV systems, and found that LDV was capable of detecting the particle displacements in the order of a few nanometers with a bandwidth of several kilohertz. The performance and limitations of LDV systems were also analyzed, and the effect of Brownian motion (i.e., thermal agitation in the fluid) on the measured data was shown to produce only negligible broadening of the spectral density of the signal of interest.[17] Vignola *et al.*[18] also derived an equation of motion of microparticles in suspension in an insonified fluid, and concluded that ''the motion of neutrally buoyant microparticles closely emulates the displacement of the surrounding insonified fluid and confirms the basic tenet associated with the laser detection of sound.''

2. Differential laser Doppler interferometry (DLDI)

DLDI is evolved from the principle of LDA and is used to measure simultaneously the out-of-plane and the in-plane velocities on the surface of a vibrating object.[19,20] The principle of DLDI is to measure the phase shift of the reflected or scattered light from the surface due to surface vibrational motion. The main component of a DLDI system is a probe head that consists of three illuminating fibers.[19,20] Prior to launching, the laser beams are frequency shifted by three acousto-optic Bragg cells. So the interference between any of the two beams will occur at three distinctly different frequency carriers. Geometrically, the first and second beams are positioned symmetrically with respect to the unit normal on the surface at an angle α, and the third beam is aligned with the first and second beams at an angle β ($\beta<\alpha$) with respect to the unit normal. In this differential configuration, one frequency carrier will be modulated by the in-plane motion, and the other two carriers will be modulated by both in-plane and out-of-plane motions, respectively. In this setup, the demodulation is done by using a combination of filters and the phase-locked loops (PLL). The PLL demodulates the signal with phase ϕ and generates an output which is proportional to the time rate of changes of ϕ. Therefore, by measuring the instantaneous frequency deviations $d\phi_{ij}/dt$ from the different frequency carriers, one can determine simultaneously the in-plane and out-of-plane components of the surface velocity. Vignola and Houston[21] and Vignola *et al.*[22] designed a three-dimensional laser vibrometer based on this principle to measure simultaneously the three components of the velocity on the surface of a vibrating structure.

The DLDI technique can be extended in principle to measurements of the microparticle velocity. Imagine that an object is surrounded by neutrally buoyant microparticles. As the object vibrates, the acoustic pressure fluctuations will excite the microparticles into oscillations. Suppose that we define a control surface which consists of microparticles and which completely enclosing the vibrating surface. In the special case in which the control surface coincides with the vibrating surface, the normal component of the microparticle displacement will be equal to that of the surface displacement, while the tangential components may be different. In any event, the microparticle displacement in the directions normal and tangential to the control surface will cause a Doppler shift in the phase ϕ_{ij} of the reflected light, which is modulated in the frequency carriers. Once the signal with phase ϕ_{ij} is demodulated, we can calculate the microparticle velocity which is proportional to the time derivative of the phase, $d\phi_{ij}/dt$.

3. Electronic speckle pattern interferometry (ESPI)

Alternatively, we can use ESPI[23,24] to measure the phase term of a microparticle, which is an established optical technique for measuring static and dynamic deformations and surface shapes for more than two decades. Specifically, we can utilize the stroboscopic technique,[25] which ''freezes'' the dynamic motion of a particle at one position so that during other times of the movement cycle, the particle is not illuminated and therefore is ''invisible'' to the imaging device. In practice, this technique can be implemented by using a pulsed laser[25] or a light shuttering device with a continuous wave laser.[26] The time interval between two consecutive pulses or shutters is typically in the range of nanoseconds, so ESPI can capture very high-frequency oscillations. Pouet and Krishnaswamy[27,28] and Wang and Krishnaswamy[29] show that by using an additive–subtractive speckle pattern interferometry, the accuracy of the phase measurement can be further enhanced.

C. Objective and significance

The objective of the present paper is to present an alternative formulation[30,31] which enables one to predict the radiated acoustic pressure directly, once the particle velocity on the control surface that encloses a vibrating surface is determined by a nonintrusive laser velocimeter. The significance of this formulation is two-fold: (1) solutions thus obtained are unique and (2) the efficiency of numerical computations is high. This is because the radiated acoustic pressure is expressed explicitly as a surface integral of the particle velocity, which can be implemented numerically using standard Gaussian quadratures. There is no need to use BEM to solve a set of simultaneous integral equations for the surface acoustic pressures at the discretized nodes, which can be time consuming for an arbitrarily shaped object. Such an approach can be desirable for analyzing the acoustic responses of the structures which require the use of a nonintrusive laser measurement technique.

I. BASIC THEORY

Consider sound radiation from a finite object immersed in an unbounded fluid medium with density ρ_0 and sound speed c. Assume that the object vibrates at a constant frequency ω, so the time dependence of an acoustic quantity can be written as $e^{-i\omega t}$. Accordingly, the complex amplitude of the radiated acoustic pressure $\hat{p}(\mathbf{x})$ at any point $\mathbf{x}$ can be written as[32]

$$\hat{p}(\mathbf{x}) = \frac{1}{4\pi} \int_S \left[\frac{\partial G(\mathbf{x}|\mathbf{x}_S)}{\partial n_S} \hat{p}(\mathbf{x}_S) - G(\mathbf{x}|\mathbf{x}_S) \frac{\partial \hat{p}(\mathbf{x}_S)}{\partial n_S} \right] dS, \tag{1}$$

where G and $\partial G/\partial n$ are the free-space Green's function and its normal derivative given, respectively, by

$$G(\mathbf{x}|\mathbf{x}_S) = \frac{e^{ikR}}{R} \quad \text{and} \quad \frac{\partial G(\mathbf{x}|\mathbf{x}_S)}{\partial n_S} = \frac{(ikR-1)e^{ikR}}{R^2} \frac{\partial R}{\partial n_S}, \tag{2}$$

where n_S stands for the outward unit normal on the surface S and $R = |\mathbf{x} - \mathbf{x}_S|$; here $\mathbf{x}$ and $\mathbf{x}_S$ represent the receiver and source position vectors, respectively.

Equation (1) is an integral representation of the Helmholtz equation in the frequency domain. The normal derivative of the acoustic pressure on the right side of Eq. (1) is related to the normal component of the surface velocity $\hat{v}_n(\mathbf{x}_S)$, which is equal to that of the particle velocity at the interface. For an acoustic radiation problem, the normal component of the surface velocity is specified. Hence, one must solve an integral equation obtained by letting the field point $\mathbf{x}$ in Eq. (1) approach the surface to determine the surface acoustic pressure $\hat{p}(\mathbf{x}_S)$

$$\hat{p}(\mathbf{x}'_S) = \frac{1}{2\pi} \int_S \left[\frac{\partial G(\mathbf{x}'_S|\mathbf{x}_S)}{\partial n_S} \hat{p}(\mathbf{x}_S) - i\omega\rho_0 \hat{v}_n(\mathbf{x}_S) G(\mathbf{x}'_S|\mathbf{x}_S) \right] dS, \tag{3}$$

where R is replaced by $R_S = |\mathbf{x}'_S - \mathbf{x}_S|$, here both $\mathbf{x}'_S$ and $\mathbf{x}_S$ are on the surface S. For an arbitrary surface, there is no closed-form solution to Eq. (3) and $\hat{p}(\mathbf{x}_S)$ must be solved numerically by using BEM. Once $\hat{p}(\mathbf{x}_S)$ and $\hat{v}_n(\mathbf{x}_S)$ are specified, the radiated acoustic pressure $\hat{p}(\mathbf{x})$ can be calculated by Eq. (1).

II. AN ALTERNATIVE FORMULATION

In this section, we present an alternative formulation for directly predicting the radiated acoustic pressure once the particle velocity at a control surface is specified by a nonintrusive laser velocimeter. For completeness, we give formulations for both exterior and interior regions, respectively.

A. Exterior problems

Derivations of the alternative formulation start from the Euler equation,[32]

$$\nabla \hat{p}(\mathbf{x}) = i\omega\rho_0 \hat{\mathbf{v}}(\mathbf{x}). \tag{4}$$

Integrating both sides of Eq. (4) along a line which connects one field point $\mathbf{x}'$ to another $\mathbf{x}$ yields

$$\int_{(\mathbf{x}' \to \mathbf{x})} \nabla \hat{p}(\mathbf{x}) \cdot \mathbf{e}\, dl = \int_{(\mathbf{x}' \to \mathbf{x})} d\hat{p}(\mathbf{x}) = i\omega\rho_0 \int_{(\mathbf{x}' \to \mathbf{x})} \hat{\mathbf{v}}(\mathbf{x}) \cdot \mathbf{e}\, dl, \tag{5}$$

where $\mathbf{e}$ is a unit vector in the direction of the line integral from $\mathbf{x}'$ to $\mathbf{x}$.

Obviously, the integral of $d\hat{p}(\mathbf{x})$ on the left side of Eq. (5) is determined by its upper and lower limits,[33] but independent of the choice of the integration path. Hence, we obtain

$$\hat{p}(\mathbf{x}) = \hat{p}(\mathbf{x}') + i\omega\rho_0 \int_{(\mathbf{x}' \to \mathbf{x})} \hat{\mathbf{v}}(\mathbf{x}) \cdot \mathbf{e}\, dl. \tag{6}$$

Equation (6) shows that the complex amplitude of the acoustic pressure at any point $\mathbf{x}$ can be expressed as the sum of the complex amplitude of the acoustic pressure at another point $\mathbf{x}'$ plus a line integral of the apparent force per unit volume over any path that connects these two points. Since there is no restriction on the selection of the integration path, we choose to let it lie on a control surface S_c. Accordingly, we can write

$$\hat{p}(\mathbf{x}_{S_c}) = \hat{p}(\mathbf{x}'_{S_c}) + i\omega\rho_0 \int_{(\mathbf{x}'_{S_c} \to \mathbf{x}_{S_c})} [\hat{v}_n(\mathbf{x}_{S_c}) dn + \hat{v}_\eta(\mathbf{x}_{S_c}) d\eta], \tag{7}$$

where $\hat{v}_n$ and $\hat{v}_\eta$ represent the normal and tangential components of the particle velocity at the surface S_c, respectively, and dn and $d\eta$ are the increments in the normal and tangential directions on S_c, respectively. In the special case where the control surface coincides with the real surface, the normal component of the particle velocity $\hat{v}_n(\mathbf{x}_S)$ is equal to that of the surface velocity, but its tangential component $\hat{v}_\eta$ may be different.

Without loss of generality, we will omit the subscript c on S_c in the following. Since the integration path remains perpendicular to the unit normal $\mathbf{n}$ at all times, $dn \equiv 0$. Hence the first term in the square brackets on the right side of Eq. (7) is identically zero,

$$\hat{p}(\mathbf{x}_S) = \hat{p}(\mathbf{x}'_S) + i\omega\rho_0 \int_{(\mathbf{x}'_S \to \mathbf{x}_S)} \hat{v}_\eta(\mathbf{x}_S) d\eta. \tag{8}$$

Substitute $\hat{p}(\mathbf{x}_S)$ into Eq. (3) and notice that $\hat{p}(\mathbf{x}'_S)$ represents the acoustic pressure at a fixed surface point $\mathbf{x}'_S$, which is independent of the integration with respect to the unprimed surface coordinates and therefore can be factored out of the integral sign. Combining the coefficients of $\hat{p}(\mathbf{x}'_S)$ on both sides of Eq. (3) yields a solution for $\hat{p}(\mathbf{x}'_S)$, which when substituted back into Eq. (8) gives a solution for $\mathbf{p}(\mathbf{x}_S)$

$$\hat{p}(\mathbf{x}_S)=i\omega\rho_0\left\{\int_S\left[\frac{\partial G(\mathbf{x}'_S|\mathbf{x}_S)}{\partial n_S}\left(\int_{(\mathbf{x}'_S\to\mathbf{x}_S)}\hat{v}_\eta(\mathbf{x}_S)d\eta\right)-\hat{v}_n(\mathbf{x}_S)G(\mathbf{x}'_S|\mathbf{x}_S)\right]dS\right\}\left[2\pi-\int_S\frac{\partial G(\mathbf{x}'_S|\mathbf{x}_S)}{\partial n_S}\,dS\right]^{-1}$$

$$+i\omega\rho_0\int_{(\mathbf{x}'_S\to\mathbf{x}_S)}\hat{v}_\eta(\mathbf{x}_S)d\eta. \tag{9}$$

Substituting Eq. (9) into Eq. (1), we obtain the following integral formulation:

$$\hat{p}(\mathbf{x})=\mathscr{L}_1(\hat{v}_\eta)+\mathscr{L}_2(\hat{v}_n), \tag{10}$$

where $\mathscr{L}_{1,2}$ represent integral operators operating on the normal and tangential components of the particle velocity, respectively,

$$\mathscr{L}_1(\hat{v}_\eta)=\frac{iw\rho_0}{4\pi}\int_S\left\{\frac{\partial G(\mathbf{x}|\mathbf{x}_S)}{\partial n}\left[\int_{S'}\frac{\partial G(\mathbf{x}_S|\mathbf{x}_{S'})}{\partial n_{S'}}\left(\int_{(\mathbf{x}'_{S'}\to\mathbf{x}_{S'})}\hat{v}_\eta(\mathbf{x}_{S'})d\eta\right)dS'\right]\left[2\pi-\int_{S'}\frac{\partial G(\mathbf{x}_S|\mathbf{x}_{S'})}{\partial n_{S'}}\,dS'\right]^{-1}\right\}dS$$

$$+\frac{i\omega\rho_0}{4\pi}\int_S\frac{\partial G(\mathbf{x}|\mathbf{x}_S)}{\partial n}\left(\int_{(\mathbf{x}'_S\to\mathbf{x}_S)}\hat{v}_\eta(\mathbf{x}_S)d\eta\right)dS, \tag{11a}$$

$$\mathscr{L}_2(\hat{v}_n)=-\frac{i\omega\rho_0}{4\pi}\int_S\left\{\frac{\partial G(\mathbf{x}|\mathbf{x}_S)}{\partial n}\left[\int_{S'}\hat{v}_n(\mathbf{x}_{S'})G(\mathbf{x}_S|\mathbf{x}_{S'})dS'\right]\left[2\pi-\int_{S'}\frac{\partial G(\mathbf{x}_S|\mathbf{x}_{S'})}{\partial n_{S'}}\,dS'\right]^{-1}\right\}dS$$

$$-\frac{i\omega\rho_0}{4\pi}\int_S\hat{v}_n(\mathbf{x}_S)G(\mathbf{x}|\mathbf{x}_S)dS. \tag{11b}$$

Equation (10) is the main result of this paper. The radiated acoustic pressure $\hat{p}(\mathbf{x})$ is shown to be expressible in terms of the particle velocity only. This alternative formulation is in contrast with the classical Kirchhoff integral formulation. There is no need to solve an integral equation for the surface acoustic pressure, given the normal component of the surface velocity. Instead, the radiated acoustic pressure can be calculated directly once the particle velocity at the surface is specified. Since the surface integrals in Eq. (11) can be readily implemented by the standard Gaussian quadratures, the efficiency of numerical computations may be significantly enhanced.

B. Interior problems

Following the same procedures as outlined above and changing the sign of the unit normal derivative $\partial/\partial n$, we obtain

$$\hat{p}(\mathbf{X})=\mathscr{L}_1^{\text{int}}\{\hat{v}_\eta\}+\mathscr{L}_2^{\text{int}}\{\hat{v}_n\}, \tag{12}$$

where $\hat{p}(\mathbf{X})$ represents the acoustic pressure at an interior point $\mathbf{X}$ enclosed by the surface S, and $\mathscr{L}_{1,2}^{\text{int}}$ are defined by

$$\mathscr{L}_1^{\text{int}}\{\hat{v}_\eta\}=\frac{i\omega\rho_0}{4\pi}\int_S\left\{\frac{\partial G(\mathbf{X}|\mathbf{X}_S)}{\partial n_S}\left[\int_{S'}\frac{\partial G(\mathbf{X}_S|\mathbf{X}_{S'})}{\partial n_{S'}}\left(\int_{(\mathbf{X}'_{S'}\to\mathbf{X}_{S'})}\hat{v}_\eta(\mathbf{X}_{S'})d\eta\right)dS'\right]\left[2\pi+\int_{S'}\frac{\partial G(\mathbf{X}_S|\mathbf{X}_{S'})}{\partial n_{S'}}\,dS'\right]^{-1}\right\}dS$$

$$-\frac{i\omega\rho_0}{4\pi}\int_S\frac{\partial G(\mathbf{X}|\mathbf{X}_S)}{\partial n_S}\left(\int_{(\mathbf{X}'_S\to\mathbf{X}_S)}\hat{v}_\eta(\mathbf{X}_S)d\eta\right)dS, \tag{13a}$$

$$\mathscr{L}_2^{\text{int}}\{\hat{v}_\eta\}=-\frac{i\omega\rho_0}{4\pi}\int_S\left\{\frac{\partial G(\mathbf{X}|\mathbf{X}_S)}{\partial n_S}\left[\int_{S'}\hat{v}_n(\mathbf{X}_{S'})G(\mathbf{X}_S|\mathbf{X}_{S'})dS'\right]\left[2\pi+\int_{S'}\frac{\partial G(\mathbf{X}_S|\mathbf{X}_{S'})}{\partial n_{S'}}\,dS'\right]^{-1}\right\}dS$$

$$+\frac{i\omega\rho_0}{4\pi}\int_S\hat{v}_n(\mathbf{X}_S)G(\mathbf{X}|\mathbf{X}_S)dS. \tag{13b}$$

C. Scattering problems

In a similar manner, we can extend this alternate formulation to the acoustic pressure field scattered from a finite object in free or half-space.[34,35] For an object in free space, the solution for the scattered acoustic pressure is given by

$$\hat{p}^{\text{sca}}(\mathbf{x})=\mathscr{L}_1\{\hat{v}_\eta^{\text{total}}\}+\mathscr{L}_2\{\hat{v}_n^{\text{total}}\}, \tag{14}$$

where $\mathscr{L}_{1,2}$ are defined by Eq. (11). The quantities $\hat{v}_\eta^{\text{total}}$ and $\hat{v}_n^{\text{total}}$ in Eq. (14) represent the components of the total particle velocity normal and tangential to the surface, respectively,

$$\hat{v}_\eta^{\text{total}}=(\hat{\mathbf{v}}^{\text{sca}}+\hat{\mathbf{v}}^{\text{inc}})\cdot\mathbf{e}_\eta, \tag{15a}$$

$$\hat{v}_n^{\text{total}}=(\hat{\mathbf{v}}^{\text{sca}}+\hat{\mathbf{v}}^{\text{inc}})\cdot\mathbf{e}_n, \tag{15b}$$

where $\hat{\mathbf{v}}^{\text{sca}}$ and $\hat{\mathbf{v}}^{\text{int}}$ are the scattered and incident components of the particle velocity, respectively, and $\mathbf{e}_\eta$ and $\mathbf{e}_n$ represent

the unit vectors in the tangential and normal directions at the surface of the object, respectively.

For an object in a half-space bounded by an infinite baffle with certain surface acoustic impedance, the effect of the acoustic pressure reflected from the baffle and that scattered from the object due to this reflected wave must all be taken into account. One way of solving for the scattered acoustic pressure in the presence of an infinite baffle is to use the image source method. The resulting formulation can be written as

$$\hat{p}^{\mathrm{sca}}(\mathbf{x}) = \mathscr{L}_3\{\hat{v}_\eta^{\mathrm{total}}\} + \mathscr{L}_4\{\hat{v}_n^{\mathrm{total}}\}, \tag{16}$$

where $\mathscr{L}_{3,4}$ are given by

$$\mathscr{L}_3\{\hat{v}_\eta^{\mathrm{total}}\} = \frac{i\omega\rho_0}{4\pi}\int_S \left\{ \frac{\partial \mathscr{G}(\mathbf{x}|\mathbf{x}_S)}{\partial n_S} \left[\int_{S'} \frac{\partial \mathscr{G}(\mathbf{x}_S|\mathbf{x}_{S'})}{\partial n_{S'}} \left(\int_{(\mathbf{x}'_{S'}\to\mathbf{x}_{S'})} \hat{v}_\eta^{\mathrm{total}}(\mathbf{x}_{S'})\, d\eta \right) dS' \right] \left[2\pi - \int_{S'} \frac{\partial \mathscr{G}(\mathbf{x}_S|\mathbf{x}_{S'})}{\partial n_{S'}} dS' \right]^{-1} \right\} dS$$
$$+ \frac{i\omega\rho_0}{4\pi}\int_S \frac{\partial \mathscr{G}(\mathbf{x}|\mathbf{x}_S)}{\partial n_S} \left(\int_{(\mathbf{x}'_S\to\mathbf{x}_S)} \hat{v}_\eta^{\mathrm{total}}(\mathbf{x}_S)\, d\eta \right) dS, \tag{17a}$$

$$\mathscr{L}_4\{\hat{v}_n^{\mathrm{total}}\} = -\frac{i\omega\rho_0}{4\pi}\int_S \left\{ \frac{\partial \mathscr{G}(\mathbf{x}|\mathbf{x}_S)}{\partial n_S} \left[\int_{S'} \hat{v}_n^{\mathrm{total}}(\mathbf{x}_{S'})\, \mathscr{G}(\mathbf{x}_S|\mathbf{x}_{S'})\, dS' \right] \left[2\pi - \int_{S'} \frac{\partial \mathscr{G}(\mathbf{x}_S|\mathbf{x}_{S'})}{\partial n_{S'}}\, dS' \right]^{-1} \right\} dS$$
$$- \frac{i\omega\rho_0}{4\pi}\int_S \hat{v}_n^{\mathrm{total}}(\mathbf{x}_S)\, \mathscr{G}(\mathbf{x}|\mathbf{x}_S)\, dS, \tag{17b}$$

where $\mathscr{G}$ is the Green's function that accounts for the effect of an image source. For a surface with arbitrary acoustic impedance, there is no closed-form solution for this Green's function. However, for an observer at a point $\mathbf{x}$ which is at least one half-wavelength away from the surface, $\mathscr{G}$ can be approximated by[36]

$$\mathscr{G}(\mathbf{x}|\mathbf{x}_S) = \frac{e^{ikR}}{R} + \left(\frac{\cos\theta_I - \beta}{\cos\theta_I + \beta} \right) \frac{e^{ikR_I}}{R_I}, \tag{18a}$$

and its normal derivative is given by

$$\frac{\partial \mathscr{G}(\mathbf{x}|\mathbf{x}_S)}{\partial n} = \frac{(ikR-1)e^{ikR}}{R^2}\frac{\partial R}{\partial n} + \left(\frac{\cos\theta_I - \beta}{\cos\theta_I + \beta} \right) \frac{(ikR_I - 1)e^{ikR_I}}{R_I^2}\frac{\partial R_I}{\partial n}, \tag{18b}$$

where R and R_I are the distances measured from the observer to the source and to the image, respectively, $\cos\theta_I = \mathbf{n}_b \cdot \mathbf{e}_I$; here $\mathbf{n}_b$ is the unit normal vector on the baffle and $\mathbf{e}_I$ is the unit vector in the direction of wave propagation from the image to the observer, β is the acoustic admittance of the baffle surface defined by

$$\beta = \frac{\rho_0 c}{Z(\omega)}, \tag{19}$$

where $Z(\omega)$ is the surface acoustic impedance. For a rigid surface, $Z(\omega)\to\infty$, so $\beta\to 0$ and $\mathscr{G} = e^{ikR}/R + e^{ikR_I}/R_I$. On the other hand, for a pressure-release surface, $Z(\omega)\to 0$, so $\beta\to\infty$ and $\mathscr{G} = e^{ikR}/R - e^{ikR_I}/R_I$.

III. UNIQUENESS OF SOLUTION

It is well known[9] that in carrying out the numerical computations for the radiated acoustic pressure in the exterior region, the surface Kirchhoff integral equation (3) may fail to yield a unique solution whenever the excitation frequency is close to one of the eigenfrequencies of the interior boundary value problem. This is because Eq. (3) shares the same eigenfrequencies as those of the corresponding integral equation in the interior region, and further, the solution to the adjoint homogeneous equation satisfies the compatibility condition for any velocity distribution $\hat{v}_n(\mathbf{x})$. Since the alternative formulation (9) is derived from the Kirchhoff integral theory, an examination of its uniqueness seems to be in order. In what follows, we follow the procedures outlined by Schenck[9] to examine the uniqueness of Eq. (9). For brevity, the theorems regarding the characteristic values of an integration kernel, and those regarding trivial and nontrivial solutions to homogeneous and inhomogeneous equations as well as their adjoints[37] are omitted.

The integral formulation governing the acoustic pressure in an interior region is given by

$$\hat{p}(\mathbf{X}) = -\frac{1}{4\pi}\int_S \left[\frac{\partial G(\mathbf{X}|\mathbf{X}_S)}{\partial n_S}\, \hat{p}(\mathbf{X}_S) - G(\mathbf{X}|\mathbf{X}_S)\frac{\partial \hat{p}(\mathbf{X}_S)}{\partial n_S} \right] dS. \tag{20}$$

Taking a normal derivative of Eq. (20) at an interior point $\mathbf{X}$ and then letting $\mathbf{X}\to\mathbf{X}'_S$ from the inside, we obtain

$$2\pi\hat{v}_n(\mathbf{X}'_S) - \int_S \hat{v}_n(\mathbf{X}_S)\, \frac{\partial G(\mathbf{X}'_S|\mathbf{X}_S)}{\partial n_S}\, dS = -\frac{1}{i\omega\rho_0}\frac{\partial}{\partial n'_S}\int_S \frac{\partial G(\mathbf{X}'_S|\mathbf{X}_S)}{\partial n_S}\, \hat{p}(\mathbf{X}_S) dS. \tag{21}$$

Substituting $\hat{p}(\mathbf{X}_S)$ on the right side of Eq. (21) by $\hat{p}(\mathbf{X}'_S)$ plus a line integral as given by Eq. (8) yields

$$2\pi\hat{v}_n(\mathbf{X}'_S)-\int_S \hat{v}_n(\mathbf{X}_S)\,\frac{\partial G(\mathbf{X}'_S|\mathbf{X}_S)}{\partial n_S}\,dS$$
$$+\frac{\partial}{\partial n'_S}\int_S \frac{\partial G(\mathbf{X}'_S|\mathbf{X}_S)}{\partial n_S}\left(\int_{(\mathbf{X}'_S\to\mathbf{X}_S)}\hat{v}_\eta(\mathbf{X}_S)\,d\eta\right)dS$$
$$=-\frac{1}{i\omega\rho_0}\frac{\partial}{\partial n'_S}\int_S \frac{\partial G(\mathbf{X}'_S|\mathbf{X}_S)}{\partial n_S}\,\hat{p}(\mathbf{X}'_S)\,dS. \tag{22}$$

For the homogeneous Dirichlet problem, $\hat{p}(\mathbf{X}'_S)\equiv 0$. Hence Eq. (22) reduces to

$$2\pi\hat{v}_n(\mathbf{X}'_S)-\int_S \hat{v}_n(\mathbf{X}_S)\,\frac{\partial G(\mathbf{X}'_S|\mathbf{X}_S)}{\partial n_S}\,dS$$
$$+\frac{\partial}{\partial n'_S}\int_S \frac{\partial G(\mathbf{X}'_S|\mathbf{X}_S)}{\partial n_S}\left(\int_{(\mathbf{X}'_S\to\mathbf{X}_S)}\hat{v}_\eta(\mathbf{X}_S)\,d\eta\right)dS=0. \tag{23}$$

The integral equation governing the surface acoustic pressure in the exterior region is given by Eq. (9). If we consider the homogeneous Neumann problem in which $\hat{\mathbf{v}}(\mathbf{X}_S)\equiv 0$ and take the complex conjugate of this homogeneous equation, we obtain

$$\hat{p}^\dagger(\mathbf{X}_S)\left[2\pi-\int_S \frac{\partial G(\mathbf{X}'_S|\mathbf{X}_S)}{\partial n_S}\,dS\right]=0. \tag{24}$$

Obviously, Eq. (24) is different from Eq. (23), so they do not share the same eigenfrequencies. Consequently, Eq. (9) has a unique solution for the radiated acoustic pressure in the exterior region. However, this uniqueness may break down for the special case of a dilating sphere in which $\hat{\mathbf{v}}\equiv\hat{v}_n\mathbf{e}_n$ and $\hat{v}_\eta\equiv 0$. Since $\hat{v}_n$ is a constant, it can be factored out of the integral sign and Eq. (23) reduces to

$$\hat{v}_n\left[2\pi-\int_S \frac{\partial G(\mathbf{X}'_S|\mathbf{X}_S)}{\partial n_S}\,dS\right]=0, \tag{25}$$

which is identical in form to Eq. (24) for the exterior problem. Therefore $\hat{p}^\dagger$ and $\hat{v}_n$ share the same eigenfrequencies ω^*.

To check if $\hat{p}^\dagger$ also satisfies the compatibility condition given by[9]

$$\int_S \hat{p}^\dagger(\mathbf{X}_S)\left[\int_{S'}\hat{v}_n(\mathbf{X}_{S'})G(\mathbf{X}_{S'}|\mathbf{X}_S)dS'\right]dS=0, \tag{26}$$

we return to the integral formulation (20) for the interior region.

Taking the limit as $\mathbf{X}\to\mathbf{X}'_S$ from the inside leads to

$$\hat{p}(\mathbf{X}'_S)\left[2\pi+\int_S \frac{\partial G(\mathbf{X}'_S|\mathbf{X}_S)}{\partial n_S}\,dS\right]$$
$$=i\omega\rho_0\int_S \hat{v}_n(\mathbf{X}_S)G(\mathbf{X}'_S|\mathbf{X}_S)dS$$
$$-i\omega\rho_0\int_S \frac{\partial G(\mathbf{X}'_S|\mathbf{X}_S)}{\partial n_S}\left(\int_{(\mathbf{X}'_S\to\mathbf{X}_S)}\hat{v}_\eta(\mathbf{X}_S)\,d\eta\right)dS. \tag{27}$$

Since $\hat{v}_\eta(\mathbf{X}_S)\equiv 0$, the second term on the right side of Eq. (27) vanishes identically. Therefore, for the interior homogeneous Dirichlet problem for which $\hat{p}(\mathbf{X}_S)\equiv 0$ and $\omega=\omega^*$, we have

$$\int_S \hat{v}_n(\mathbf{X}_S)G(\mathbf{X}'_S|\mathbf{X}_S)dS=0. \tag{28}$$

Because of the equivalence of $\hat{v}_n(\mathbf{X}_S)$ and $\hat{p}^\dagger(\mathbf{X}_S)$, we can rewrite Eq. (28) as

$$\int_S \hat{p}^\dagger(\mathbf{X}_S)G(\mathbf{X}'_S|\mathbf{X}_S)dS=0. \tag{29}$$

Interchanging the order of integrations in Eq. (26) yields

$$\int_{S'}\hat{v}_n(\mathbf{X}_{S'})\left[\int_S \hat{p}^\dagger(\mathbf{X}_S)G(\mathbf{X}_{S'}|\mathbf{X}_S)dS\right]dS'=0. \tag{30}$$

Since the square-bracket term in Eq. (30) is identically zero [see Eq. (29)], the compatibility condition (26) is satisfied. Consequently, for a dilating sphere Eq. (9) may fail to yield a unique solution for the surface acoustic pressure whenever the excitation frequency is close to one of the corresponding interior Dirichlet eigenfrequencies. However, the ill-conditioning in the numerical computation of Eq. (9) is much less severe than that in the BEM-based Kirchhoff integral equation, as shown below.

For a dilating sphere, the solution for the surface acoustic pressure given by Eq. (9) takes the following form

$$\hat{p}=-i\omega\rho_0\hat{v}_n\int_S \frac{e^{ikR_S}}{R_S}\,dS$$
$$\times\left[2\pi-\int_{S'}\frac{(ikR_{S'}-1)e^{ikR_{S'}}}{R^2_{S'}}\,\frac{\partial R_{S'}}{\partial n_{S'}}\,dS'\right]^{-1}. \tag{31}$$

The corresponding interior Dirichlet eigenfrequencies are determined by the roots of the spherical Bessel function of the first kind and order zero, $\mathcal{J}_0(k^*a)=0$, which yields $k^*a=m\pi$, where m is an integer and a is the radius of the sphere. Using the spherical coordinates, it is easy to show[38] that $R_S=2a\cos(\theta/2)$, $\partial R_S/\partial n_S=\cos(\theta/2)$, and $dS=a^2\sin\theta d\theta d\phi$, where θ and ϕ vary from 0 to π and 0 to 2π, respectively.

Substituting R_S, $\partial R_S/\partial n_S$, and dS into Eq. (31), we obtain

$$\hat{p}=-\frac{i\rho_0 c\hat{v}_n ka}{(1-ika)}\times\frac{(1-e^{i2ka})}{(1-e^{i2ka})}. \tag{32}$$

When $ka=m\pi$, Eq. (32) reduces to $\hat{p}=-i\rho_0 c\hat{v}_n m\pi(1-im\pi)^{-1}\times(0/0)$. However, the computer cannot take the limit of (0/0). Hence any round-off error in the numerical computation may lead to an erroneous result.

In using the BEM-based Kirchhoff integral equation, the surface acoustic pressure is obtained by solving a set of simultaneous integral equations. When the excitation frequency is close to one of the interior eigenfrequencies, the diagonal terms become very small and the matrix becomes ill-conditioned. Table I exhibits this trend in detail for a dilating sphere around $ka=\pi$ and 2π. Here numerical computations are carried out by a general BEM code with the

TABLE I. Dimensionless acoustic pressures on the surface of a dilating sphere around $ka=\pi$ and 2π.

ka	Exact values	BEM	Errors (%)	Eq. (9)	Errors (%)
2.0	8.944 27$E-01$	8.393 06$E-01$	6.16	8.944 27$E-01$	<0.01
3.0	9.486 83$E-01$	4.731 09$E-01$	50.13	9.486 82$E-01$	<0.01
3.14	9.528 46$E-01$	3.861 43$E+01$	3 952.52	9.527 24$E-01$	0.01
3.1415	9.528 88$E-01$	1.894 15$E+02$	19 777.99	9.510 50$E-01$	0.14
3.141592	9.528 90$E-01$	8.624 46$E+01$	8 950.85	6.683 42$E-01$	29.82
3.1416	9.528 91$E-01$	6.572 39$E+01$	6 797.32	9.784 23$E-01$	2.68
3.15	9.531 24$E-01$	8.962 01$E+00$	840.28	9.531 45$E-01$	<0.01
4.0	9.701 42$E-01$	1.022 09$E+00$	5.35	9.701 43$E-01$	<0.01
5.0	9.805 81$E-01$	9.261 71$E-01$	5.55	9.805 82$E-01$	<0.01
6.0	9.863 94$E-01$	5.108 93$E-01$	48.21	9.864 01$E-01$	<0.01
6.28	9.875 58$E-01$	4.386 67$E+01$	4 341.94	9.879 23$E-01$	0.04
6.2830	9.875 70$E-01$	1.339 35$E+02$	13 462.08	9.942 31$E-01$	0.67
6.283185	9.875 71$E-01$	1.365 84$E+02$	13 730.30	3.579 67$E+00$	262.47
6.2832	9.875 71$E-01$	1.513 51$E+02$	15 225.58	9.948 20$E-01$	0.73
6.30	9.876 36$E-01$	1.014 60$E+01$	927.30	9.875 78$E-01$	<0.01
7.0	9.899 50$E-01$	1.082 05$E-00$	9.30	9.899 59$E-01$	<0.01
8.0	9.922 78$E-01$	5.347 67$E-01$	46.11	9.923 07$E-01$	<0.01
9.0	9.938 84$E-01$	2.207 37$E-01$	77.79	9.939 71$E-01$	<0.01

spherical surface discretized into 48 quadratic quadrilaterals and 130 nodes. The results thus obtained are compared with those of Eq. (9).

With Eq. (9) one can carry out surface integrals using Gaussian quadratures directly, without the need of solving a set of simultaneous equations. In evaluating these integrations, we make use of the axisymmetry of the acoustic pressure distribution and discretize the spherical surface into 12 rings along the generator. The integrations with respect to the polar angle θ within each ring are carried out by the Gaussian quadrature formula with three interior points.[39] The integration over the azimuthal angle ϕ can be done independently, yielding a factor of 2π.

Numerical results in Table I show that both BEM and Eq. (9) fail to yield unique solutions at the interior Dirichlet eigenfrequencies. However, the BEM results show the sign of ill-conditioning over a large frequency range around the interior Dirichlet eigenfrequencies, whereas the results obtained by Eq. (9) are more or less correct until ka hits the eigenfrequencies almost directly.

As a second example, we calculate the dimensionless acoustic pressures on the surface of an oscillating sphere. The corresponding interior eigenfrequencies are determined by the spherical Bessel function of the first kind and order one, $\mathcal{J}_1(k^*a)=0$, which yields $k^*a=4.493\,409$, 7.725 233,... . Since in this case the tangential component of the particle velocity is not zero and the normal component of the velocity is not constant, the complex conjugate of the homogeneous equation (24) for the exterior region does not share the same eigenfrequencies as those of the homogeneous equation (23) for the interior problem. Hence, the solution given by Eq. (9) is unique, while those obtained by the BEM-based Kirchhoff integral formulation are not as $ka\rightarrow k^*a$.

Table II lists the dimensionless acoustic pressures obtained by BEM and Eq. (9), respectively, on the surface of an

TABLE II. Dimensionless acoustic pressures on the surface of an oscillating sphere at $\theta=45°$ around $ka=4.493\,409$ and 7.725 233.

ka	Exact values	BEM	Errors (%)	Eq. (9)	Errors (%)
2.0	7.071 07$E-01$	7.130 86$E-01$	0.85	7.046 91$E-01$	0.34
3.0	7.276 07$E-01$	7.069 85$E-01$	2.83	7.121 35$E-01$	2.13
4.0	7.232 40$E-01$	5.916 81$E-01$	18.19	7.250 03$E-01$	0.24
4.49	7.208 93$E-01$	1.914 77$E+01$	2 556.11	7.216 53$E-01$	0.11
4.4934	7.208 78$E-01$	3.206 28$E+01$	4 347.74	7.216 33$E-01$	0.10
4.493409	7.208 78$E-01$	3.109 66$E+01$	4 213.71	7.216 33$E-01$	0.10
4.4935	7.208 77$E-01$	3.073 48$E+01$	4 163.53	7.216 32$E-01$	0.10
4.5	7.208 48$E-01$	1.018 60$E+01$	1 313.06	7.215 94$E-01$	0.10
5.0	7.188 14$E-01$	8.438 12$E-01$	17.39	7.189 25$E-01$	0.02
6.0	7.157 57$E-01$	7.203 51$E-01$	0.64	7.126 63$E-01$	0.43
7.0	7.136 91$E-01$	2.935 55$E-01$	58.87	7.153 55$E-01$	0.23
7.72	7.126 13$E-01$	2.573 10$E+01$	3 510.80	7.133 62$E-01$	0.11
7.7252	7.126 06$E-01$	1.263 46$E+02$	17 630.13	7.133 51$E-01$	0.10
7.725233	7.126 06$E-01$	1.340 69$E+02$	18 713.90	7.133 50$E-01$	0.10
7.7253	7.126 06$E-01$	1.150 68$E+02$	16 047.49	7.133 50$E-01$	0.10
7.73	7.126 00$E-01$	2.616 64$E+01$	3 571.96	7.133 40$E-01$	0.10
8.0	7.122 62$E-01$	1.217 48$E+00$	70.93	7.127 89$E-01$	0.07
9.0	7.112 41$E-01$	9.115 00$E-01$	28.16	7.103 50$E-01$	0.13

TABLE III. Maximum relative errors of the magnitudes and phases of the dimensionless acoustic pressures on the surface of a finite cylinder by using Eq. (9) around $ka=2.8724$ and 3.9563.

ka	Exact mag.	Eq. (9)	Errors (%)	Exact phase	Eq. (9)	Errors (%)
1.5	$0.1083E-05$	$0.1083E-05$	0.035	29.04	29.02	0.056
2.0	$0.1759E-05$	$0.1760E-05$	0.057	40.32	40.34	0.043
2.5	$0.2498E-05$	$0.2502E-05$	0.158	53.36	53.44	0.145
2.8	$0.2798E-05$	$0.2806E-05$	0.288	70.57	70.62	0.077
2.87	$0.2868E-05$	$0.2877E-05$	0.331	74.58	74.63	0.061
2.872	$0.2869E-05$	$0.2879E-05$	0.332	74.70	74.74	0.061
2.8723	$0.2870E-05$	$0.2879E-05$	0.332	74.71	74.76	0.061
2.8724	$0.2870E-05$	$0.2879E-05$	0.332	74.72	74.76	0.061
2.873	$0.2870E-05$	$0.2880E-05$	0.333	74.75	74.80	0.060
2.88	$0.2877E-05$	$0.2887E-05$	0.337	75.16	75.20	0.059
2.9	$0.2897E-05$	$0.2908E-05$	0.351	76.30	76.34	0.054
3.0	$0.2997E-05$	$0.3010E-05$	0.427	82.04	82.06	0.030
3.5	$0.3497E-05$	$0.3534E-05$	1.053	110.71	110.63	0.073
3.9	$0.3897E-05$	$0.3951E-05$	1.397	133.65	133.64	0.005
3.95	$0.3947E-05$	$0.4001E-05$	1.384	136.51	136.53	0.011
3.956	$0.3953E-05$	$0.4007E-05$	1.381	136.86	136.86	0.013
3.9563	$0.3953E-05$	$0.4007E-05$	1.381	136.88	136.89	0.013
3.9564	$0.3953E-05$	$0.4008E-05$	1.381	136.88	136.90	0.014
3.957	$0.3954E-05$	$0.4008E-05$	1.381	136.92	136.93	0.014
3.96	$0.3957E-05$	$0.4011E-05$	1.380	137.09	137.11	0.015
4.0	$0.3997E-05$	$0.4051E-05$	1.359	139.38	139.42	0.028

oscillating sphere at $\theta=45°$. In using Eq. (9) we divide the spherical surface into eighteen rings along its generator in order to compare the surface acoustic pressures at the same locations as those of BEM with 130 nodes. Numerical results demonstrate that the BEM results show the sign of ill-conditioning over a large frequency range around the interior eigenfrequencies, whereas the results of Eq. (9) are accurate and unique.

In the last example, we use Eq. (9) to calculate the acoustic pressure on the surface of a right cylinder of finite length. The aspect ratio of the cylinder is $b/a=1$, where a and b are the radius and half length of the cylinder, respectively. The eigenfrequencies of the corresponding boundary value problem in the interior region are given by[40]

$$k^*_{mnq}=\sqrt{\left(\frac{m\pi}{2b}\right)^2+\left(\frac{\alpha_{ng}}{a}\right)^2}, \tag{33}$$

where m is a positive integer and α_{nq} is the qth root of the nth Bessel function

$$J_n(\alpha_{nq})=0, \quad n=0,1,2,\ldots . \tag{34}$$

Thus for $a=b=1(m)$ the first two eigenfrequencies are $k^*_{101}a=2.8724...$ and $k^*_{201}a=3.9563...$.

In the following, we examine the uniqueness of the numerical solutions given by Eq. (9) around $ka=2.8724$ and 3.9563, respectively. For convenience sake, the numerical results of Eq. (9) are compared with those of a point source of radius a_0 ($a_0=0.001a$) located at the center of the cylinder. The procedures are described as follows. First, we calculate the acoustic pressure distribution on a cylindrical surface due to a point source. Next, we determine the particle velocity $\hat{\mathbf{v}}(\mathbf{x}_S)$ on the cylindrical surface via Eq. (4). Once $\hat{\mathbf{v}}(\mathbf{x}_S)$ is specified, the surface acoustic pressure is recalculated using Eq. (9). The results thus obtained are compared with the surface acoustic pressure due to the point source.

In carrying out the numerical integrations in Eq. (9), we uniformly divide the surface into 48 rings along its generator, i.e., 12 rings on each of the 2 flat ends and 24 rings on the side wall. Each of these 48 rings is further divided into 48 equal segments along the circumference. Numerical integrations over each segment are carried out by Gaussian quadratures with nine interior points.[39] Table III lists the maximum relative errors of the magnitudes and phases of the dimensionless surface acoustic pressures given by Eq. (9) as compared with those from a point source around $ka=2.8724$ and 3.9563. It is seen that Eq. (9) shows no signs of the nonuniqueness difficulties around these characteristic frequencies. The relative error in the magnitude of the dimensionless surface acoustic pressure increases monotonically with ka, which is to be expected for a fixed grid size. The relative error in the phase of the surface acoustic pressure remains essentially the same however.

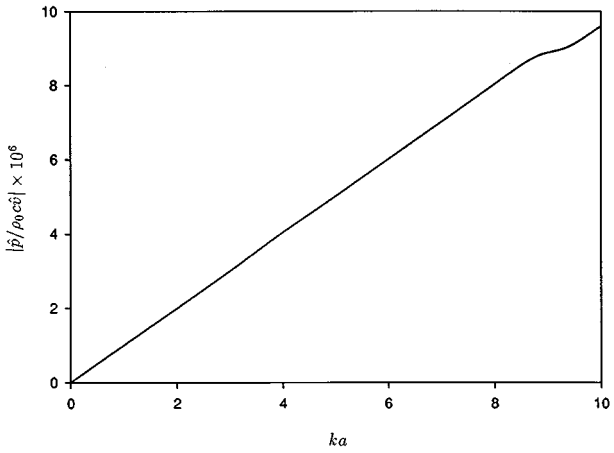


FIG. 1. Dimensionless surface acoustic pressure obtained by Eq. (9) at the center of a flat end of the cylinder as ka varying from 0 to 10 with an increment of $\Delta ka=0.01$.

Figure 1 displays the magnitude of the dimensionless surface acoustic pressure evaluated at the center of a flat end of the cylinder using Eq. (9) with ka varying from 0 to 10 at an interval of $\Delta ka = 0.01$. Within this frequency range, there could be many characteristic frequencies for the corresponding interior boundary value problem [see Eq. (34)]. Nevertheless, the numerical solution given by Eq. (9) is always unique.

In summary, solutions given by Eq. (9) are unique, except for the special case of a dilating sphere. Even under this circumstance, the ill-conditioning in the numerical computation is much less severe than that of the BEM-based Kirchhoff integral formulation.

IV. VALIDATIONS OF THE ALTERNATIVE FORMULATION

In this section, we demonstrate the validations of the alternative formulation derived in this paper on sound radiation from both separable and nonseparable geometry.

A. A dilating sphere

In the first example, we consider acoustic radiation from a sphere of radius a vibrating radially at a constant frequency ω in a free field. Since the sphere dilates uniformly in all directions, the particle velocity $\hat{\mathbf{v}} \equiv \hat{v}_n \mathbf{e}$, where $\hat{v}_n$ is a constant. Substituting $\hat{v}_n$ into Eq. (10) yields

$$\hat{p}(\mathbf{x}) = -\frac{i\omega\rho_0 \hat{v}_n}{4\pi} \int_S \left\{ \frac{(ikR-1)e^{ikR}}{R^2} \frac{\partial R}{\partial n} \left[\int_{S'} \frac{e^{ikR_{S'}}}{R_{S'}} dS' \right] \right.$$
$$\left. \times \left[2\pi - \int_{S'} \frac{(ikR_{S'}-1)e^{ikR_{S'}}}{R_{S'}^2} \frac{\partial R_{S'}}{\partial n_{S'}} dS' \right]^{-1} dS \right.$$
$$\left. - \frac{i\omega\rho_0 \hat{v}_n}{4\pi} \int_S \frac{e^{ikR}}{R} dS. \right. \tag{35}$$

The surface integrals in the square brackets on the right side of Eq. (35) can be evaluated independently and the results are

$$\left[\int_{S'} \frac{e^{ikR_{S'}}}{R_{S'}} dS' \right] = \frac{i2\pi}{k} (1 - e^{i2ka}), \tag{36a}$$

$$\left[2\pi - \int_{S'} \frac{(ikR_{S'}-1)e^{ikR_{S'}}}{R_{S'}^2} \frac{\partial R_{S'}}{\partial n_{S'}} dS' \right]$$
$$= \frac{i2\pi}{ka} (1 - ika)(1 - e^{i2ka}). \tag{36b}$$

Substituting Eq. (36) into (35) leads to

$$\hat{p}(\mathbf{x}) = \frac{i\omega\rho_0 \hat{v}_n}{4\pi} \int_0^{2\pi} \int_0^{\pi} \left[\left(\frac{a}{1-ika} \right) \frac{(1-ikR)}{R} \frac{\partial R}{\partial n} - 1 \right]$$
$$\times \frac{e^{ikR}}{R} a^2 \sin\theta_S \, d\theta_S \, d\phi_S, \tag{37}$$

where

$$R = \sqrt{r^2 + a^2 - 2ar(\sin\theta \cos\phi \sin\theta_S \cos\phi_S + \sin\theta \sin\phi \sin\theta_S \sin\phi_S + \cos\theta \cos\theta_S)}, \tag{38a}$$

$$\frac{\partial R}{\partial n} = -R^{-1}[(r \sin\theta \cos\phi - a \sin\theta_S \cos\phi_S)\sin\theta_S \cos\phi_S + (r \sin\theta \sin\phi - a \sin\theta_S \sin\phi_S)\sin\theta_S \sin\phi_S$$
$$+ (r \cos\theta - a \cos\theta_S)\cos\theta_S], \tag{38b}$$

where r is the distance from the center of the sphere to the observation point in the field, θ and ϕ are the polar and azimuthal angles of the field point, and θ_S and ϕ_S are the polar and azimuthal angles of a surface point.

Since the radiated acoustic pressure is spherically symmetric, we can set the field point to lie on any axis, say, $\theta = \pi$ and $\phi = 0$. Thus R and $\partial R/\partial n$ can be simplified to

$$R = \sqrt{r^2 + a^2 + 2ar \cos\theta_S}$$

and (39)

$$\frac{\partial R}{\partial n} = \frac{a + r \cos\theta_S}{\sqrt{r^2 + a^2 + 2ar \cos\theta_S}}.$$

Substituting Eq. (39) into (37), we obtain

$$\hat{p}(\mathbf{x}) = -\frac{\rho_0 c \hat{v}_n a}{2r} \left\{ \frac{(1+ika)e^{ik(r-a)} - (1-ika)e^{ik(r+a)}}{(1-ika)} \right.$$
$$\left. - [e^{ik(r-a)} - e^{ik(r+a)}] \right\}$$
$$= -\frac{i\rho_0 c \hat{v}_n ka}{(1-ika)} \left(\frac{a}{r} \right) e^{ik(r-a)}, \tag{40}$$

which agrees exactly with the analytic solution.[32]

B. An oscillating sphere

The second example concerns sound radiation from a sphere oscillating back and forth along the z axis at a constant frequency ω in an unbounded fluid medium. To check the validity of the alternative formulation, we substitute the analytic solution for the particle velocity[32]

$$\hat{v}_n(\mathbf{x}_S)=\hat{v}_c\cos\theta,\quad \hat{v}_\theta(\theta)=\frac{\hat{v}_c(1-ika)\sin\theta}{(2-k^2a^2-i2ka)},\qquad (41)$$

into the alternate formulation and then compare the resulting acoustic pressure with that of the analytic solution. In Eq. (41), θ is the angle between the unit outward normal and the z-axis direction, $\hat{v}_c$ is the magnitude of the velocity at the center of the sphere.

Because of the presence of $\hat{v}_\theta$, evaluations of the integrals in Eq. (10) become a bit lengthy. Without loss of generality, we demonstrate the evaluation of the surface acoustic pressure given by Eq. (9). To this end, we first carry out the line integral involved in Eq. (9). Using the axisymmetry, we can set $\mathbf{x}'_S$ at $(a,\pi,0)$ and $\mathbf{x}_S$ at (a,θ,ϕ). Substituting Eq. (41) into the line integral on the right side of Eq. (9) then yields

$$\int_\pi^\theta \hat{v}_\theta(\theta')a\,d\theta'=\int_\pi^\theta \frac{\hat{v}_c(1-ika)\sin\theta'}{(2-k^2a^2-i2ka)}\,a\,d\theta'$$
$$=-\frac{a\hat{v}_c(1-ika)(1+\cos\theta)}{(2-k^2a^2-i2ka)}.\qquad (42)$$

Substituting Eqs. (36) and (42) into (9) leads to

$$\hat{p}(\mathbf{x}_S)=-\frac{\rho_0 c\hat{v}_c(k^2a)}{2\pi(1-ika)(1-e^{i2ka})}\int_S\left[\frac{a(1-ika)}{(2-k^2a^2-i2ka)}\frac{(1+\cos\theta)(ikR_S-1)}{R_S}\frac{\partial R_S}{\partial n_S}+\cos\theta\right]\frac{e^{ikR_S}}{R_S}dS$$
$$-\frac{i\rho_0 c\hat{v}_c ka(1-ika)(1+\cos\theta)}{(2-k^2a^2-i2ka)}$$
$$=-\frac{i\rho_0 c\hat{v}_c}{ka(1-ika)(1-e^{i2ka})}\left\{\frac{(1-ika)[2-e^{i2ka}(i2k^3a^3-4k^2a^2-i4ka+2)]}{(2-k^2a^2-i2ka)}\right.$$
$$\left.-[1+k^2a^2+e^{i2ka}(k^2a^2+i2ka-1)]\right\}-\frac{i\rho_0 c\hat{v}_c ka(1-ika)(1+\cos\theta)}{(2-k^2a^2-i2ka)}$$
$$=\frac{i\rho_0 c\hat{v}_c ka(1-k^2a^2-i2ka)}{(1-ika)(2-k^2a^2-i2ka)}-\frac{i\rho_0 c\hat{v}_c ka(1-ika)(1+\cos\theta)}{(2-k^2a^2-i2ka)}$$
$$=-\frac{i\rho_0 c\hat{v}_c ka(1-ika)\cos\theta}{(2-k^2a^2-i2ka)},\qquad (43)$$

which agrees perfectly with the analytic solution.[32]

C. Right circular cylinders

In the previous two cases, the source geometry is separable so the analytic solutions exist and are well-known. In what follows, we consider sound radiation from right circular cylinders whose solutions can be obtained only numerically. In particular, we choose a slender cylinder with two flat ends. The aspect ratio of the cylinder is $b/a=10$, where a and b are the radius and half-length of the cylinder, respectively. As before, we compare the acoustic pressures given by Eqs. (9) and (10) with those of a monopole and a dipole, respectively. The procedures involved in these comparisons are the same as those described in Sec. III.

First, we consider the case in which the particle velocity field is generated by a monopole located at the center of the cylinder. Substituting these velocity components into Eqs. (9) and (10) yields the surface and field acoustic pressures, which are then compared with those from the monopole source. Excellent agreements are obtained for all the cases tested. For brevity, we only show the results of $kb=50$ below. The surface integrals in Eqs. (9) and (10) are implemented with the cylindrical surface discretized into 480 rings on the side wall and 24 rings on each end. These rings are further discretized into 48 segments along the azimuth. Numerical integrations over each segment are carried out using Gaussian quadratures with nine interior points.

Figure 2 demonstrates the comparisons of the real and imaginary parts of the dimensionless acoustic pressure obtained by using Eq. (9) with those of a monopole at $kb=50$ along the generator of the cylindrical surface. Here the abscissa represents a dimensionless distance s/a along the generator of the cylindrical surface. In particular, $s/a=0$ indicates the center of the side wall, $s/a=\pm 10$ the edges that separate the side wall from the flat end, and $s/a=\pm 11$ the centers of the two ends. Figure 3 shows the comparison of the radiation pattern of the dimensionless field acoustic pressure given by Eq. (10) at a radial distance of $r=20$ (m) with that of the monopole source.

Next, we repeat the same procedures and compare the calculated surface and field acoustic pressures with those of a dipole at $kb=100$. Because of the increase in the excitation frequency, we double the number of rings along the generator of the cylindrical surface as well as along the azimuth angle. Figure 4 depicts the comparison of the calculated real and imaginary parts of the surface acoustic pressure with the exact solution. The corresponding radiation patterns at a radial distance of $r=20$ (m) are shown in Fig. 5.

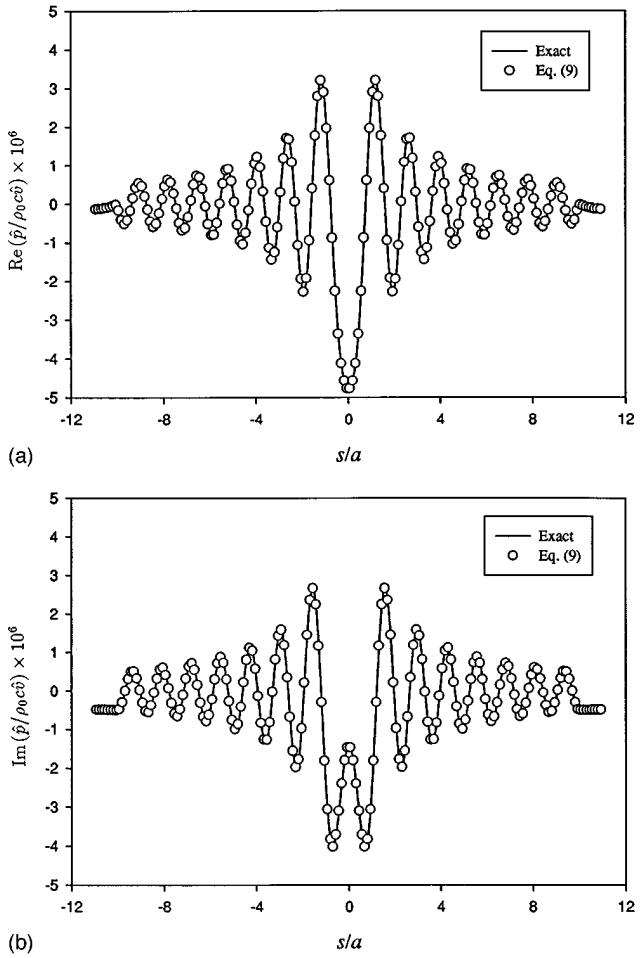


FIG. 2. Comparison of the dimensionless acoustic pressure given by Eq. (9): ○, and that due to a monopole: ———, along the generator of a slender cylinder at $kb=50$. (a) Real part; (b) imaginary part.

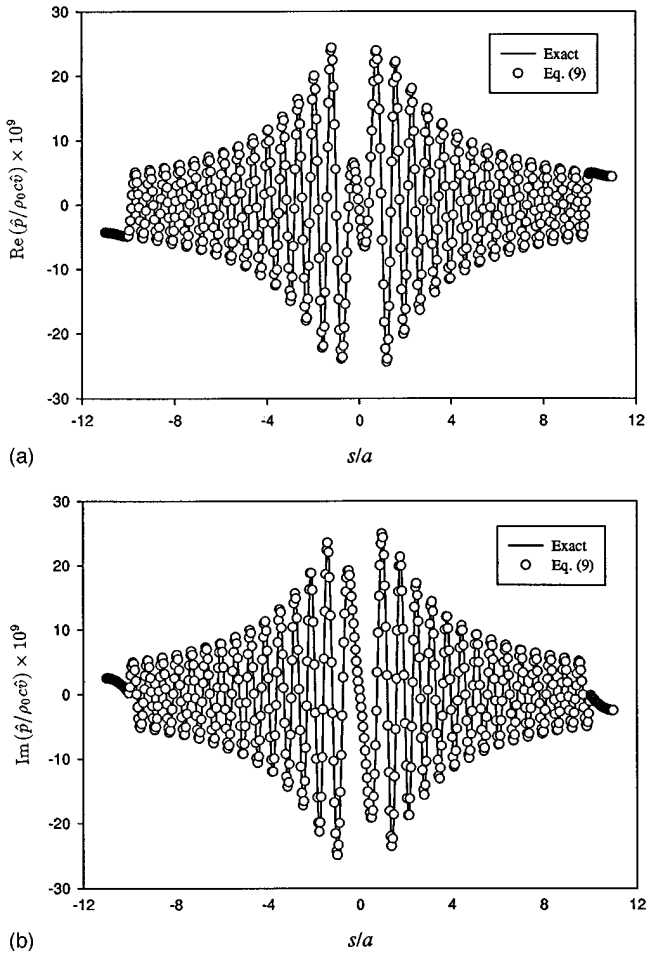


FIG. 4. Comparison of the dimensionless acoustic pressure given by Eq. (9): ○, and that due to a dipole: ———, along the generator of a slender cylinder at $kb=100$. (a) Real part; (b) imaginary part.

V. CONCLUSIONS

An alternative formulation is derived which enables one to predict the radiated acoustic pressure directly once the particle velocity on a control surface is specified. Solutions thus obtained are unique, except for the special case of a dilating sphere in which the particle velocity is uniform over the entire spherical surface in the radial direction. Even under this circumstance, the ill-conditioning in the numerical

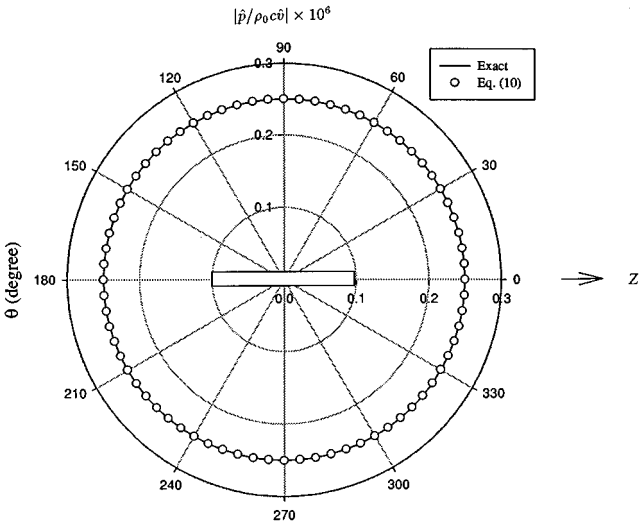


FIG. 3. Comparison of the radiation pattern of the dimensionless field acoustic pressure given by Eq. (10): ○, and that due to a monopole: ———, at a radial distance of $r=20$ (m) at $kb=50$.

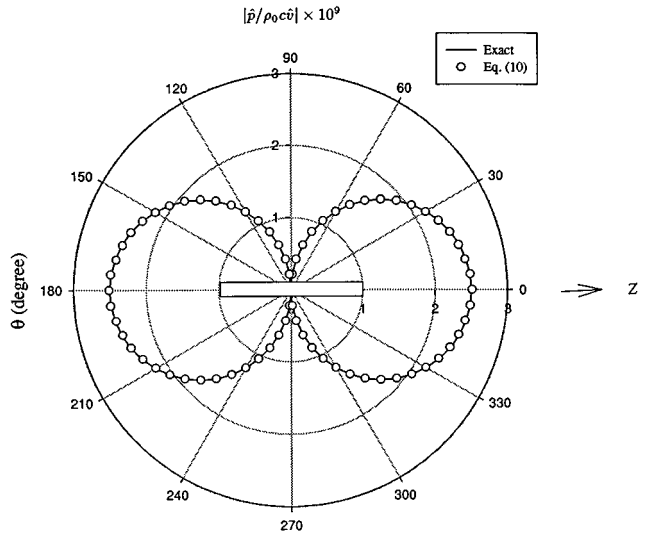


FIG. 5. Comparison of the radiation pattern of the dimensionless field acoustic pressure given by Eq. (10): ○, and that due to a monopole: ———, at a radial distance of $r=20$ (m) at $kb=100$.

computation is much less severe than that in the BEM-based Kirchhoff integral formulation. Moreover, the surface integral involved in this formulation can be readily implemented numerically using Gaussian quadratures. Hence the efficiency of numerical computation may be significantly enhanced.

ACKNOWLEDGMENTS

The authors wish to thank Professor Allan D. Pierce for many enlightening discussions and suggestions. Special thanks are also extended to Professor Yves H. Berthelot and Dr. Joe Vignola for providing many helpful materials on the theories and applications of nonintrusive laser Doppler velocimeter that may be used in conjunction with this alternative formulation. This work was supported by a grant from National Science Foundation, Grant No. CMS-9414424.

[1] A. J. Burton and G. F. Miller, ''The application of the integral equation method to the Numerical solution of some exterior boundary value problems,'' Proc. R. Soc. London, Ser. A **323**, 202–210 (1971).
[2] W. L. Meyer, W. A. Bell, B. T. Zinn, and M. P. Stallybrass, ''Boundary integral solutions of three-dimensional acoustic radiation problems,'' J. Sound Vib. **59**, 245–262 (1978).
[3] T. Terai, ''On calculation of sound around three dimensional objects by integral equation methods,'' J. Sound Vib. **69**, 71–100 (1980).
[4] G. H. Koopmann and H. Benner, ''Method for computing the sound power of machines based on the Helmholtz integral,'' J. Acoust. Soc. Am. **71**, 78–89 (1982).
[5] A. F. Seybert, B. Soenarko, F. J. Rizzo, and D. J. Shippy, ''Application of the BIE method to sound radiation problems using an isoparametric element,'' ASME Trans. J. Vib. Acoust. Stress Reliab. Design **106**, 414–420 (1984).
[6] A. F. Seybert, B. Soenarko, F. J. Rizzo, and D. J. Shippy, ''An advanced computational method for radiation and scattering of acoustic waves in three dimensions,'' J. Acoust. Soc. Am. **77**, 362–368 (1985).
[7] M. A. Latcha and A. Akay, ''Application of the Helmholtz integral in acoustics,'' ASME Trans. J. Vib. Acoust. Stress Reliab. Design **108**, 447–453 (1986).
[8] T. W. Wu, ''A direct boundary element method for acoustic radiation and scattering from mixed regular and thin bodies,'' J. Acoust. Soc. Am. **97**, 84–91 (1995).
[9] H. A. Schenck, ''An improved integral formulation for acoustic radiation problems,'' J. Acoust. Soc. Am. **44**, 41–58 (1968).
[10] H. Yeh and H. Z. Cummins, ''Localized fluid flow measurements with a He-Ne laser spectrometer,'' Appl. Phys. Lett. **4**, 176–178 (1964).
[11] L. E. Drain, *The Laser Doppler Technique* (Wiley, Chichester, 1980).
[12] J. Jarzynski, D. Lee, J. Vignola, Y. H. Berthelot, and A. D. Pierce, ''Fiberoptic Doppler systems for remote sensing of fluid flow,'' Proc. SPIE **925**, 250–254 (1988).
[13] K. J. Taylor, ''Absolute measurement of acoustic particle velocity,'' J. Acoust. Soc. Am. **59**, 691–694 (1976).
[14] K. J. Taylor, ''Absolute calibration of microphones by a laser-Doppler technique,'' J. Acoust. Soc. Am. **70**, 939–945 (1981).
[15] P. S. Dubbelday and H. C. Schau, ''Laser Doppler anemometry detection of hydroacoustic particle velocity,'' J. Acoust. Soc. Am. **86**, 891–894 (1989).
[16] J. Vignola, Y. H. Berthelot, and J. Jarzynski, ''Nonintrusive absolute measurements of acoustic particle velocity in fluids,'' in *Proceedings of the 13th International Congress on Acoustics*, edited by P. Pravica (Dragan Srnic, Sabac, Yugoslavia, 1989), Vol. 4, pp. 45–48.
[17] J. Vignola, Y. H. Berthelot, and J. Jarzynski, ''Laser detection of sound,'' J. Acoust. Soc. Am. **90**, 1275–1286 (1991).
[18] J. Vignola, Y. H. Berthelot, S. Jones, and J. Jarzynski, ''Equation of motion of microparticles in suspension in an insonified medium,'' J. Acoust. Soc. Am. **92**, 332–334 (1992).
[19] Y. H. Berthelot, M. Yang, and J. Jarzynski, ''Recent progress on laser-Doppler measurements in structural acoustics,'' in *Proceedings of the 4th International Congress on Intensity Techniques*, edited by G. Pavic (CETIM, Senlis, 1993), pp. 199–206.
[20] Y. H. Berthelot, J. Jarzynski, and M. Yang, ''Method and apparatus for detecting surface wave vector dynamics using three beams of coherent light,'' U.S. Patent No. 5,477,324 (19 December 1995).
[21] J. Vignola and B. Houston, ''The design of a three dimensional laser vibrometer,'' ASME Winter Annual Meeting, 93-WA/NCA-10, November 28 to December 3, Atlanta, Georgia, 1993.
[22] J. Vignola, H. Simpson, and B. Houston, ''Measurement of internal and external three-dimensional velocities of a submerged shell surface by optical vibrometry,'' Proc. SPIE **2868**, 71–81 (1996).
[23] R. Jones and C. Wykes, *Holographic and Speckle Pattern Interferometry* (Cambridge U.P., Cambridge, 1989), 2nd ed.
[24] J. N. Butters and J. A. Leendertz, ''Holographic and video techniques applied to engineering measurement,'' J. Measurement Control **4**, 349–354 (1971).
[25] F. M. Santoyo, M. C. Shellabear, and J. R. Tyrer, ''Whole field vibration analysis using pulsed phase-stepped ESPI,'' Appl. Opt. **30**, 717–721 (1991).
[26] J. D. Valera, J. D. C. Jones, and A. F. Doval, ''Whole field vibration phase measurement with ESPI,'' Proc. SPIE **2342**, 201–208 (1994).
[27] B. Pouet and S. Krishnaswamy, ''Additive-subtractive decorrelated electronic speckle pattern interferometer,'' Opt. Eng. (Bellingham) **32**, 1360–1369 (1994).
[28] B. Pouet and S. Krishnaswamy, ''Additive-subtractive phase modulated electronic speckle interferometry: analysis and fringe visibility,'' Appl. Opt. **33**, 6609–6616 (1994).
[29] L.-S. Wang and S. Krishnaswamy, ''Additive-subtractive speckle interferometry: Extraction of phase data in noisy environments,'' Opt. Eng. (Bellingham) **35**, 794–801 (1996).
[30] Q. Hu and S. F. Wu, ''Sound radiation from a finite cylinder,'' Proc. ASME Noise Control Acoust. Div. NCA 22, **1**, 39–46 (1996).
[31] S. F. Wu and Q. Hu, ''Predicting sound radiation from a vibrating object based on the particle velocity specified on the surface,'' to appear in the *Theoretical and Computational Acoustics '97*, edited by Lee, Pao, Schultz, and Teng (World Scientific, Singapore, 1998).
[32] A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (McGraw-Hill, New York, 1981), Chap. 4, pp. 165–194; Chap. 5, pp. 220–234.
[33] O. D. Kellogg, *Foundations of Potential Theory* (Dover, New York, 1954).
[34] A. F. Seybert and B. Soenarko, ''Radiation and scattering of acoustic waves from bodies of arbitrary shape in a three-dimensional half space,'' ASME Trans. J. Vib. Acoust. Stress Reliab. Design **110**, 112–117 (1988).
[35] A. F. Seybert and T. W. Wu, ''Modified Helmholtz integral equation for bodies sitting on an infinite plane,'' J. Acoust. Soc. Am. **85**, 19–23 (1989).
[36] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton U.P., Princeton, NJ, 1986).
[37] F. Smithies, *Inteqral Equations* (Cambridge U.P., Cambridge, England, 1985), pp. 51–52.
[38] S. F. Wu, ''Transient sound radiation from impulsively accelerated bodies,'' J. Acoust. Soc. Am. **94**, 542–553 (1993).
[39] G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers* (McGraw-Hill, New York, 1968), pp. 773–777.
[40] G. C. Gaunaurd, E. Tanglis, and H. Überall, ''Surface wave interpretation of the eigenfrequencies of a finite-length fluid cylinder,'' J. Acoust. Soc. Am. **67**, 764–769 (1980).

Autophasing of free volume oscillations of air cavities in water

V. V. Bredikhin, Yu. A. Kobelev, and N. I. Vasilinenko
Institute of Applied Physics, Russian Academy of Sciences, Nizhny Novgorod, Russia

(Received 3 April 1996; accepted for publication 22 April 1997)

This paper deals with theoretical and experimental investigations on free oscillations of air cavities formed in a metal plate lying on a water surface. The cavity is both an oscillator produced by the air elasticity in combination with the water mass, and a source of sound. Unlike gas bubbles in water, a cavity allows one to control the cubic nonlinearity coefficient up to a change of sign and to excite free oscillations rather easily. Measurements of the eigenfrequencies, damping factor, and the cubic nonlinearity coefficient of oscillations, taken for a variety of cavity volumes, are reported. Two- and three-cavity systems have been investigated to determine the eigenfrequencies of oscillations of the monopole and dipole modes. For the first time, the effect of autophasing of oscillations generated by ensembles of two and of three oscillators involved in a linear interaction via the near field was observed. The energy of the dipole oscillations of the system was transferred in this process to the monopole oscillations that effectively radiated acoustic waves. A possibility is discussed of using such oscillators as active elements for acoustic masers. © *1998 Acoustical Society of America.* [S0001-4966(97)05608-7]

PACS numbers: 43.25.Yw [MAB]

INTRODUCTION

Autophasing of oscillations is a common effect in various fields of physics. Despite this generality, though, it varies greatly from case to case, depending on a specific problem or a practical interest. For example, in electrical engineering it is phasing of self-oscillations of two oscillators to have the frequency of one oscillator stabilized by the other. In quantum electronics, the greatest interest is in the problems of phasing the oscillations of a multitude of excited particles with a view to developing electromagnetic oscillators at wavelengths ranging from centimeters to fractions of a millimeter. A class of such devices known as gyrotrons is based on the principle of phasing oscillations of free electrons in a magnetic field. Having proved classic by nature,[1] this effect challenged the acoustic physicists seeking to create an acoustic maser, a device for conversion of mechanical oscillations (possibly random) of single particles into a sound wave. One of the first efforts in this area was, apparently, the work[2] in which a positive solution to the problem of acoustic wave amplification in a medium with gas bubbles excited by some extraneous forces was obtained theoretically. However, such issues as how to prepare the medium and how to excite the bubbles remained unsolved. In Ref. 3 the authors offered a model of an active medium, essentially, a hard surface bounding a liquid with gas cavities. By randomly exciting the gas cavities on the surface it is, in principle, possible to generate an acoustic wave and amplify the surface-reflected sound. The gas cavities in this case serve as active elements. This seems to be a better model for a bubbly liquid, at least, for experimental validation of the possibility of phasing free oscillations of a cavity. The present paper seeks a qualitative rather than quantitative solution to this problem.

The only experimental work in this direction of which we are aware is Ref. 4, which reports an attempt at an experimental estimation of the cubic nonlinearity coefficient for a gas bubble in liquid by measuring the bubble resonance frequency. The experiment showed that a resonant bubble with a resonance frequency of about 10 kHz has a very low Q factor, so it breaks up with an increasing amplitude of oscillations before the cubic nonlinearity becomes sufficient for the oscillations' phasing process. Therefore, at this stage we believe it is more important to follow, within one experiment, the entire process—from excitation of the cavities to observation of the autophasing, at least qualitatively—rather than some aspect of the problem. Experimentally, the problem of cavity excitation is largely simplified, since a three-cavity system, even if only two cavities are excited, provides a nearly complete description of the oscillations' phasing effect for a multitude of oscillators (cavities).

Let us briefly consider the physics underlying the effect of phasing of natural oscillations. Specifically, we examine gas bubbles in liquid. Assume identical bubbles separated by much less than the acoustic wavelength at their natural frequency in the surrounding liquid. The distance separating the bubbles is assumed to be much larger than their radius. Either bubble, as an oscillator, is described by its eigenfrequency, Q factor, and a degree of nonlinearity, and as a radiator by the level of the radiated acoustic field. Due to the bubbles interaction via this field, the bubble region can radiate acoustic waves as a monopole, dipole, etc. By exciting the bubbles in phase one can make the whole of the bubbly region radiate as a monopole, at a frequency lower than a bubble eigenfrequency and with a damping factor larger than that of a single bubble. The change in frequency and damping factor in this case is proportional to the number of the bubbles. If one half of the bubbles is excited in phase and the other in antiphase, we shall have a dipole radiation, the frequency and the damping factor changing, again, in proportion to the number of particles excited in phase. Then, dividing the bubbles into four groups according to the phases of excited particles we obtain a quadrupole radiation, etc. Thus

 0001-4966/98/103(4)/1775/12/$10.00

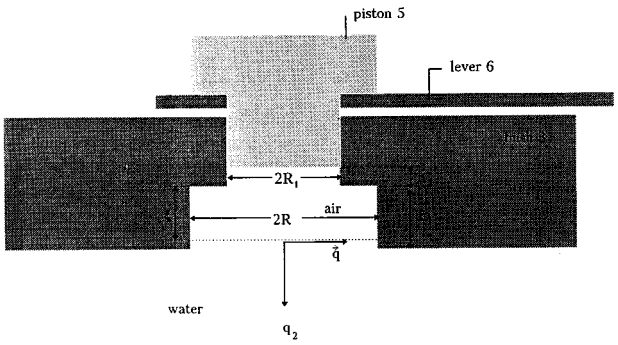


FIG. 1. Design of a cavity.

phase distribution of the initial amplitudes of bubble oscillations determines the type, frequency, and damping rate of the acoustic radiation from a bubbly region. Nonlinearity of bubble oscillations facilitates interaction between individual multipoles. The most interesting process here is when a multipole of the order approaching the number of particles, excited at some initial time, loses its energy at a rate close to that of the energy loss by a single bubble. As a result of the interaction between this and the monopole mode, the energy is transferred to the latter at a rate much higher than the rate of energy loss by radiation in the initially excited mode. This process is known as phasing of free oscillations.

I. THEORETICAL VALIDATION OF THE EXPERIMENT

We search for an equation to describe oscillations of a gas cavity in a hard unbounded plate lying on a water surface. The problem geometry is shown in Fig. 1. Unlike a similar problem solved in Ref. 3, the cavity here is homogeneous in height, and we also introduce a moving pump. It is assumed that the cavity size is small as compared to the length of the acoustic wave at the frequency ω_0 of the cavity's natural oscillations (the incompressible liquid approximation), and the water–air interface is not deformed under oscillations (the pump approximation).

A change in the impulse of the liquid column inside the cavity between the liquid–gas and plate–water ($q_2=0$) interfaces can be written as

$$\sigma\rho\,\frac{d}{dt}\,[(h_1-x)\dot{x}]=\sigma P_g-\int P(\mathbf{q},q_2=0)\,d\sigma, \tag{1}$$

where ρ is the water density, P_g is the gas pressure, $P(\mathbf{q},q_2=0)$ is the pressure at an arbitrary point with the coordinate $\mathbf{q}$ of the cross section $q_2=0$ (Fig. 1), $\sigma=\pi R^2$ is the area of the cross section $q_2=0$, $d\sigma$ the element of this cross-section area. The pressure $P(\mathbf{q},q_2=0)$ in our approximation is represented by the formula in Ref. 5:

$$\begin{aligned}P(\mathbf{q},q_2=0)&=P-\rho\left(\frac{\partial\varphi}{\partial t}+\frac{1}{2}\,\mathbf{v}^2\right)\\&=P+\frac{\rho}{2\pi}\int\ddot{\xi}\left(t-\frac{r}{c}\right)\frac{d\sigma}{r}-\frac{\rho\dot{\xi}^2}{2}.\end{aligned} \tag{2}$$

In (2) $\varphi=\varphi(\mathbf{q},q_2=0)$ is the liquid potential, $\mathbf{v}=\xi$ is the velocity of the liquid particles motion in the cavity, r is the distance between the point $\mathbf{q}$ and the point with the coordinate used for integration, P is the external pressure from water on the cavity (assumed homogeneous over cross section $q_2=0$), c is the sound velocity in liquid. From (1) and (2) using the equality in Ref. 5

$$\int\int\ddot{\xi}\left(t-\frac{r}{c}\right)\frac{d\sigma\,d\sigma'}{r}=\frac{8\pi R^3}{3}\,\ddot{\xi}-i\,\frac{\pi^2R^4}{2c}\,\dddot{\xi}$$

we find

$$\left(\frac{8R}{3\pi}+h_1-\xi\right)\ddot{\xi}-\frac{R^2}{2c}\,\dddot{\xi}-\frac{3}{2}\,\dot{\xi}^2=\frac{(P_g-P)}{\rho}. \tag{3}$$

For the adiabatic oscillations of gas in the cavity

$$P_gV^\gamma=P_0V_0^\gamma, \tag{4}$$

where V is the cavity volume, V_0 its unperturbed value, and P_0 is the static pressure in the liquid. After expansion of (4) into series in terms of the small parameter $\Delta V/V_0=u_f$ up to the 4th component inclusive we can obtain from Eq. (3) the equation for the cavity oscillations:

$$\begin{aligned}&\ddot{u}_1-\beta\dddot{u}_1+\alpha_1(u_1\ddot{u}_1+\tfrac{3}{2}\,\dot{u}_1^2)+\omega_0^2(1+\alpha_2u_f+\alpha_3u_f^2)u_f\\&\quad=-\frac{\omega_0^2\Delta P}{\gamma P_0},\end{aligned} \tag{5}$$

where $u_1=2\pi R^2\Delta\xi/V_0$ ($\Delta\xi$ is the differential of ξ under oscillations), $\beta=\pi R^2/2c[8R/3+\pi(h_1-\xi_{00})]$ is the radiation loss factor, ξ_{00} is the value of ξ upon the damping of oscillations (it differs from the initial value ξ_0 by a quantity proportional to the displacement Δh_{20} of the piston: $\xi_{00}=\xi_0-R_1^2\Delta h_{20}/R^2$), and ΔP is the variation of the pressure in the liquid affecting the cavity and

$$\begin{aligned}&\alpha_1=-\frac{V_0\pi}{[8R/3+\pi(h_1-\xi_{00})]R^2},\quad \alpha_2=-\frac{(\gamma+1)}{2},\\&\alpha_3=\frac{(\gamma+1)(\gamma+2)}{6},\quad \omega_0^2=\frac{\pi^2R^2\gamma P_0}{\rho V_0[8R/3+\pi(h_1-\xi_{00})]}.\end{aligned} \tag{6}$$

As follows from the definition of the variables u_1 and u_f,

$$u_f=u_1+\frac{\pi R_1^2\Delta h_2}{V_0}. \tag{7}$$

The cavity is excited by varying Δh_2 from zero to Δh_{20} over the time Δt which is shorter than the period of the cavity oscillations, $2\pi/\omega_0$. The quantity Δh_2 remains constant afterwards. Since the definitions for β, α_1, and ω_0^2 account for the piston displacement, we may assume $u_f=u_1=u$ and the correction for u_f from Eq. (7) to be assigned initially.

In (5) the influence of other cavities is taken into account in the value for pressure ΔP in the right-hand part of the equation. This quantity represents the sum of the fields scattered by these cavities, which can be written as[5]

$$\Delta P_j=\sum_{k\neq j}\frac{\rho V_{0k}}{2\pi r_{kj}}\,\ddot{u}_k\left(t-\frac{r_{kj}}{c}\right). \tag{8}$$

Here subscript k indicates that the given quantity refers to the cavity with the number k and r_{kj} is the distance between the centers of the cavities j and k. The summation is taken

over all cavities except for the one labeled j. For identical cavities spaced at equal distances ($r_{kj}=r_0$), if the delay r_0/c is shorter than the oscillation period of the cavities $2\pi/\omega_0$, such that $\ddot{u}(t-r_0/c)\cong\ddot{u}(t)-(r_0/c)\dddot{u}(t)$, we find from (5) and (8) that

$$\ddot{u}_j+\kappa\dot{u}_j-\beta\dddot{u}_j+\alpha_1(u_j\ddot{u}_j+\tfrac{3}{2}\dot{u}_j^2)+\omega_0^2(1+\alpha_2u_j+\alpha_3u_j^2)u_j$$
$$=-\nu\sum_{k\neq j}\left(\ddot{u}_k-\frac{r_0}{c}\dddot{u}_k\right), \tag{9}$$

where $\nu=c\beta/r_0$ is the cavities' coupling coefficient. Equation (9) accounts for the losses κ through heat exchange in the cavity and the liquid viscosity.

We now make a substitution $\omega_0 t=\tau$ and introduce another variable, y_j, assuming it is slowly varying from τ by the formula $u_j=\frac{1}{2}(y_je^{i\tau}+\text{c.c.})$. Then, on a further condition $du_j/d\tau=(i/2)(y_je^{i\tau}-\text{c.c.})$, we find from (9) the expression for the amplitudes y_j:

$$\frac{dy_j}{d\tau}+(\delta-ia|y_j|^2)y_j=-(i\nu+\delta_r)\sum_{k\neq j}y_k, \tag{10}$$

where $\delta=\delta_t+\delta_r$, $\delta_t=\kappa/2\omega_0$, $\delta_r=\beta\omega_0/2$ and $a=(3\alpha_3/2-5\alpha_2^2/3-\alpha_1^2/3)/4$ is the nonlinearity coefficient. We shall investigate the oscillators' behavior in terms of Eqs. (10).

First let us consider free oscillations of one cavity. The right-hand part of Eq. (10) is zero, and its solution is expressed by the formula

$$y=y_0\exp\left\{-\delta\tau+i\left[S_0+\frac{ay_0^2}{2\delta}(1-e^{-2\delta\tau})\right]\right\}. \tag{11}$$

Here S_0 is the initial phase of the oscillators, and the amplitude $y_0=u_0$ is defined by Eq. (7). Formula (11) allows determination of the nonlinearity coefficient a from the spectrum of the natural oscillations. We now expand $y(\tau)$ into the Fourier series:

$$y(\tau)=\int_{-\infty}^{+\infty}F(\Omega)e^{i\Omega\tau}\,d\tau. \tag{12}$$

Then, for $F(\Omega)=1/2\pi\int_0^\infty y(\tau)e^{-i\Omega\tau}\,d\tau$ one can easily derive the expression

$$F(\omega)=\frac{y_0}{2\pi\delta(1+i(\Omega/\delta))}\left\{1+\sum_{n=1}^{\infty}\eta^n\prod_{m=1}^{n}A_m\right\}, \tag{13}$$

where $A_m=1/[\Omega/\delta-i(2m+1)]$ $\eta=ay_0^2/\delta$. The spectral component $F(f)$ for the dimensional circular frequency f is described by the formula resulting from (13):

$$F(f)=\frac{y_0\Delta}{2\pi(\delta+i\Omega)f_0}\left\{1+\sum_{n=1}^{\infty}\eta^n\prod_{m=1}^{n}A_m\right\}. \tag{14}$$

Here $\Omega=2\pi(f-f_0)/f_0$ and $f_0=\omega_0/2\pi$, Δ is the width of discretization of the spectrum analyzer, expressed in Hz. From the spectrum of module F, yielded by spectrum analyzers, using Eq. (14) we can find the amplitude u_0, the nonlinearity coefficient a, the frequency f_0, and the damping factor δ:

$$\delta=\frac{\Delta f}{2f_0}=\frac{1}{2\Omega}, \tag{15}$$

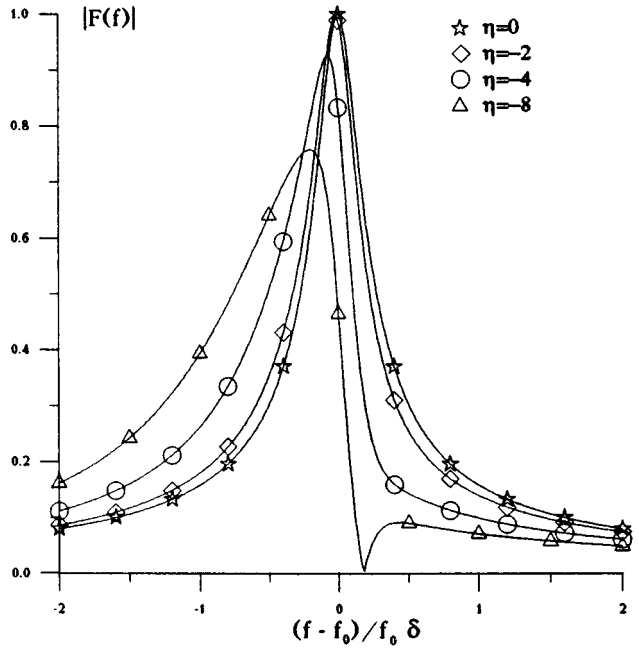


FIG. 2. Variation of the natural oscillations spectrum of a cavity versus the parameter η.

where Δf is the width of the resonance curve at $1/\sqrt{2}$ amplitude, Q is the quality of oscillations.

In the experiment we studied the following parameter values: $V_0=0.2$ cm^3, $R=0.38$ cm, $R_1=0.25$ cm, $h_1=0.4$ cm, and $\xi_{00}=0.2$ cm. Using expressions (6) and the definitions that go with Eq. (10) we find

$$f_0=390\ \text{Hz},\quad a=-0.8,\quad \delta_r=10^{-3}. \tag{16}$$

By way of example, for a gas bubble at this frequency $\delta_r=kR/2\cong7\times10^{-3}$. The real losses in the cavity are much higher: $\delta\cong1.5\times10^{-2}$. This will be discussed in greater detail in the experimental part of the paper. Figure 2 shows variable $|F(f)|$ normalized to its maximum value versus the parameter $(f-f_0)/\delta f_0$ for different values $\eta=au_0^2/\delta$ ($\eta=0,-2,-4,-8$). The quantity η will further be determined by comparison with the experimental curve.

Let us examine two cavities separated by the distance r_0 (in the experiment $r_0=2$ cm). In Eq. (10), y_1 and y_2 are replaced with z and w using the formula

$$z=(y_1+y_2)/2,\quad w(y_1-y_2)/2. \tag{17}$$

The parameter z describes the monopole component of the radiation, w is used for the dipole component; substitution in (10) yields

$$\dot{z}+\delta_z z+i[\nu-a(2|w|^2+|z|^2)]z=iaz^*w^2,$$
$$\dot{w}+\delta_w w-i[\nu+a(2|z|^2+|w|^2)]w=iaw^*z^2, \tag{18}$$

where $\delta_z=\delta_t+2\delta_r$ (the monopole mode damping factor increases), $\delta_w=\delta_t$ (the dipole mode damping factor decreases). It follows from (18) that while the monopole mode frequency decreases by ν, the dipole mode frequency increases by just the same quantity. For $r=2$ cm and the above parameters of the cavity, $\nu=0.07$.

Now consider two limiting cases. In one, it is the monopole mode that is excited initially, and the question is whether it will be able to transfer a part of its energy to the dipole mode. Assume both cavities to be excited in phase with the amplitudes $y_0(\tau=0)$. Then, ignoring the nonlinear components with w in the equation for z and w, and, for simplicity, the dissipative component in the equation for z, we have

$$z=y_0e^{-i\omega_n\tau}, \quad w=w_0e^{(\lambda-i\omega_n)\tau}, \tag{19}$$

where $\omega_n=\nu-ay_0^2$, and

$$\lambda=-\delta_w\pm 2\sqrt{-\nu(\nu+ay_0^2)}. \tag{20}$$

At $\lambda>0$ the dipole oscillations increase exponentially. This is possible only when $a<0$. The threshold for this process $(\lambda=0)$ is determined by the excitation intensity from (20):

$$-ay_0^2>\nu+\frac{\delta_w^2}{4\nu}. \tag{21}$$

If this condition is met, the monopole and the dipole modes oscillate at one frequency (19), and the energy flows from the monopole to the dipole mode. As follows from (19), the oscillation frequency near the threshold at first is close to the double frequency of the monopole mode oscillations. This seems to be the major result of the experimental demonstration of the process that can be referred to as antiphasing of cavities.

Let us address the other case. We shall seek the conditions that permit phasing of particles. To this effect, the cavities are excited in antiphase with amplitudes y_0. Using the same assumptions as in the first case, the only difference being that z and w exchanged positions, for the solutions

$$z=z_0e^{(\lambda+i\omega_f)\tau}, \quad w=y_0e^{+i\omega_f\tau}, \tag{22}$$

where $\omega_f=\nu+a|y_0|^2$, we find the expression for λ:

$$\lambda=-\delta_z\pm 2\sqrt{-\nu(\nu-ay_0^2)}. \tag{23}$$

The instability here is at $a>0$, and its threshold is defined, again, by expression (21), where we need to change the sign of the left-hand part. The oscillation frequency is equal to the double frequency of dipole oscillations, naturally, at the onset of the oscillations. We reduce our investigation only to the two limiting cases considered above, as they are most interesting and in line with the objectives of our study.

Now we shall focus on the process of oscillation phasing in a three-cavity system. To identify the oscillation modes, we label each cavity with a number: $k=0,1,2$, and represent the oscillation amplitude y_k of cavity k as a series expansion of the complete orthogonal system of functions $\exp(i2\pi kl/3)$, where l takes the values 0, 1, 2:

$$y_k=\sum_{l=0}^{2} A_l\exp(i2\pi kl/3). \tag{24}$$

This expansion, first suggested in Ref. 3, proved very convenient in studies of oscillation processes. Using the orthogonality property,

$$\frac{1}{3}\sum_{k=0}^{2}\exp(i2\pi km/3)=\begin{cases}1 & \text{if } m=0,\pm 3,\pm 6,...\\ 0 & \text{if } m\neq 0,\pm 3,\pm 6,...\end{cases}. \tag{25}$$

From (24) we find the feedback

$$A_m=\frac{1}{3}\sum_{k=0}^{2} y_k\exp(-i2\pi km/3). \tag{26}$$

By substituting (24) in Eq. (10) via (25) for the variables

$$b_m=A_m\exp\left(-i2a\sum_{k=0}^{2}|A_k|^2\tau\right) \tag{27}$$

we derive the equations:

$$\dot{b}_0+[\delta+2\delta_r+i(2\nu+a|b_0|^2)]b_0 = ia(2b_0^*b_1b_2+b_1^2b_2^*+b_2^2b_1^*), \tag{28}$$

$$\dot{b}_1+[\delta-\delta_r+i(-\nu+a|b_1|^2)]b_1 = ia[b_0(2b_1^*b_2+b_2^*b_0)+b_2^2b_0^*], \tag{29}$$

$$\dot{b}_2+[\delta-\delta_r+i(-\nu+a|b_2|^2)]b_2 = ia[b_0(2b_2^*b_1+b_1^*b_0)+b_1^2b_0^*]. \tag{30}$$

It is clear from the above equations that in the linear approximation these are independent modes of oscillations. The shift of the mode eigenfrequency b_0 is twice that of the monopole oscillations of two cavities, and the radiation loss factor here is three times as much for the oscillations of a single cavity. The question of to which oscillations of the cavity system these modes correspond can be understood from Eq. (24). Assuming mode A_0 (the monopole one) excited, we find from (24) that $y_k=A_0$, i.e., all cavities oscillate in phase. For the mode A_1, $y_k=A_1\exp(i2\pi k/3)$, which implies that the cavities' oscillations are shifted in phase by $2\pi/3$ relative to A_1, and that y_0 and A_1 are in phase. Finally, mode A_2 is again in phase with amplitude y_0, while y_1 and y_2 have exchanged positions as opposed to the case when mode A_1 is excited. By excitation of mode A_0, the three identical cavities radiate as a monopole. With excitation of modes A_1 and A_2 we have two dipoles that can rotate about each other scanning the directivity diagram. Another noteworthy finding is that the oscillation frequencies of a system of identical cavities are determined, along with the geometrical sizes of the cavities, by the value of mode-mode coupling.

Now consider a few more of the oscillation stability problems. We assume mode A_0 excited (all cavities initially oscillate in phase), and modes A_1 and A_2 small. Again, for simplicity ignoring the dissipative term in (28), and the components quadratic in b_1 and b_2 in all the equations, we derive

$$b_0=y_0\exp(i\omega_{00}\tau), \quad \omega_{00}=-(2\nu+a|y_0|^2), \tag{31}$$

$$\dot{b}+(\delta_g-i\nu)b=iab_0^2b^*, \tag{32}$$

where $b=b_1+b_2$ and $\delta_g=\delta-\delta_r$. The solution to Eq. (32) is written in the form: $b=(b_{00}^2/|b_{00}|)B\exp(\lambda+i\omega_{00})\tau$, $(|b_{00}|=y_0)$, and λ is defined as

$$\lambda=-\delta_g\pm 3\sqrt{-\nu(\nu+\tfrac{2}{3}ay_0^2)}. \tag{33}$$

At $a>0$ the monopole oscillations are stable. Instability sets in only at $a<0$. The excitation intensity threshold ($\lambda=0$) is defined by the formula

$$-ay_0^2=\frac{1}{2}\left(3\nu+\frac{\delta_g^2}{3\nu}\right). \tag{34}$$

The near-threshold oscillations arise at a frequency shifted downwards by the quantity

$$\Delta\omega=-2\nu+ay_0^2. \tag{35}$$

Let us consider another situation when two modes, b_{10} and b_{20}, are excited initially. Assuming these modes assigned, i.e., ignoring the damping and the terms proportional to b_0 in Eqs. (29) and (30), we have

$$b_1=b_{10}e^{i\omega_1\tau},\quad b_2=b_{20}e^{i\omega_2\tau},\quad \omega_{1,2}=\nu-a|b_{10,20}|^2. \tag{36}$$

We further seek b_0 in the form

$$b_0=\frac{b_{10}b_{20}}{|b_{10}||b_{20}|}\,Te^{i(\omega_1+\omega_2)\tau/2}. \tag{37}$$

Substitution of (37) in Eq. (28) yields the expression for T:

$$\dot{T}+\left\{\delta+2\delta_r+i\left[3\nu-\frac{a}{2}\left(|b_{10}|^2+|b_{20}|^2\right)\right]\right\}T$$
$$=2ia|b_{10}||b_{20}|T^*+\left[ia\,\frac{b_{10}b_{20}^*}{b_{20}}\,e^{i(2\omega_1-\omega_2)\tau}\right.$$
$$\left.+ia\,\frac{b_{20}b_{10}^*}{b_{10}}\,e^{i(2\omega_2-\omega_1)\tau}\right]|b_{10}||b_{20}|. \tag{38}$$

The last two components in the right-hand part of Eq. (38) represent an applied force that can induce growth of T, but in our approximation this term can be zero. The fact is that two modes, b_1 and b_2, can be excited given small b_0 only if certain relations between the initial amplitudes and the cavities' oscillation phases are fulfilled. From (26) we obtain

$$\begin{aligned}
A_0&=\tfrac{1}{3}(y_0+y_1+y_2),\\
A_1&=\tfrac{1}{3}(y_0+y_1e^{-i2\pi/3}+y_2e^{i2\pi/3}),\\
A_2&=\tfrac{1}{3}(y_0+y_1e^{i2\pi/3}+y_2e^{-i2\pi/3}).
\end{aligned} \tag{39}$$

Let $A_0=0$; then, for the case of excitation of only two cavities ($y_{20}=0$) we derive from (39):

$$y_{10}=-y_{00},\quad |b_{10}|^2=|b_{20}|^2=\tfrac{1}{3}y_{00}^2. \tag{40}$$

The last two components in the right-hand part of Eq. (38) are zero. Nevertheless, as in the case of instability, any violation of the condition (40) will lead to appearance of the applied force in (38). If $y_{10}=-y_{00}+\delta y$ and $\delta y\ll y_{00}$, it follows from (39) that

$$A_{00}=b_{00}=\tfrac{1}{3}\delta y. \tag{41}$$

Using (39) and (40) we represent Eq. (38) in the form

$$\dot{T}=[\delta_m+i\varphi]T=\frac{2ia}{3}\,y_{00}^2\left(T^*-\frac{1}{9}\,\delta y e^{i\omega_1\tau}\right). \tag{42}$$

Here $\delta_m=\delta+2\delta$, and $\varphi=3\nu-(a/3)y_{00}^2$. The solution of a homogeneous (without the 2nd component in the right-hand part) equation (42) is written as

$$T=c_1e^{\lambda\tau}+c_2e^{\lambda^*\tau}. \tag{43}$$

Substituting the latter in (42) we find the definitions for λ and c_2:

$$\lambda_{1,2}=-\delta_m\pm[\tfrac{4}{9}a^2y_{00}^4-\varphi^2]^{1/2}, \tag{44}$$

$$c_2=\frac{2iay_{00}^2}{3(\lambda^*+\delta_m+i\varphi)}\,c_1^*. \tag{45}$$

Since λ has two values, there must be two solutions too:

$$T=c_1e^{\lambda_1\tau}+\frac{2iay_{00}^2}{3(\lambda_1^*+\delta_m+i\varphi)}\,c_1^*e^{\lambda_1^*\tau}+c_2e^{\lambda_2\tau}$$
$$+\frac{2iay_{00}^2}{3(\lambda_2^*+\delta_m+i\varphi)}\,c_2^*e^{\lambda_2^*\tau}. \tag{46}$$

By applying the variation method to the constants in (46) we can find a particular solution of (42), but as it is quite clumsy, we do not present it here. When the nonlinearity level is low, there is no resonance with the applied force oscillating at the frequency ω_1, which is why an increase in T is limited by a quantity of the order δy. For a small τ the particular solution increases linearly:

$$T\cong-\frac{i2ay_{00}^2}{27}\,\tau. \tag{47}$$

For any excitation level near the instability threshold the particular solution may be of the order of the fundamental oscillations. The instability region $\mathrm{Re}\,\lambda>0$ of the eigensolutions to (46), regardless of the damping term, is defined by the inequality that follows from Eq. (44):

$$y_{00}^2>\frac{9\nu}{5a}\begin{cases}1{,}4 & \text{if } a>0\\ -3{,}4 & \text{if } a<0.\end{cases} \tag{48}$$

Consider the excitation of three cavities whose oscillations are shifted in phase by $2\pi/3$. The formulas (39) here yield:

$$A_{00}=b_{00}\ll y_{00},\quad A_{10}=y_{00},\quad A_{20}=b_{20}\ll y_{00}. \tag{49}$$

Neglecting in (28)–(30) the components quadratic in b_0 and b_1, and the damping of mode b_1, we find

$$\begin{aligned}
&b_1=y_{00}e^{i\omega_1\tau},\quad \omega_1=\nu-ay_{00}^2,\\
&\dot{b}_0+(\delta_m+i2\nu)b_0=iay_{00}^2e^{2i\omega_1\tau}b_2^*,\\
&\dot{b}_2+(\delta_g-i\nu)b_2=iay_{00}^2e^{2i\omega_1\tau}b_0^*.
\end{aligned} \tag{50}$$

For solutions of the form

$$b_0=b_{00}e^{(\lambda+i\omega_1)\tau},\quad b_2=b_{20}e^{(\lambda^*+i\omega_1)\tau} \tag{51}$$

the expression for λ is found from (50):

$$[\lambda+\delta_m+i(3\nu-ay_{00}^2)](\lambda+\delta_g+iay_{00}^2)-a^2y_{00}^4=0, \tag{52}$$

where from, ignoring the damping factor, we have

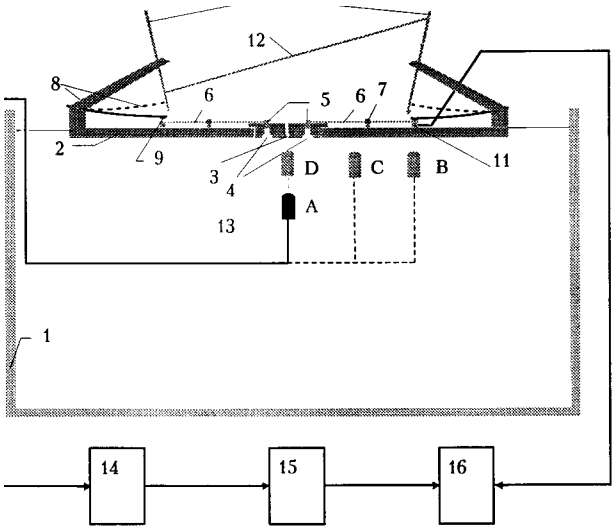


FIG. 3. Block diagram of the setup (see the text for detailed description).

$$\lambda = \frac{3\nu}{2}\left(-i \pm \sqrt{-1+\frac{4ay_0^2}{3\nu}}\right). \tag{53}$$

At $a<0$ the specified mode of oscillations is stable, and at $a>0$, starting with the exciting quantity

$$y_0^2 > 3\nu/4a, \tag{54}$$

we observe an increase in oscillations of the monopole mode.

The problems discussed here are most interesting from the viewpoint of a linear interaction between the free oscillations of the cavities. Major difficulties are involved in finding an optimal oscillator. Such an oscillator must show stability to oscillations (see the following section); low intrinsic losses relative to the loss by radiation; low nonlinear losses; a large and real nonlinearity coefficient a with a sign specified for the problem, etc. For a gas cavity, the above theory indicates that the nonlinearity coefficient is negative and of a magnitude nearly the same as for a bubble, which makes oscillation phasing practically impossible. Next we shall consider the problem experimentally.

II. EXPERIMENTAL SETUP

A block diagram of the setup is shown in Fig. 3. Measurements were taken in a tub 150×120 cm, the height of the water column was 47 cm. We placed a 30-cm diam, 1-cm-thick metal disk 2 with a 3-cm-diam hole in the center at the water surface. A 1.85-cm-thick bushing 3 with three cavities 4 was fitted into the hole. The cavities are arranged in the angles of an equilateral triangle with a 2-cm side. The cavity dimensions are specified above. A polytetrafluorothylene (PTFE) piston 5 is tightly fitted into the upper part of either cavity. Air is forced into the cavities with a syringe of 1-cm^3 volume and 0.01-cm^3 value of scale division. The cavity is broadened at the bottom to increase gas bubble ''durability'' against a sudden move of the piston. As was discovered in the experiment, bubble durability is also increased when the piston is moved upwards to excite oscillations. The usual break-up mechanism for a bubble is by one or more smaller bubbles breaking away into the water, but sometimes foam is produced. It should be noted that in our experiment a bubble could withstand up to 100 impacts of the piston, displaying a remarkable stability and identity of oscillations, but sometimes it broke up after a few shocks. The reason for this remains unclear.

Lever 6 is fastened to axis 7. One end of the lever is fitted between the piston ''head'' and the bushing, the other one is adjusted so as to be hit by a spring 8 with a ''hammer'' 9. The spring essentially is a steel plate 10 cm long, 3.2 cm wide, and 0.06 cm thick, clamped over 2.5-cm length at one end. The other end of the spring holds the ''hammer'' 9. This end is elevated to 0.5–2 cm until caught on lever 10; by turning this lever one can release the spring. The average speed of the hammer-driven piston was 0.07–0.15 cm/s. The motion of lever 6 subject to the hammer strokes is restricted by a screw stopper with a synchronizing contact 11 activating the oscillation analyzer 16 (IWATSU signal analyzer SM-21008). The second cavity has a similar design. Its lever 10 is connected to lever 10 of the other cavity via a rigid bar 12. By moving the lever of one cavity, it is possible to activate both springs. Despite the system rigidity, the moments of the shocks from the two hammers were shifted by the value varying from zero to plus/minus several milliseconds. However, this system has shown better performance than a number of electromagnetic piston drivers and spring triggers which produced either an insignificant movement of the piston of about 0.01 cm (with a low speed of motion due to friction) or a considerable time delay (5–10 ms) of the strokes. And still, for all the benefits of the above system, excitation of cavities in the specified phase is an extraordinary phenomenon which complicates the experimental procedure.

The signal sent by the oscillating cavities was received by a hydrophone either at point A directly beneath the cavities at a distance of 5 cm, where dipole radiation is nearly zero, or at points B, C on the dipole axis at a 5-cm distance from the center of the cavities-formed triangle, and at 2 cm from the surface of the metal disk, respectively, where the monopole and dipole radiation of the cavities is present. The hydrophone 13 transmitted the signal to amplifier 14 and filter 15, and then to the oscillation analyzer 16. We chose the signal pickup points and a one-octave filter so as to reduce the level of reverberation.

III. MEASUREMENTS

At first we measured the free oscillation parameters of one cavity for various air volumes: $V_0=0.4$, 0.35, 0.25, and 0.2 cm^3. It was arranged that for volume 0.4 cm^3 the air was partly left outside the cavity. The air volume 0.35 cm^3 filled the cavity to the metal–liquid interface, and the volume 0.25 cm^3 came to half of the height h_1. The air–gas interface for the 0.2-cm^3 gas volume was a fraction of a millimeter lower than the transition from the cavity radius R_1 to the radius R.

The hydrophone was located under the cavity at a depth $r=5$ cm. The sensitivity of the overall receiving path was $\mu=1$ Pa/mV$=10$ g cm^{-1} s^{-2} mV^{-1}. Using (8) and (14) we can define the modulus of the Fourier component $F(f)$:

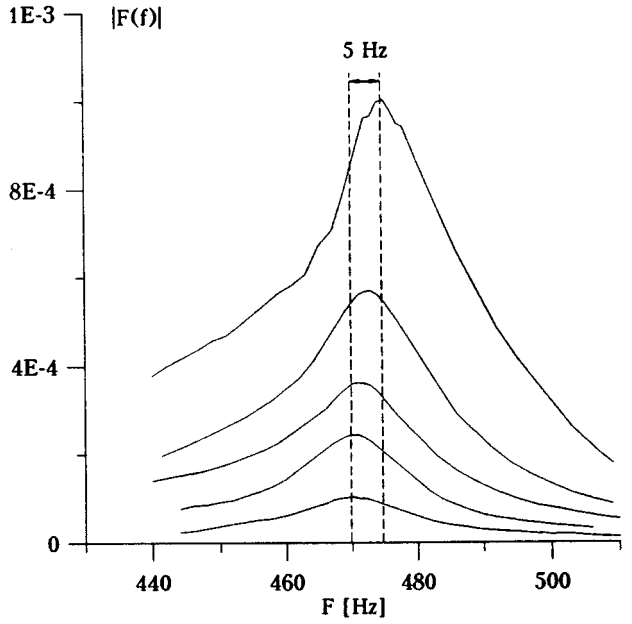


FIG. 4. The natural oscillations spectra for a cavity of volume 0.4 cm^3 for various oscillation amplitudes. Dashed line indicates the spectrum maxima bias.

$$|F(f)| = \frac{r\mu |E(f)|}{2\pi\rho V_0 f^2}, \tag{55}$$

where $E(f)$ is the Fourier component of the electric signal spectrum. From (14) it is possible to derive the amplitude of oscillations, y_0. For $\Omega = 0$ we have

$$y_0 = 2\pi F(f = f_0)\,\delta f_0 / q\Delta, \tag{56}$$

where Δ is the frequency width of discretization of the analyzer (Hz). The quantity

$$q = \left| \left(1 + \sum_{n=1}^{\infty} \eta^n \prod_{m=1}^{n} A_m \right)_{f=f_n} \right|$$

describes the nonlinear correction for the amplitude; its value is taken at the frequency f_n corresponding to the nonlinear resonance. Computer processing of the positions of frequencies f_n and the values of q from (14) yields the formulas for q and η:

$$q = 1 + (3{,}7 - 6{,}8|\eta| + 0{,}3|\eta|^2)|\eta| \times 10^{-3}, \tag{57}$$

$$|\eta| = \left(65 - 158\frac{|f_n - f_0|}{f_0} + 161\frac{(f_n - f_0)^2}{f_0^2} \right) \frac{|f_n - f_0|}{f_0}. \tag{58}$$

The sign of η is determined by the sign of the difference $f_n - f_0$. Figure 4 demonstrates the oscillation spectra of the 0.4-cm^3 cavity for various amplitudes. The maximum frequency shift for the largest amplitude is about 5 Hz, or 1%. A qualitative comparison with Fig. 2 shows that the nonlinearity coefficient is really positive. In Table I we provide the results of processing the resonance curves by formulas (56)–(58). The quantity δ was defined by Eq. (15) for a resonance curve with the minimum amplitude:

$$\delta = 0.015. \tag{59}$$

TABLE I. The nonlinear parameters of free oscillations for 0.4-cm^3 cavity. $V_0 = 0.4$; $\delta = 0.015$.

$f_n - f$ [Hz]	1	2	3	4
y_0	0.014	0.020	0.032	0.056
η	0.03	0.16	0.30	0.55
a	8	7	6	3
ay_0^2	0.002	0.004	0.008	0.012

The frequency width of discretization of the analyzer, $\Delta = 1$ Hz. One can see from the table that the nonlinear shift $f_n - f_0$ nearly equals the nonlinear correction for the frequency. The nonlinearity coefficient falls with a higher intensity, which is most likely to relate to the breakaway of small bubbles that induce nonlinear damping. This is also confirmed by the observation of spectral broadening with an increasing amplitude, as well as by a visible discharge of bubbles from the cavity, as the oscillation amplitude y_0 increases up to $(6 \text{ to } 7)\times 10^{-2}$. Since $\eta = ay_0^2/\delta$ is less than unity, no nonlinear effects due to cavity synchronization can take place in this situation. In Fig. 5 the cavity oscillations are shown as a function of time. The resonance frequency of the 0.4-cm^3 cavity was 470 Hz, whereas in the 0.35-cm^3 volume cavity it was 473 Hz, i.e., the frequency is shifted by 3 Hz. We failed to observe a nonlinear shift of the oscillation spectrum for the cavity of 0.35-cm^3 volume, although the range of the oscillation amplitudes was the same as for the 0.4-cm^3 cavity. Figure 6 shows the spectral curves for three amplitudes of the oscillations for this cavity. We find that only δ varies from 0.014 to 0.021 with a higher amplitude. There is no evidence of cubic nonlinearity here.

In Fig. 7 we provide two resonance curves demonstrating the sign of nonlinearity ($a > 0$) and the Q factor ($\delta = 0.015$) for oscillations of the 0.25-cm^3 cavity. Figure 8 shows the maximum value of spectral component of modulus F versus its frequency for multiple oscillations. It is seen clearly enough that the nonlinearity has a threshold pattern. The straight lines in the figure indicate the extrapolation data and are described by the formula

$$f_0 = 448 \text{ Hz}, \tag{60}$$

$$\Delta f = \begin{cases} 0, & \text{if } 0 < F < 5\times 10^{-4}, \\ 6{,}4\times 10^4 F - 32, & \text{if } F > 5\times 10^{-4}. \end{cases}$$

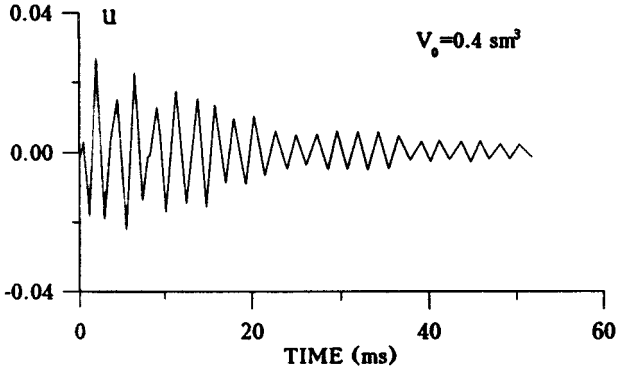


FIG. 5. The time evolution of the oscillations for the 0.4-cm^3 cavity.

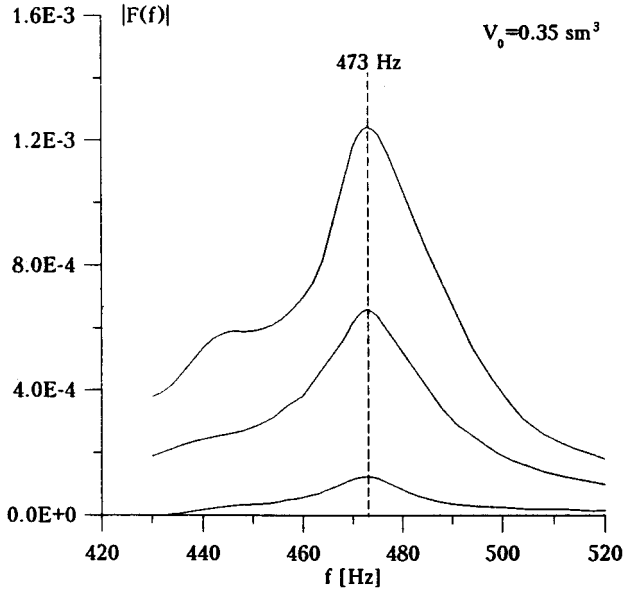


FIG. 6. The natural oscillations spectra for a cavity of volume 0.35 cm^3 for various oscillation amplitudes. Dashed line indicates the absence of bias in the spectrum maximum.

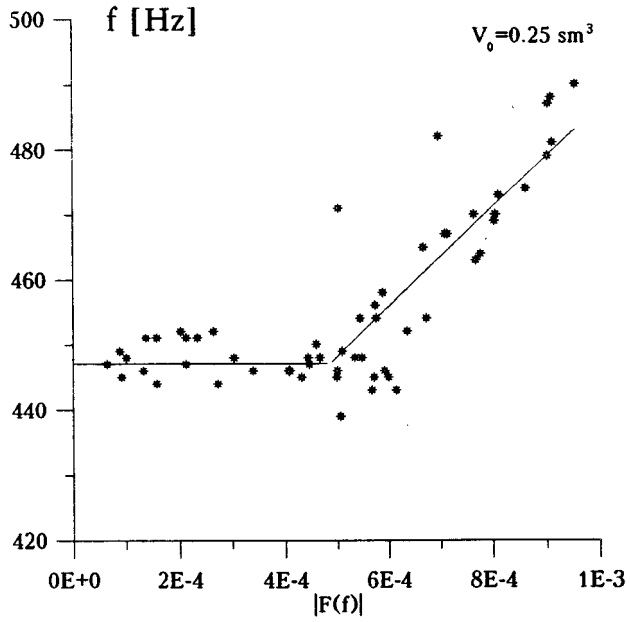


FIG. 8. The frequency of the peak position in the spectrum of free oscillations for a 0.25-cm^3 cavity versus amplitude of the spectral component in the maximum.

Unlike the cavities $V=0.4$ and 0.35 cm^3 those with smaller volumes display instability of oscillations, which is seen in Fig. 8. This instability manifested itself in the difference of both the successively excited oscillations of one cavity, and the oscillations excited in this cavity after we changed air in it. Nevertheless, in this case oscillations can also be described using the nonlinearity coefficient a and Eqs. (10). The calculation results obtained by formulas (56)–(58), where $\Delta=1$ Hz, are given in Table II. Unlike the 0.4-cm^3 volume cavity, the nonlinearity coefficient here first increases and then decreases. Its increase is determined by the nonlinearity threshold, and its fall by the cavity breakup.

Figure 9 demonstrates the spectrum of natural oscilla

TABLE II. The nonlinear parameters of free oscillations for a 0.25-cm^{-1} cavity. $V=0.25$; $\delta=0.015$.

f_n-f [Hz]	6	19	32	50
y_0	0.025	0.035	0.046	0.058
η	0.9	2.5	3.9	5.1
a	22	30	28	23
ay_0^2	0.014	0.037	0.059	0.077

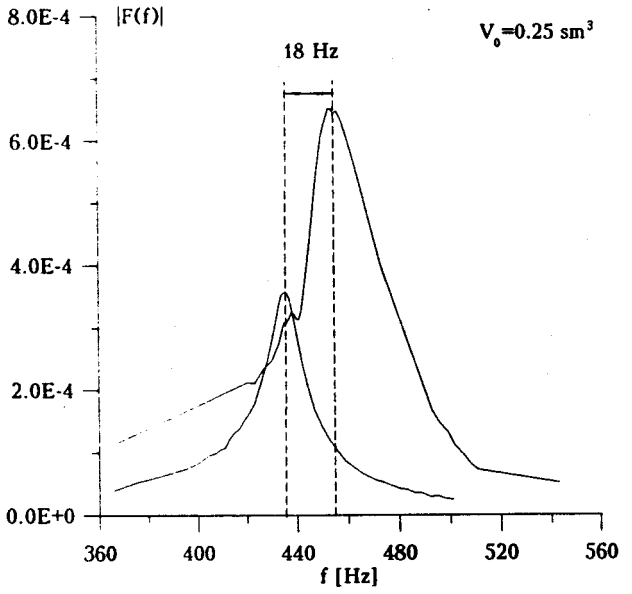


FIG. 7. The natural oscillations spectra for a 0.25-cm^3 air volume cavity. Dashed lines show the bias of the spectrum maxima.

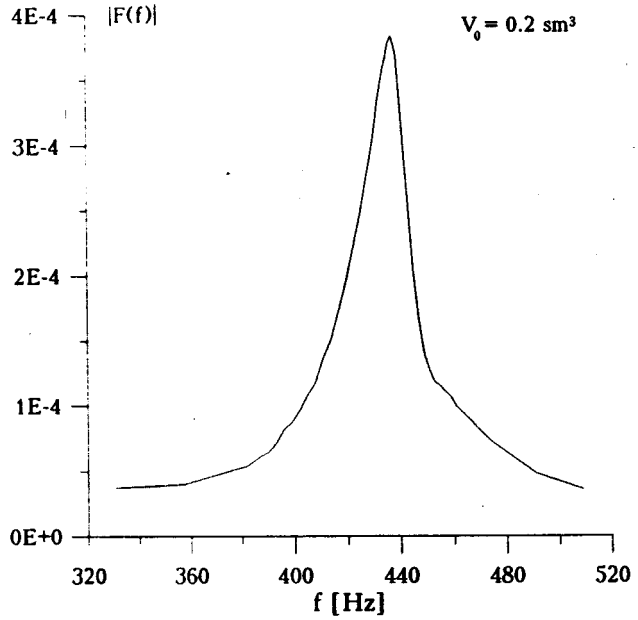


FIG. 9. The natural oscillations spectrum for a 0.2-cm^3 volume cavity, demonstrating the negative sign of nonlinearity.

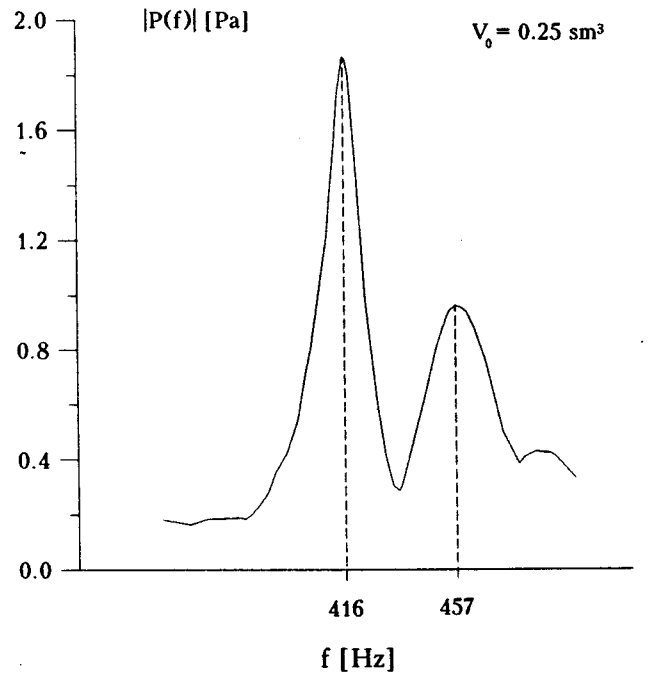


FIG. 10. The natural oscillations spectra for two 0.25-cm^3 cavities by the excitation of only one of them (the linear regime).

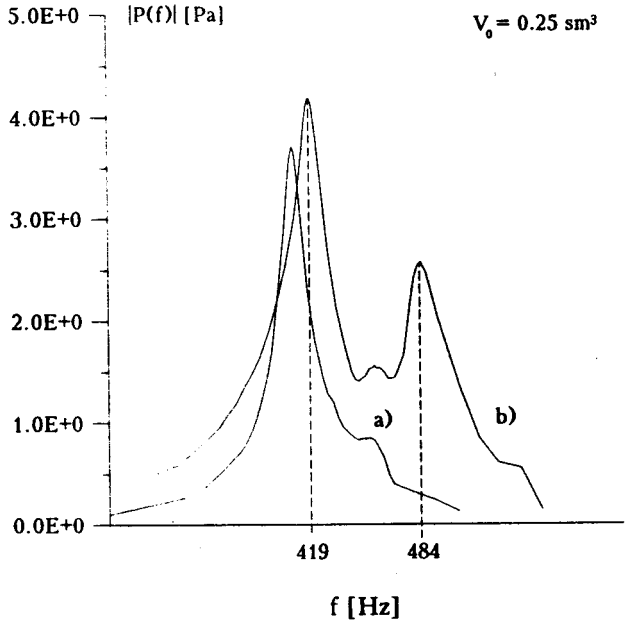


FIG. 11. The natural oscillations spectra for two 0.25-cm^3 cavities by the excitation of one (the nonlinear regime).

tions of a cavity of $V=0.2$ cm^3. In contrast to Fig. 7, here one can see that the nonlinearity coefficient has a negative sign. On the whole, this cavity exhibits both instability and a threshold character of nonlinearity, the nonlinearity coefficients are close to the values listed in Table II, only with the opposite sign. The eigenfrequency of the oscillations, $f_0=440$ Hz ($\delta\cong 0.015$), i.e., it remained practically unchanged with a 20% variation of the volume. A comparison with the theoretical data (16) shows that the frequencies differ by more than 10% (for example, for $V=0.25$ cm^3 formula (6) yields 390 Hz), and the nonlinearity coefficients have the same sign but differ by an order of magnitude. The losses by radiation are small relative to the total losses.

Now consider the oscillations of two cavities. It follows from the above that the only cavity where phasing is possible is $V=0.25$ cm^3. We evaluate the coupling coefficient of the cavities using the expression for β from (5): it yields $v=0.085$. Figure 10 offers the spectrum of the pressure field $P(f)$ scattered by the natural oscillations of two cavities following excitation of just one of them. The spectrum is taken at point B (Fig. 3) on the dipole axis at a distance of 5 cm from the dipole center and 2 cm from the plate. At this point we could observe both the monopole and the dipole oscillations. By placing a hydrophone directly at point A under the cavities one can detect a lower-frequency resonance, which in Fig. 10 is shown to occur at 416 Hz. This means that the second resonance, at 457 Hz, can be associated only with the dipole mode. The contributed nonlinearity, given the intensity indicated in the figure, is small though nonzero (low Q factor of the dipole oscillations). Therefore the frequency difference between the resonances provides for a double value of the coupling coefficient [see Eqs. (25)]. Here $\nu=0.05$, which is nearly half the theoretical value. As is shown in Table II, only two values of ay_0^2 are larger than the obtained result for ν, which indicates the feasibility of phasing for two-cavity oscillations [see (21) at $a>0$] by excitation of all cavities (54). Moreover, it proved possible to enable phasing of three cavities by exciting just two of them. This process has the highest threshold at $a>0$ (48).

In Fig. 11(a) and (b) we provide spectral curves for the natural oscillations of two cavities by exciting one of them when the oscillation amplitudes are large enough. The measurements shown in Fig. 11(a) were taken at point A, in Fig. 11(b) at point B. The resonance of monopole oscillations is shifted downwards in frequency so that the resonance curve ''captures'' the frequency range of the dipole oscillations. Figure 11(b) shows the sum of dipole and monopole oscillations. One can see the bias in resonances and the region where they fully overlap. The time evolution of the oscillations of two cavities excited in antiphase is shown in Fig. 12; Fig. 12(b) is the spectrum of these oscillations. A hydrophone is located at point A where no dipole oscillations are recorded. The oscillations are small over more than a period, then follow an increase and an apparently nonexponential decay. The oscillations spectrum in Fig. 12(b) is shown to broaden to 520 Hz as compared to that in Fig. 11(a), which can be accounted for by the instability of dipole oscillations that excite monopole oscillations at their eigenfrequency. This suggestion is proved by the oscillations recorded at point B (on the dipole axis), which are given in Fig. 13. The first figure shows the monopole mode buildup against the dipole background and, further, the competition of the two modes. The oscillation spectrum in Fig. 13(c) exhibits three maxima: 520, 440, and 390 Hz. The 520-Hz peak pertains to the dipole oscillations, which is confirmed by the equation for w (18) at $z=0$; hence, the nonlinear frequency shift $ay_0^2\cong 3\nu$. It should be noted here that the amplitudes of the collective oscillations increase considerably without destroy-

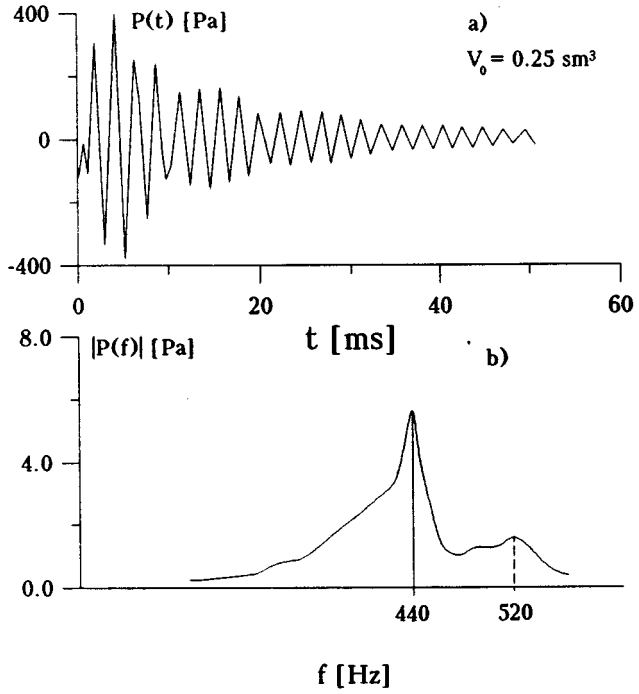


FIG. 12. (a) The time evolution of the oscillations of two cavities 0.25-cm^3 air volume, excited in antiphase; (b) the spectrum of this oscillation recorded at point A.

ing the cavities, in contrast to the one-cavity oscillations. The spectrum near the 440-Hz peak relates to the monopole oscillations. The 390-Hz maximum is most likely connected with the dipole oscillations, which is in conformity with Eq. (18) for w, whose right-hand part can be regarded as the extraneous force at the 390-Hz frequency. To validate this in

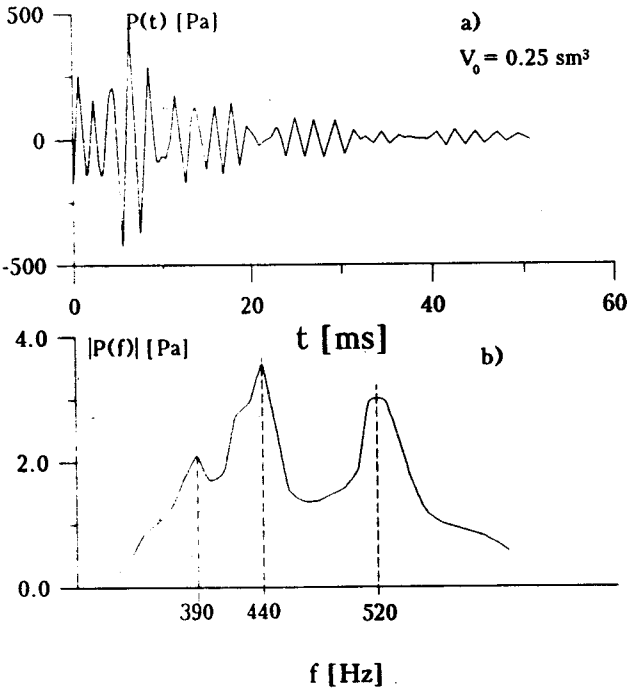


FIG. 13. The spectrum and time evolution of the antiphase-excited oscillations of two 0.25-cm^3 cavities at point B.

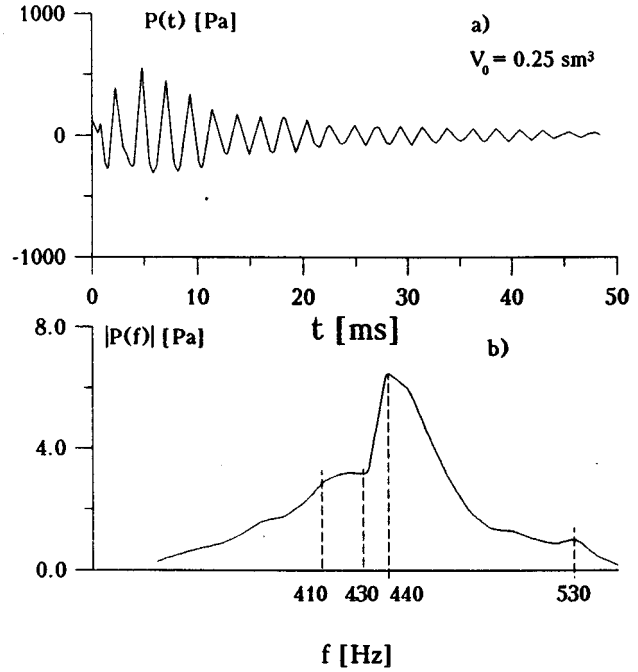


FIG. 14. The same as in Figs. 12 and 13 recorded at point D.

Fig. 14(a) we provide the time evolution, and in Fig. 14(b), the spectrum of two-cavity oscillations excited in antiphase, which are recorded at point D spaced 5 cm from the cavity and 2 cm from the surface in the direction perpendicular to the dipole axis. The spectrum does not have a peak at the frequency of 390 Hz. There is no resonance at this frequency at point *A*, either.

Figure 15 shows the time evolution and spectrum of the oscillations of two cavities excited in antiphase, which were

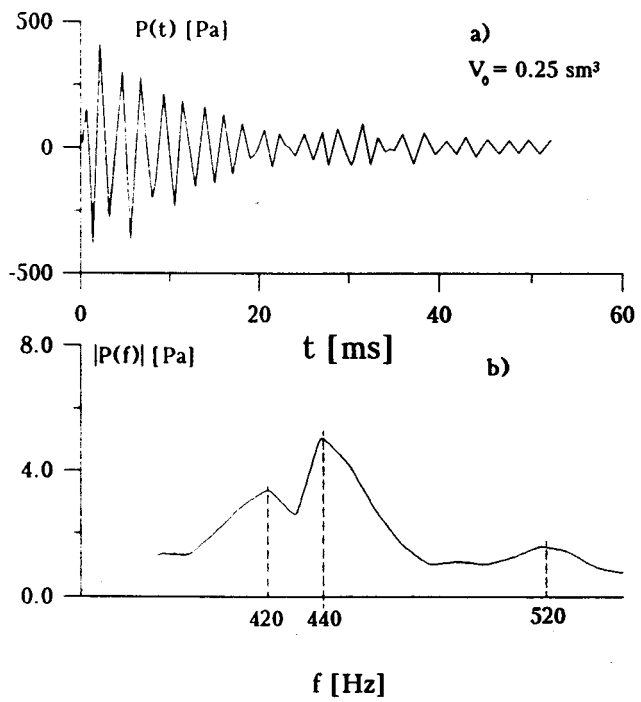


FIG. 15. (a) The time evolution of the oscillations of two 0.25-cm^3 cavities, excited in phase; (b) the spectrum of this oscillation recorded at point D.

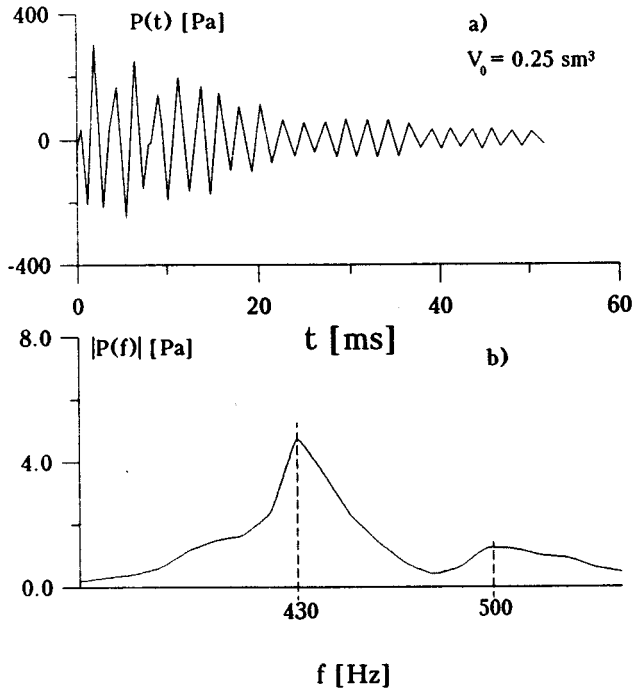


FIG. 16. The same as in Fig. 15, observed at point B.

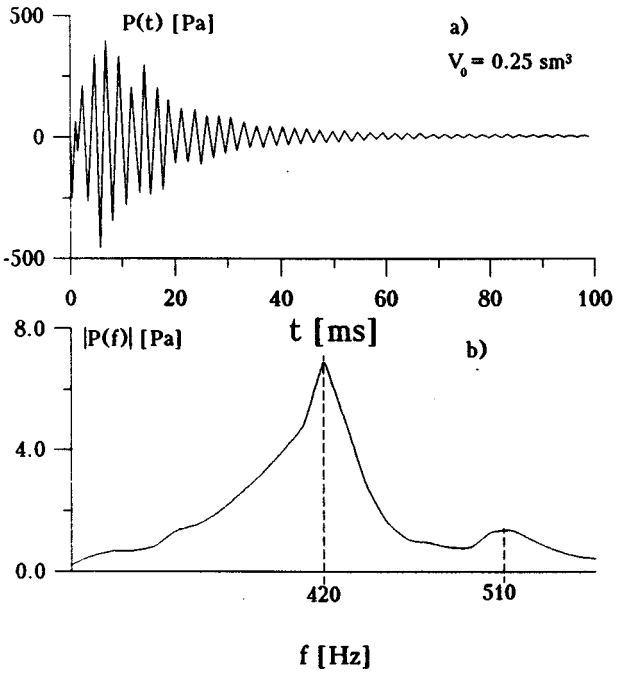


FIG. 18. (a) time evolution, (b) the spectrum of oscillations of three cavities by excitation of two of them in antiphase. The data is taken at point A.

observed at point D. Besides the 440-Hz resonance which also occurs by the oscillations involved in the phasing process (Fig. 13), there is a resonance at 420 Hz, i.e., near the frequency of the monopole linear oscillations. In Fig. 16 one can see oscillations of the initially phased cavities, recorded at point B on the dipole axis. Absence of resonance by the dipole oscillations testifies to stability of phased oscillations. There is one more noteworthy fact. The acoustic waves from dipole oscillations, as follows from Eq. (8), attenuate over distance by the law r^{-2} so that at point B they become 0.1 of the monopole oscillations. However, it is practically impossible to discriminate these oscillations. The above discussion leads us to a conclusion about the feasibility of phasing for free oscillations of two gas cavities excited in antiphase in water.

Now consider oscillations of three gas cavities in water.

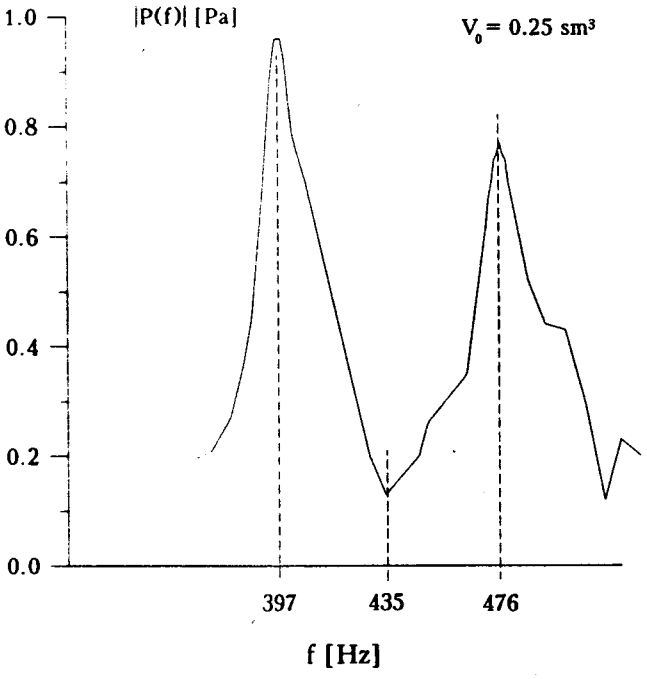


FIG. 17. The spectrum of linear natural oscillations of three 0.25-cm^3 volume cavities by the excitation of one cavity, taken at point C.

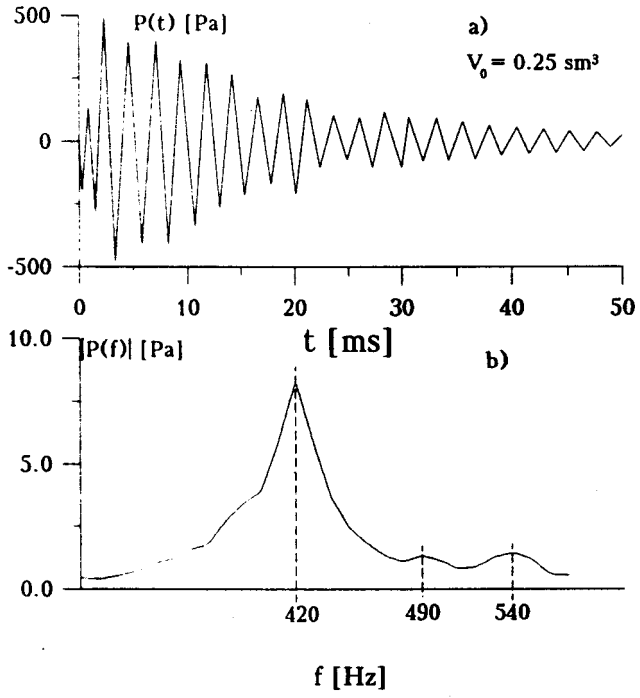


FIG. 19. The oscillations of a three-cavity system, when two cavities are excited in phase, point A.

Figure 17 demonstrates the spectra of linear oscillations of three cavities following excitation of just one of them, which were taken at point C lying on the axis of a dipole formed by the excited cavities and by two passive cavities (this axis passes through the center of the triangle and the excited cavity) at 5 cm from the center of the triangle and 2 cm away from the plate surface. The spectrograms show two oscillation modes, which is in conformity with Eqs. (28)–(30), the value ν being in good agreement with the result obtained for two cavities (see Fig. 10). Figure 18 represents the time evolution and spectrogram of the oscillations of three cavities when two of them are excited in antiphase, recorded at point A. A small peak in the vicinity 510 Hz [Fig. 18(b)] points to the primary growth of monopole oscillations at the dipole frequency. This maximum is absent in Fig. 19(b) which shows the oscillations of three cavities, two of them excited in phase, observed at the same point. The new peaks at frequencies 490 and 540 Hz are most likely to come from the monopole oscillations of the system under the influence of dipole oscillations arising due to an initial slight phase mismatch of the excited oscillations.

IV. CONCLUSION

This paper describes the linear and nonlinear properties of gas cavities in water in detail sufficient to evaluate these cavities in terms of application of the autophasing effect in the development of an acoustic maser. The benefits offered by a cavity as the acoustic maser element include ease of fabrication and excitation of oscillations, as well as high nonlinearity. A major drawback, however, is the weakness of the liquid–gas interface, which leads to oscillation instability and, at large amplitudes, to formation of a gas–liquid foam.

ACKNOWLEDGMENTS

This work is supported by the U.S. Office of Naval Research and by the Russian Foundation for Basic Research, Grant No. 96-02-19481.

[1] A. V. Gaponov, ''Instability of an excited oscillator system to electromagnetic perturbations,'' Zh. Eksp. Teor. Fiz. **39**, 326 (1960), in Russian.
[2] Yu. A. Kobelev, L. A. Ostrovsky, and I. A. Soustova, ''Autosynchronization of nonlinear oscillators in acoustics,'' Izv. Vyssh. Uchebn. Zaved. Radiofiz. **29**, 1129 (1986).
[3] Yu. A. Kobelev, L. A. Ostrovsky, and I. A. Soustova, ''The nonlinear model of autophasing for classical oscillators,'' Zh. Eksp. Teor. Fiz. **99**, 470 (1991) [Sov. Phys. JETP **72**, 262 (1991)].
[4] V. V. Bredikhin, Yu. A. Kobelev, A. I. Potapov, and N. I. Vasilinenko, ''A method for measuring the parameters of acoustic waves scattered by gas bubbles in liquid,'' Sov. Phys. Acoust. **41**, 390 (1995), in Russian.
[5] Lord Rayleigh, *The Theory of Sound* (Dover, New York, 1945), Vol. 2.

Migration of backscatter data from the Mid-Atlantic Ridge

Alistair J. Harding, Michael A. H. Hedlin, and John A. Orcutt
Institute of Geophysics and Planetary Physics, A-0225, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0225

(Received 16 September 1996; accepted for publication 31 December 1997)

In studies of low-frequency reverberation within the marine environment, a central concern is the relationship between reverberation events and morphological features of the seafloor. A time-domain migration algorithm for the reverberation intensity field is developed that produces scattering coefficient maps coregistered with a bathymetry database. The algorithm is tailored to broadband transient sources with good range resolution, and was developed to analyze an extensive set of reverberation records from a 200–255 Hz source collected on the flanks of the Mid-Atlantic ridge. The precise, sample-by-sample, tracking of wavefronts across elements of the bathymetry database that forms the foundation of the algorithms implementation results in reverberation maps that show a clear and detailed correlation between scattering and morphology with narrow scarp slopes consistently highlighted. Environmentally induced asymmetries in transmission loss and incidence angle are exploited to break the inherent left–right ambiguity of the receiver array. Iterative migration, assuming a dominant dependence of backscatter on grazing angle, produces images, even from individual records, that show good ambiguity resolution. Results from multiple records corroborate the effectiveness of the ambiguity resolution and demonstrate the stability of the scattering coefficient estimates and the acoustic system. © *1998 Acoustical Society of America.* [S0001-4966(98)03104-X]

PACS numbers: 43.30.Gv, 43.30.Vh, 43.30.Pc [JHM]

INTRODUCTION

In July of 1993, a coordinated set of low-frequency acoustics experiments were conducted on the western flank of the Mid-Atlantic Ridge within a corridor that lay just north of the Kane fracture zone. These experiments were designed to investigate and quantify the relationship between acoustic reverberation and seafloor morphology and constituted the Main Acoustics Experiment (MAE) of a Special Research Program (SRP)[1] on bottom reverberation sponsored by the Office of Naval Research. Included within the experimental suite were monostatic and bistatic reverberation measurements by the research vessels CORY CHOUEST and ALLIANCE, and near bottom measurements of low grazing angle interactions using vertical arrays of hydrophones. One element of the reverberation experiments that makes them unique is the availability of detailed bathymetric and geophysical data for the area, collected on a pair of SRP sponsored large-scale and small-scale geophysics cruises. The bathymetric database includes a map of the entire SRP experimental corridor gridded at 200 m and finer scale surveys of selected areas with resolution on the order of 10 m or better. It is the availability of these databases that permits the investigation, in detail, of the correspondence between reverberation returns and bathymetry.

The process of mapping reverberation returns back onto the seafloor scattering sites that produced them can be referred to as charting.[2–4] Here we prefer to use the term migration, because of the strong similarity of the methods employed here to the Kirchoff migration method used in seismic reflection processing. The principal difference between the two is that the reverberation data—beamformed, time series of acoustic intensity—are assumed to be the result of incoherent scattering rather than coherent reflections. Also, as is the case for seismic migration, we restrict attention to selected primary paths and ignore multiple reflections in the formulation of the method.

In outline, migration is accomplished in two stages; first the seafloor is mapped into the time-beam coordinates of the data, then the reverberation signal from a given beam and small time interval is distributed over the corresponding ensonified area of the seafloor. Migration is complicated by the fact that data were recorded by a horizontal line array (HLA) and each beam is associated with a pair of directions that are oriented symmetrically with respect to the array axis, thus a reverberation return could originate from one of a pair of distinct ensonified areas. Simple mapping schemes that ignore variations in seafloor depth produce reverberation maps that are perfectly symmetric with respect to the array axis and provide no means of resolving, from a single observation, the inherent ''left–right ambiguity'' introduced by a HLA. Accounting for the bathymetry introduces ''environmental symmetry breaking'' [3,4]—bathymetry induced variations in transmission loss, acoustic shadowing—which can be exploited by algorithms, such as the one presented here, to reduce substantially the left–right ambiguity, even for a single ping. The difference between the environmental symmetry breaking technique and the migration technique presented here is that the former assigns all energy to either the left or right side if the difference in transmission loss reaches a predefined threshold, while the current technique partitions energy according to a set of weights that vary continuously as a function of transmission loss and other factors such as expected backscatter strength.

An additional problem with simpler schemes is that the

migrated images are in the time-beam coordinates of individual transmissions rather than in global coordinates of the bathymetry, complicating the comparison of multiple pings and studies of the relationship between scattering and bathymetry. The current migration algorithm avoids this problem by being formulated directly in terms of the backscatter strength of individual elements of the bathymetry grid.

The algorithm is based on a time-domain formulation of the incoherent scattering process which relates the scattering strength to the expected acoustic intensity via a large sparse matrix. This formulation means that we could treat the migration as a linear inverse problem and invert multiple pings simultaneously for scattering strength. Combining pings with diverse look angles would eliminate left–right ambiguity from the solutions. A linear inverse approach would also allow regularization constraints to be employed on the solutions.[5]

In this paper, we solve the scattering equations approximately using a version of iterative backprojection under the assumption that backscatter strength is spatially isotropic and a function only of the local ensonification and backscatter angles, angles which are equal for the monostatic, direct paths considered here. The advantage of using iterative backprojection is that it is computationally fast and guaranteed to produce migrated images that satisfy the data exactly while only weakly enforcing assumptions about the behavior of scattering strength. Iterative backprojection could potentially be used as a basis for producing well resolved, migrated images in near real time. We use it here to rapidly check the consistency of multiple images from successive reverberation records of the MAE and to produce preliminary estimates of backscattering strength as a function of grazing angle. The linear inverse approach outlined above will only be effective if the individual reverberation images are basically consistent. Uncorrected errors in, for example, array heading or source location would substantially degrade the result of inverting multiple pings.

We concentrate on the analysis of a subset of the data, namely half-convergence zone (1/2 CZ), monostatic reverberations from broadband, 200–255 Hz, linearly frequency modulated (LFM) transmissions recorded by the R/V CORY CHOUEST. We chose this subset because it has the best range resolution and simplest propagation paths, and thus affords the best opportunity of investigating, quantitatively, the relationship between backscattering strength and bathymetry. Previous examinations of the SRP acoustics data have, for the most part, concentrated on the longer range, continuous wave (cw) data, and have established that there is a broad correspondence between high amplitude reverberations and well ensonified features at 1/2 and 1 1/2 CZ, with the highest returns from backfacing ridges. However, the detailed investigation and quantification of this relationship is as yet incomplete; this paper represents one step in that process.

I. OVERVIEW OF DATA AND EQUIPMENT

The MAE was conducted within an approximately 4 by 2° corridor on the flanks of the Mid-Atlantic ridge extending from 45° W to 49° W and from 25°30′ N to 27°30′ N. The corridor lies within a larger area designated as an ONR natural laboratory,[6] which straddles the Mid-Atlantic ridge itself and is bounded to the south by the Kane fracture zone. The seafloor fabric within the corridor is dominated by lightly sedimented, ridge parallel abyssal hills, but there are also large sedimented ponds and the area is dissected by deep corridors that are the fossilized, off-axis expression of spreading segment boundaries at the ridge axis. The feature whose backscatter we will examine in detail here is a $\sim$30 km long abyssal hill, designated B′, that lies at the western end of one of the segment boundary corridors. In particular we will concentrate on a set of 1/2 CZ, monostatic reverberation returns recorded by the R/V CORY CHOUEST that span the intersection of a pair of the experimental runs, run 5a and run 6 (Fig. 1 and Table I). This intersection lies slightly northwest of a large set of intersecting tracks that constitutes the focus of an extensive set of bistatic and monostatic experiments that had B′ as the primary target. We have chosen the first intersection here because the associated transmissions provide better ensonification of the front slope of B′ at 1/2 CZ.

The basic features of the R/V CORY CHOUEST source and receiver arrays are shown in Fig. 2. The vertical source array consisted of 10 elements spaced at 2.29 m and centered at 101 m, while the horizontal receiving array consisted of 128 hydrophone groups spaced at 2.5 m. Beamforming of the reverberation data was performed aboard ship using a delay and sum beamformer with a Hamming window, resulting in a nominal broadside beamwidth of $\sim$1.3°, measured at the -3 dB point. The presence of dead or improperly gained phones in the HLA resulted in nearly uniform sidelobe level of approximately -30 dB.[1]

The schedule of source transmissions for the experiment was conceptually arranged in a hierarchical fashion with the bottom two levels in the hierarchy being transmission segments and individual source wavetrains or pings. Segments were numbered consecutively and each one lasted 12 min. Ping numbers were also numbered consecutively with typically six source wavetrains being transmitted during each segment. For the portions of runs 5a and 6 examined here, 2 out of every 3 segments were used for R/V CORY CHOUEST transmissions and the remaining segment for Alliance transmissions. One of the standard, usually the first, transmissions within a CORY CHOUEST segment was a 5-s duration linearly frequency modulated (LFM) signal chirped over the band 200–255 Hz. It had the largest bandwidth of any of the source wavetrains and hence the best range resolution after matched filtering, $\Delta r = 13.6$ m; it is thus the best choice for examining, in detail, the spatial structure of the scattering process.

The structure of the LFM source wavefield out to beyond a 1/2 CZ is shown in Fig. 3. The broadband transmission loss was calculated by integrating across the band, the transmission losses predicted at individual frequencies by a wide-angle parabolic equation (PE) code.[7] At times during the experiment the source array was steered down at various angles with respect to the surface, but for the transmissions of interest and the calculation, the steering angle was set at 0°. The sound speed profile, displayed in Fig. 3(a), was es-

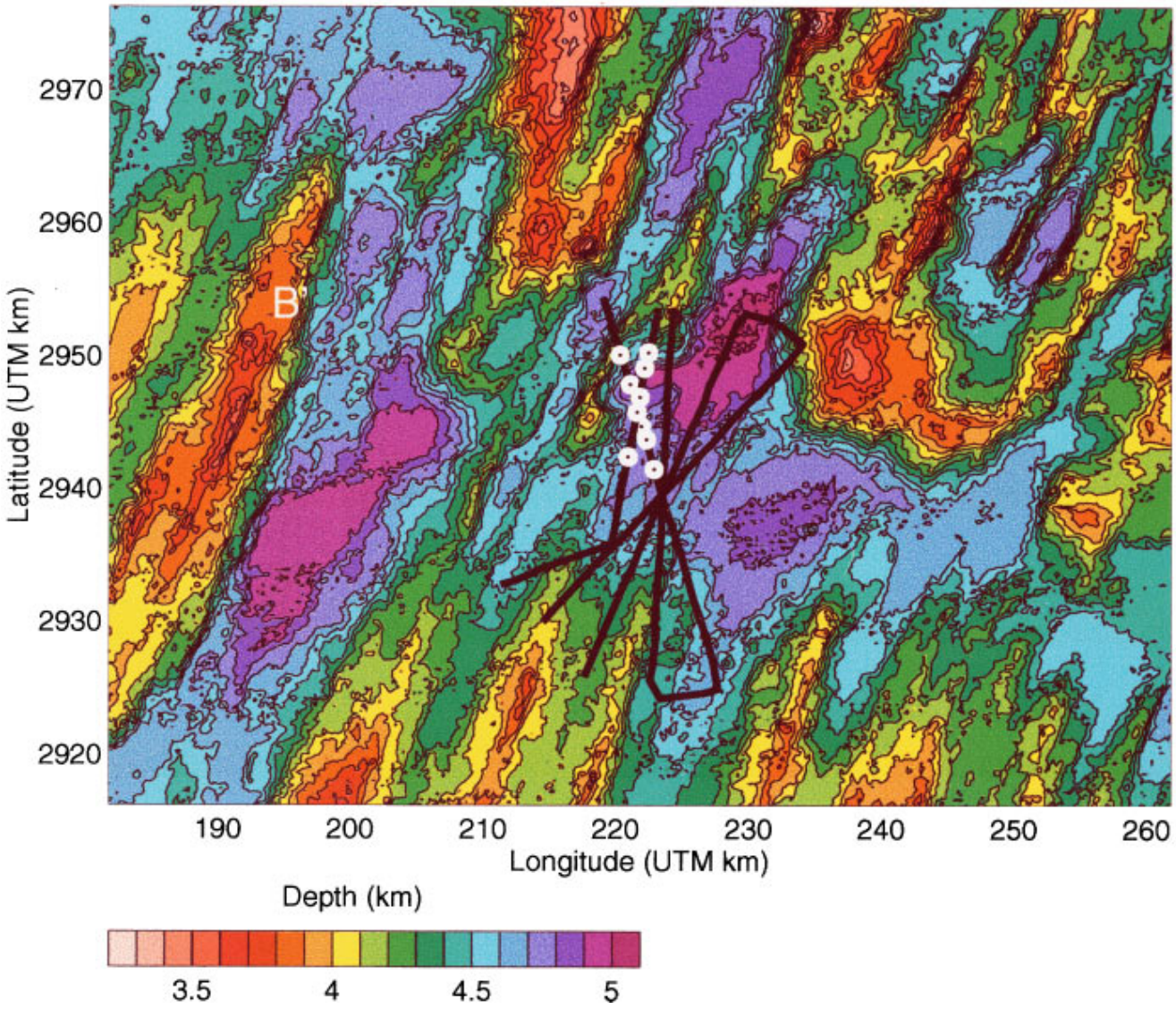


FIG. 1. Bathymetry and a portion of R/V CORY CHOUEST track lines (heavy black lines) in the vicinity of target B′, an elongated ridge at the western edge of the figure. In this paper we analyze reverberation data from Pings at the start of runs 5a and 6 of the MAE (heavy white circles).

sentially the same throughout the experiment and the sea state was calm. The center of the source array at 181 m was in the upper part of the waveguide with the conjugate depth being at 3800 m and the turning range, 1/2 CZ distance, being around 33 km. The core of the main acoustic beam contains a pair of high amplitude fringes which result from the interference between energy that was up and downgoing at the source level. Behind the core, the two components are sufficiently separated in time, up to ~30 ms, that there is no significant interference effect on the amplitude. Furthermore

TABLE I. List of reverberation records, pings, analyzed in this paper.

Run	Segment	Ping	Location (Deg)		Location (UTM km)		HLA array heading
			Lat	Lon	Lat	Lon	
5a	415	1313	26° 38.6′N	−47° 48.6′W	2950.1	220.3	168.6
5a	417	1318	26° 37.4′N	−47° 48.1′W	2947.8	221.1	166.1
5a	420	1327	26° 35.7′N	−47° 47.5′W	2944.7	221.9	169.5
5a	421	1331	26° 35.2′N	−47° 47.3′W	2943.7	222.3	169.1
5a	423	1336	26° 34.0′N	−47° 46.9′W	2941.4	222.9	167.6
6	547	1683	26° 38.7′N	−47° 47.3′W	2950.2	222.5	197.3
6	548	1687	26° 38.1′N	−47° 47.4′W	2949.1	222.2	195.9
6	550	1692	26° 36.9′N	−47° 47.6′W	2946.9	221.8	194.8
6	551	1696	26° 36.3′N	−47° 47.7′W	2945.7	221.6	194.6
6	554	1705	26° 34.5′N	−47° 48.1′W	2942.4	220.9	195.3

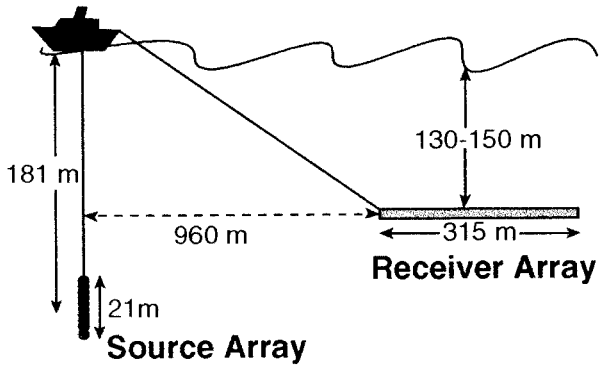


FIG. 2. Sketch of the source and receiver array of the R/V CORY CHOUEST. The 10 element source array had a spacing of 2.29 m while the 126 element receiver array had a spacing of 2.5 m.

in this region the ray propagation angles diverge by about 2° (Fig. 4). Beyond the turning range, most of the structure in the main beam is due to caustics in either the up or downgoing source field. The source field also has a number of sidelobes that intersect the seafloor at ranges less than 15 km.

II. MIGRATION ALGORITHM

The goal of the migration algorithm is to produce a map of scattering strength registered on the same grid as the bathymetry data. The basic observational data are the intensity time series $I(\varphi_b,t)$ associated with beam angle φ_b that result from squaring the matched filtered and beamformed pressure series. For direct, reciprocal paths between source/receiver and the scattering site, the relationship between the expected intensity of the backscattered signal and the scattering coefficient is taken as the following integral over the seafloor

$$I(\varphi_b,t)=\int\int dA\ \gamma m(x,\theta,\phi)g(x,\varphi_b,\varphi_a,t-T_r-T_s), \tag{1}$$

where

$$\gamma=\gamma_{mf}\gamma_r\Psi_r\gamma_s\Psi_s p_T^2 T. \tag{2}$$

$T_{r,s}$ are the ray theoretical travel times to and from the scattering location. γ is the product of the transmission losses, $\Psi_{s,r}$, to and from the seafloor; the gains, γs, γ_r, and γ_{mf}, associated with the source array, receiver array and matched filtering; and the source level, a product of the transmission level, p_T, and signal duration, T. Since we are dealing with broadband transient signals, gains for individual components of the system are defined in terms of the ratio of integrals over the input and output pressure fields[8]

$$\gamma=\int_0^{T_0} dt\ p_0^2(t)\ \Big/\ \int_0^{T_i} dt\ p_i^2(t). \tag{3}$$

The scattering coefficient, $m(x,\theta,\phi)$ in Eq. (1), is a function of position, x, and the monostatic incidence/backscatter direction at the seafloor (θ,ϕ). The quantity $g(x,\varphi_b,\varphi_a,t-T_r-T_s)$ can be termed the scattering function for the system, it depends on the details of the scattering process, as well as quadratically on the effective source function at the seafloor and quadratically on the time domain response of the beamformer for a given steering angle φ_b and arrival angle φ_a at the receiver array. The partition of scattering between the scattering coefficient and the scattering function is defined by normalizing $g(t)$ so that its time integral is unity.

For direct paths, Eq. (1) is a time domain variant of the standard equation relating the expected reverberation intensity to scattering coefficients for cw signals.[8,9] The additional element in the time domain formulation is the introduction of g, which can be regarded as representing the resolution function of the acoustic imaging system. Analytic expressions or empirical forms could be found for g (and also m), given explicit assumptions about the nature of the surface, for example small amplitude roughness.[8] In practice, we assume that the details of the time response will not be important provided that the resulting scattering coefficients estimates represent averages over a sufficiently large area. We thus use a simple factored boxcar form for g

$$\begin{aligned} g(\varphi_b,\varphi_a,t)&=g^*(t)b(\varphi_b,\varphi_a)\\ &=B(t/t_w)B((\varphi_b-\varphi_a)/\Delta\varphi), \end{aligned} \tag{4}$$

where is $B(t)$ is a unit boxcar centered at zero and width one, and $b(\varphi_b,\varphi_a)$ is the beamformer response, which is in turn simplified to a boxcar of width $\Delta\varphi$, the angular separation between adjacent beams. In the limit $t_w=0$, the time dependence, $g^*(t)$, reduces to a δ-function. A potentially more accurate representation of the angular dependence would be to use the response function of the beamformer at a representative frequency rather than a boxcar.

We can derive a sonar equation from Eq. (1) by substituting for the scattering function from Eq. (4). With $g^*(t)$ set equal to a δ-function, integrating from t_1 to t_1+T_a yields

$$\overline{p_R^2}T_a=\Psi_r\Psi_s\cdot\gamma_{mf}\gamma_s\gamma_r\cdot p_T^2 T\cdot m\cdot A, \tag{5}$$

where

$$\overline{p_R^2}=\frac{1}{T_a}\int_{t_1}^{t_1+T_a} I(\varphi_b,t) \tag{6}$$

and A is the ensonified area corresponding to the integration limits in time and beam angle, and it is assumed that transmission losses, etc., are constant over A. This is equivalent to the following sonar equation for the reverberation level $R=20\log p_R$

$$\begin{aligned} R=&-\mathrm{TL}_r-\mathrm{TL}_s+G+M+10\log(A)+S\\ &-10\log(T_a/T), \end{aligned} \tag{7}$$

where $\mathrm{TL}_{r,s}$ are the transmission losses, G is the combined system gains, M the scattering strength, and S the source strength, all defined in the obvious way from Eq. (5). The final term corrects for differences between the averaging time and the signal duration. Frequently for cw signals the averaging time is chosen equal to the pulse duration and there is no correction. However, for the LFM pulse, choosing an averaging time equal to the width of the matched filtered pulse, 18 ms, corresponds to a correction of 24 dB. A problem with choosing a short averaging time is that it is then

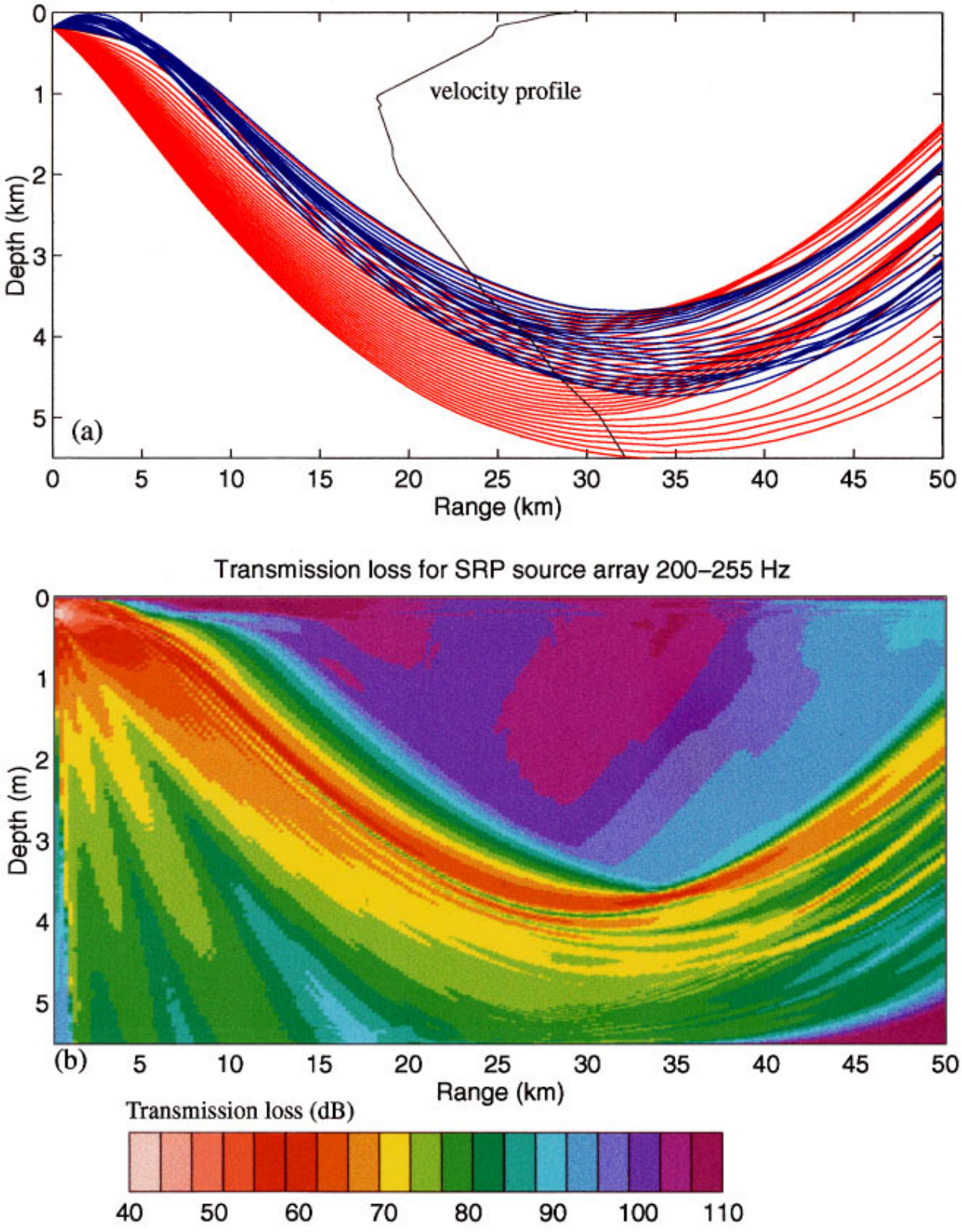


FIG. 3. (a) Representative up (blue) and downgoing (red) rays paths within the main acoustics beam from a point source at 181 m, the center depth of the vertical source array. The reference sound speed (black) has a surface velocity of 1540 m/s and a minimum velocity of 1494 m/s at 1.1 km. The downgoing ray paths are simple but caustics form beyond the 1/2 CZ at 30–35 km. (b) PE calculations of the transmission loss for a broadband, 200–255 Hz, source from the 10 element source array. Transmission loss variations near the axis of the downgoing main beam are the result of interference of the downgoing and surface refracted energy. Transmission loss is more variable beyond a 1/2 CZ due to the presence of caustics.

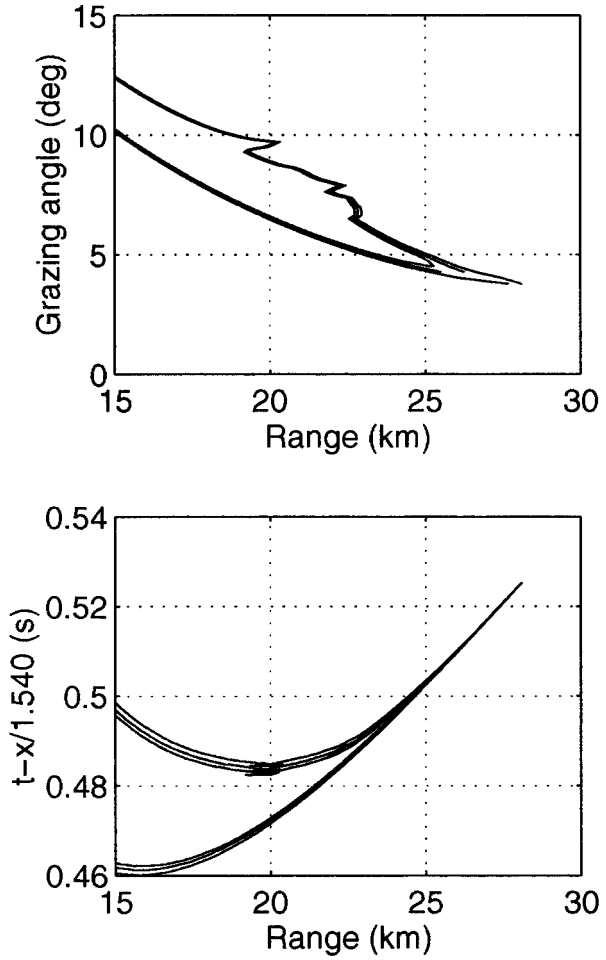


FIG. 4. (a) Grazing angle of rays at a depth of 3.5 km. Arrivals are for rays from the top, middle and bottom elements of the source array. There is approximately a 2° difference in ray angle between surface reflected/refracted rays, upper cluster, and rays leaving the source traveling downward. At this depth, the penumbra of the main beam starts at 25 km and extends to about 28 km. (b) Difference in travel times at 3.5 km for up and downgoing rays at the source.

hard to justify the use of a δ-function for the scattering function, since effectively the convolutional nature of the reverberation process is being ignored.

The above result is most naturally viewed as producing a scattering strength map in the local time-angle coordinates of the beamformer: The scattering strength estimate obtained from Eq. (7) is assigned to area A bounded by a pair of isochrons and a pair of beam boundaries. We wish, instead, to produce a scattering strength map in terms of a global coordinate system and thus we will use Eq. (1) to reexpress the reverberation intensity in terms of contributions from individual bathymetry elements. To this end, we again use the factored form of the scattering function, Eq. (4), and assume that all quantities, except g^*, associated with a small patch, k, of the bathymetric grid are constant. The scattering response of patch k can then be approximated as

$$I_k(t) = \gamma_k m_k g^*(t)^* \frac{dA_k}{dt} = \gamma_k m_k G_k(t), \tag{8}$$

where, for notational convenience, we have incorporated the beamformer factor $b(\varphi_b, \varphi_a)$ into the general gain factor γ_k. Here, dA_k/dt is the rate at which the patch area is swept out by the isochrons; it is zero for $t < T_{\min,k}$ and $t > T_{\max,k}$, the minimum and maximum travel times associated with the patch. Convolution of g^* with the sweep rate function acts to smooth g^*, suggesting, once again, that the knowledge of the details of the scattering is not critical provided that estimates are averages over large enough areas, a necessary condition for which is that $T_{\max,k} - T_{\min,k} \gg t_w$.

The total scattered intensity of a single beam is the sum of the individual patches responses, i.e.,

$$I(t) = \sum_k I_k(t) = \sum_k \gamma_k m_k G_k(t). \tag{9}$$

The summation implicitly includes the left–right ambiguity of the receiving array through the patch weighting γ_k, since the beamformer factor is nonzero only along the conjugate beam directions. As before Eq. (9) could be explicitly averaged by integration over a time intervals T_a which would serve to stabilize the individual intensity values and reduce the size of the subsequent matrix equations. If the scattering function is taken as a δ-function, averaging yields

$$\overline{p_R^2}(t) T_a = \sum_k \gamma_k m_k A_k f_k(t), \tag{10}$$

where A_k is the area of the kth patch and f_k is the fraction of the area of the kth patch ensonified during the averaging interval. If the fractional contributions are rounded to 0 or 1, then Eq. (10) is equivalent to Eq. (5) of Ref. 2. In the latter, a synthetic test case was considered with the source wavetrain taken as 2-s long cw pulse and the averaging time taken equal to the pulse length, resulting in multiple seafloor patches contributing to a single averaged intensity measurement. In practical applications, though, it would be difficult to justify the use of a δ-function when the duration of the source pulse is long relative to the size of the individual patches, since implicitly, as noted above, it ignores the convolutional nature of the reverberation process. A scattering function whose duration matched that of the cw pulse would be more appropriate, but even so the structure of the function would potentially be important for good scattering estimates.

A direct discretization of Eq. (9) for time samples at $t = j\Delta t$ is

$$I_j = \sum_k I_{jk} = \sum_k G_{jk} \gamma_k m_k. \tag{11}$$

Individual beam responses from multiple pings can be combined to form a single, large matrix equation relating reverberation to the scattering coefficients of the patches. For the datasets considered below, the number of scattering coefficients would be $\sim$30 000, and the number of intensity samples $\sim$20 000, a value which increases proportionately if multiple pings are combined. However, the matrix is sparse since $K(j)$, the number of patches illuminated at time sample, j, is on the of order 10 for a single ping.

The matrix equation could be solved as a least squares problem with the addition of regularization constraints on the scattering coefficients, using any suitable sparse matrix solver such as conjugate gradients[10] or LSQR,[11] an approach

investigated elsewhere.[5] Here we choose to solve the inverse problem approximately using a form of iterative backprojection. The advantage of backprojection is that it is simple, fast and robust, while at the same time capable of providing a fair degree of ambiguity resolution; we use it here to check the ping-to-ping consistency of the reverberation data. Although, backprojection does not have the potential resolution of more complete inversion methods, especially for multiple pings, these methods will not perform optimally unless the data are essentially consistent and we can be assured that there are no unresolved problems with ship position and array orientation.[4]

For a single beam, the intensity contribution assigned to a patch k at time sample j by the backprojection is proportional to the weighting factor for that patch

$$I_{jk}=w_k I_j, \tag{12}$$

where

$$w_k=\frac{\gamma_k \cdot G_{jk}}{\Sigma_{k\in K(j)}\gamma_k \cdot G_{jk}}. \tag{13}$$

The backprojection thus incorporates a degree of ambiguity resolution, assigning intensity based on such factors as transmission loss to the ensonified patches and also on the basis of the area swept out, a face pointing more directly into the acoustic beam will attract proportionately higher assigned intensity.

The result of the backprojection is an estimate of the scattering intensity function $I_k(t)$ associated with each patch k. An estimate of the scattering coefficient, m_k, is found by integrating the scattering intensity, Eq. (9), and using the fact that the integral of $g^*(t)$ is unity and thus the integral of $G(t)$ is A_k, the area of patch k. The discrete time version of the estimate is

$$m_k=\frac{\Sigma_j I_{jk}}{\gamma_k A_k}, \tag{14}$$

or substituting for γ_k

$$m_k=\frac{1}{A_k}\frac{(\gamma_{mf}\gamma_r\Psi_r)^{-1}\Sigma_j I_{jk}}{\gamma_s\Psi_s p_T^2 T}. \tag{15}$$

This latter form shows explicitly that the scattering coefficient estimate is, as it should be, the ratio of the scattered intensity to the incident intensity, scaled by the area.

The summation over the individual intensity contributions helps to stabilize the scattering coefficient estimate for each patch. A rough estimate of its expected variance can be obtained by considering the variance of the contributing intensity values, assuming the w_k are approximately constant. The time bandwidth product for the LFM pulse and 200-m patches is approximately 15, thus the expected variance of the sum is approximately 1 dB.[12,4]

If a patch is illuminated by more than one beam, then the scattering coefficient estimate is modified appropriately to become a weighted sum of the individual beam contributions,

$$m_k=\frac{1}{A_k}\frac{\Sigma_n\Sigma_j I_{jkn}}{\Sigma_n \mu_{kn}\gamma_{kn}}, \tag{16}$$

where μ_{kn} is the fractional area of patch k illuminated by beam n and γ_{kn} is the appropriate gain factor for the beam.

In the examples examined later, we have employed an iterative form of the backprojection in which the weighting function in Eq. (12) is modified to

$$w_k=\frac{\gamma_k \cdot \hat{m}_k \cdot G_{k,j}}{\Sigma_{k\in K(j)}\gamma_k \cdot \hat{m}_k \cdot G_{k,j}}, \tag{17}$$

where $\hat{m}_k$ is the scattering coefficient estimate from the previous iteration. The numerator is now the predicted scattering intensity for each patch, and the effect of the backprojection, when there is a mismatch between the predicted and recorded intensities, is to modify each of the assigned intensities by an equal decibel increment.

The backprojection method has similarities to the ''environmental symmetry breaking'' approach of Ref. 3, particularly on the first iteration when the factor controlling the assignment of energy is usually transmission loss. The difference between the two is that Ref. 3 requires the difference in transmission loss to reach a threshold before assigning all energy to one side, while here assignment of energy changes continuously as a function of the weighting factors.

The above formulation is essentially a local one, relating a deterministic reverberation signal, consisting of sections of elevated intensity, to scattering produced by direct paths to and from the seafloor. Local volume scattering can be considered to be included as part of the scattering signal via the scattering function g. All other contributions to the recorded reverberations, including nonlocal, multiple scattering, and long-range, subseafloor refraction paths are grouped together as noise. Forward modeling using a PE approach indicates that multipath contributions to the reverberation field are not significant for ranges below a 1/2 CZ except as diffuse arrivals that fill in the noise floor. However, compact multipath arrivals do become significant in the time interval immediately beyond the direct path 1/2 CZ returns, where they fill what otherwise would be an acoustic shadow zone.[13] In principle, it would be possible to incorporate multiple scattering in the above formulation by including suitable ray paths. However, such an approach would quickly become unwieldy if many paths were included, and would itself require an *a priori* parameterization of the scattering process, which is one of the objectives of the study. Full wave solutions, such as PE, automatically handle the multipath problem but do not circumvent the need for prior parameterization. The validity of the local scattering assumption and the possible influence of multipaths can be examined, *a posteriori*, by determining whether there are any regions of high scattering coefficient not associated with identifiable bathymetric features or that are not stable between migration images.

III. IMPLEMENTATION DETAILS

In implementing the migration algorithm for the SRP data, we used as the grid for the scattering coefficient map the 200 m by 200 m swath bathymetry grid from the Hydrosweep survey. This grid size is smaller than the nominal cross track resolution of the beamformed data, which is ~600 m at broadside at 1/2 CZ, but expressed as an ensoni-

fication duration, 270–380 ms, it is considerably greater than the 18-ms duration of the compressed LFM pulse. For such a broadband signal, the compressed pulse width yields an optimistic estimate of range resolution, since the actual resolution would depend strongly on the duration of the scattering response, including the surface reflection delay at the source/receiver. Whether the 200-m grid is large with respect to the scattering function duration can be checked *a posteriori* by assessing the stability of the scattering strength maps.

The quantities needed for backprojection of a given ping are calculated in two stages: First relevant propagation quantities such as travel time and ray angles are found at the vertices of the grid; then the sweep function $G(t)$ is estimated for each 200 by 200 m grid element, which we refer to here as a patch. The spatial homogeneity of the sound speed profile during the MAE, ignoring small-scale fluctuations, permits a considerable reduction in computational effort: Propagation quantities are not found directly by three-dimensional ray tracing but are instead interpolated from fields calculated on regular two-dimensional, range–depth, grids. This simplification to a one-dimensional profile also makes it computationally feasible to use PE calculations, rather than less accurate ray theory, to estimate broadband transmission losses. Two sets of PE results were calculated, one with a ten element source array for the forward direction, the other with a single element at the HLA depth for the reverse direction, assuming reciprocity. The single frequency results were then integrated across the source band to obtain the broadband TL estimates (cf. Fig. 3). The effects of the free surface reflection are included in the amplitude terms of the migration via the TL calculation. However, only one set of travel times for the downgoing ray paths are calculated, thus the theoretical resolution is degraded by at least the time difference in the up and downgoing ray paths, $\sim$20 ms (cf. Fig. 4).

The properties of the individual patches are calculated from the vertex values. The patches are assumed planar, and their orientation is found by a least squares fit to the vertex depths. Similarly the isochron surfaces of two-way travel time are assumed to be locally planar and are found by a least squares fit to the travel times at the vertices. In the following examples, we take the scattering function, $g(t)$ to be a δ-function, and thus the sweep function, $G(t)$ for a patch is comprised of a series of straight line segments with vertices corresponding to times when the wavefront passes a patch vertex (Fig. 5). From Fig. 5, it is evident that using a more complex short duration function rather than a δ function, for example a 18-ms boxcar, would have only a minor effect on the scattering coefficient estimates after averaging.

Geometric shadowing calculations were performed prior to migration using a mean bathymetric profile along each beam direction. A small transition region of 300 m was added to the edge of each ensonified region in order to help prevent small numerical artifacts or local roughness from producing shadow zones. Thus individual patches that point away from the local ray direction were considered ensonified, and it was left to the migration to determine whether they produced small backscatter. During backprojection, the scattering strengths of patches in the shadow zones were adjusted at each iteration in order to produce an expected backscattered intensity equal to the 10th percentile of the ensonified patches. The purpose of this was to ensure that the preponderance of energy would be assigned to an ensonified area if the conjugate side was in shadow, while at the same time permitting energy assignment if both sides were in shadow. In this way it was possible to keep track of unexpected scattering sites in the reverberation map.

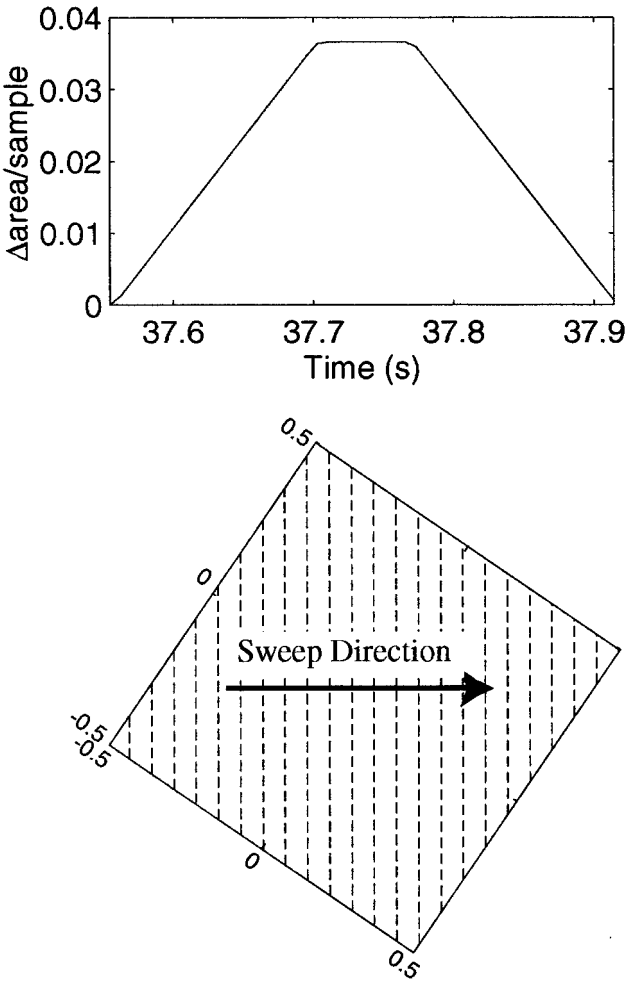


FIG. 5. Example sweep area function for a seafloor patch. The sweep function consists of linear segments with joints corresponding to the passage of the beam wavefront (dash lines) past the patch vertices. For the calculation both the patch and wavefront are assumed locally planar. (Patch area is normalized to unit area.)

The migration algorithm was implemented primarily in MATLAB but with the sweep functions for individual patches calculated by an external C function. The first stage of the algorithm, the calculation of propagation quantities at the vertices of the grid took 6 min of CPU, and the second stage took 7 min per iteration on a HP 715/75 workstation (SPECINT95 3.1, SPECFP95 3.6). MATLAB is an interpreted environment and we conservatively estimate that these times could be reduced by a factor of 4 if the algorithm was implemented as compiled code.

IV. MIGRATION EXAMPLE FROM B′

We take as an initial migration example, ping 1313 of segment 415, for which the R/V CORY CHOUEST was located

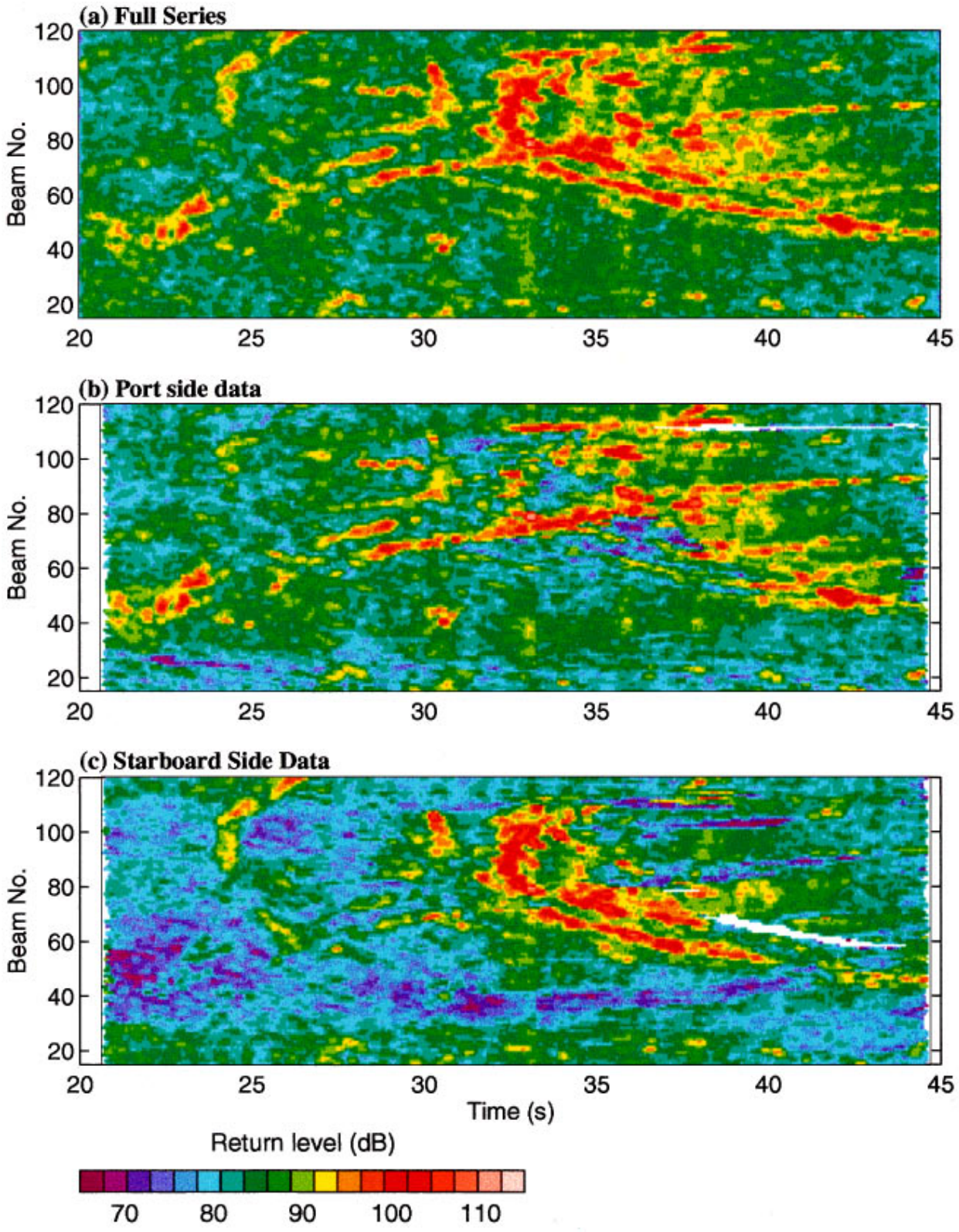


FIG. 6. (a) Reverberation data for ping 1313, broadband LFM transmission of s415. The displayed data spans the time interval of reverberation returns from 1/2 CZ and beam numbers of 15–120, which avoids endfire (Beam Nos. 64/65 are the broadside beams). To facilitate comparision with the reverberation maps, the data have been averaged over an interval of 270 ms and corrected for the average number of ensonified patches, a correction of −11.1 dB. The prominent returns between 30 and 45 s are a combination of returns from scarps on the front face of B′ on the starboard side of the array and ridges on the port side that point almost directly into the beams. (b), (c) Estimated port and starboard time series after iterative migration. The split is plausible and for the most part clean, although some residual energy appears to be misplaced.

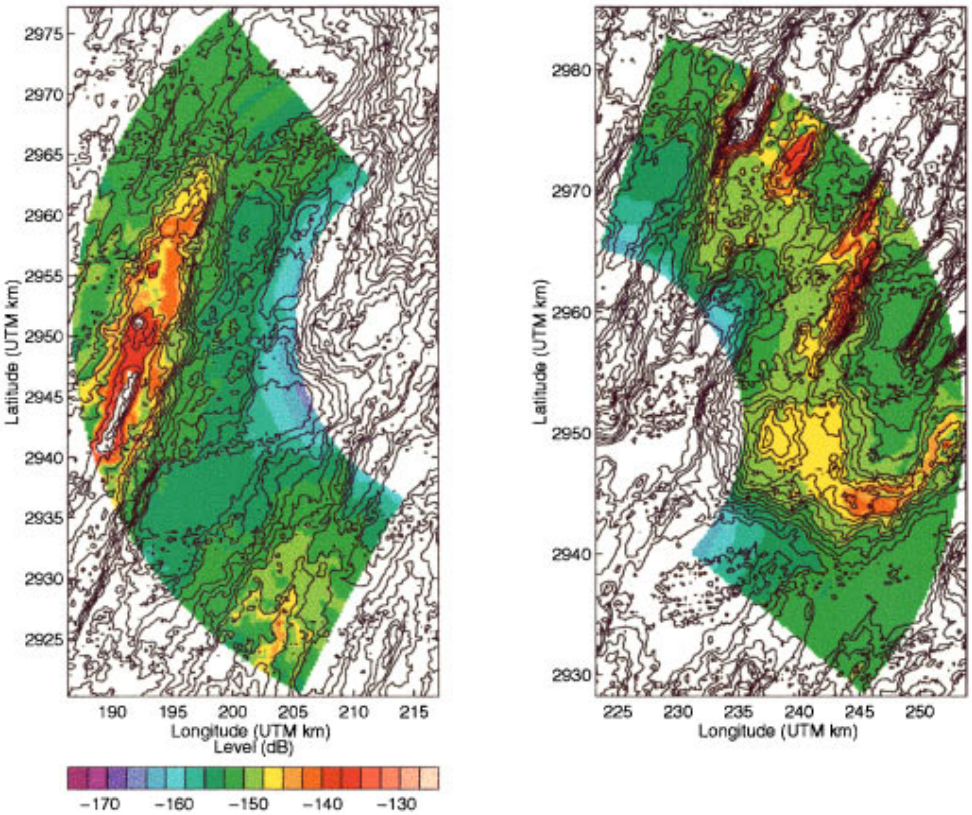


FIG. 7. Two-way transmission loss for ping 1313 projected onto the port and starboard side bathymetry. Strongly ensonified features include the top of B′ on the starboard side, and the high ridges pointing approximately into the beams on the port side. The uppermost parts of the two prominent ridges on the port and starboard side colored white project up above the acoustic beam and are thus not included in the calculation. The current figure does not take into account geometric shadowing.

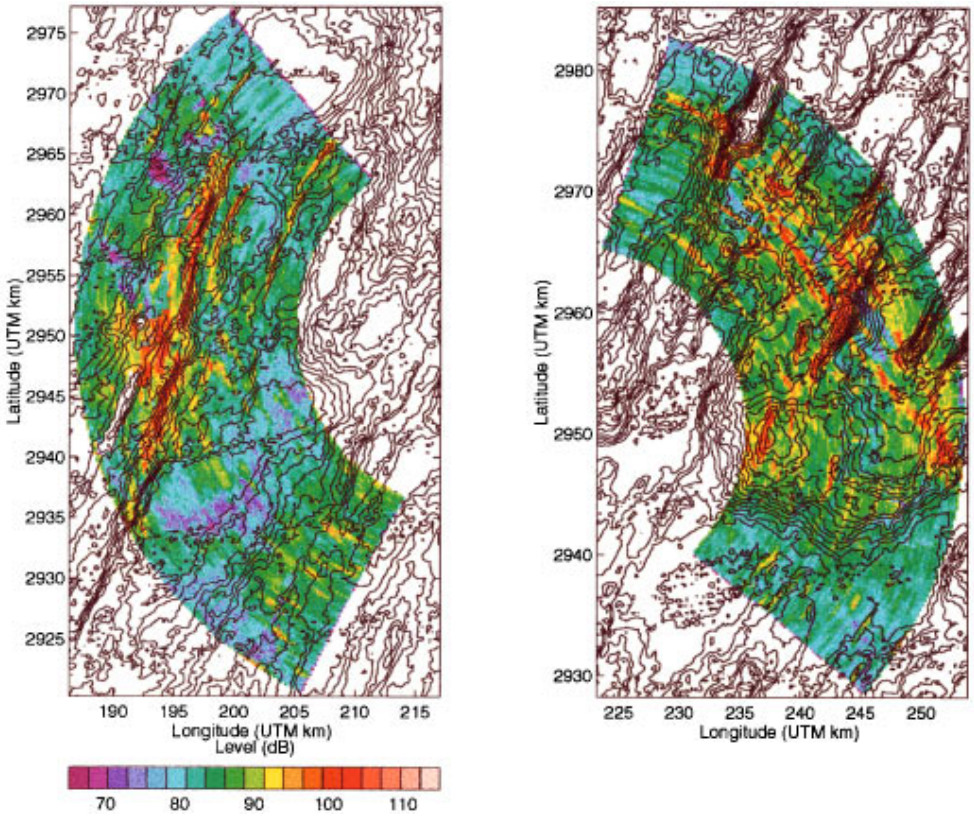


FIG. 8. Reverberation map from the initial migration of ping 1313. The left–right ambiguity resolution is based primarily on ensonification level, although some dependence on local incidence angle is included via the sweep rate function. High amplitude returns due to local slope such as those from the scarp slopes of B′ are not fully resolved and also appear in the conjugate image.

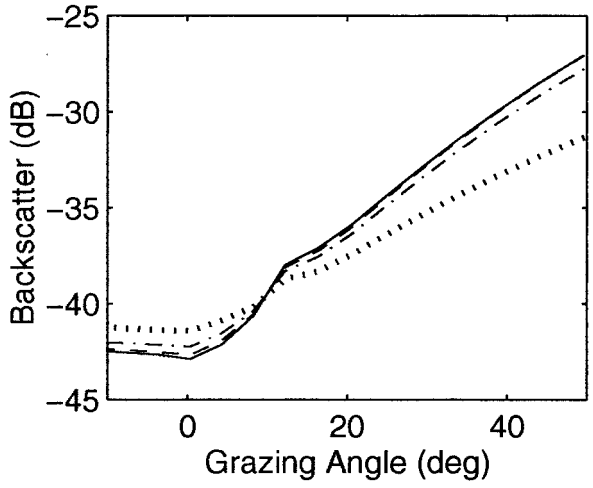


FIG. 9. Estimate of backscatter strength as a function of grazing angle for the four migration iterations of ping 1313. Iteration 1—dotted; 2—dash-dot; 3—dashed; 4—solid. For iteration 1 there is no *a priori* assumption of increased backscatter with grazing angle, yet the mean backscatter strength shows a correlation with grazing angle down to the noise floor at about 10 °. For subsequent iterations the mean backscatter of the previous iteration is used and convergence occurs in 3–4 iterations.

just north and west of the center of the track star near ridge B′, Fig. 1, and had a HLA heading of 169°, Table I. At a 1/2 CZ on the starboard side is B′ with its backfacing scarps oriented almost perpendicular to the beam directions, while at the same range on the port side, a series of N25°E trending ridges point almost directly into the beam directions (Fig. 1). The reverberation returns from these ridges and the scarp of B′ interfere to produce a set of crossing, high amplitude events in the data between 30 and 45 s (Fig. 6). When we examine the two-way transmission loss to the conjugate areas, we find that only the port side ridges and the upper part of B′ are shallow enough to project into the center of the acoustics beam, although a broad platform on the port side is also well ensonified (Fig. 7).

Figure 8 displays the reverberation map obtained after the first iteration of the migration using data from 20 s out to times corresponding to the 1/2 CZ turning ranges. The reverberation map is an intermediate backprojection output and is calculated from the sum of the backscattered energy associated with each patch of the grid, Eq. (12). The reverberation map is a stable estimate in the sense that backprojection is guaranteed to be intensity preserving, and the partition of energy between simultaneously ensonified patches is affected only by relative not absolute errors in, for example, transmission loss. From Fig. 8 we can see that the high amplitude backscatter events are predominantly located within areas that were strongly ensonified. However, a notable exception is the strong scattering associated with a line of small scarps on B′ in the region (2955–2970, 205–210) UTM km, which lies well outside the main acoustic beam. Furthermore, it is reassuring that within the strongly ensonified regions the reverberation events tend to be localized along bathymetric features with high slopes, such as the scarps covering the upper parts of B′. This impression is supported by Fig. 9, which shows that above a cutoff of about 10° the mean scattering strength increases with local grazing angle, a dependence that arises even though the *a priori* assumption is that the scattering coefficient is constant. The cutoff angle of about 10° represents the point at which the scattering strength curve reaches the noise floor, approximately 30 dB below the maximum scattering strength. The primary source of noise is cross-talk between the beamformed channels and thus the usable dynamic range of the ARSRP reverberation data is set by the −30 dB sidelobe level of the beamformer. The influence of the cross-talk noise on the ARSRP reverberation data is nicely demonstrated by finite-difference modeling studies.[13]

Although, the initial migration produces a fair degree of left–right ambiguity resolution, it is clearly not perfect if local slope is the sole or primary determinant of backscatter strength. For example, energy that on the starboard side is associated with the scarps of B′, also appears on the port side where it cuts across the mostly shallow slope bathymetric fabric. We can test the degree to which slope accounts for strong backscatter by incorporating the estimated angular dependence of the scattering coefficient into the partition coefficients for the backprojection, Eq. (17). All events associated with high slope features will migrate solely to the appropriate left or right image; only anomalous events will remain ambiguous and appear on both images. For this dataset, the migration effectively converged after three iterations, with only a barely perceptible change in the mean backscatter coefficient between iterations 3 and 4 (Fig. 9). The perceived ambiguity in the backscattered energy images is significantly reduced and only a few small isolated events appear split between the two images (Fig. 10). The effectiveness of the ambiguity resolution is more readily assessed in the original time-beam coordinates (Fig. 6). For the most part this view confirms the impression that the energy is cleanly split between the left and right side. However, energy near the edges of some prominent reverberation events leaks through to the other side, suggesting that there may be a slight error in the array orientation or that the migrated image could be improved by incorporating some cross talk between beams in the beamformer factor b, Eq. (4).

The scattering strength map for the fourth iteration is displayed in Fig. 11. As expected the elongated scarp slopes on B′ are highlighted in the image, but compared to the energy image, the prominence of the port ridges is reduced, their previous prominence being due solely to strong ensonification. As was to be expected from the mean behavior, there is an overall correspondence between the grazing angle and scattering strength with the normalization emphasizing the scattering strength of the lower slopes. Also, predictably, the scattering strength map is noisier than the reverberation one, since now global as well as local errors in estimated quantities are important. For example, any mismatch between the predicted and actual pattern of transmission loss will be reflected in the scattering strength map.

V. COMPOSITE IMAGES OF B′

We can assess the validity of the single ping ambiguity resolution and simultaneously gain insight into the general

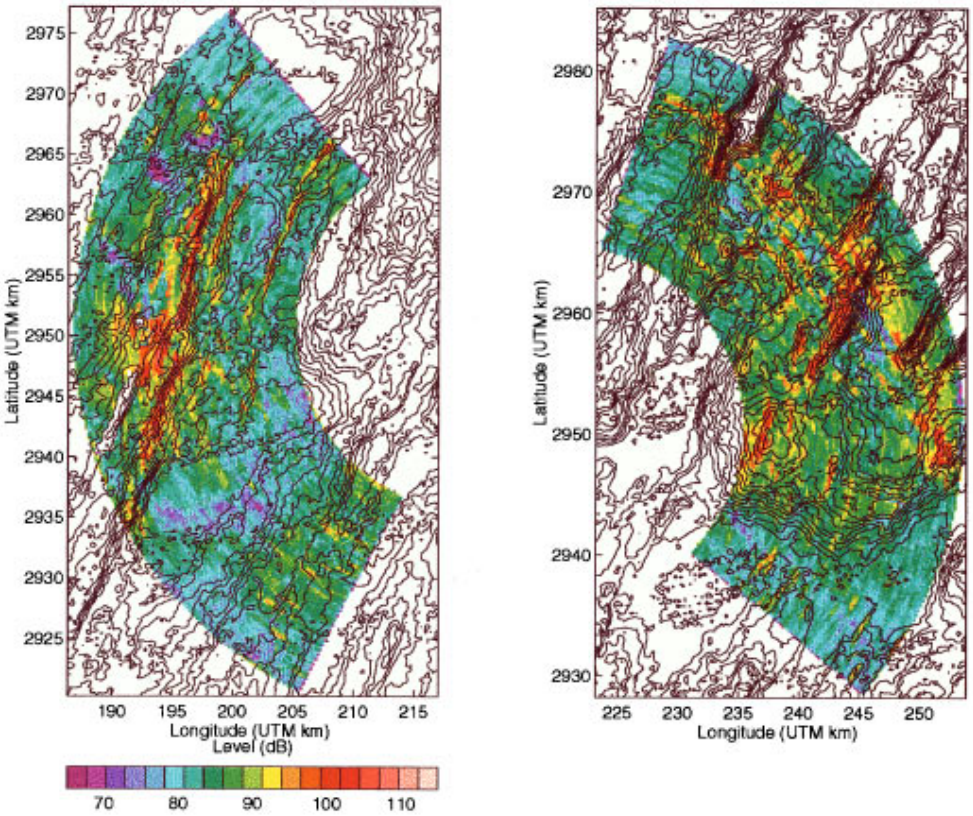


FIG. 10. Reverberation map of ping 1313 after four iterations. Compared to iteration 1, the left–right ambiguity resolution is significantly improved, with a much weaker ghosting of the B′ scarps in the port side image.

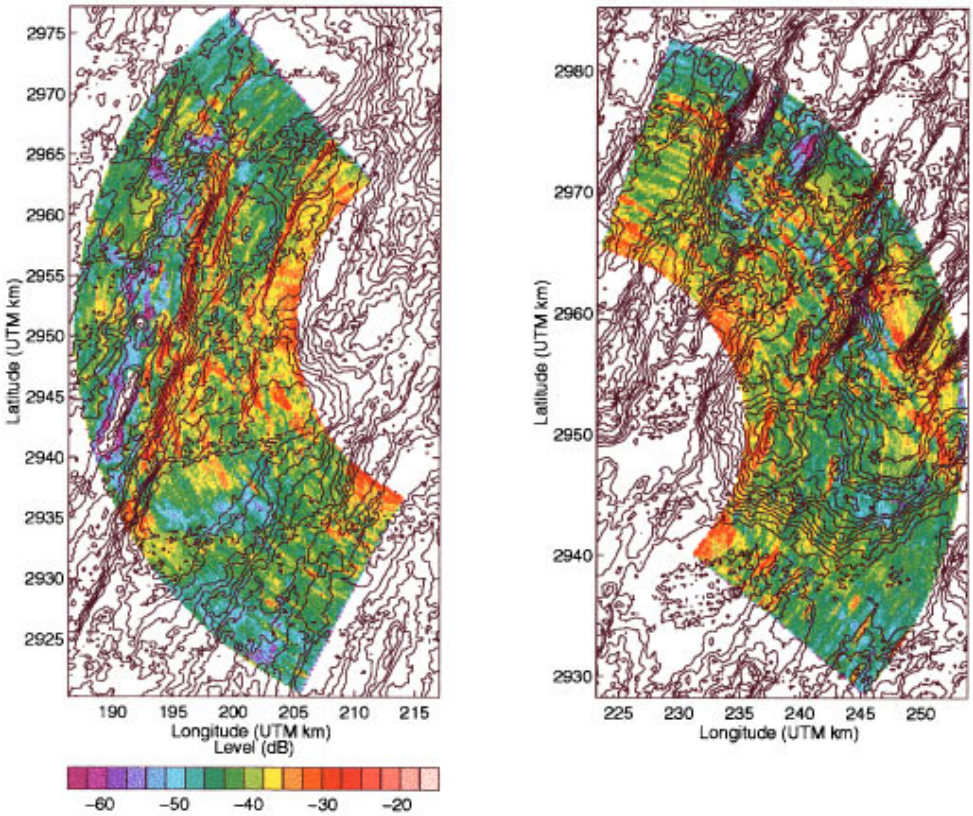


FIG. 11. Scattering strength maps derived from the reverberation map (Fig. 10) by correcting for transmission loss, area, etc. The previous prominence of the top of B′ and the port side ridges is revealed as primarily a consequence of high ensonification levels, while the B′ scarp slopes are identified clearly as regions of high backscatter.

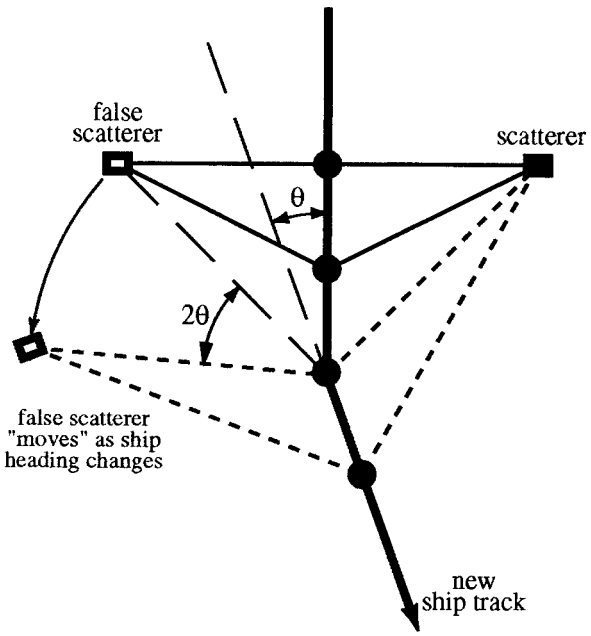


FIG. 12. With incomplete ambiguity resolution, energy from a scatterer will also be migrated to a false scatterer location; that is, the mirror image in the line array of the true location. These locations will remain paired for different pings (black circles) if the ship heading remains constant, but the location of the false scatterer will rotate through 2Θ when the array heading changes by Θ.

quality of the data by comparing the migration results from multiple ensonifications of the same target with nearly the same ship location but different headings. The standard means of obtaining multiple looks at a target is to use reverberation records from crossing paths and this we do by comparing images from runs 5a and 6 of the MAE which cross at an angle of approximately 25°. But we can also gain useful information from multiple pings from a single run, taking advantage of small changes in the heading of the receiver array between pings. If a reverberation event is migrated to its true origin on the seafloor, then its location will not change when the array heading changes by θ. Conversely a false scatterer that resulted from migrating energy to the wrong side of the array will rotate through 2θ since it is located at the mirror image of true scatterer in the array (Fig. 12). Thus the incorrectly migrated energy will move in a predictable fashion in step with the changes in array orientation. This movement can most readily be appreciated by combining the migration results from successive pings into an animation sequence. For the five consecutive LFM pings from run 5a considered here, the maximum difference in array heading is 3° which is sufficient to move false scatterers across 3–4 beams in the migrated image.

We have investigated the consistency of the migrated images using five consecutive broadband LFM pings from both run 5a and run 6 spanning their intersection at 26°36′N, −47°47′W [(2946, 222), Fig. 1]. Figure 13 displays the scattering strength maps for the northern foreslopes of B′ for nine of the pings. Perhaps the most striking aspect of the images is that relatively narrow scarp slopes appear consistently, with only relatively minor ping-to-ping variation in strength and location. The images from run 6 (pings 1683–1696) are cleaner than the ones from run 5a (pings 1313–1336) due to the fact that the conjugate area on run 6 is the source of much less backscatter. For run 5a, the lower portion of the images below 2950 includes residual mislocated energy from the port side ridges. A port side origin is supported by the fact that it does not appear in the images from run 6 and that its location moves between the successive run 5a images. Similarly energy that appears in the run 5a images near the center of a small basin at (2960, 202) can also be attributed to a partial failing of the ambiguity resolution.

Averaging the individual images effectively suppresses false scatterers attributable to energy leakage and multipaths and emphasizes the consistency of direct path, high backscatter from the narrow ridges (Fig. 14). All the high backscatter regions correspond to scarp features that are ensonified at high local grazing angles. However, the converse is not true, there are features, most notably the front edge of a small platform centered (2955, 198) that although illuminated at high angle does not produce high backscatter. Spots of high backscatter do appear along the front edge of the platform in the individual migration images, perhaps indicative of backscatter speckle that fluctuates in response to changes in the ensonification. However, since the strongest backscatter events appear in the noisier run 5a images it is more likely that the high ensonification angles succeed in capturing port side energy during migration.

For each of the ten pings, we have used the migration results to estimate the variation of mean backscatter strength with grazing angle (Fig. 15). The backscatter curve was found using the Loess algorithm,[14] assuming that the local grazing angles were correct. The Loess algorithm finds a point on the backscatter curve for a given grazing angle by fitting a line to a subset of the data centered on the grazing angle using weighted least squares. Formal confidence intervals for the mean were estimated from 200 bootstrap samples[15] for each ping, and were typically 1–2 dB at the 95% confidence level, increasing at high grazing angles. Over the interval from 10° to 45°, the mean scattering strength curves for run 5a are roughly comparable to a Lambert's law type behavior with a Mackenzie parameter of −27 dB, although the curves are distinct based on the formal confidence intervals. The backscatter estimates for a given grazing angle are distributed approximately normally with a standard deviation of 4–5 dB (Fig. 16).

The split is probably a consequence of differences in the distribution of the relatively small number of patches that are ensonified at high grazing angles. For run 5a, 73% of the 700–900 patches ensonified at angles greater than 25° (approximately 3% of the total) come from the B′ scarps, while for run 6, this percentage drops to 63% with a greater number of high angle targets on the starboard side that are small and relatively isolated. When the backscatter curves are split into port and starboard contributions, the gap between the starboard side curves that includes B′ is reduced, although not eliminated. Moreover, the port side targets display a weaker dependence on angle (Fig. 17) and are consistent

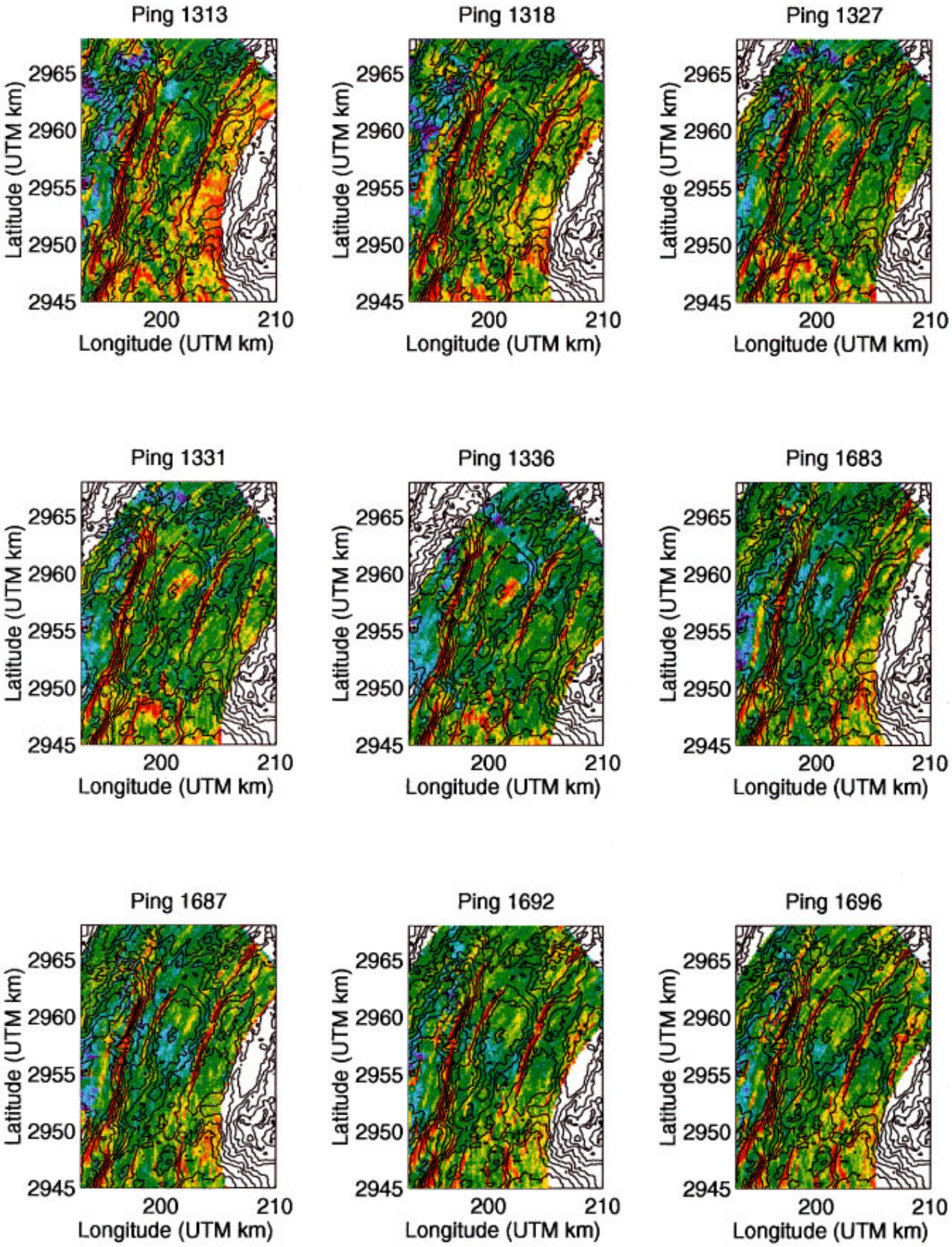


FIG. 13. Scattering coefficient maps of a subarea of the B′ abyssal hill for 5 consecutive broadband LFM pings of run 5a, 1313–1336, and 4 consecutive pings for run 6, 1683–1696. Most of the scarp slopes in the area show up consistently as regions of high backscatter. A notable exception is the edge of a small platform near (2955, 200) which is only intermittently highlighted possibly only as the result of being mistakenly assigned energy from the conjugate area.

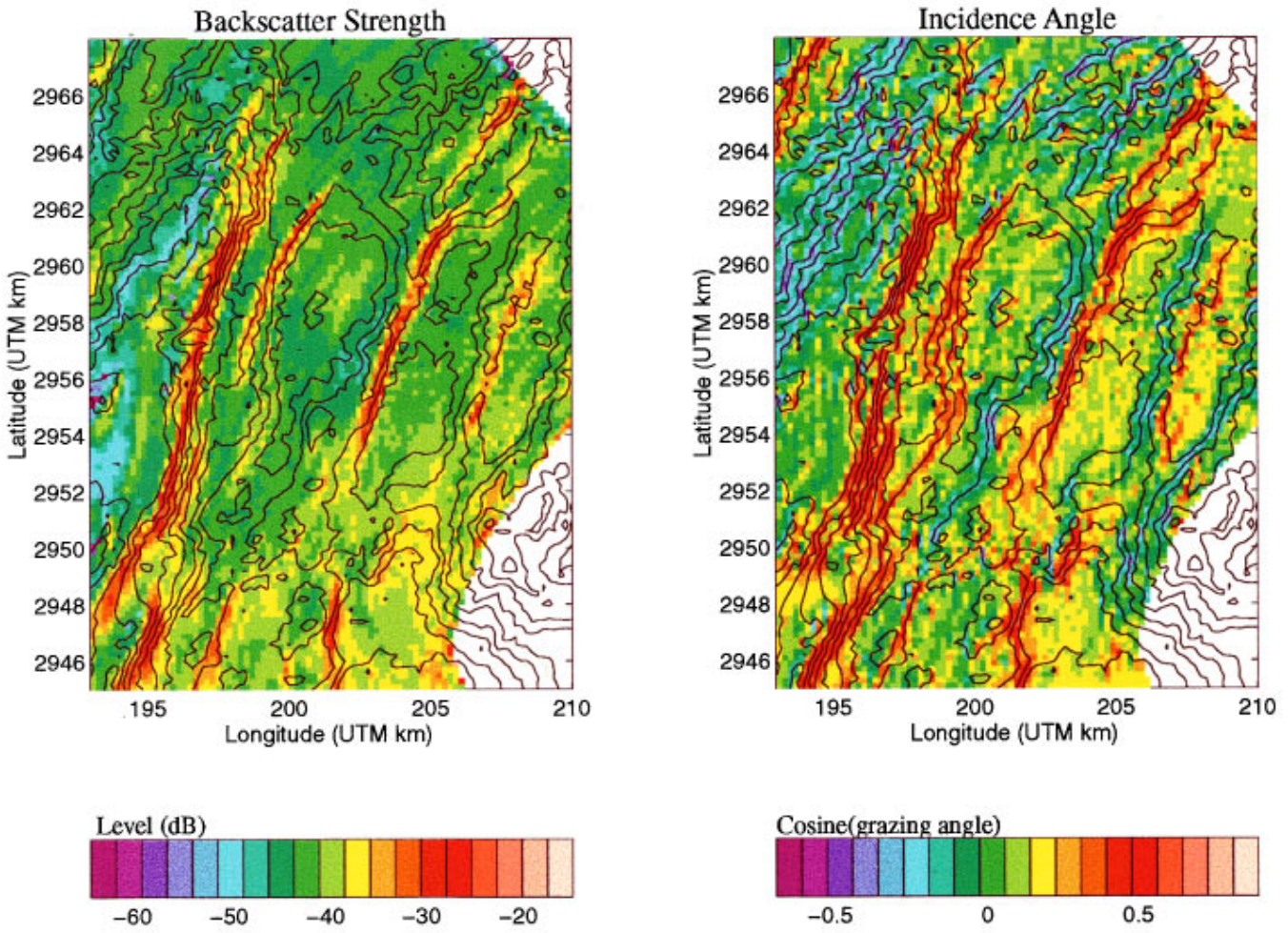


FIG. 14. Comparison of the mean backscatter strength, left panel, found by stacking the ten migrated pings, with the mean illumination angle, right panel, expressed in terms of the cosine of the local grazing angle. For the most part, there is a close correspondence between high backscatter targets and local grazing angle: all prominent backscatter targets correspond to scarp features ensonified at high grazing angles. However, there are some features, most notably the front slope of a small platform centered (2955, 198) that, although illuminated at high angle, does not produce a high backscatter.

over the two runs. The difference in scattering strength may indeed reflect an intrinsic difference in the backscatter level between the areas. However, it may also be a consequence of processing artifacts and of the limitations of the bathymetry database. The size of the port side targets tend to be near or below the angular resolution of the system, and thus the intrinsic averaging of the beamforming coupled with any small errors in the heading calculations will tend to mute the influence of high angle patches. Also the number of high angle targets is small enough that imperfections in the bathymetry database such as small noise spikes that produce erroneous high angle targets would bias the scattering strength estimates downward.

VI. CONCLUSIONS

We have developed an algorithm for migrating reverberation data that is based upon time-domain expressions for the expected intensity of the reverberation field. The time dependent aspects of the field are parameterized by a function referred to as the scattering function, which includes both stochastic components dependent on the details of the seafloor scattering and deterministic ones due to the characteristics of the acoustic imaging system. We argue that the details of the scattering function are not critical provided that scattering coefficients are estimated over patches of the seafloor with response times that are long compared with the characteristic time of the scattering function, or equivalently whose area is large compared to the correlation scales of the scattering process.[9] This approach is comparable to reducing speckle in optics by averaging over a scanning aperture.[16] Time averaging has the additional advantage of producing stable estimates from individual reverberation records, the trade-off is that spatial resolution is reduced. However, even with averaging, the broadband LFM pulse of the SRP experiments is capable of yielding consistent scattering strength estimates at the resolution of the main bathymetry database.

A key feature of the algorithm is that the results are computed in the global coordinate system of the bathymetry database rather than the beam-travel time coordinates of individual pings, facilitating the comparison of results from multiple pings. The basis for the coordinate conversion is the precise, sample-by-sample, tracking of the acoustic wavefronts across the individual patches and the assumption that the scattering coefficient is constant within individual elements, patches, of the bathymetry database. These two as-

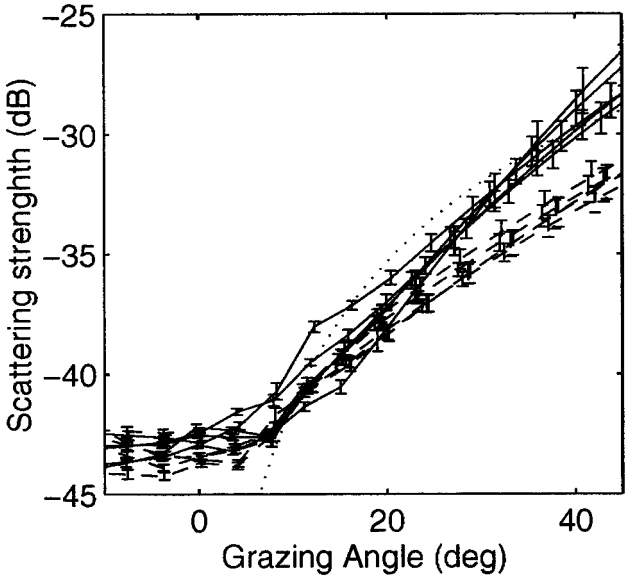


FIG. 15. Estimate of scattering strength versus grazing angle for pings from run 5a (solid) and run 6 (dashed) with bootstrap estimates of ± 2 s errors in the mean. For reference, Lamberts law curve using a Mackenzie parameter of -27 dB is included, dotted curve.

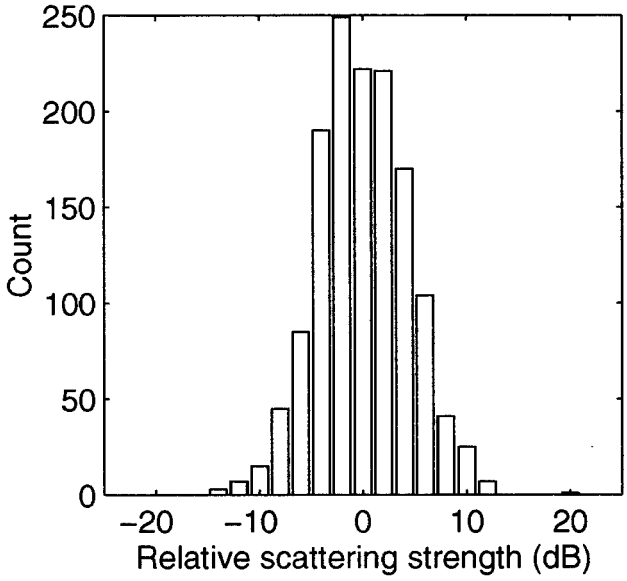


FIG. 16. Representative histogram of scattering strength relative to mean trend from ping 1313 for a bin extending from 21° to 29°. There are 1385 patches in the bin. The distribution is approximately gaussian with $\sigma = 4.4$ dB.

sumptions allow us to reexpress the reverberation response as a large sparse matrix equation with the scattering coefficients as the unknowns. The matrix formulation enables multiple records to be combined simply into a single linear inversion problem.

In this paper we have taken a processing oriented approach taking advantage of the sparsity to solve the matrix equation approximately using iterative backprojection rather than attempting to solve the equation directly as part of a linear inversion. We describe this process as migration since the method is analogous to the Kirchoff migration procedure of seismic processing, the principal difference being that here it is applied to incoherent scattering rather than coherent reflections. It has the processing virtue of producing reverberation and scattering coefficient maps rapidly thus facilitating efficient analysis of multiple records. The reverberation maps, which are produced by backprojection as an intermediate output, represent an exact partition of the record intensity and are thus guaranteed to be stable and allow us to assess directly the effectiveness of the left–right ambiguity resolution.

The performance of the migration algorithm is illustrated using two sets of broadband monostatic reverberation records from a pair of crossing runs of the SRP experiment on the flanks of the Mid-Atlantic ridge. At a 1/2 CZ, iterative migration produces well resolved maps of reverberation and scattering strength. In the reverberation maps, there is a clear correspondence between the location of high amplitude reverberation events and bathymetric features with either low transmission loss, high local grazing angle or both. In particular, narrow scarps less than 1 km wide are highlighted consistently. The migration also produces a plausible resolution of the left–right ambiguity, dividing the majority of events in a reverberation record cleanly between the two ensonified areas. The migration algorithm can be regarded as

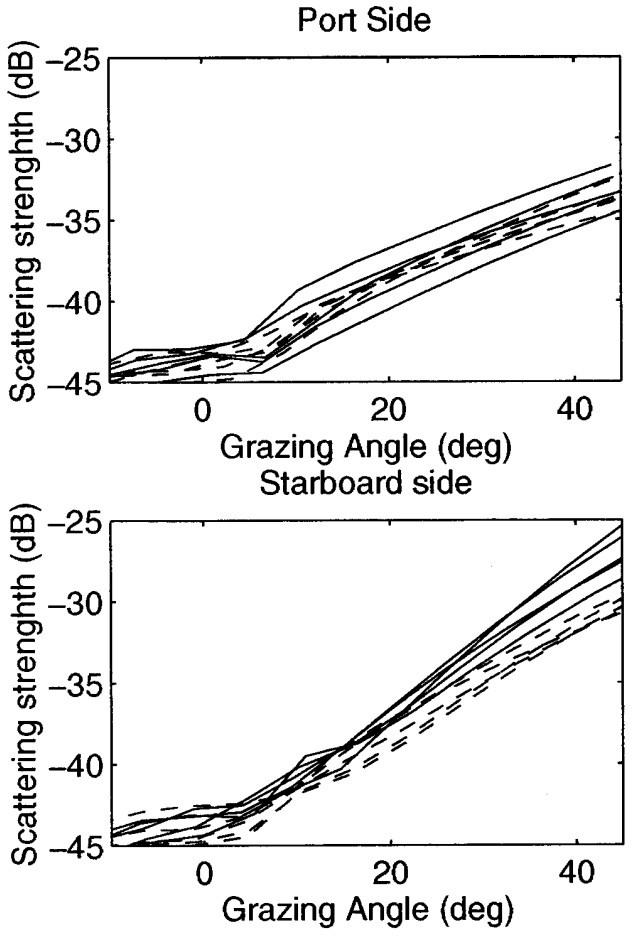


FIG. 17. Port and starboard estimates of mean scattering strength for the ten pings. Estimates from run 5a are solid lines and those from run 6 are dashed. The estimates from the starboard side that ensonify B′ show a consistently stronger dependence on grazing angle and there is less of a gap between run 5a and 6 estimates.

an environmental symmetry breaking method that uses seafloor induced asymmetries to resolve the inherent left–right ambiguity of the receiving array. Although the migration images are by no means unique, the improvement of the event separation with iteration and the clear correspondence between events and bathymetric features provides strong *a posteriori* justification for the inclusion of mean backscatter strength as a factor in the backprojection weights.

The results of multiple migrations from a single run and from crossing runs corroborates the overall effectiveness of the left–right ambiguity resolution for individual records, and supports the idea that scattering strength is primarily a function of grazing angle. The majority of high scattering strengths areas, primarily associated with scarp slopes, appear consistently in all migration images. A few potentially anomalous areas move in step with changes in array heading within a single run and tend to be absent in images from the crossing run. These areas are identified as energy misassigned to the wrong side of the array, and are attenuated when multiple maps are stacked together. Employing multiple heading directions from crossing runs is the standard means of reducing left–right ambiguity. The present results demonstrate that even small heading changes (e.g., 2°–3°) from a single run can be profitably employed to reduce ambiguity. The consistency of the migration images demonstrates the stability of the acoustic imaging system used for the SRP experiments and is a good indication that linear inversion could be applied successfully to the data.

The migration results indicate that transmission loss and grazing angle are the dominant factors controlling reverberation response in the study area. A dependence, in the mean, of backscatter strength on grazing angle appears after the first iteration of the migration when the backprojection weights contain no bias toward patches with high grazing angles. We thus conclude, as have previous investigators,[3,8,17] that the correlation, in the mean, between backscatter strength and grazing angle is a robust feature of the data.

Weaker dependencies of scattering strength on spatial and azimuthal variations in scattering strength were not explicitly sought for in the present analysis. However, the difference in the mean scattering curves between B′ and the conjugate sides (Fig. 17) indicate that there may be an azimuthal dependence resolvable in the data. The prominent ridges of B′ are ensonified almost perpendicular to the structural grain, whereas the starboard side ridges with the weaker backscatter are ensonified parallel to the grain. Resolution of questions such as azimuthal dependence of backscattering or the influence of sediments on backscattering will require further, more detailed analysis of the data.

Above a noise floor of about 5°, the estimated dependence of the mean scattering strength on grazing angles is approximately equal to, but formally statistically distinct from, that of Lambert's law with a Mackenzie parameter of −27 dB. Although the general magnitude of the increase in backscatter strength appears robust, the detailed dependence on grazing angle must be treated with some caution as the accuracy of the slopes derived from the swath bathymetry is itself questionable, especially when the high slopes are associated with smaller features near the resolution limit of the system.[18]

ACKNOWLEDGMENTS

We wish to thank Art Baggeroer, the scientific party, and the crew of the CORY CHOUEST for their efforts in collecting the data on the SRP acoustics experiment. Similarly we wish to acknowledge the efforts of Brian Tucholke, Marty Kleinrock, the science party, and crew of the R/V EWING in collecting the bathymetry database. We would also like to thank Paul Henkart for maintaining the SRP data archive and ensuring its accessibility. Finally we would like to thank Kevin Smith and Johan Robertsson for thoughtful reviews of the original manuscript.

[1] *Acoustic Reverberation Special Research Program, Initial Report*, edited by J. A. Orcutt (Scripps Institution of Oceanography, San Diego, CA, 1993).
[2] N. C. Makris, ''Imaging ocean-basin reverberation via inversion,'' J. Acoust. Soc. Am. **94**, 983–993 (1993).
[3] N. C. Makris and J. M. Berkson, ''Long-range backscatter from the Mid-Atlantic ridge,'' J. Acoust. Soc. Am. **95**, 1865–1881 (1994).
[4] N. C. Makris, L. Z. Avelino, and R. Menis, ''Deterministic reverberation from ocean ridges,'' J. Acoust. Soc. Am. **97**, 3547–3574 (1995).
[5] J. C. Van Decar and R. Snieder, ''Obtaining smooth solutions to large, linear, inverse problems,'' Geophysics **59**, 818–829 (1994).
[6] B. E. Tucholke, K. C. Macdonald, and P. J. Fox, ''ONR seafloor natural laboratories on slow- and fast-spreading Mid-Ocean ridges,'' EOS Trans. Am. Geophys. Union **72**, 268–270 (1991).
[7] M. D. Collins, ''A split-step Pade solution for the parabolic equation method,'' J. Acoust. Soc. Am. **93**, 1736–1742 (1993).
[8] C. S. Clay and H. Medwin, *Acoustical Oceanography: Principles and Applications* (Wiley, New York, 1977).
[9] L. M. Brekhovskikh and Y. P. Lysanov, *Fundamentals of Ocean Acoustics* (Springer-Verlag, New York, 1991), 2nd ed.
[10] G. H. Golub and C. F. Van Loan, *Matrix Computations* (Johns Hopkins Univ., Baltimore, 1989), 2nd ed.
[11] C. C. Paige and M. A. Saunders, ''LSQR: An algorithm for sparse linear equations and sparse least squares,'' ACM Trans. Math. Softw. **8**, 43–72 (1982).
[12] I. Dyer, ''Statistics of sound propagation in the ocean,'' J. Acoust. Soc. Am. **48**, 337–345 (1970).
[13] J. O. A. Robertson, A. Levander, and K. Holliger, ''Modeling of the acoustic reverberation special research program deep oceanseafloor scattering experiments using a hybrid wave propagation simulationtechnique,'' J. Geophys. Res. **101**, 3085–3101 (1996).
[14] W. S. Cleveland, ''Robust Locally Weighted Regression and Smoothing Scatter Plots,'' J. Am. Stat. Assoc. **74**, 829–836 (1979).
[15] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman & Hall, New York, 1993).
[16] B. R. Frieden, *Probability, Statistical Optics, and Data Testing: A Problem Solving Approach* (Springer-Verlag, New York, 1991), 2nd ed.
[17] K. B. Smith, W. S. Hodgkiss, and F. D. Tappert, ''Propagation and analysis issues in the prediction of long-range reverberation,'' J. Acoust. Soc. Am. **99**, 1387–404 (1996).
[18] M. C. Kleinrock, ''Capabilities of some systems used to survey the deep sea floor,'' in *CRC Handbook of Geophysical Exploration at Sea*, edited by R. A. Geyer (CRC, Boca Raton, 1992), 2nd ed.

Fast and explicit Wentzel–Kramers–Brillouin mode sum for the bottom-interacting field, including leaky modes

Michael A. Ainslie, Mark N. Packman, and Christopher H. Harrison
BAeSEMA Ltd., Apex Tower, 7 High Street, New Malden, Surrey KT3 4LH, England

(Received 15 October 1996; revised 22 October 1997; accepted 16 January 1998)

A fast and explicit method is presented for calculating the bottom-interacting field by means of a mode sum, using the Wentzel–Kramers–Brillouin approximation for the eigenfunctions and the Tindle–Weston formula for eigenvalue spacing. Both trapped and leaky modes are automatically included, requiring no numerical inversions for either eigenfunctions or eigenvalues. Examples are provided which demonstrate the speed and accuracy of the approach by comparison with reference solutions. The importance of leaky modes, especially in range-dependent environments with variable water depth, is demonstrated. A simple criterion for neglecting leaky modes is developed. [S0001-4966(98)05004-8]

PACS numbers: 43.30.Bp, 43.30.Ma [SAC-B]

INTRODUCTION

The normal mode sum is a popular and efficient method for solving the acoustic Helmholtz equation,[1] especially at low frequency or in shallow water. At high frequency, conventional numerical methods become less viable because of the increasingly large number of modes. For practical purposes, it is therefore often necessary to revert to approximations either to the eigenvalues[2] or eigenfunctions.[3] Here we do both, using Wentzel–Kramers–Brillouin (WKB) eigenfunctions[4] and a simple explicit eigenvalue formula first suggested for trapped modes by Tindle and Weston.[5,6] Our implementation demonstrates the computational speed of their method, which we apply to all bottom-interacting modes, including leaky modes. The WKB approximation is used for bottom-interacting modes only, avoiding the need for turning point corrections.[3,7]

Following Zhang and Tindle,[8] we use the term ''leaky'' modes rather than ''virtual'' or ''improper'' modes to describe those whose equivalent ray grazing angle exceeds the sea bed critical angle (continuous spectrum), and ''trapped'' modes for the discrete spectrum of proper modes, corresponding to totally internally reflected ray paths. The term ''discrete'' modes, often used synonymously with ''trapped'' or ''proper'' modes, is deliberately avoided, because all of the eigenvalues correspond to discrete poles in the complex plane. The reader is referred to Chapman and Ward[9] for a discussion of how the leaky modes are related to branch line integrals associated with the commonly used Pekeris[10] and EJP[11] branch cuts.

It is well known[12,13] that the leaky modes can be extremely important at short ranges (less than a few water depths) or low frequency (below the waveguide cutoff frequency), and these cases are illustrated here. As demonstrated by Fawcett *et al.*,[14] we find that, for a sloping bottom and isovelocity water, leaky modes can also be significant at *long range*, because after multiple reflections down slope the ray grazing angle is reduced so that some leaky modes become trapped; the same happens in reverse for upslope propagation. In addition to confirming the conclusions of Ref. 14, we also show how the influence of a sound-speed profile can result in (initially) leaky modes making significant contributions at very long ranges (beyond 100 km).

The theory is described in Sec. I, followed by numerical examples in Sec. II, including comparisons with reference solutions where available. This is followed by a discussion of the impact of leaky modes (Sec. III), including a criterion for neglecting them, and a discussion of computation times (Sec. IV).

I. NORMAL MODE THEORY

A. Formulation

Denoting the sound-speed profile in the water by $c_0(z)$, the Helmholtz equation for a point source at range $r=0$ and depth $z=z_0$ (see Fig. 1) is (Brekhovskikh and Lysanov,[15] p. 122)

$$[\nabla^2+\omega^2/c_0^2(z)]p(r,z)=-(2/r)\,\delta(z-z_0)\,\delta(r), \tag{1}$$

valid for depths z, z_0 less than the water depth H. The solution can be written, neglecting the (Pekeris) branch line integral,[10] as an infinite sum over the pressure eigenfunctions $\psi_n(z)$ with eigenvalues ξ_n

$$p(r,z)=\sum_{n=1}^{\infty} a_n\psi_n(z)H_0^{(1)}(\xi_n r), \tag{2}$$

where $H_0^{(1)}(x)$ is a zeroth order Hankel function of the first kind[16] and the constant a_n is equal to $\pi i\psi_n(z_0)$. For a density profile $\rho(z)$, the trapped mode eigenfunctions are normalized such that

$$\int_0^{\infty}[\rho_0/\rho(z)]\psi_n(z)\psi_m(z)dz=\delta_{nm}, \tag{3}$$

where ρ_0 is the density of water (assumed uniform) and δ_{nm} is the Kronecker delta function. The integral in Eq. (3) is unbounded for the leaky modes, and for these we use a normalization based on the cycle distance (Sec. I B below).

Consider initially a sequence of eigenvalues ξ_n that are either real or imaginary (so that ξ_n^2 is always real). Depending on the value of ξ_n^2, each mode can be categorized as one

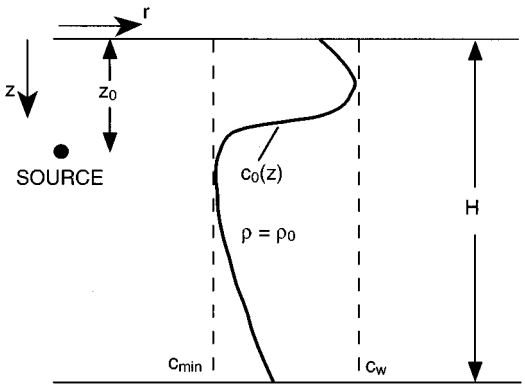


FIG. 1. Geometry and typical sound-speed profile illustrating the definition of c_w.

of: evanescent (imaginary eigenvalue, ξ_n^2 less than zero); bottom-interacting (real eigenvalue, ξ_n^2 between zero and ω^2/c_w^2), including both trapped and leaky modes; or trapped water borne (real eigenvalue, ξ_n^2 greater than ω^2/c_w^2). Here c_w is the maximum sound speed in the water, as illustrated by Fig. 1.

Let N_0 be the total number of nonevanescent modes, and N_w the number of water-borne modes, such that, allowing now for complex ξ^2 (see Fig. 2)

$$\mathrm{Re}(\xi_{N_0+1}^2)<0\leqslant\mathrm{Re}(\xi_{N_0}^2), \tag{4a}$$

$$\mathrm{Re}(\xi_{N_w+1}^2)<\omega^2/c_w^2\leqslant\mathrm{Re}(\xi_{N_w}^2). \tag{4b}$$

The evanescent modes are insignificant at horizontal ranges exceeding a few wavelengths from the source because the eigenvalues are largely imaginary, resulting in exponentially decaying terms of the form $\exp(-|\xi_n|r)$ from the asymptotic (large argument) behavior of the Hankel function. Ignoring these evanescent modes, we can write the pressure field $p(r,z)$ as a sum over water-borne modes p_w and another over bottom-interacting modes p_B

$$p(r,z)\approx p_w(r,z)+p_B(r,z), \tag{5}$$

where

$$p_w(r,z)=\sum_{n=1}^{N_w} a_n\psi_n(z)H_0^{(1)}(\xi_n r) \tag{6}$$

and

$$p_B(r,z)=\sum_{n=N_w+1}^{N_0} a_n\psi_n(z)H_0^{(1)}(\xi_n r). \tag{7}$$

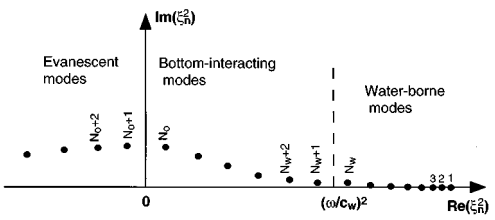


FIG. 2. Eigenvalue spectrum.

B. Bottom-interacting modes

As defined above, the bottom-interacting modes are all those (nonevanescent) modes with horizontal wave number less than ω/c_w. For these modes we can use uncorrected WKB eigenfunctions because (by our definition) none have equivalent ray turning points between the surface and bottom of the water. In other words, we have[5]

$$\psi_n(z)\approx N_n\gamma_n^{-1/2}(z)\sin\int_0^z\gamma_n(\zeta)d\zeta \quad (0\leqslant z\leqslant H) \tag{8}$$

with the normalization constant written in terms of the cycle distance D_n

$$N_n^2=4\kappa_n/D_n. \tag{9}$$

Here κ_n and γ_n are the horizontal and vertical wave numbers, respectively, given by

$$\kappa_n=\mathrm{Re}(\xi_n) \tag{10}$$

and

$$\gamma_n(z)=[\omega^2/c_0^2(z)-\kappa_n^2]^{1/2}. \tag{11}$$

The cycle distance can be written[5]

$$D_n=2\kappa_n\int_0^H dz/\gamma_n(z)+\Delta(\kappa_n), \tag{12}$$

where $\Delta(\kappa_n)$ is the beam displacement,[17] equal to (minus) the rate of change of the phase of the plane wave reflection coefficient $R(\kappa_n)$ at the sea bed. Specifically if

$$R(\kappa_n)=|R|e^{i\varphi} \tag{13}$$

then

$$\Delta(\kappa_n)=-\partial\varphi/\partial\kappa_n. \tag{14}$$

Approximate relationships between eigenvalue spacing, modal decay rates and cycle distances, derived by Tindle and Weston,[5] can be written as

$$\kappa_n\approx\kappa_{n-1}-2\pi/D_{n-1} \quad (\kappa_n>0) \tag{15}$$

and

$$\alpha_n=\mathrm{Im}(\xi_n)\approx-[\ln|R(\kappa_n)|]/D_n. \tag{16}$$

Given a suitable ''seed'' (initial values of κ_{n-1} and D_{n-1} for the recursive series), Eqs. (15) and (16) can be used to generate a good approximation to the spectrum of complex eigenvalues. The reflection coefficient $R(\kappa_n)$ can be calculated either analytically[18,19] or numerically.[20]

The approximate nature of Eq. (15) in particular means that eigenvalues are known only approximately and fine interference structure is therefore not correctly predicted by this method. However, such fine structure is usually smeared by the slightest nonzero bandwidth,[21] and an accurate calculation of the eigenvalue *spacing* $\delta\kappa$ is more important. This is because $\delta\kappa$ (or $\partial\kappa/\partial n$) and its derivatives $\partial^2\kappa/\partial n^2$, etc. determine the structure of the reflected field, as can be shown by means of a simple stationary phase calculation.[22]

An alternative approximation for the decay rate[6,23] is

$$\alpha_n\approx\frac{\rho_0\omega^2}{2\pi\kappa_n}\int_H^\infty\frac{\beta(z)\psi_n^2(z)dz}{\rho(z)c^2(z)}, \tag{17}$$

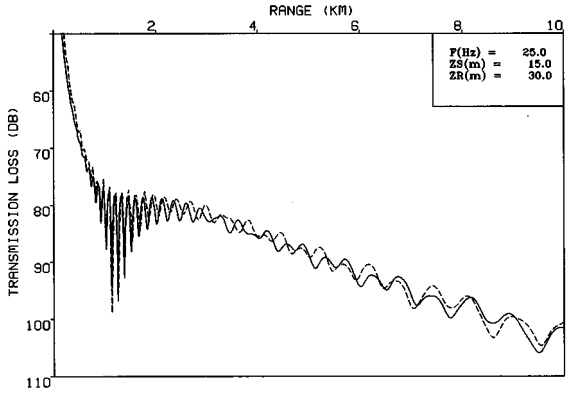


FIG. 3. Short range transmission loss for reflecting bottom (see Table II) predicted by the WKB mode sum (solid line) and by SAFARI (dotted line). The WKB sum comprises 37 modes, all leaky.

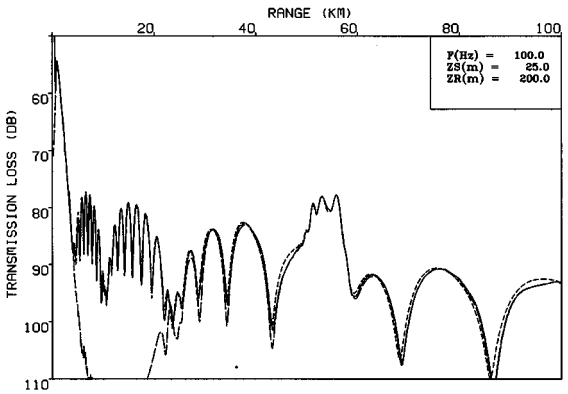


FIG. 5. Long range transmission loss for *N*. Pacific profile (Table III) predicted by the WKB mode sum and by SAFARI (dotted line). The solid line comprises 538 modes (of which 190 are trapped, including 95 water-borne modes) and the dashed line comprises the 190 trapped modes only.

where $\beta(z)$ is the volume attenuation coefficient in the sediment, in units of nepers per wavelength. Tindle and Weston have shown[5,6,24] that Eqs. (16) and (17) are approximately equivalent, and for trapped modes we are at liberty to choose either. For leaky modes, however, Eq. (17) is unbounded. In practice, for the following examples, we use Eq. (17) for modes trapped either by total internal reflection at the water–sediment boundary or by refraction in the sediment. Equation (16) is used for all other WKB modes, including leaky modes.

In the above derivation we have ignored evanescent modes on the grounds that their contribution is negligible except at very short range. There is no reason in principle why they could not be included if desired, although a slight modification to Eqs. (10) and (15) is required in order to generate the necessary imaginary values of the horizontal wave number κ_n (see the Appendix).

C. Application to weakly range-dependent media

So far we have restricted attention to range-independent environments. If we ignore coupling between modes, we can allow for weak range dependence by means of the adiabatic approximation (Brekhovskikh and Lysanov,[15] p. 141). The main differences are that the product $\xi_n r$ in Eqs. (6) and (7) above is replaced with an integral, and the wave functions $\psi_n(z)$ become functions of range as well as depth. The details are described in Sec. II C.

II. EXAMPLES

We proceed with a number of examples illustrating different features of our method. In each of Figs. 3–8 below there are up to three curves: a solid curve evaluated using the full WKB mode sum (including leaky modes); a partial WKB sum, excluding leaky modes (long dashes); and, where available, an accurate reference solution or ''benchmark'' (short dashes, or ''dotted'' line). Two of the graphs (Figs. 5 and 7) include contributions from water-borne modes as well, and these are calculated using the method of Porter and Reiss.[25]

A. Short range examples

The first two examples, from Ref. 22, are chosen to illustrate the processes of reflection and refraction at the sea bed. Refraction in the water column is not considered and this is justified by restricting attention to short ranges. The first case has uniform sound speed everywhere, with a den-

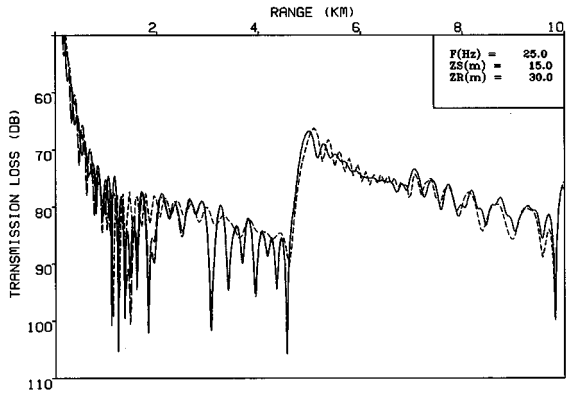


FIG. 4. Short range transmission loss for reflecting and refracting bottom (see Table II) predicted by the WKB mode sum including leaky modes (solid line, with 64 modes) and excluding them (dashed line, with 42 modes). The dotted line is SAFARI.

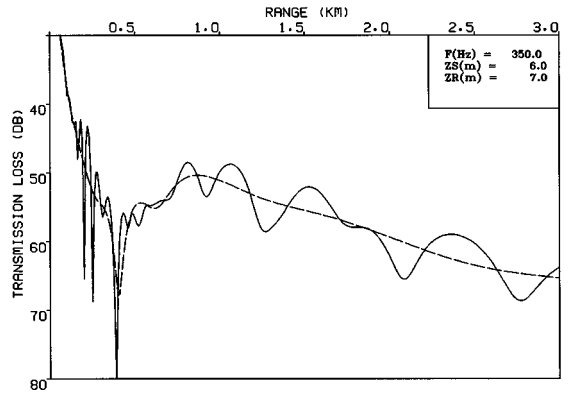


FIG. 6. Short range transmission loss for Tindle–Deane down slope environment (Table IV) predicted by the WKB mode sum. The solid line comprises 26 modes (of which initially 9 are trapped) and the dashed line comprises 9 modes (all trapped).

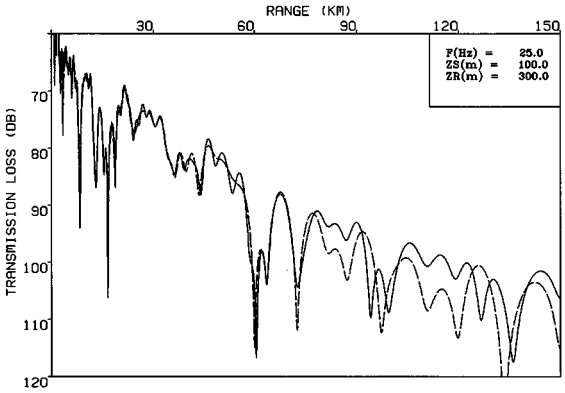


FIG. 7. Long range transmission loss for Iceland-Faeroes Ridge environment (Table V) predicted by the WKB mode sum. The solid line comprises 19 modes (of which initially 7 are trapped) and the dashed line comprises 7 modes (all trapped). The number of water-borne modes increases from 1 at the start to 12 at the end.

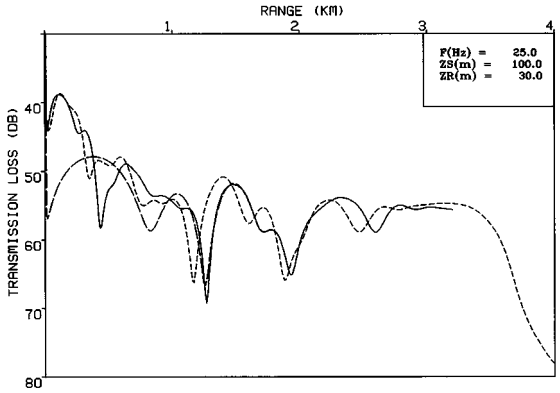


FIG. 8. Transmission loss for penetrable wedge (Table VI) predicted by the WKB mode sum and by COUPLE (dotted line). The solid line comprises eight modes (of which initially three are trapped) and the dashed line comprises three modes (initially all trapped).

sity step (ρ_0 to ρ_1) across the water–sediment boundary. The purpose of this rather unrealistic model of the sea bed is to illustrate a case for which there are no trapped modes and can therefore not be attempted by conventional normal mode theory. All we have are bottom-interacting leaky modes for which [see Eq. (12) and Table I] $D_n = 2H\kappa_n/\gamma_n$. It follows from Eqs. (2), (8), and (9), using the asymptotic form of the Hankel function, that

$$p(r,z) \approx p_B(r,z) = \frac{1}{H} \sum_{n=1}^{N_0} \left(\frac{8\pi}{\kappa_n r} \right)^{1/2} \sin\ \gamma_n z_0\ \sin\ \gamma_n z \times \exp[i(\kappa_n r + \pi/4) - \alpha_n r], \tag{18}$$

where

$$\kappa_n = \kappa_{n-1} - \pi\gamma_{n-1}/H\kappa_{n-1}, \tag{19}$$

and

$$\alpha_n = \frac{\gamma_n}{2H\kappa_n} \ln\left| \frac{\rho_1 + \rho_0}{\rho_1 - \rho_0} \right|. \tag{20}$$

Figure 3 shows transmission loss ($\mathrm{TL} = -20\log_{10}|p|$) plotted versus range for a frequency of 25 Hz, source depth 15 m, and receiver depth 30 m, calculated using Eq. (18) with the parameters of Table II. The water depth H is 1000 m and the sediment density ρ_1 is $1.222\rho_0$. Also shown is a reference solution (from Fig. 5 of Ref. 22) calculated using the fast field program SAFARI.[20] The agreement at short range is excellent, deteriorating slightly at longer ranges.

The effect of introducing a sediment sound-speed gradient c', in addition to the density step, is shown in Fig. 4, which shows TL calculated with and without the leaky modes. The bottom parameters are described in Table II. For this more realistic example the trapped modes alone (dashed line) provide a reasonable approximation to the SAFARI reference curve (dotted line, from Fig. 11 of Ref. 22) because of the high maximum sediment sound speed (2500 m/s), although there are still discrepancies at short range even with the leaky modes included (solid line) which deserve further investigation. For this example we have

$$p(r,z) \approx p_B(r,z) = \frac{1}{H} \sum_{n=1}^{N_0} \left(\frac{8\pi}{\kappa_n r} \right)^{1/2} \left[1 + \frac{\gamma_n \Delta(\kappa_n)}{2H\kappa_n} \right]^{-1} \times \sin\ \gamma_n z_0\ \sin\ \gamma_n z\ \exp[i(\kappa_n r + \pi/4) - \alpha_n r], \tag{21}$$

$$\kappa_n = \kappa_{n-1} - \frac{\pi\gamma_{n-1}}{H\kappa_{n-1}} \left[1 + \frac{\gamma_{n-1}\Delta(\kappa_{n-1})}{2H\kappa_{n-1}} \right]^{-1}. \tag{22}$$

TABLE I. Reflection coefficient and beam displacement for reflecting bottom (density step $\Delta\rho$), refracting bottom (sound-speed gradient c'), and combined reflecting and refracting bottom ($\Delta\rho + c'$). The angle θ_n is the grazing angle of the ray equivalent to mode n, such that $\tan\theta_n = \gamma_n/\kappa_n$.

	Density step $\Delta\rho$	Sound speed gradient c'	Combined ($\Delta\rho$ and c')
Reflection coefficient	$R_{\Delta\rho} = \dfrac{\rho_1 - \rho_0}{\rho_1 + \rho_0}$	$R_{c'} = \exp\left[-\dfrac{\omega\beta}{\pi c'} \ln\tan\left(\dfrac{\theta_n}{2} + \dfrac{\pi}{4} \right) + i\varphi_{c'} \right]$	$\dfrac{R_{\Delta\rho} + R_{c'}}{1 + R_{\Delta\rho}R_{c'}}$
Phase	$\varphi_{\Delta\rho} = \begin{cases} 0 & (\rho_1 > \rho_0) \\ \pi & (\rho_1 < \rho_0) \end{cases}$	$\varphi_{c'} = \dfrac{2\omega}{c'} \left[\ln\tan\left(\dfrac{\theta_n}{2} + \dfrac{\pi}{4} \right) - \sin\theta_n \right] - \dfrac{\pi}{2}$	$2\tan^{-1}\left[\dfrac{\tan(\varphi_{c'}/2)}{\rho_1/\rho_0} \right]$
Beam displacement	$\Delta_{\Delta\rho} = 0$	$\Delta_{c'} = (2c_0/c')\tan\theta_n$	$\dfrac{\rho_1\rho_0\Delta_{c'}}{\rho_1^2\cos^2(\varphi_{c'}/2) + \rho_0^2\sin^2(\varphi_{c'}/2)}$

TABLE II. Geoacoustic parameters for Figs. 3 and 4. The sediment thickness for Fig. 4 is 1000 m, beneath which is a uniform fluid substrate with parameters identical to those at the bottom of the sediment ($c=2500$ m/s, $\rho=1.222\rho_0$).

Bottom type	ρ_1/ρ_0	$c'(/s)$	$(20\log e)\beta$
Reflecting (Fig. 3)	1.222	0	0
Reflecting +Refracting (Fig. 4)	1.222	1.0	0.3 dB/λ

The beam displacement is approximately zero for leaky modes, and for trapped modes is given by (see Table I)

$$\Delta(\kappa_n)=\frac{2c_0\gamma_n}{c'\kappa_n}\,\frac{\rho_1\rho_0}{\rho_1^2\cos^2(\varphi_{c'}/2)+\rho_0^2\sin^2(\varphi_{c'}/2)}. \quad (23)$$

Equation (23) above is essentially the same as Eq. (A11) from Ainslie and Weston,[26] and is valid for low sediment attenuation (small β). For moderate or large β, their more general Eq. (A1) should be used instead. The oscillatory nature of the beam displacement (through the $\cos^2$ and $\sin^2$ terms in the denominator) is crucial for the correct determination of the bottom-interacting field for a layered sediment. It is this periodic behavior which gives rise to the rich structure of ray multipaths.[26]

B. Long range example

At long ranges (beyond a few water depths), refraction in the water becomes important and the sound speed profile must be taken into account. A deep water example for the North Pacific Ocean is shown in Fig. 5 illustrating a clear convergence zone at 55 km (from the p_w contribution), as well as the bottom reflected paths; see Table III for the sound speed profile (from Ref. 4) and sea bed parameters. For this case, Eq. (5) becomes

$$p(r,z)\approx p_w(r,z)+\sum_{n=N_w+1}^{N_0}\left(\frac{2\pi\kappa_n}{r}\right)^{1/2}\times\frac{4\sin\int_0^{z_0}\gamma_n(\zeta)d\zeta\,\sin\int_0^{z}\gamma_n(\zeta)d\zeta}{D_n[\gamma_n(z_0)\gamma_n(z)]^{1/2}}\times\exp[i(\kappa_n r+\pi/4)-\alpha_n r], \quad (24)$$

with D_n and κ_n given by Eqs. (12) and (15), respectively. A total of 538 modes are included in the full sum (solid line in Fig. 5), including 95 water-borne modes and 348 leaky modes. The importance of the leaky modes is clearly seen by comparison with the dashed line, which excludes them. The dotted line is a reference solution calculated using SAFARI. All three curves have been smoothed by assuming a 1% bandwidth, using the method of Harrison and Harrison.[21]

The parameters for Fig. 5 have been chosen for comparison with Fig. 6.2.13 of Ref. 4, and the same sound speed profile appears to have been used also by Henrick *et al.*[3] References 3 and 4 both ignore interactions with the sea bed.

TABLE III. North Pacific environment for Fig. 5 (SSP from Boyles, Ref. 4).

	z(m)	c(m/s)	ρ/ρ_0	$(20\log e)\beta$
	0	1507.2	1	0
	150	1498.1	1	0
	305	1491.7	1	0
	533	1480.7	1	0
	610	1478.9	1	0
Water	686	1478.0	1	0
	762	1478.6	1	0
	1372	1483.2	1	0
	1829	1488.6	1	0
	3048	1507.5	1	0
	4000	1523.0	1	0
Sediment	>4000	1600.0	2	1 dB/λ

C. Downslope examples

A simple generalization of the method to weakly range-dependent media is possible by means of the adiabatic approximation.[15] For isovelocity water, assuming cylindrical symmetry about the source, and using the superscript (r) to denote evaluation at range r, the result is[27]

$$p(r,z)\approx p_B(r,z)=\sum_{n=1}^{N_0^{(r)}}\left(\frac{2\pi\kappa_n^{(0)}}{r}\right)^{1/2}\frac{4\sin\gamma_n^{(0)}z_0\,\sin\gamma_n^{(r)}z}{[D_n^{(0)}D_n^{(r)}\gamma_n^{(0)}\gamma_n^{(r)}]^{1/2}}\times\exp\left[i\left(\int_0^r\kappa_n^{(x)}\,dx+\pi/4\right)-\int_0^r\alpha_n^{(x)}\,dx\right]. \quad (25)$$

This formula is evaluated in Fig. 6 for the downslope environment of Table IV, previously considered by Tindle and Deane,[28] who observed the need to ''accommodate the appearance of additional modes as the water depth increases.'' Equation (25) does this automatically because both trapped modes (those whose phase speed is less than the sediment sound speed, corresponding to totally internally reflected ray paths) and leaky modes (phase speed greater than the sediment sound speed) are included in the sum. From a ray point of view, on each bottom reflection the ray grazing angle is reduced by twice the bottom slope. As a result, in isovelocity water the product γH ($H=$water depth) is approximately constant.[29] As H increases, the vertical wave numbers (and therefore the phase speeds) are gradually reduced, providing the required mechanism for leaky modes to become trapped (i.e., cut on). In the example of Fig. 6 there are 26 (initially nonevanescent) modes in all, of which 17 start out as leaky modes. After 3 km of down slope conversion, all 17 leaky

TABLE IV. Down slope environment used for Fig. 6. The water depth H increases linearly from 50 m at the source to 260 m at a range of 3 km. The attenuation of 0.181 dB per wavelength corresponds to 0.013 Np/(m kHz), from Tindle and Deane (Ref. 28).

	c (m/s)	ρ/ρ_0	$(20\log e)\beta$
Water	1508	1	0
Sediment	1605	1.25	0.181 dB/λ

TABLE V. Idealized Iceland-Faeroes Ridge environment with deep water sound channel. The water depth H increases linearly from 500 m at the source to 2500 m at a range of 150 km.

	z (m)	c (m/s)	ρ/ρ_0	$(20 \log e)\beta$
	0	1494	1	0
Water	900	1463	1	0
	2500	1485	1	0
Sediment	>2500	1600	1.7	0.5 dB/λ

modes are trapped, and it is clear that they make a significant contribution, because of the difference between the solid line (which includes them) and the dashed line (which does not). The corresponding calculations by Tindle and Deane[28] (their Fig. 3, 4° slope) are comparable with the dashed line. They do not predict the interference seen in our solid line because their ray and normal mode sums are both restricted to totally internally reflected ray angles, thus neglecting important contributions from the continuous spectrum.

The inclusion of a sound-speed profile can produce significant effects at much longer ranges, and now we consider the trapping of sound in a deep sound channel, typical of downslope propagation.[30] We therefore have a contribution from both water-borne and bottom-interacting modes so that

$$p(r,z) \approx p_w(r,z) + p_B(r,z), \tag{26}$$

where

$$p_B(r,z) = \sum_{n=1}^{N_0^{(r)}} \left(\frac{2\pi\kappa_n^{(0)}}{r} \right)^{1/2} \times \frac{4 \sin \int_0^{z_0} \gamma_n^{(0)}(\zeta)\, d\zeta \, \sin \int_0^{z} \gamma_n^{(r)}(\zeta)\, d\zeta}{[D_n^{(0)} D_n^{(r)} \gamma_n^{(0)}(z_0)\, \gamma_n^{(r)}(z)]^{1/2}} \times \exp\left[i \left(\int_0^r \kappa_n^{(x)}\, dx + \pi/4 \right) - \int_0^r \alpha_n^{(x)}\, dx \right]. \tag{27}$$

Figure 7 shows the result of evaluating Eqs. (26) and (27) for the down slope environment of Table V, an idealized representation of the Iceland-Faeroes Ridge.[30] The frequency is lower than for Fig. 6 and here there are 19 modes in all, of which at the start 7 are trapped (including 1 water borne) and 12 are leaky. As with the previous example, and for the same reason, all leaky modes eventually become trapped. At the end there are 12 water-borne modes of which 5 (modes 8–12) started life as leaky modes, and it is the inclusion of these 5 modes in particular which causes the significant departures of the solid line in Fig. 7 (from the dashed line, which excludes them) at ranges beyond 75 km. It is therefore misleading to ascribe only short range effects to leaky modes. The same propagation problem is considered by Packman *et al.*,[31] who show that propagation is dominated by bottom duct paths (bounded above by refraction and below by bottom reflection) up to a range of 120 km, beyond which the deep sound channel (refracted above and below) takes over.

TABLE VI. Up slope environment used for ASA wedge Benchmark 3. The water depth decreases linearly from 200 m at the source to zero at a range of 4 km.

	c (m/s)	ρ/ρ_0	$(20 \log e)\beta$
Water	1500	1	0
Sediment	1700	1.5	0.5 dB/λ

D. Up slope example

Our final example is the ASA wedge benchmark for a lossy and penetrable sea bed (see Table VI). Figure 8 shows two predictions (solid and dashed lines) using Eq. (25), as well as the reference result (dotted line) using COUPLE (2-way) from Fig. 9(a) of Ref. 32. Because of the shrinking water depth, the total number of modes $N_0^{(r)}$ reduces with increasing range up the slope. The solid line starts with eight modes (including five leaky ones), and ends with only two (one trapped and one leaky) at $r = 3.2$ km; and the dashed line starts with three modes (all trapped), ending with the same two.

The discrepancy between our eight mode prediction (solid line) and the benchmark (dotted line) is attributed to our use of approximate eigenvalues [Eq. (15)] which leads to errors in the interference pattern, and this is confirmed by the improved agreement in Fig. 9 where we use the method of Porter and Reiss[25] for all trapped modes. Even better agreement for this case has been achieved by others[8,14,33] using a complex effective depth method[34] for all eigenvalues (trapped and leaky). We conclude that improvements to Eq. (15), for the eigenvalue, are desirable for low frequency calculations. This contrasts with previous examples at moderate frequencies (e.g., Fig. 5) where the consequences of such eigenvalue errors are shown to be unimportant.

E. Implementation details

In each of the above six examples there is a need to initialise the recursive sequence of eigenvalues Eq. (15). Whenever possible we use the Porter–Reiss[25] algorithm for the first two bottom-interacting eigenvalues κ_{Nw+1} and

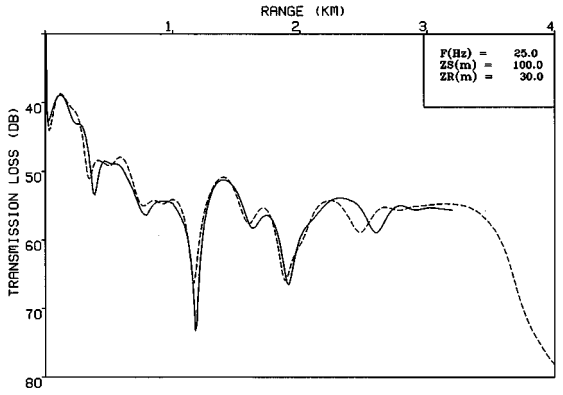


FIG. 9. Transmission loss for the same case as Fig. 8. The solid line is calculated using Porter–Reiss eigenvalues for all trapped modes and Tindle–Weston eigenvalues for leaky modes. The dotted line is the COUPLE solution from Fig. 8.

κ_{Nw+2}, and Eq. (15) is then used for $n > N_w + 2$. In exceptional cases, alternative seeds are necessary (for example in Fig. 3 where there are no trapped modes); and in Fig. 9 we use Porter–Reiss eigenvalues for *all* trapped modes. Full implementation details are described in Ref. 35.

III. IMPORTANCE OF LEAKY MODES

It is well known that leaky modes can make a significant contribution to the underwater pressure field at ranges up to a few water depths from the source, but at longer ranges it is commonly assumed that leaky modes can be neglected. If the number of trapped modes is always sufficiently large then this is reasonable. However, there are circumstances for which the contribution from trapped modes is negligible and the inclusion of leaky modes becomes essential. An obvious example is at low frequency below the cutoff of the first mode. For a less obvious case, consider the classic downslope conversion problem: a bottom-interacting ray in shallow water travels down a slope, becoming less and less steep at each interaction until it eventually becomes waterborne, propagating thereafter in a deep water sound channel. For a sufficiently gradual bottom slope, the ray invariant[29] $\int \gamma \, dz$ is conserved and there is a direct relationship between the steepness of the ray in shallow water and its proximity to the channel axis in deep water. A moments thought shows that an initially steep ray (representing a leaky mode in shallow water) will be less tightly bound to the channel axis than one which was less steep (representing a trapped mode). Under these circumstances a hydrophone close to a sound speed maximum in deep water may only receive contributions from modes which started life as leaky modes in shallow water. This has important implications for modelling of long range ambient noise which are discussed further below. We acknowledge that for a steep bottom slope the ray invariant may not be conserved and this complicates our simple picture, but it remains the case that if an initially leaky mode becomes water borne in deep water, it has the potential to make a significant contribution at long range, despite initially high losses (e.g., Fig 7). To be sure this cannot happen the number of trapped modes in shallow water must exceed the number of deep water water-borne modes.

A. Criterion for neglecting leaky modes

The number of water-borne modes N_w can be related to c_w by means of the WKB formula $N\pi \approx \int \gamma dz$:

$$N_w \pi = \int_0^H \left(\frac{\omega^2}{c_0(z)^2} - \frac{\omega^2}{c_w^2} \right)^{1/2} dz. \tag{28a}$$

Similarly the number of trapped modes N_T can be written in terms of the maximum sediment sound speed $c_{\max}$ and sediment thickness h

$$N_T \pi = \int_0^{H+h} \left(\frac{\omega^2}{c(z)^2} - \frac{\omega^2}{c_{\max}^2} \right)^{1/2} dz. \tag{28b}$$

The following inequalities follow from Eqs. (28):

$$N_w \pi < \omega H (c_{\min}^{-2} - c_w^{-2})^{1/2}, \tag{29a}$$

$$N_T \pi > \omega H (c_w^{-2} - c_{\max}^{-2})^{1/2}, \tag{29b}$$

where $c_{\min}$ is the axial sound speed (see Fig. 1). We require the value of N_T (at range zero) to exceed that of N_W (at range r). It follows that a sufficient condition for neglecting leaky modes at a receiver in deep water (range r) for a source in shallow water is

$$H^{(0)} \left(\frac{1}{c_w^{(0)^2}} - \frac{1}{c_{\max}^2} \right)^{1/2} > H^{(r)} \left(\frac{1}{c_{\min}^{(r)^2}} - \frac{1}{c_w^{(r)(2)}} \right)^{1/2}. \tag{30}$$

The same condition can be used for up slope propagation but with the (0) and (r) superscripts interchanged. In fact Eq. (30) remains applicable to a completely arbitrary bathymetry provided that the left (right) hand side is evaluated in shallow (deep) water. Specifically the left- and right-hand sides should be evaluated at whatever range they are minimized and maximized, respectively.

B. Implications for long range noise modeling

The shallow to deep propagation discussed above is often relevant to ambient noise calculations where distant shallow water shipping, through its wide azimuthal spread, contributes as much as local sources in deep water. In particular, shipping noise can enter the sound channel and fill the angle range that would otherwise constitute a ''noise-notch.'' [36,37]

TABLE VII. Summary of CPU times (microVAX 3100-80) for the six examples of Sec. II. For range-dependent environments, the water depth and number of computed modes at start and end are given (end values in parentheses).

			Trapped modes only		All modes	
Fig.	Water depth (m)	Frequency (Hz)	No. of modes	CPU (s)	No. of modes	CPU (s)
3	1000	25	⋯	⋯	37	9
4	1000	25	42	7	68	9
5	4000	100	190[a]	62	538[a]	124
6	50 (260)	350	9(9)	15	26(26)	19
7	500(2500)	25	7[b](7[c])	17	19[b](19[d])	29
8	200 (40)	25	3(2)	7	8(2)	8

[a]Including 95 water-borne modes.
[b]Including 1 water-borne mode.
[c]All water borne.
[d]Including 12 water-borne modes.

TABLE VIII. Summary of CPU times (in seconds, rounded to the nearest 10 s) for the North Pacific case ($H=4000$ m). The VAX 11/780 figures are from Ref. 3.

	microVAX 3100-80							
	Full spectrum models			Trapped spectrum models		VAX 11/780		
Frequency	Present method (inc. leaky modes)	SAFARI	EFEPE	Present method (exc. leaky modes)	KRAKEN	WKB	Normal mode (Stickler)	Parabolic equation
100 Hz	120[a]	440	11 400	60[b]	100[b]	30[c]	850[c]	1320
750 Hz	1200[a]	3520[d]	⋯	990[b]	5010[b]	⋯	⋯	⋯

[a]Including water-borne modes, trapped bottom-interacting modes, and leaky modes.
[b]Including water-borne modes and trapped bottom-interacting modes only.
[c]Water-borne modes only.
[d]Estimated value.

This phenomenon is usually handled adequately by noise models such as RANDI-2[38] using adiabatic mode theory (i.e., trapped normal modes calculated by SUPERSNAP[25]). However, the above argument makes it clear that the trapped mode sum is not necessarily adequate, and the noise could be underestimated if the leaky modes in the shallow water are ignored.

IV. COMPUTATION TIMES

Computation times for the six examples presented in Figs. 3–8 are shown in Table VII. When making comparisons with other models it should be remembered that, with rare exceptions (e.g., Ref. 14), conventional normal mode programs[39] usually compute the trapped modes only.

For the case of Fig. 5 (the North Pacific profile from Boyles[4]), a direct comparison (Table VIII) is available with a number of other methods, three of which are from Henrick *et al.*[3] (for a VAX11/780). The other five (for a micro VAX 3100-80) are from the present work; of these, the results from KRAKEN[40] and EFEPE[41] are not shown in Fig. 5. The large CPU time quoted for EFEPE is the result of a very fine range-depth grid (2 m×2 m) found to be necessary to obtain agreement with SAFARI. Notice the speed advantage over alternative computational methods: the Tindle-Weston method (with leaky modes included) is comparable in accuracy with SAFARI and EFEPE, and 3.7 and 95 times faster, respectively. Also included in Table VIII is a comparison with run times for the higher frequency of 750 Hz, showing how this advantage varies with frequency.

An alternative approach to fast normal mode calculations is described by Levinson *et al.*,[42] who assume a piecewise linear variation in wave number squared. The wave functions for trapped modes are then found as linear combinations of Airy functions, but leaky modes are not considered. They present a three-way comparison of speed and accuracy between their method, KRAKEN and a WKB method similar to that of Refs. 3 and 4.

V. CONCLUSIONS

The proposed algorithm, using WKB eigenfunctions and Tindle–Weston eigenvalues, is shown to be a fast and accurate method for predicting the bottom-interacting field, especially at moderate to high frequencies. In particular the phase error associated with the use of Tindle–Weston eigenvalues is shown to be unimportant, except at low frequencies with a very small number of trapped modes. The eigenvalue *spacing* however is important, as it is this that determines cycle distances and multipath structure.

We confirm earlier findings[12,13] that the contribution from leaky modes (i.e., the ''continuous'' spectrum) is important at short range (less than a few water depths) or below mode cutoff (e.g., Fig. 3, for which there are no trapped modes), and at slightly longer ranges for a sloping sea bed.[14] In addition we find the presence of a deep water sound channel to be especially important, allowing previously leaky modes to become trapped and thus propagate to distances beyond 100 km from the source.

ACKNOWLEDGMENTS

The authors are indebted to Dr. F. B. Jensen for providing the COUPLE data shown in Fig. 8 and to Dr. D. E. Weston for many useful discussions. Programming was by H. Parker. This work has been carried out with the support of the Ministry of Defence, D. Sc. (Sea), UK.

APPENDIX: EIGENVALUE RECURSION FORMULA FOR EVANESCENT MODES

We define ''evanescent'' modes as those whose squared complex eigenvalue (ξ_n^2) has a negative real part. As pointed out in Sec. I B, Eqs. (10) and (15) do not apply as they stand to evanescent modes because they do not allow for the required imaginary values of κ_n. The purpose of this Appendix is to consider the conditions of validity of Eq. (15), and to describe the changes required in order to incorporate evanescent modes. First we redefine κ_n using

$$\kappa_n^2 \equiv \mathrm{Re}(\xi_n^2) \tag{A1}$$

and rearrange Eq. (15) in the form

$$(\kappa_n - \kappa_{n-1})(\kappa_n + \kappa_{n-1}) \approx -2\pi(\kappa_n + \kappa_{n-1})/D_{n-1}. \tag{A2}$$

Making the substitution $\kappa_n \approx \kappa_{n-1}$ in the right hand side *only*, this becomes

$$\kappa_n^2 \approx \kappa_{n-1}^2 - 4\pi\kappa_{n-1}/D_{n-1}, \tag{A3a}$$

or [equivalently, using Eq. (9) for N_n]

$$\kappa_n^2 \approx \kappa_{n-1}^2 - \pi N_{n-1}^2. \tag{A3b}$$

This version of the eigenvalue recursion formula can be shown to work for the complete mode spectrum including trapped, leaky and evanescent modes. Equation (15) can be used for trapped modes and leaky modes, but only when the eigenvalues are large compared with the eigenvalue spacing, i.e.,

$$\kappa_n^2 \gg \kappa_{n-1}^2 - \kappa_n^2. \tag{A4a}$$

Equation (A4a) can be rearranged as

$$D_n \gg 4\pi/\kappa_n, \tag{A4b}$$

showing that the physical requirement is for the cycle distance to be large compared with the horizontal wavelength.

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

[2] D. M. F. Chapman, P. D. Ward, and D. D. Ellis, ''The effective depth of a Pekeris ocean waveguide, including shear wave effects,'' J. Acoust. Soc. Am. **85**, 648–653 (1989).

[3] R. F. Henrick, J. R. Brannan, D. B. Warner, and G. P. Forney, ''The uniform WKB modal approach to pulsed and broadband propagation,'' J. Acoust. Soc. Am. **74**, 1464–1473 (1983).

[4] C. A. Boyles, *Acoustic Waveguides: Applications to Oceanic Science* (Wiley, New York, 1984).

[5] C. T. Tindle and D. E. Weston, ''Connection of acoustic beam displacement, cycle distances, and attenuations for rays and normal modes,'' J. Acoust. Soc. Am. **67**, 1614–1622 (1980).

[6] C. T. Tindle, ''The equivalence of bottom loss and mode attenuation per cycle in underwater acoustics,'' J. Acoust. Soc. Am. **66**, 250–255 (1979).

[7] R. E. Langer, ''On the connection formulas and the solutions of the wave equation,'' Phys. Rev. **51**, 669–676 (1937).

[8] Z. Y. Zhang and C. T. Tindle, ''Complex effective depth of the ocean bottom,'' J. Acoust. Soc. Am. **93**, 205–213 (1993).

[9] D. M. F. Chapman and P. D. Ward, ''The normal-mode theory of air-to-water sound transmission in the ocean,'' J. Acoust. Soc. Am. **87**, 601–618 (1990).

[10] C. L. Pekeris, ''Propagation of explosive sounds in shallow water,'' Geol. Soc. Am. Mem. **27**, Sec. 2, 1–117 (1948).

[11] W. Ewing, W. S. Jardetsky, and F. Press, *Elastic Waves in Layered Media* (McGraw-Hill, New York, 1957), Chap. 4.

[12] D. C. Stickler, ''Normal-mode program with both the discrete and branch line contributions,'' J. Acoust. Soc. Am. **57**, 856–861 (1975).

[13] C. T. Tindle, A. P. Stamp, and K. M. Guthrie, ''Virtual modes and the surface boundary condition in underwater acoustics,'' J. Sound Vib. **49**, 231–240 (1976).

[14] J. A. Fawcett, E. K. Westwood, and C. T. Tindle, ''A simple coupled wedge mode propagation method,'' J. Acoust. Soc. Am. **98**, 1673–1681 (1995).

[15] L. Brekhovskikh and Yu. Lysanov, *Fundamentals of Ocean Acoustics* (Springer-Verlag, New York, 1982).

[16] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. Stegun (Dover, New York, 1965).

[17] D. E. Weston, ''Wave shifts, beam shifts and their role in modal and adiabatic propagation,'' J. Acoust. Soc. Am. **96**, 406–416 (1994).

[18] M. A. Ainslie, ''Reflection and transmission coefficients for a layered fluid sediment overlying a uniform solid substrate,'' J. Acoust. Soc. Am. **99**, 893–902 (1996).

[19] A. J. Robins, ''Plane wave reflection from a solid layer with nonuniform density, sound speed and shear speed,'' J. Acoust. Soc. Am. **103**, 1337–1345 (1998).

[20] H. Schmidt, ''Seismo-Acoustic Fast field Algorithm for Range Independent environments—User's Guide,'' unpublished report (1988).

[21] C. H. Harrison and J. A. Harrison, ''A simple relationship between frequency and range averages for broadband sonar,'' J. Acoust. Soc. Am. **97**, 1314–1317 (1995).

[22] M. A. Ainslie, ''Stationary phase evaluation of the bottom interacting field in isovelocity water,'' J. Acoust. Soc. Am. **94**, 1496–1509 (1993); erratum, J. Acoust. Soc. Am. **95**, 3670 (1994).

[23] F. Ingenito, ''Measurements of mode attenuation coefficients in shallow water,'' J. Acoust. Soc. Am. **53**, 858–863 (1973).

[24] D. E. Weston and C. T. Tindle, ''Reflection loss and mode attenuation in a Pekeris model,'' J. Acoust. Soc. Am. **66**, 872–879 (1979).

[25] M. B. Porter and E. L. Reiss, ''A numerical method for bottom interacting ocean acoustic normal modes,'' J. Acoust. Soc. Am. **77**, 1760–1767 (1985).

[26] M. A. Ainslie and D. E. Weston, ''Reflection with beam displacement from simple barriers,'' Acust. Acta Acust. **82**, 607–614 (1996).

[27] M. B. Porter, ''Adiabatic modes for a point source in a plane-geometry ocean,'' J. Acoust. Soc. Am. **96**, 1918–1921 (1994).

[28] C. T. Tindle and G. B. Deane, ''Sound propagation over a sloping bottom using rays with beam displacement,'' J. Acoust. Soc. Am. **78**, 1366–1374 (1985).

[29] D. E. Weston, ''Guided propagation in a slowly varying medium,'' Proc. Phys. Soc. London **73**, 365–384 (1959).

[30] F. B. Jensen, G. Dreini, and M. Prior, ''Acoustic effects of the Iceland-Faeroe Front,'' in *Ocean Variability and Acoustic Propagation*, edited by J. Potter and A. Warn-Varnas (Kluwer, Dordrect, 1991), pp. 359–374.

[31] M. N. Packman, C. H. Harrison, and M. A. Ainslie, ''Rapid calculations of acoustic propagation loss in range dependent oceans,'' IEE Proc., Radar, Sonar Navig. **143**, 184–189 (1996).

[32] F. B. Jensen and C. M. Ferla, ''Numerical solutions of range-dependent benchmark problems in ocean acoustics,'' J. Acoust. Soc. Am. **87**, 1499–1510 (1990).

[33] C. T. Tindle and Z. Y. Zhang, ''An adiabatic normal mode solution for the benchmark wedge,'' J. Acoust. Soc. Am. **101**, 606–609 (1997).

[34] P. Balasubramanian and M. M. Muni, ''A note on 'The effective depth of a Pekeris waveguide including shear wave effects' [Chapman *et al.*, J. Acoust. Soc. Am. **85**, 648–653 (1989)],'' J. Acoust. Soc. Am. **88**, 564–565 (1990).

[35] M. N. Packman and H. T. Parker, ''SUPERSNAP Speed-up: Final Report,'' BAeSEMA Report C3548/14/TR-1 (1994).

[36] R. A. Wagstaff, ''Low-frequency ambient noise in the deep sound channel-The missing component,'' J. Acoust. Soc. Am. **69**, 1009–1014 (1981).

[37] R. M. Hamson, ''The modeling of ambient noise due to shipping and wind sources in complex environments,'' Appl. Acoust. **51**, 251–287 (1997).

[38] R. M. Hamson and R. A. Wagstaff, ''An ambient noise model that includes coherent hydrophone summation for sonar system performance in shallow water,'' SACLANTCEN Report SR 70 (1983).

[39] P. C. Etter, *Underwater Acoustic Modeling: Principles, Techniques and Applications* (Chapman-Hall, London, 1996), 2nd ed.

[40] M. Porter, ''The KRAKEN normal mode program,'' Unpublished memorandum (September 1991).

[41] M. D. Collins, ''A split-step Padé solution for the parabolic equation method,'' J. Acoust. Soc. Am. **93**, 1736–1742 (1993).

[42] S. J. Levinson, E. K. Westwood, R. A. Koch, S. K. Mitchell, and C. V. Sheppard, ''An efficient and robust method for underwater acoustic normal-mode computations,'' J. Acoust. Soc. Am. **97**, 1576–1585 (1995).

A unified framework for mode filtering and the maximum *a posteriori* mode filter

John R. Buck[a)]
Department of Electrical and Computer Engineering, University of Massachusetts Dartmouth, 285 Old Westport Road, North Dartmouth, Massachusetts 02747-2300

James C. Preisig
Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Kathleen E. Wage
Research Lab of Electronics, MIT, Cambridge, Massachusetts 02139-4207 and Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 12 May 1997; accepted for publication 9 December 1997)

A unified framework is presented for examining the performance of linear mode filtering algorithms. Two common mode filters, samples of the mode shapes and the pseudo-inverse of the mode shapes, are presented in this framework as a tradeoff between sensitivity to other modes and sensitivity to white noise. The maximum *a posteriori* mode filter is presented as an alternative which gracefully transitions between these extremes, and attains the minimum mean squared error when the modes to be estimated are well modeled as samples of a Gaussian random process. Numerical simulations in both shallow and deep water environments confirm the analytically derived properties of these mode filters. © *1998 Acoustical Society of America.* [S0001-4966(98)01204-1]

PACS numbers: 43.30.Bp, 43.30.Pc, 43.30.Vj [DLB]

INTRODUCTION

The acoustic pressure field in many underwater environments is well described by a superposition of normal modes. Many oceanographic problems, including the characterization of the acoustic propagation through an ocean volume and internal wave tomography, rely on estimates of the normal modes propagating at a given location. In order to obtain these estimates, the pressure field must be sampled using a hydrophone array. These pressure samples are then inverted to estimate the constituent modes of this field at the array. Ideally, this mode estimate should be robust to the presence of environmental or sensor noise. Several different techniques have been proposed for solving this estimation problem, often referred to in the ocean acoustics literature as mode filtering. This paper presents a unified framework for examining the subclass of mode filtering problems in which the mode estimates are linear functions of the observed pressure samples from a vertical hydrophone array at a single time. The results presented here are easily extended to situations with multiple observations of the same pressure field over time.

The first mode filtering algorithm published in the literature was the sampled mode shape mode filter.[1,2] The sampled mode shape filter is the optimal linear mode filter for the detection and estimation of any single mode in spatially white noise.[3] However, this filter generally provides poor rejection of interference from the other propagating modes. The pseudo-inverse mode filter[4] rejects interference from other modes but at a cost of increased sensitivity to white noise. When the mode coefficients are considered to be complex Gaussian random variables (CGRV's)[5–8] neither of these filters attains the bound on the minimum mean square error (MMSE) given by the Fisher Information Matrix (FIM). If an adequate statistical model exists for the observation noise then this error bound is attained by the maximum *a posteriori* (MAP) mode filter. This paper develops the MAP filter and demonstrates that this mode filter transitions gracefully between the extremes defined by the sampled mode shape and pseudo-inverse mode filters. In addition, we observe that the MAP mode filter provides a theoretical justification for the empirically motivated mode filter proposed by Yang,[9] which drops small eigenvalues from the computation of the inverse.

The deterministic model for mode coefficients is a common and familiar framework for acoustic propagation.[10–13] Recent work has investigated the role of stochastic models for characterizing acoustic propagation, allowing the application of new classes of signal processing algorithms to detection and estimation problems for ocean acoustic parameters. Specifically, the complex Gaussian distribution is used to model uncertainty in the absolute phase of the signal due to a variety of causes, including source–receiver range uncertainty.[14,15] This distribution has also been shown to be the asymptotic limit of the received field for long-range propagation through random media under specific conditions.[16,17] It is beyond the scope of the present paper to discuss the validity of the deterministic and stochastic propagation models. Rather, we address the implications of each model for a specific estimation problem: the linear estimate of the complex mode coefficients observed at a vertical line array at a single time.

[a)]Formerly at Research Lab of Electronics, MIT, and Dept. of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution.

The remaining subsections of the Introduction review the normal mode equations, common models for noise in the underwater acoustic environment, and the singular value decomposition, an orthogonal matrix factorization useful in characterizing the behavior of mode filters. Section I describes previously proposed mode filters and develops the MAP mode filter. Section II presents simulations comparing the MAP mode filter to other mode filters in a shallow water environment, while Sec. III presents similar results for a deep water sound speed profile. Finally, Sec. IV draws conclusions about the appropriateness of different mode filters based on the simulation results.

A. Normal mode equations

The normal modes of a single-frequency (CW) acoustic pressure field are the solutions to the homogeneous Helmholtz equation, which in a region of constant density is

$$\nabla^2 p(\mathbf{r}) + k^2(\mathbf{r}) p(\mathbf{r}) = 0,$$

where $k(\mathbf{r})$ is the local acoustic wave number. The local wave number is defined to be $\omega/c(\mathbf{r})$, the ratio of the angular acoustic frequency to the sound speed. If the solution is assumed to be separable in range and depth and cylindrically symmetric, the resulting vertical (depth) eigenfunction equation is

$$\frac{d^2}{dz^2}\,\Psi_m(z) + k_{zm}^2(z)\Psi_m(z) = 0. \tag{1}$$

The solutions of this equation which satisfy the auxiliary conditions are the normal modes.[11,13] In Eq. (1), $k_{zm}(z) = \sqrt{k^2(z) - k_{rm}^2}$ is the vertical wave number of the mth mode and k_{rm}^2 is the separation constant for that mode. The square root of the separation constant, k_{rm}, is the horizontal wave number of the mode. By convention, the modes are normalized such that $\int |\Psi_m(z)|^2 \rho^{-1}(z)\,dz = 1$, which simplifies to $\int |\Psi_m(z)|^2\,dz = \rho_0$ for our constant density assumption. In a realistic scenario, where the environment is range varying, the solution to the wave equation is not separable in general. However, we continue to use the normal modes $\Psi(r,z)$ computed using $k(r,z)$ as a basis for the field at range r.

The pressure field can be written as a weighted superposition of these local normal modes

$$p(r,z) = \sum_m d_m(r)\Psi_m(z;r),$$

where the $d_m(r)$ are the mode coefficients at range r. We parametrize $\Psi(z;r)$ in this fashion because k and Ψ vary more rapidly in depth than range for most ocean environments. Only a finite set of modes in any environment have predominantly real horizontal wave numbers k_{rm}. These M modes are known as the propagating or trapped modes for the channel. In the far field of the acoustic source, the modes propagate in range as $\exp(ik_{rm}r)/\sqrt{k_{rm}r}$,[12,13,18] therefore any mode with a significant imaginary part to k_{rm} decays quickly with range, and is known as an evanescent mode. Evanescent modes are usually not excited in the far field, unless range inhomogeneities couple energy from the trapped modes into the evanescent ones.

The spatial samples of the pressure field observed at a vertical array of N hydrophones can be written as

$$\begin{bmatrix} p(z_1) \\ \vdots \\ p(z_N) \end{bmatrix} = \begin{bmatrix} \Psi_1(z_1) & \cdots & \Psi_M(z_1) \\ \vdots & \ddots & \vdots \\ \Psi_1(z_N) & \cdots & \Psi_M(z_N) \end{bmatrix} \begin{bmatrix} d_1 \\ \vdots \\ d_M \end{bmatrix} + \begin{bmatrix} n(z_1) \\ \vdots \\ n(z_N) \end{bmatrix},$$

or in vector notation

$$\mathbf{p} = \Psi\mathbf{d} + \mathbf{n}, \tag{2}$$

where $\mathbf{n}$ is the vector of observation noise at the hydrophone locations, and $z_1,\ldots,z_N$ are the depths of the hydrophones. The goal of mode filtering is to estimate the mode coefficients ($\hat{\mathbf{d}}$) from the observed pressure samples ($\mathbf{p}$) as accurately as possible in the presence of the noise ($\mathbf{n}$).

B. Observation noise

The noise $\mathbf{n}$ in the observed pressure $\mathbf{p}$ can be due to several causes: two common sources are instrumentation noise and sea-surface noise. In estimating the mode coefficients $\mathbf{d}$ from the observed pressures $\mathbf{p}$, the noise is generally considered to be a CGRV with zero mean and spatial covariance $\mathbf{K_{nn}}$. The structure of $\mathbf{K_{nn}}$ depends on the geometry of the array, the ocean conditions during the observations, and the source of the noise. Two common models for $\mathbf{K_{nn}}$ are the spatially white (SW) noise model and the Kuperman–Ingenito (KI) surface noise model.[19] The SW noise model assumes that the noise at each hydrophone is equal in power and uncorrelated with the noise at all the other hydrophones, so $\mathbf{K_{nn}} = \sigma_n^2\mathbf{I}$. This model is most appropriate under high signal-to-noise ratio conditions when the noise that is present consists predominantly of instrumentation noise.

The KI noise model proposes that the noise generated by the sea surface couples into each mode independently with varying power such that

$$K_{\mathbf{nn}} = \Psi \begin{bmatrix} \sigma^2_{\tilde{d}_1} & 0 & \cdots & 0 \\ 0 & \sigma^2_{\tilde{d}_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma^2_{\tilde{d}_M} \end{bmatrix} \Psi^H = \Psi K_{\tilde{d}\tilde{d}} \Psi^H, \tag{3}$$

where $\sigma^2_{\tilde{d}_1},\ldots,\sigma^2_{\tilde{d}_M}$ are functions of the mode profiles and surface noise processes. The elements of the vector $\tilde{\mathbf{d}}$ are the mode coefficients of the noise process at the array.

In some scenarios, both surface generated and instrumentation noise contribute significantly to $\mathbf{n}$, and the covariance matrix $\mathbf{K_{nn}}$ contains components of both forms.

C. Singular value decomposition

The singular value decomposition (SVD) is an orthogonal matrix factorization that is helpful in evaluating and understanding the performance of mode filters. For an $N \times M$ matrix $\mathbf{A}$, the SVD factors the matrix such that

$$\mathbf{A} = \mathbf{U_A}\Sigma_\mathbf{A}\mathbf{V_A}^H,$$

where $\mathbf{U_A}$ and $\mathbf{V_A}$ are $N \times N$ and $M \times M$ unitary matrices, respectively.[20] The matrix $\Sigma_\mathbf{A}$ is a nearly diagonal real ma-

trix. When $N>M$, as will generally be true in mode filtering problems, $\Sigma_{\mathbf{A}}$ has the form

$$\Sigma_{\mathbf{A}}=\left[\begin{array}{cccc} \sigma_{\mathbf{A}1} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{\mathbf{A}M} \\ \hline & & 0 & \end{array}\right], \tag{4}$$

where $\sigma_{\mathbf{A}1}\geqslant\sigma_{\mathbf{A}2}\geqslant\cdots\geqslant\sigma_{\mathbf{A}M}\geqslant 0$ are called the singular values of $\mathbf{A}$.

I. MODE FILTERS

This section reviews several common mode filters and analyzes their performance. For each filter we first evaluate its performance in terms of its bias and covariance when the mode coefficients $\mathbf{d}$ are considered to be deterministic but unknown quantities. The performance is also evaluated in terms of the mean squared error (MSE) when the mode coefficients are considered to be CGRV's. In each case, the performance is compared against the bound derived from the FIM. Following this discussion of the common mode filters, we then derive the MAP mode filter. As noted in the previous section, we restrict our attention to linear mode filters.

The complex Gaussian random process plays an important role in the study of linear estimators of random variables because the optimal estimator for a complex Gaussian probability distribution is the optimal linear estimator for all probability distributions with the equivalent first and second moments.[3] Thus the performance bound derived for the complex Gaussian case is the bound for all linear estimates of any probability distribution with the same mean and covariance.

For linear mode filters, the estimated mode coefficient vector $\hat{\mathbf{d}}$ can be written as

$$\hat{\mathbf{d}}=\mathbf{H}\mathbf{p}=\mathbf{H}\Psi\mathbf{d}+\mathbf{H}\mathbf{n},$$

where the matrix $\mathbf{H}$ represents the linear mode filter. The performance of the mode filter depends on the choice of the linear function represented by $\mathbf{H}$. The estimator error $\mathbf{e}$ is $\hat{\mathbf{d}}-\mathbf{d}=(\mathbf{H}\Psi-\mathbf{I})\mathbf{d}+\mathbf{H}\mathbf{n}$. When $\mathbf{d}$ is considered to be a nonrandom unknown parameter, the bias $B(\mathbf{d})$ of the estimator is the expected value of the error, $E\{\mathbf{e}\}=(\mathbf{H}\Psi-\mathbf{I})\mathbf{d}$. The covariance of $\hat{\mathbf{d}}$, $\mathbf{K}_{\hat{\mathbf{d}}\hat{\mathbf{d}}}=E\{\hat{\mathbf{d}}\hat{\mathbf{d}}^H\}-E\{\hat{\mathbf{d}}\}E\{\hat{\mathbf{d}}\}^H$ is

$$\mathbf{K}_{\tilde{\mathbf{d}}\tilde{\mathbf{d}}}=\mathbf{H}\mathbf{K}_{\mathbf{nn}}\mathbf{H}^H, \tag{5}$$

where $\mathbf{K}_{\mathbf{nn}}$ is the spatial covariance of the noise vector as discussed in the Introduction and the covariance of the estimator error $\mathbf{K}_{\mathbf{ee}}=\mathbf{K}_{\hat{\mathbf{d}}\hat{\mathbf{d}}}$. Thus the covariance of the mode coefficient estimate vector depends entirely on the noise process covariance $\mathbf{K}_{\mathbf{nn}}$ and the mode filter $\mathbf{H}$.

The FIM[3] provides a method of computing the bound on the covariance of the error that can be achieved by an unbiased estimator of a nonrandom parameter. The FIM is defined to be

$$\mathbf{J}_D=-E\{\nabla_{\mathbf{d}}[\nabla_{\mathbf{d}}\ln p_{\mathbf{P}|\mathbf{D}}(\mathbf{p}|\mathbf{d})]^H\},$$

where ∇_d is the gradient operator with respect to the mode coefficients $\mathbf{d}$, and $p_{\mathbf{P}|\mathbf{D}}(\mathbf{p}|\mathbf{d})$ is the conditional probability density function[21] of the observed pressure $\mathbf{p}$ given the mode coefficients $\mathbf{d}$. The variance of any estimate error e_i, $\sigma_{e_i}^2$, is bounded from below by the corresponding diagonal element of $\mathbf{J}_D^{-1}$, i.e., $\sigma_{e_i}^2\geqslant[\mathbf{J}_D^{-1}]_{ii}$. This error covariance bound is known as the Cramer–Rao lower bound (CRLB). An efficient estimator is one whose variance attains the CRLB. Because the estimate covariance equals the error covariance, this CRLB also applies to $\sigma_{\hat{d}_i}^2$, so $\sigma_{\hat{d}_i}^2\geqslant[\mathbf{J}_D^{-1}]_{ii}$. For the data model given in Eq. (2),

$$\mathbf{J}_D=\Psi^H\mathbf{K}_{\mathbf{nn}}^{-1}\Psi.$$

For the SW noise model, this yields $\mathbf{J}_D^{-1}=\sigma_n^2(\Psi^H\Psi)^{-1}$ as the bound on the error covariance. Generalizing $\mathbf{K}_{\mathbf{nn}}^{-1}$ to the pseudo-inverse[20] for the KI noise model gives a bound of

$$\mathbf{J}_D^{-1}=\mathbf{K}_{\tilde{\mathbf{d}}\tilde{\mathbf{d}}}, \tag{6}$$

where $\mathbf{K}_{\tilde{\mathbf{d}}\tilde{\mathbf{d}}}$ is defined as in Eq. (3).

When $\mathbf{d}$ is a random parameter with zero mean and covariance $\mathbf{K}_{\mathbf{dd}}$, the covariance of the error vector is

$$\mathbf{K}_{\mathbf{ee}}=\mathbf{H}\Psi\mathbf{K}_{\mathbf{dd}}\Psi^H\mathbf{H}^H+\mathbf{K}_{\mathbf{dd}}-\mathbf{H}\Psi\mathbf{K}_{\mathbf{dd}}-\mathbf{K}_{\mathbf{dd}}\Psi^H\mathbf{H}^H+\mathbf{H}\mathbf{K}_{\mathbf{nn}}\mathbf{H}^H. \tag{7}$$

The definition of the bias can be extended to the case when $\mathbf{d}$ is a random variable by letting the bias be the conditional expectation of the error, i.e., $B(\mathbf{d})=E\{\mathbf{e}|\mathbf{d}\}=(\mathbf{H}\Psi-\mathbf{I})\mathbf{d}$. The covariance $\mathbf{K}_{\mathbf{ee}}$ can then be interpreted as the sum of two components. The first component, $E_{\mathbf{d}}\{B(\mathbf{d})B(\mathbf{d})^H\}$ is independent of the noise power, but depends entirely on $\mathbf{H}$, Ψ, and $\mathbf{K}_{\mathbf{dd}}$. The notation $E_{\mathbf{d}}\{\cdot\}$ indicates taking the expectation of the argument only with respect to the subscripted variable. The second component of $\mathbf{H}\mathbf{K}_{\mathbf{nn}}\mathbf{H}^H$ is due to the observation noise and equal to the estimator covariance in Eq. (5) when $\mathbf{d}$ is an unknown deterministic quantity. By extension, we call an estimator of a random variable $\mathbf{d}$ unbiased when $B(\mathbf{d})=0$ for all $\mathbf{d}$. One pleasing feature of this definition of bias is that if an estimator $\mathbf{H}$ is unbiased for deterministic $\mathbf{d}$, it is also unbiased for random $\mathbf{d}$, and vice versa. Moreover, if $\mathbf{H}$ is an unbiased estimator, $\mathbf{K}_{\mathbf{ee}}$ is equal for both scenarios, as confirmed by comparing Eqs. (5) and (7).

The FIM can also be extended to incorporate the *a priori* information available in the probability density function (PDF) of $\mathbf{d}$ for the scenario when $\mathbf{d}$ is a random variable. This information is

$$\mathbf{J}_P=-E\{\nabla_{\mathbf{d}}[\nabla_{\mathbf{d}}\ln p_{\mathbf{D}}(\mathbf{d})]^H\},$$

where $p_{\mathbf{D}}(\mathbf{d})$ is the PDF of the mode coefficients. Assuming that $p_{\mathbf{D}}(\mathbf{d})$ is a CGRV, the resulting bound on the error covariance is

$$\mathbf{J}_T^{-1}=[\mathbf{J}_D+\mathbf{J}_P]^{-1}=[\Psi^H\mathbf{K}_{\mathbf{nn}}^{-1}\Psi+\mathbf{K}_{\mathbf{dd}}^{-1}]^{-1},$$

where the total information $\mathbf{J}_T$ is the sum of the information in the data $\mathbf{J}_D$ and the prior information $\mathbf{J}_P$. Note that one consequence of this definition is that $\mathbf{J}_T^{-1}\leqslant\mathbf{J}_D^{-1}$ in the positive definite sense, i.e., $\mathbf{x}^H\mathbf{J}_T^{-1}\mathbf{x}\leqslant\mathbf{x}^H\mathbf{J}_D^{-1}\mathbf{x}$ for all $\mathbf{x}$. Thus an

unbiased estimator of deterministic $\mathbf{d}$ which is efficient, i.e., $\mathbf{K_{ee}}=\mathbf{J}_D^{-1}$, will generally not attain the MMSE bound for estimators of a random $\mathbf{d}$ unless $\mathbf{J}_T^{-1}=\mathbf{J}_D^{-1}$. Since the CRLB guarantees that no unbiased estimator of a deterministic unknown $\mathbf{d}$ can have $\mathbf{K_{ee}}<\mathbf{J}_D^{-1}$, and $\mathbf{K_{ee}}$ is the same for an unbiased estimator regardless of whether $\mathbf{d}$ is random or deterministic, this implies that the estimator attaining the MMSE for a random $\mathbf{d}$ cannot be unbiased unless $\mathbf{J}_T^{-1}=\mathbf{J}_D^{-1}$.

It is instructive to examine $\mathbf{J}_T^{-1}$ for the case when the mode coefficients are uncorrelated, as will be assumed in the subsequent simulations. If $\mathbf{K_{dd}}=\sigma_d^2\mathbf{I}$, the bound on the error covariance for the SW noise model is

$$\begin{aligned}\mathbf{J}_T^{-1}&=[\sigma_n^{-2}\Psi^H\Psi+\sigma_d^{-2}\mathbf{I}]^{-1}\\&=\sigma_d^2\mathbf{I}-\sigma_d^2\Psi^H[\sigma_n^2\mathbf{I}+\sigma_d^2\Psi\Psi^H]^{-1}\Psi\sigma_d^2,\end{aligned}$$

where the second step follows from the matrix inversion lemma. This expression can be put into the following form:

$$\mathbf{J}_T^{-1}=\mathbf{V}\begin{bmatrix}\sigma_d^2-\dfrac{(\sigma_d^2)^2}{\sigma_d^2+\left(\dfrac{\sigma_n^2}{\sigma_{\Psi 1}^2}\right)} & 0 & \cdots & 0\\ 0 & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & \sigma_d^2-\dfrac{(\sigma_d^2)^2}{\sigma_d^2+\left(\dfrac{\sigma_n^2}{\sigma_{\Psi M}^2}\right)}\end{bmatrix}\mathbf{V}^H, \tag{8}$$

where $\sigma_{\Psi i}$ is the ith singular value of Ψ as defined in Eq. (4). For the KI noise model,

$$\begin{aligned}\mathbf{J}_T^{-1}&=[\mathbf{K_{dd}}^{-1}+\mathbf{K}_{\tilde{\mathbf{d}}\tilde{\mathbf{d}}}^{-1}]^{-1}\\&=\begin{bmatrix}\sigma_d^2-\dfrac{(\sigma_d^2)^2}{\sigma_d^2+\sigma_{\tilde{d}_1}^2} & 0 & \cdots & 0\\ 0 & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & \sigma_d^2-\dfrac{(\sigma_d^2)^2}{\sigma_d^2+\sigma_{\tilde{d}_M}^2}\end{bmatrix},\end{aligned} \tag{9}$$

is the bound on the error covariance. These bounds form a useful basis for comparison of the different mode filters. Moreover, the similar structure of Eqs. (8) and (9) provides insight into the nature of the CRLB for mode filtering. Both noise models have diagonal matrices with terms of the form $\sigma_d^2-(\sigma_d^2)^2/(\sigma_d^2+a^2)$. This is the classic form of the variance reduction for a linear estimate of a random variable with variance σ_d^2 in uncorrelated noise with variance a^2. The variance of the estimate is the *a priori* variance of the variable minus the information gained by the observation. The spatially white case [Eq. (8)] includes the unitary matrices $\mathbf{V}$ and $\mathbf{V}^H$ to transform the estimates from the basis where the estimates are uncorrelated to the physical basis of the acoustic modes. These matrices are not necessary for the KI noise model, since the noise is uncorrelated in the basis for the acoustic mode by definition in this model.

The uncorrelated mode coefficient model is not proposed as a realistic model of ocean propagation: For many ocean environments the mode coefficients will be correlated. However, any covariance matrix $\mathbf{K_{dd}}$ may be diagonalized by an appropriate similarity transform.[20] Thus there is always some basis isomorphic with the acoustic modes in which the estimation problem is uncorrelated. Consequently, there always exists some basis in which the structure of the mode estimators matches those given above, modulo an additional unitary matrix implementing the diagonalizing similarity transform. Thus the intuition gained by studying the case above when the standard acoustic modes are uncorrelated transfers easily to scenarios with more realistic covariance matrices.

One issue examined for all the linear mode filters in this paper is their performance under conditions of poor spatial sampling. We distinguish between two types of poor sampling. In the first, undersampling, the number of hydrophones N is fewer than the number of modes to be estimated M. In this scenario, Eq. (2) is an underdetermined least squares problem, and thus lacks a unique solution. While it is possible to find the minimum norm solution to this equation, there is no reason to believe that the modes propagating are the set giving the minimum norm for $\mathbf{d}$. For this reason, it is crucial to insure that the number of hydrophones in the array exceeds the number of modes that can reasonably be expected to be observed at the array for the frequency of propagation.

The second kind of poor sampling occurs when the number of hydrophones exceeds the number of propagating modes, but the locations of the hydrophones are such that they poorly sample the mode shapes. We refer to this scenario as poorly conditioned sampling. In poorly conditioned sampling, some of the singular values of Ψ grow disparately small compared to $\sigma_{\Psi 1}$, resulting in a large condition number $\sigma_{\Psi 1}/\sigma_{\Psi M}$ for the matrix Ψ.[20] Poorly conditioned sampling has varying consequences for different mode filters, as we show in the following sections.

A. Sampled mode shapes mode filter

One common choice for $\mathbf{H}$ in mode propagation experiments is Ψ^H, the sampled mode shape (SMS) filter. The motivation for this choice is that the mode functions are orthogonal when considered as continuous functions of depth. The SMS filter may also be interpreted as a spatially matched filter. As such, it is optimal for detecting a single mode in spatially white noise.[3] However, spatially sampling the modes with the hydrophones does not in general preserve the orthogonality, i.e., $\Psi^H\Psi\neq\mathbf{I}$. This lack of orthogonality appears as contamination or cross-talk when estimating several modes simultaneously. Assuming the mode coefficient vector $\mathbf{d}$ is a nonrandom, unknown quantity to be estimated in the presence of random noise $\mathbf{n}$, we can characterize the performance of the estimator $\mathbf{H}=\Psi^H$ in terms of its bias and covariance.[3] The bias of the SMS filter is $B(\mathbf{d})=(\mathbf{H}\Psi-\mathbf{I})\mathbf{d}$, so the lack of orthogonality of Ψ can introduce a significant bias into the mode estimate. For the SMS

filter, the rows of $\mathbf{H}$ should be normalized such that the diagonal of $(\mathbf{H}\Psi-\mathbf{I})$ is zero, and the remaining bias is due to cross-talk between the modes and not normalization. The correct normalization is $\mathbf{H}=\mathbf{B}\Psi^H$, where $\mathbf{B}$ is the $M\times M$ diagonal matrix diag$(\|\Psi_m\|^{-2})$, and Ψ_m is the mth column of Ψ, i.e., the mth mode shape sampled at the hydrophone locations.

Theoretically, as more hydrophones are added to sample the water column more finely in depth, the sampled modes matrix Ψ becomes arbitrarily close to orthogonal, making the effect of the bias negligible. This is unrealistic in practice, as there are shallow water scenarios where the ocean bottom contains significant energy. It is impractical to deploy a vertical array of hydrophones spanning the entire sediment layer. Thus even in the limiting case of a continuous array of hydrophones spanning the water column, a bias may still exist due to the unsampled pressure field in the bottom.[22] Many experiments including Ferris,[1] and Clay and Huang[23] used the SMS filter, assuming the samples of the orthogonal mode functions are themselves orthogonal without examining the potential bias. However, Tindle *et al.*[24] and Gazanhes and Garnier[25] both used the pseudo-inverse mode filter and examined the cross-talk introduced by sampling to verify that the bias was negligible for the purposes of their experiments.

When the array geometry results in poorly conditioned sampling, the bias can grow so large as to make it impossible to obtain reliable estimates of $\mathbf{d}$. Rewriting the bias $B(\mathbf{d})$ with the SVD for Ψ yields

$$B(\mathbf{d})=\mathbf{V}_\Psi\left(\begin{bmatrix}\sigma_{\Psi 1}^2 & & 0\\ & \ddots & \\ 0 & & \sigma_{\Psi M}^2\end{bmatrix}-\mathbf{I}\right)\mathbf{V}_\Psi^H\mathbf{d}. \tag{10}$$

The presence of the identity matrix in this equation indicates that decreasing any of the $\sigma_{\Psi m}$s below one increases the bias of the estimator. In the extreme when one of the singular values is zero, the component of the bias in that direction equals the projection of the mode coefficient vector in that direction.

The covariance of the SMS filter can be found for both the SW and KI noise models. For the SW case,

$$K_{\hat{\mathbf{d}}\hat{\mathbf{d}}}=\sigma_n^2\Psi^H\Psi=\sigma_n^2\sum_{m=1}^{M}\sigma_{\Psi m}^2\mathbf{v}_{\Psi m}\mathbf{v}_{\Psi m}^H, \tag{11}$$

where the vectors $\mathbf{v}_{\Psi m}$ are the columns of $\mathbf{V}_\Psi$. The KI noise model gives

$$\begin{aligned}K_{\hat{\mathbf{d}}\hat{\mathbf{d}}}&=\Psi^H\Psi\mathbf{K}_{\tilde{\mathbf{d}}\tilde{\mathbf{d}}}\Psi^H\Psi\\&=\left(\sum_{m=1}^{M}\sigma_{\Psi m}^2\mathbf{v}_{\Psi m}\mathbf{v}_{\Psi m}^H\right)\begin{bmatrix}\sigma_{\tilde{d}_1}^2 & 0 & \cdots & 0\\ 0 & \sigma_{\tilde{d}_2}^2 & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & \sigma_{\tilde{d}_M}^2\end{bmatrix}\\&\quad\times\left(\sum_{m=1}^{M}\sigma_{\Psi m}^2\mathbf{v}_{\Psi m}\mathbf{v}_{\Psi m}^H\right).\end{aligned} \tag{12}$$

For both noise models, it can be seen that while decreasing $\sigma_{\Psi M}$ may initially decrease the covariance slightly, once $\sigma_{\Psi M}$ is insignificant compared to the other singular values of Ψ, further decreases in $\sigma_{\Psi M}$ do not change $\mathbf{K}_{\hat{\mathbf{d}}\hat{\mathbf{d}}}$ significantly. This is true to a lesser extent for other singular values, but we focus on $\sigma_{\Psi M}$, which is the smallest singular value by definition from Eq. (4). This limit on the deterioration of the performance of the SMS filter is intuitively sensible because the norms of the rows of $\mathbf{H}=\Psi^H$ are limited by the maximum amplitudes of the modes $\Psi_m(z)$ no matter how poorly conditioned the sampling is. Consequently, the error covariance in Eq. (5) is limited in its growth. As the sampling grows even more poorly conditioned, the covariance reaches its upper limit, and the estimate $\hat{\mathbf{d}}$ is mainly corrupted by the bias shown in Eq. (10). For both noise models, if the sampling is such that the covariance is still a significant factor, the covariance will increase as the noise powers (either σ_n^2 or $\sigma_{d_i}^2$) increase. This is in contrast to the bias which is independent of the noise power.

At the other extreme, consider the scenario where $N\to\infty$ for an array spanning the entire water column. In this highly oversampled case, all the singular values $\sigma_{\Psi m}$ approach 1, so

$$\sum_{m=1}^{M}\sigma_{\Psi m}^2\mathbf{v}_{\Psi m}\mathbf{v}_{\Psi m}^H\to\mathbf{V}_\Psi\mathbf{V}_\Psi^H=\mathbf{I}$$

and Eqs. (11) and (12) simplify to $K_{\hat{\mathbf{d}}\hat{\mathbf{d}}}=\sigma_n^2 I$ and $K_{\hat{\mathbf{d}}\hat{\mathbf{d}}}=K_{\tilde{\mathbf{d}}\tilde{\mathbf{d}}}$, respectively. The bias also becomes negligible, with only a small contribution remaining due to the unsampled energy in the bottom sediments. Thus when the array oversamples the mode shapes, using $\mathbf{H}=\Psi^H$ can give mode estimates with only a small bias and a covariance reflecting the underlying noise process of the observations.

The covariance for the SMS filter does not equal the CRLB for either noise model. For some Ψ it is possible that the filter will have lower variances on some mode coefficients. The bound given by the FIM applies only to unbiased estimators, and the SMS filter is not unbiased except in the limit when $\Psi^H\Psi\to\mathbf{I}$. In this limit, the SMS filter achieves the bound for both noise models.

When $\mathbf{d}$ is considered to be a random vector, the error covariance for the SMS filter is

$$\begin{aligned}\mathbf{K}_{\mathbf{ee}}=&\Psi^H\Psi\mathbf{K}_{\mathbf{dd}}\Psi^H\Psi+\mathbf{K}_{\mathbf{dd}}-\Psi^H\Psi\mathbf{K}_{\mathbf{dd}}-\mathbf{K}_{\mathbf{dd}}\Psi^H\Psi\\&+\Psi^H\mathbf{K}_{\mathbf{nn}}\Psi.\end{aligned}$$

This does not attain the MMSE bound. It is a straightforward extension of proofs given in Ref. 3 to show that a necessary and sufficient condition for $\mathbf{K}_{\mathbf{ee}}=\mathbf{J}_T^{-1}$ is that $\mathbf{e}$ can be written in the form

$$\mathbf{e}=\mathbf{A}[\nabla_{\mathbf{d}}\ln p_{\mathbf{P},\mathbf{D}}(\mathbf{p},\mathbf{d})], \tag{13}$$

where $p_{\mathbf{P},\mathbf{D}}(\mathbf{p},\mathbf{d})$ is the joint PDF of the mode coefficients $\mathbf{d}$ and observed pressure $\mathbf{p}$, and $\mathbf{A}$ is a constant matrix independent of $\mathbf{p}$ or $\mathbf{d}$. From Eq. (2), it can be shown that for the CGRV case

$$\nabla_{\mathbf{d}}\ln p_{\mathbf{P},\mathbf{D}}(\mathbf{p},\mathbf{d})=\Psi^H\mathbf{K}_{\mathbf{nn}}^{-1}\mathbf{n}-\mathbf{K}_{\mathbf{dd}}^{-1}\mathbf{d}.$$

The error **e** for the SMS filter cannot be put in the form of Eq. (13) in general, and thus the SMS does not achieve the bound on the error covariance or MSE.

B. Pseudo-inverse mode filter

The pseudo-inverse (PI) mode filter results from choosing $\hat{\mathbf{d}}$ to minimize the squared error between $\Psi\hat{\mathbf{d}}$ and **p**. Intuitively, the PI filter can be thought of as removing all cross-talk between mode estimates, but at a cost of higher sensitivity to noise. Tindle *et al.*[4] appears to be the first reference in the ocean acoustics literature to formulate the mode estimation problem in this least squares sense. The resulting mode filter $\mathbf{H}=(\Psi^H\Psi)^{-1}\Psi^H$, denoted $\Psi^\dagger$, is called the pseudo-inverse or Penrose–Moore inverse of Ψ.[20,26] This name results from the fact $\Psi^\dagger\Psi=\mathbf{I}$. If the mode coefficient vector **d** is considered to be a nonrandom but unknown vector, the PI mode filter is unbiased for both the SW and KI noise models.

The covariance of the estimated mode coefficient vector depends on the noise model. For the SW noise model, Eq. (5) yields

$$K_{\tilde{\mathbf{d}}\tilde{\mathbf{d}}}=\sigma_n^2\mathbf{V}_\Psi\Sigma_\Psi^\dagger(\Sigma_\Psi^\dagger)^H\mathbf{V}_\Psi^H = \sigma_n^2\sum_{m=1}^{M}\sigma_{\Psi m}^{-2}\mathbf{v}_{\Psi m}\mathbf{v}_{\Psi m}^H. \tag{14}$$

If the array gives poorly conditioned sampling of the mode shapes, some of the singular values approach zero and consequently the corresponding $\sigma_{\Psi m}^{-2}$ terms in Eq. (14) dominate the sum, giving a very large covariance. Alternatively, if the array grossly oversamples the channel, the singular values approach one, and $\mathbf{K}_{\hat{\mathbf{d}}\hat{\mathbf{d}}}$ approaches $\sigma_n^2\mathbf{I}$. For the SW Gaussian noise case, the PI mode filter can be shown to be the maximum-likelihood (ML) estimator, as well as efficient.

If the KI noise model is substituted into Eq. (5), the estimator covariance is

$$\mathbf{K}_{\hat{\mathbf{d}}\hat{\mathbf{d}}}=\Psi^\dagger\Psi K_{\tilde{\mathbf{d}}\tilde{\mathbf{d}}}\Psi^H\Psi^{\dagger H}=\mathbf{K}_{\tilde{\mathbf{d}}\tilde{\mathbf{d}}}, \tag{15}$$

which is intuitively sensible, as $K_{\tilde{\mathbf{d}}\tilde{\mathbf{d}}}$ is the covariance of the noise process as it is coupled into the channel by the modes. Theoretically, this covariance is unchanged by reductions in the array aperture. Practically, the mode filter $\Psi^\dagger$ is usually based on an estimate of Ψ computed by numerical integration of an observed or estimated sound speed profile. As the array aperture decreases and the singular values of the true Ψ grow smaller, the PI filter may become very sensitive to errors between the Ψ obtained by numerical integration and the actual Ψ of the ocean channel. These errors can introduce a bias and increase the covariance above $\mathbf{K}_{\tilde{\mathbf{d}}\tilde{\mathbf{d}}}$.

The PI filter is also the ML estimator for the Kuperman–Ingenito noise model. This conclusion is not surprising since the definition of the noise model assumes **n** is in the range of Ψ. The existence of the ML estimate depends on the Kuperman–Ingenito model perfectly describing the noise process, since if **n** contains any component in the orthogonal complement to the range of Ψ, the conditional probability density $p_{\mathbf{P}|\mathbf{D}}(\mathbf{p}|\mathbf{d})=0$ for any **d**, and the ML estimate is meaningless since no set of mode coefficients **d** could have produced the observed signal. Equation (6) demonstrated that the CRLB for the KI noise model is Eq. (15). The PI mode filter attains this bound on the variance, and is an efficient estimator for the KI noise model. If **d** is considered to be a random parameter, the error covariance bound is reduced and the unbiased PI filter no longer achieves the bound, as noted earlier. This can be confirmed by observing that the error signal $\mathbf{e}=\Psi^\dagger\mathbf{n}$ does not have the form required by Eq. (13). Thus the error covariances of the PI mode filter, $\sigma_n^2(\Psi^H\Psi)^{-1}$ for SW noise and $\mathbf{K}_{\tilde{\mathbf{d}}\tilde{\mathbf{d}}}$ for the KI noise model, do not meet the bound specified by $\mathbf{J}_T^{-1}$. In fact, the error variance for the PI filter is independent of the actual modal energy propagating. This confirms the earlier statement that the PI mode filter removes all mode cross-talk, so all remaining error is due to noise.

C. Diagonal weighting

The diagonally weighted (DW) mode filter attempts to compensate for situations when the array yields poorly conditioned sampling of the modes. As noted above, such sampling causes one or more of the singular values of Ψ to be very small. As a result of this sampling $\Psi^H\Psi$ is singular or nearly singular, and the computation of the inverse of this matrix becomes numerically sensitive. One method of compensating for this sensitivity is to modify the error function being minimized to include a term proportional to the magnitude squared of the estimated mode coefficient vector $\hat{\mathbf{d}}$.[27,28] The quantity to be minimized is then

$$e=\|\mathbf{p}-\Psi\hat{\mathbf{d}}\|^2+\beta\|\hat{\mathbf{d}}\|^2,$$

where β is a scale factor indicative of the relative importance of the two terms in the error expression. The estimator minimizing this quantity is

$$\hat{\mathbf{d}}_{\mathrm{DW}}=(\Psi^H\Psi+\beta\mathbf{I})^{-1}\Psi^H\mathbf{p}.$$

This expression is very similar to the PI filter, except for a small diagonal matrix $\beta\mathbf{I}$ which has been added to $\Psi^H\Psi$ before inversion to alleviate conditioning problems. The $\beta\mathbf{I}$ term is often referred to as the white noise sensitivity term. The addition of this term places a lower bound of β on the singular values of $(\Psi^H\Psi+\beta\mathbf{I})$. For the SW noise model, this limits the covariance of the estimator shown in Eq. (14), since no $\sigma_{\Psi m}^{-2}$ for the diagonally weighted inverse can exceed β^{-2}. For this reason, this approach is often referred to as diagonal loading or weighting. While this estimator does not possess many of the nice theoretical properties of the pseudo-inverse mode filter, it is computationally more stable for poorly conditioned sampling. As discussed for the undetermined mode filtering problem, there are many propagation environments where there is no reason that the propagating modes should minimize $\|\mathbf{d}\|^2$. Consequently, when choosing β for diagonal weighting, there is a compromise between minimizing the filter's numerical sensitivity and overemphasizing the somewhat artificial criterion of minimizing $\|\mathbf{d}\|^2$.

D. Maximum *a posteriori* mode filters

The immediate motivation for the MAP mode filter is the fact that none of the mode filters examined so far are

efficient for the scenario when $\mathbf{d}$ is considered to be a random variable. The MAP mode filter chooses $\hat{\mathbf{d}}_{\mathrm{MAP}}$ to maximize the probability of the conditional probability density function for $\mathbf{d}$ conditioned on the observed pressure $\mathbf{p}$, i.e., $p_{\mathbf{D}|\mathbf{P}}(\mathbf{d}|\mathbf{p})$. When both $\mathbf{d}$ and $\mathbf{n}$ are CGRV's, the MAP filter is equivalent to the MMSE filter.[3]

For the case when the mode coefficients are well modeled by a CGRV with zero mean and covariance $\mathbf{K_{dd}}$ and the noise is also well modeled by a zero-mean CGRV with covariance $\mathbf{K_{nn}}$ and uncorrelated with $\mathbf{d}$, the MAP mode filter can be solved in closed form. Specifically,

$$\hat{\mathbf{d}}_{\mathrm{MAP}}=\mathbf{K_{xx}}\Psi^{H}\mathbf{K_{nn}}^{-1}\mathbf{p}, \tag{16}$$

where

$$\mathbf{K_{xx}}^{-1}=\mathbf{K_{dd}}^{-1}+\Psi^{H}\mathbf{K_{nn}}^{-1}\Psi.$$

The error signal $\mathbf{e}$ can be shown to be $\mathbf{K_{xx}}(\Psi^{H}\mathbf{K_{nn}}^{-1}\mathbf{n}-\mathbf{K_{dd}}^{-1}\mathbf{d})$, which satisfies Eq. (13). Consequently, the MAP filter achieves the bound on the error covariance when $\mathbf{d}$ is a random parameter satisfying the assumptions stated above.

Some insight into the performance of this mode filter may be gained by considering the somewhat unrealistic case when the modes are independent and identically distributed, i.e., $\mathbf{K_{dd}}=\sigma_d^2\mathbf{I}$, and the noise is spatially white with $\mathbf{K_{nn}}=\sigma_n^2\mathbf{I}$. Assuming there are more hydrophones than modes $(N>M)$, Eq. (16) reduces to

$$\hat{\mathbf{d}}_{\mathrm{MAP}}=\mathbf{V}_{\Psi}\left[\begin{array}{cccc|c}\frac{\sigma_d^2\sigma_{\Psi 1}}{\sigma_d^2\sigma_{\Psi 1}^2+\sigma_n^2} & 0 & \cdots & 0 & \\ 0 & \frac{\sigma_d^2\sigma_{\Psi 2}}{\sigma_d^2\sigma_{\Psi 2}^2+\sigma_n^2} & \ddots & \vdots & 0 \\ \vdots & \ddots & \ddots & 0 & \\ 0 & \cdots & 0 & \frac{\sigma_d^2\sigma_{\Psi M}}{\sigma_d^2\sigma_{\Psi M}^2+\sigma_n^2} & \end{array}\right]\mathbf{U}_{\Psi}^{H}\mathbf{p}, \tag{17}$$

where $\sigma_{\Psi i}$ is the ith singular value of Ψ as defined in Eq. (4). Equation (17) has an appealing interpretation as a generalization of the discrete spatial Wiener filter (DSWF).[5] Multiplying $\mathbf{p}$ by $\mathbf{U}_{\Psi}^{H}$ rotates the problem into the coordinate frame where the spatial components are uncorrelated. Each component is then weighted by the Wiener gain for the ratio of the mode power to the noise power for that component $\sigma_{\Psi m}^2\sigma_d^2/(\sigma_{\Psi m}^2\sigma_d^2+\sigma_n^2)$. These estimates of the components are then multiplied by the inverse singular values $\sigma_{\Psi m}^{-1}$ before being transformed from the uncorrelated basis into mode coefficients by $\mathbf{V}_{\Psi}$. For the case when all the singular values are 1, $\mathbf{V}_{\Psi}=\mathbf{I}$ and $\mathbf{U}_{\Psi}$ is the appropriate set of samples of complex exponentials, Eq. (17) reduces exactly to the DSWF.

The mode filter proposed by Yang[9] can be interpreted as an asymptotic result of Eq. (17). For practical reasons, Yang proposed setting very small eigenvalues of $\Psi^{H}\Psi$ to zero before inverting this matrix in the process of computing the PI mode filter. Yang's motivation for this modification was rough *a priori* knowledge of the mode coefficients expected and considerations of numerical stability. Consider Eq. (17) when the array gives a poorly conditioned sampling of the channel. Some of the $\sigma_{\Psi i}$ grow small compared to σ_d and σ_n, causing some of the diagonal terms to go to zero. In the limit, this results in the same mode filter proposed by Yang. Based on this argument, the MAP mode filter provides a theoretical justification for Yang's *ad hoc* mode filter as the asymptotic case of a poorly sampled MAP mode filter.

It is also instructive to consider the MAP mode filter for the case when $\mathbf{K_{dd}}=\mathrm{diag}(\sigma_{d_1}^2,\ldots,\sigma_{d_M}^2)$ and $\mathbf{K_{nn}}$ is given by the KI noise model and these Gaussian random processes are considered to be independent. Under these conditions, the MAP mode filter becomes

$$\begin{aligned}\hat{\mathbf{d}}_{\mathrm{MAP}}&=[\mathbf{K_{dd}}^{-1}+\mathbf{K}_{\tilde{\tilde{\mathbf{d}}}\tilde{\tilde{\mathbf{d}}}}^{-1}]^{-1}\mathbf{K}_{\tilde{\tilde{\mathbf{d}}}\tilde{\tilde{\mathbf{d}}}}^{-1}\Psi^{\dagger}\mathbf{p}\\&=\left[\begin{array}{cccc}\frac{\sigma_{d_1}^2}{\sigma_{d_1}^2+\sigma_{\tilde{d}_1}^2} & 0 & \cdots & 0\\ 0 & \frac{\sigma_{d_2}^2}{\sigma_{d_2}^2+\sigma_{\tilde{d}_2}^2} & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & \frac{\sigma_{d_M}^2}{\sigma_{d_M}^2+\sigma_{\tilde{d}_M}^2}\end{array}\right]\Psi^{\dagger}\mathbf{p}.\end{aligned} \tag{18}$$

Intuitively, this is sensible because the pressure components due to both the noise process and the modes fall entirely within the range of Ψ if the KI noise model and mode propagation model are accurate. If this is the case, $\mathbf{p}$ contains no projection in the orthogonal complement of Ψ and thus no information is lost by transforming the pressure vector $\mathbf{p}$ back into mode coordinates by $\Psi^{\dagger}\mathbf{p}$. Seen another way, $\Psi^{\dagger}$ is the spatial Karhunen–Loeve transform, since it decorrelates the observed modes and noise processes so that they are spatially white, i.e., the covariance of $\Psi^{\dagger}\mathbf{p}$ is diagonal. Unlike the SW noise model, the physical basis of interest (mode space) coincides with the mathematical basis in which the underlying processes are uncorrelated. Once the problem has been whitened this way, the standard Wiener gains shown in Eq. (18) yield the MMSE estimate of $\mathbf{d}$. Given that the Gaussian density is symmetric about a maximum at its mean, the MMSE solution is equivalent to the MAP solution.[3]

Under many conditions, the MAP mode filter matches or exceeds the performance of either the SMS or PI mode filters. As discussed in the previous sections, the variance of the SMS filter is relatively insensitive to decreases in $\sigma_{\Psi M}$ as the sampling becomes poorly conditioned. Contrastingly, the PI filter's covariance increases rapidly as the singular values approach zero. Consequently, the PI mode filter generally performs better when the array samples the mode shapes adequately, but as the array aperture decreases and the sampling becomes poorly conditioned, the PI filter's performance deteriorates such that the sampled mode shape filter is preferable. One desirable feature of the MAP mode filter is that it performs like the PI mode filter when the sampling is well conditioned, like the SMS mode filter when the sampling is poorly conditioned, and in between these regimes the MAP mode filter transitions smoothly with a performance exceeding that of either mode filter.

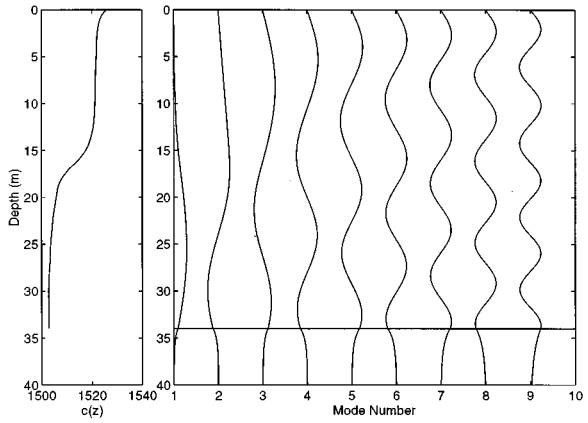


FIG. 1. Shallow-water sound-speed profile and propagating modes at 200 Hz.

II. SHALLOW WATER SIMULATIONS

This section presents the results of simulations using both the KI and SW noise models in a typical shallow-water (34-m depth) environment measured on the North Continental Shelf of North America around 41 °N 71 °W. All of the simulations use 200 Hz as the propagation frequency. For this frequency and water depth, the channel supports nine trapped modes. Figure 1 shows the observed downwardly refracting sound-speed profile, along with the nine trapped modes. The simulations use a series of vertical receiving arrays whose apertures vary between spanning the full depth (34 m) and the bottom half of the water column ($z=17-34$ m), and all of which have 19 hydrophones. Each mode filtering algorithm is evaluated at three different noise levels for both noise models. The noise levels of 0, 20, and 40 dB SNR refer to the ratio of the power in the propagating modes to the power in the noise at the hydrophones for the full aperture array, i.e.,

$$\mathrm{SNR}=10\log_{10}\left(\frac{E\{\|\Psi\mathbf{d}\|^2\}}{E\{\|\mathbf{n}\|^2\}}\right).$$

By decreasing the aperture, we are able to examine the performance of the algorithms as the mode filtering problem transitions from well conditioned [$\mathrm{cond}(\Psi)=1.04$ for the fully spanning array] to poorly conditioned [$\mathrm{cond}(\Psi)=10^5$ for the half spanning array].

For each noise level and aperture, 500 trials were run using independent choices for the mode coefficients and noise. In each trial, the mode coefficient vector $\mathbf{d}$ was chosen as a CGRV with zero mean and covariance $\mathbf{K_{dd}}=\mathbf{I}$, while the noise vector $\mathbf{n}$ was also modeled as a CGRV with statistics appropriate to the noise model under evaluation. The observed pressure field $\mathbf{p}$ was determined from $\mathbf{d}$ and $\mathbf{n}$ using Eq. (2), and then used as the input to the mode filtering algorithms. The total squared error $\|\hat{\mathbf{d}}-\mathbf{d}\|^2$ was computed for each mode filter at each trial, and then averaged over all trials to obtain the mean total squared error (MTSE) for each mode filter. Note that for $M=9$, choosing $\hat{\mathbf{d}}=\mathbf{0}$ regardless of $\mathbf{p}$ results in a $\mathrm{MTSE}=10\log_{10}9=9.5$ dB. This gives a rough bound on the worse case performance. Thus any estimator with MTSE greater than 9.5 dB exhibits worse performance than the *a priori* mean estimator $\hat{\mathbf{d}}=0$.

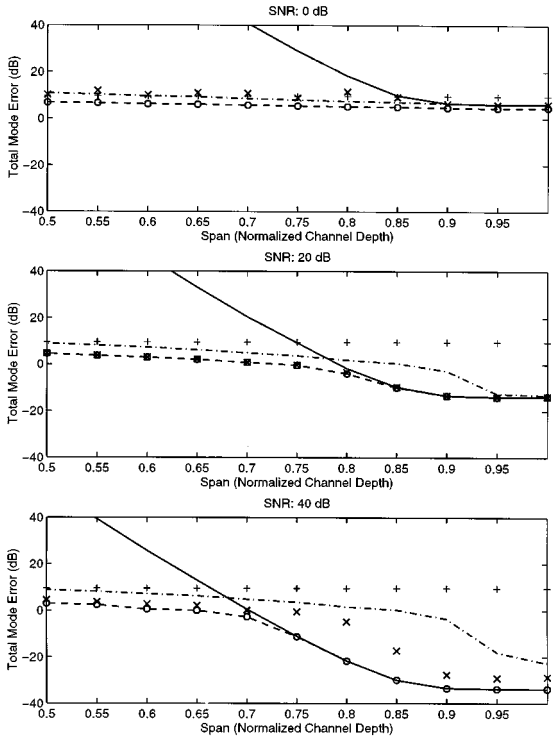


FIG. 2. Comparison of the performance of common mode filtering algorithms in spatially white noise at three different SNRs. The algorithms compared are the pseudo-inverse (solid), sampled mode shape (dash-dot), diagonal-weighting (×'s), MAP (dashed), and mismatched MAP (circles). The crosses ('+') mark the rough bound on worst case performance ($10\log_{10}9=9.5$ HF) that results from ignoring the observed data.

Figure 2 plots the performance of the algorithms for the SW noise model. In Fig. 2, each of the subplots compares the MTSE as a function of array span for each mode filtering algorithm at different noise levels. As predicted in Sec. I, the PI mode filter (solid line) does well for well conditioned sampling (full span), but the error increases dramatically as aperture shrinks and the condition number rises. As derived in Eq. (14), changing the noise level (σ_n^2) does not change the shape of the curve, but only its offset. Even at 40-dB SNR, the PI filter performs worse than choosing $\hat{\mathbf{d}}=\mathbf{0}$ (''+'' signs at 9.5 dB) before the aperture has decreased to 65% of the water column. The SMS filter exhibits reasonably good performance for the 20- and 40-dB SNR cases when the array spans the full water column. However, the performance of this algorithm deteriorates more quickly than the PI as the aperture is reduced. The bias introduced in the estimate due to cross-talk among the modes initially grows more quickly than the MTSE of the noise boosted by the PI filter. As the aperture continues to decrease, the PI mode filter eventually overtakes the SMS filter. At small apertures, the MTSE for the SMS filter is dominated by the bias of the estimator, Eq. (10), since decreasing σ_n^2 by 20 dB does not improve the performance commensurately. Equation (10) also predicts the relative insensitivity of the MTSE of the SMS filter to decreases in $\sigma_{\Psi M}$ after an initial deterioration. This effect is visible in the abrupt initial increase in the MTSE as the normalize aperture decreases to about 0.9, due to the decrease in

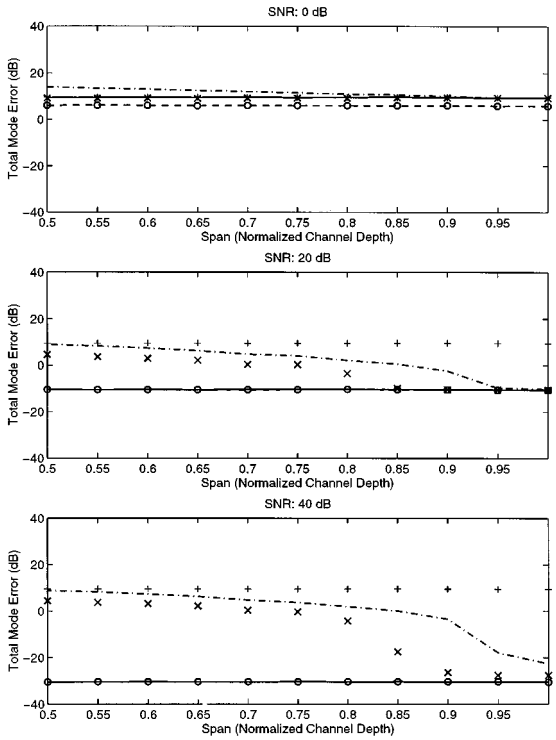


FIG. 3. Comparison of the performance of common mode filtering algorithms for the Kuperman–Ingenito noise model at three different SNRs. The algorithms compared are the pseudo-inverse (solid), sampled mode shape (dash-dot), diagonal-weighting ($\times$'s) MAP (dashed), and mismatched MAP (circles). The crosses ('+') mark the rough bound on worst case performance ($10 \log_{10} 9 = 9.5$ dB) that results from ignoring the observed data.

$\sigma_{\Psi M}$. The MTSE curve grows much more slowly as the aperture decreases further, indicating that the further decrease of additional singular values does not cause the performance to deteriorate as quickly as the initial decrease of $\sigma_{\Psi M}$.

At modest (20-dB) SNR, the DW ($\times$'s) and MAP (dashed line) mode filter follow the PI filter at full aperture and transition gracefully to behavior similar to, but still better than, the sampled mode shape filter as the conditioning of the sampling becomes poor. For the simulations shown here, β was chosen so that the condition number of $\Psi^H\Psi + \beta\mathbf{I}$ never exceeded 200. The advantage of the MAP mode filter over the DW becomes clearer at low and high SNR. For the former, the DW transitions from the PI to the variance of the mode process at 9.5 dB, while the MAP stays a few dB better than this worse case. Admittedly, this slightly better than worse case performance is not in itself impressive at 0-dB SNR, but the MAP filter does appear to match or exceed the best performance of the other filters for each aperture and SNR. The high SNR experiments also reveal that the DW filter does not match the PI or MAP filters at full aperture. As the aperture decreases, the MAP algorithm does not track the PI filter, but smoothly transitions to performance better than either the DW or SMS filters. Thus at modest SNRs, there may be little difference between the MAP and DW filters, but the MAP filter is clearly superior at either extreme of high or low SNR.

A common criticism of MAP algorithms is that they assume prior knowledge of the statistics of the unknown process to be estimated, in this case $\mathbf{K_{dd}}$. This assumption is not always realistic, as the knowledge of the covariance matrix may not be available. As part of these simulations, the sensitivity of the MAP filter to mismatch was evaluated. Specifically, the performance of the algorithm was evaluated for $\mathbf{K_{dd}} = 1.2\mathbf{I}$, a 20% mismatch in the variance of the process. The results of the simulations using this erroneous value for $\mathbf{K_{dd}}$ are shown as circles on top of the dashed line for the MAP filter in Fig. 2. Even at 0-dB SNR, when the mismatch is most significant since σ_n^2 is largest, the difference is almost imperceptible, a fraction of a dB. At higher SNR, there is no practical difference in the performance of the filters at all. Thus for this application it appears even rough estimates of the power in the process to be estimated are sufficient to allow the MAP filter to outperform the others.

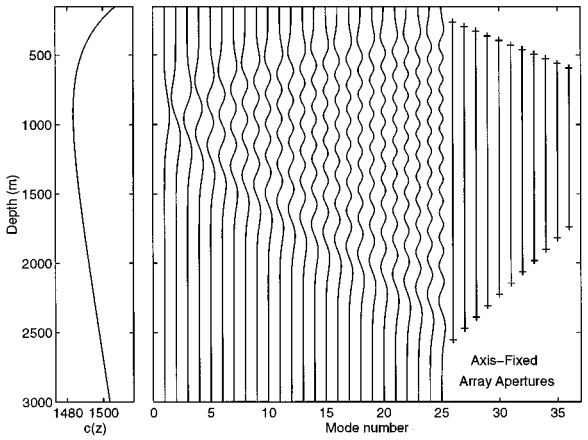


FIG. 4. Deep water sound-speed profile, first 25 propagating modes at 75 Hz, and the axis-fixed array apertures. The sound speed is a canonical Munk profile with an axis depth and speed of 923 m and 1483.5 m/s, respectively. Note that the plot shows only the upper 3000 m of the 5426-m deep waveguide.

The second set of simulations, whose results are depicted in Fig. 3, compare the mode filters in the same shallow water environment except the KI noise model is used instead of the SW noise model. Again, the noise level is determined by the ratio of the power in the mode field to the power in the noise field observed by the hydrophones for the fully spanning array. The filters are represented by the same line types as in Fig. 2: PI (solid), SMS (dash-dot), DW ($\times$'s), MAP (dashed), mismatched MAP (circles). As expected from Eqs. (10) and (12), the performance of the SMS filter is consistent with the SW noise scenario. Once again we see an initial increase of MTSE as the aperture decreases followed by a leveling of this curve at still smaller apertures. As observed for the SW noise case, this performance is due to an initial decrease in $\sigma_{\Psi M}$, and then domination by the bias term at smaller apertures.

The performances of the MAP and PI filters are closely linked in the KI noise model simulations. As shown in Eq. (18), the PI filter can be interpreted as the whitening preprocessor for the MAP filter. Consequently, the differences between the solid and dashed lines in Fig. 3 are due to the Wiener gains $[\sigma_{di}^2/(\sigma_{di}^2 + \sigma_{\tilde{d}i}^2)]$ in Eq. (18). As the SNR increases, $\sigma_{\tilde{d}i}^2$ decreases and the Wiener gain matrix approaches unity. The performance of these two algorithms

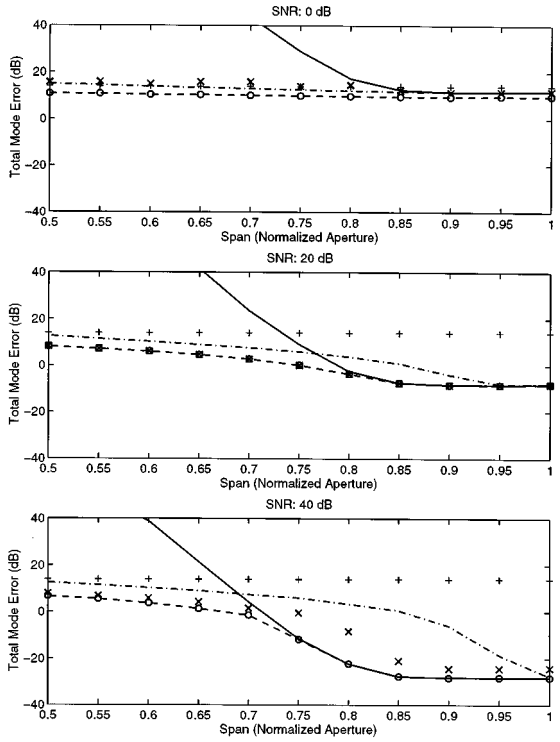


FIG. 5. Comparison of the performance of common mode filtering algorithms in spatially white noise at three different SNRs for a deep water environment, using an axis-fixed array. The algorithms compared are the pseudo-inverse (solid), sampled mode shape (dash-dot), diagonal-weighting ($\times$'s), MAP (dashed), and mismatched MAP (circles). The crosses (''+'') mark the rough bound on worst case performance ($10\log_{10} 25 = 14$ dB) that results from ignoring the observed data.

become almost identical as the SNR increases. This is reflected in the solid and dashed lines being plotted on top of each other for the 20- and 40-dB SNR cases in Fig. 3. The mismatched MAP filter again tracks the true MAP filter very closely, indicating that even rough estimates of $\mathbf{K_{dd}}$ suffice to give good performance with the KI noise model.

The DW filter displays a similar transition between the PI filter at small condition numbers to the SMS filter at large condition numbers. As a result, the DW filter's performance is far worse than the MAP at higher SNRs for the smaller aperture arrays.

III. DEEP WATER SIMULATIONS

This section presents the results of simulations using the SW noise model in a typical deep water environment modeled by a canonical Munk sound speed profile.[29] The channel is 5426-m deep with a minimum sound speed of 1483.5 m/s at 923 m. Figure 4 shows the sound-speed profile and the first 25 modes at a propagation frequency of 75 Hz which is used for all of the examples in this section. We consider only the SW noise model since the KI model is not applicable to deep ocean environments. The simulations use a series of 40-element vertical receiving arrays of varying apertures. In the shallow-water case, full aperture is defined to be the entire water column, however this is an impractical definition in deep water scenarios. Instead, we define the span of the full aperture array to be between the upper and lower turning

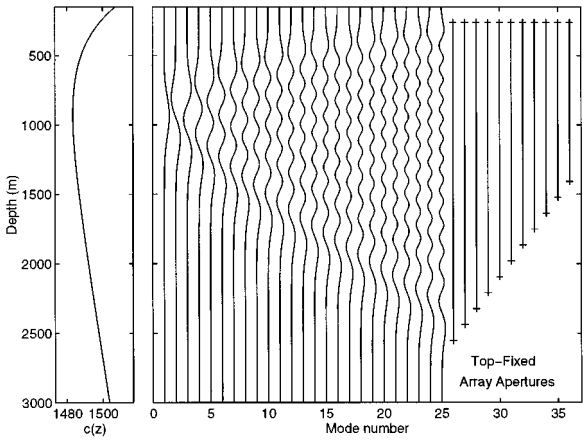


FIG. 6. Top-fixed array apertures. The deep water sound-speed profile and the first 25 modes at 75 Hz are shown for reference.

points of the highest mode we wish to estimate. In order to obtain a well-conditioned problem for the full aperture 40-element array, we must limit ourselves to estimating only the first 25 modes of the waveguide. For the purposes of these simulations, we assume that modes higher than 25 are not present in the received field. The definitions of SNR and MTSE are identical to those in the previous section.

The first set of deep water simulations uses a series of receiving arrays whose aperture varies from full to half span between the upper and lower turning points of mode 25. As the aperture shrinks, the sensor locations are chosen so that the number of hydrophones above and below the sound channel axis remains constant: 12 above and 28 below. Figure 4 shows the spans of these axis-fixed arrays. As the aperture decreases, the conditioning of the estimation problem worsens [$\mathrm{cond}(\Psi) = 1$ for the full span and $\mathrm{cond}(\Psi) = 8 \times 10^4$ for the half-span]. The specifications of the deep water simulations are identical to the shallow water case: 500 independent trials using complex Gaussian data and noise processes were run for each aperture at three different noise levels (0 dB, 20 dB, and 40 dB). Figure 5 shows the MTSE results for the axis-fixed arrays. Note that for $M = 25$, the bound on worst case performance is $10\log_{10} 25 = 14$ dB and is marked with crosses ('+') in the plots. As the plots indicate, the estimators exhibit the same type of behavior as in the shallow-water SW example. The PI mode filter performs well for the full aperture, but degrades rapidly as the aperture shrinks. By contrast, the SMS filter does not deteriorate severely as the conditioning worsens, but it suffers from bias errors due to lack of orthogonality in the sampled mode shapes at small apertures. The MAP filter provides a graceful transition between the SMS and PI filters as aperture increases. In fact the simulations show that its performance matches or exceeds that of the other estimators for the apertures and noise levels examined. For the deep water simulations the β parameter for the DW filter was chosen so that the condition number of Ψ was limited to 200 in the worst case (half-aperture array). As shown in the plots, this choice of β works well in the 20-dB SNR case (MAP and DW results are almost identical), but does not fare as well for the extreme low or high SNR cases. In a similar manner to the

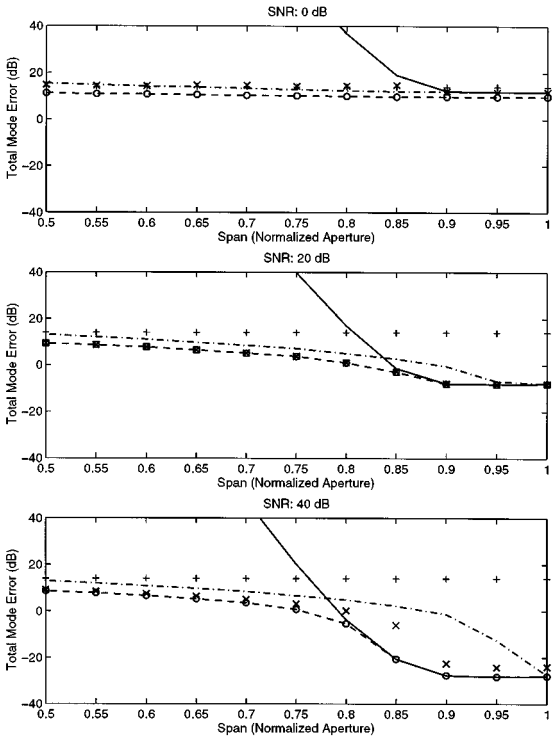


FIG. 7. Comparison of the performance of common mode filtering algorithms in spatially white noise at three different SNRs for a deep water environment using a top-fixed array. The algorithms compared are the pseudo-inverse (solid), sampled mode shape (dash-dot), diagonal-weighting ($\times$'s), MAP (dashed), and mismatched MAP (circles). The crosses (''+'') mark the rough bound on worst case performance ($10 \log_{10} 25 = 14$ dB) that results from ignoring the observed data.

shallow-water simulations, the robustness of the MAP estimator was tested by evaluating the errors assuming a 20% mismatch in the data process variance. For the examples considered here, the mismatched MAP results agree closely with the true MAP results.

The second set of simulations for the deep water environment also uses 40-element vertical arrays, but the sensor locations are set so that the shallowest element is always at the upper turning point of mode 25. These top-fixed array spans are shown in Fig. 6. For this series of apertures, the condition number of Ψ varies from 1 (full aperture) to 2×10^5 (half aperture). Figure 7 shows the MTSE results for this case. The curves show the expected behavior, but it is clear that the conditioning deteriorates much more rapidly as the aperture shrinks for the top-fixed arrays than for the axis-fixed arrays. These results indicate that the position in the water column, relative to the sound channel axis, is an important consideration in the design of arrays for deep water environments.

IV. CONCLUSIONS

This paper presents and compares several common linear mode filtering algorithms, and derives the MAP mode filter. The MAP filter is shown to be a generalization of the mode filter proposed by Yang, where the latter filter is the asymptotic bound of the MAP filter in spatially white noise and poorly conditioned sampling.

The performances of the mode filters are compared in typical shallow and deep water environments. For the shallow water simulations, the SMS filter suffers from the bias introduced by the lack of orthogonality in Ψ. The PI mode filter performs very well when the array samples the channel well, but it deteriorates rapidly as the array aperture decreases and the sampling becomes poorly conditioned. In many situations an aperture spanning less than roughly 75% of the water column could render the PI filter useless. The DW filter is a modification of the PI filter which limits this deterioration. The simulations also demonstrate that the MAP filter generally matches or exceeds the performance of the other filters under a wide range of noise levels and apertures. When the PI filter is well-suited to the current conditions, the MAP filter converges asymptotically to this solution. When the SMS filter is more appropriate, the MAP converges to a form similar to but slightly better than the SMS filter. The shallow-water simulations also demonstrate that the MAP filter is relatively insensitive to mismatch under a variety of SNRs and apertures, making it preferable to the DW filter for mode filtering in many shallow-water experiments. The set of deep water simulations confirms the shallow-water results, thereby indicating that the unified framework developed in this paper is applicable to a variety of ocean environments. In addition, the comparison of the axis-fixed and top-fixed arrays highlights the importance of the absolute positioning of the array within the water column for deep water experiments.

ACKNOWLEDGMENTS

The Lockheed Sanders/Army Fed. Lab Grant No. QK-8819, and the Advanced Research Projects Agency Grants No. MDA972-92-J-1041 and No. N00014-93-1-0686 supported JRB during various stages of this work. JCP received support from the Office of Naval Research Grant No. N00014-95-1-0153 for this research. KEW wishes to acknowledge the support of the DARPA/SEDRP University of California-Scripps ATOC Agreement with MIT, and the GE Fund Faculty for the Future Fellowship. The Office of Naval Research Grant No. N00014-95-1-0362 supported the computer facilities used to do the simulations in this paper. JRB and KEW wish to thank Dr. Andrew C. Singer for several helpful discussions during the course of this research. This paper is Woods Hole Oceanographic Institution contribution #9456.

[1] R. H. Ferris, ''Comparison of measured and calculated normal-mode amplitude functions for acoustic waves in shallow water,'' J. Acoust. Soc. Am. **52,** 981–988 (1972).
[2] R. H. Ferris, F. Ingenito, and A. L. Faber, ''Experimental separation and identification of acoustic normal modes in shallow water,'' Technical Report 7174, Naval Research Laboratory, Washington, DC, October 1970.
[3] H. L. Van Trees, *Detection, Estimation, and Modulation Theory* (Wiley, New York, 1968).
[4] C. T. Tindle, K. M. Guthrie, G. E. J. Bold, M. D. Johns, D. Jones, K. O. Dixon, and T. G. Birdsall, ''Measurements of the frequency dependence of normal modes,'' J. Acoust. Soc. Am. **64**, 1178–1185 (1978).
[5] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*, Prentice-Hall Signal Processing Series (Prentice-Hall, Englewood Cliffs, NJ, 1993).
[6] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice-Hall, Englewood Cliffs, NJ, 1993).

[7] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1992).
[8] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part III* (Wiley, New York, 1971).
[9] T. C. Yang, ''A method of range and depth estimation by modal decomposition,'' J. Acoust. Soc. Am. **82**, 1736–1745 (1987).
[10] C. L. Pekeris, ''Theory of propagation of explosive sound in shallow water,'' The Geological Society of America Memoirs, Vol. 27 (1948).
[11] C. S. Clay and H. Medwin, *Acoustical Oceanography: Principles and Applications* (Wiley, New York, 1977).
[12] I. Tolstoy and C. S. Clay, *Ocean Acoustics: Theory and Experiment in Underwater Sound* (American Institute of Physics, New York, 1966).
[13] G. V. Frisk, *Ocean and Seabed Acoustics: A Theory of Wave Propagation* (Prentice-Hall, Englewood Cliffs, NJ, 1994).
[14] A. B. Baggeroer, ''Sonar signal processing,'' in *Applications of Digital Signal Processing*, edited by A. V. Oppenheim (Prentice-Hall, Englewood Cliffs, NJ, 1978), pp. 331–437.
[15] A. B. Baggeroer, W. A. Kuperman, and P. N. Michalevsky, ''An overview of matched field methods in ocean acoustics,'' IEEE J. Ocean Eng. **18**, 401–424 (1993).
[16] 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).
[17] J. A. Ritcey, S. D. Gordon, and T. E. Ewart, ''A probability distribution for the complex field of waves propagating in random media,'' J. Acoust. Soc. Am. **100**, 237–244 (1996).
[18] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics*, AIP Series in Modern Acoustics and Signal Processing (AIP Press, Woodbury, NY, 1994).
[19] W. A. Kuperman and F. Ingenito, ''Spatial correlation of surface generated noise in a stratified ocean,'' J. Acoust. Soc. Am. **67**, 1988–1996 (1980).
[20] G. H. Golub and C. F. Van Loan, *Matrix Computations* (The Johns Hopkins University Press, Baltimore, 1989).
[21] A. W. Drake, *Fundamentals of Applied Probability Theory* (McGraw-Hill, New York, 1967).
[22] J. F. Lynch, D. K. Schwartz, and K. Sivaprasad, ''On the use of focused horizontal arrays as mode separation and source location devices in ocean acoustics. Part II: Theoretical and numerical modeling results,'' J. Acoust. Soc. Am. **78**, 575–586 (1985).
[23] C. S. Clay and K. Huang, ''Single mode transmission and acoustic backscattering measurements in a laboratory waveguide,'' J. Acoust. Soc. Am. **67**, 792–794 (1980).
[24] C. T. Tindle, H. Hobaek, and T. G. Muir, ''Normal mode filtering for downslope propagation in a shallow water wedge,'' J. Acoust. Soc. Am. **81**, 287–294 (1987).
[25] C. Gazanhes and J. L. Garnier, ''Experiments on single mode excitation in shallow water propagation,'' J. Acoust. Soc. Am. **69**, 963–969 (1981).
[26] S. L. Campbell and C. D. Meyer, Jr., *Generalized Inverses of Linear Transformations* (Pitman, London, 1979).
[27] W. Menke, *Geophysical Data Analysis: Discrete Inverse Theory* (Academic, New York, 1989).
[28] A. G. Voronovich, V. V. Goncharov, A. Yu. Nikol'tsev, and Yu. A. Chepurin, ''Comparative analysis of methods for the normal mode decomposition of a sound field in a waveguide: Numerical simulation and full-scale experiment,'' Sov. Phys. Acoust. **38**, 365–370 (1992).
[29] W. H. Munk, ''Sound channel in an exponentially stratified ocean, with application to SOFAR,'' J. Acoust. Soc. Am. **55**, 220–226 (1974).

The detection and dimension of bubble entrainment and comminution

T. G. Leighton,[a] P. R. White, and M. F. Schneider
Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom

(Received 12 May 1997; accepted for publication 22 December 1997)

Data on bubble entrainment and comminution are gathered in three experiments, involving the breakup of a disk of air trapped between two plates, and bubble cloud generation under a waterfall, and a plunging jet. In the second two cases, an automated acoustic system for characterizing the entrainment is employed. The data sets are compared with an existing theory for bubble fragmentation, in which a key parameter is the number of spatial dimensions associated with the insertion of randomly positioned planes which are used to divide up the bubble. While an appropriate best-fit theoretical curve can be obtained for the bubble population histograms generated by air disk comminution, waterfalls and plunging jets produce multimodal distributions which the theory cannot model. The differing roles of shape oscillations and surface waves in bubble fragmentation, and the issues involved with incorporating these into the model, are examined. [S0001-4966(98)04804-8]

PACS numbers: 43.30.Gv, 43.25.Yw [DB]

INTRODUCTION

Bubbles may be generated through a number of common processes, all of which relate to changes (formation, rupture, closure, merging, etc.) of one or more gas/liquid interfaces. Bubbles may, for example, be entrained from an approximately planar free surface; or a population might be generated through comminution of an existing gas pocket. Energy from such processes, channeled through mechanisms involving surface tension, hydrostatic, and hydrodynamic forces, etc., can generate bubble oscillation and consequently acoustic emission. Minnaert[1] predicted that the natural frequency should vary inversely with bubble size. The bubble, behaving to first order as a lightly damped single degree of freedom oscillator,[2] emits on entrainment an acoustic ''signature'' resembling an exponentially decaying sinusoid.[3,4] Identification of these has enabled size distribution spectra from the acoustic emissions of bubbles entrained in brooks, streams, and waterfalls to be obtained.[5] The passive acoustic emissions associated with bubble formation, including such signatures, have been examined in a large number of scenarios. Examples include the formation of bubble clouds by injection[6] or by the impact of bodies of water.[7–10] Several studies have examined entrainment by liquid jets.[11–16] Acoustic emissions and ''signatures'' have been identified with other forms of liquid drop impact, including rainfall,[17–20] and wave breaking.[21,22] Such studies have in turn led to studies of the underlying fluid dynamics[23,24] and investigations of the possibility of using the acoustic signals for monitoring purposes.[25–29]

Longuet Higgins[30] produced an analysis which demonstrated how the problem of predicting the number and size of daughter bubbles might be approached and, noting the dearth of laboratory data, compared the theory with the measurements of earlier workers.[5,31–33]

The entrainment is modeled by the dissection of a cubical air body through the insertion of planes. The newly formed bubbles (''daughters'') are represented by the rectangular sub-blocks thereby generated. Three sets of planes exist, each set being parallel to a face of the cube and containing m planes. The dimensionality (D) of the fragmentation reflects the number of sets used. The sets are inserted independently to one another (Fig. 1). For comparison of theory and experiment, the dimensionality and the number of planes are chosen to best fit the standard deviation and skewness of the data, and the height of the theoretical curve is adjusted to relate to the total number of bubbles and the bin width.

In the natural world Longuet Higgins suggested that a one-dimensional event may occur when water is running over a smooth stone and entrains a cylindrical volume of air. The one-dimensional model considers a line segment of unit length, which is divided by m points spaced randomly into $m+1$ subsegments of typical length X. The probability density $p(X)$ of the length of this subsegment is

$$p(X)=m(1-X)^{(m-1)} \quad (\text{provided } 0<X<1 \text{ and } m>1), \tag{1}$$

and the cumulative probability function $P(X)$ is given by

$$P(X)=\int p(X)dX. \tag{2}$$

The derivative of the cumulative probability function $P(X)$ with respect to $\ln(X)$ gives the density of the distribution:

$$\frac{dP}{d\{\ln(X)\}}=\frac{dP}{dX}\frac{dX}{d\{\ln(X)\}}=p(X)X=m(1-X)^{(m-1)}X. \tag{3}$$

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

 0001-4966/98/103(4)/1825/11/$10.00

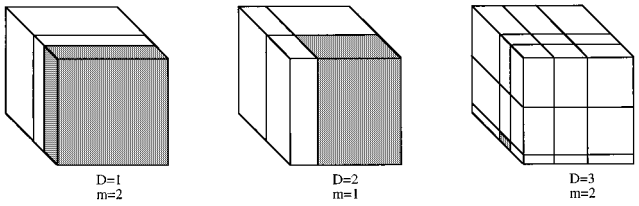


FIG. 1. The model of bubble fragmentation: a cube which is split by m planes inserted from D dimensions (each coordinate axis being perpendicular to the respective cube face). One of the daughter bubbles produced in each case is shaded. (a) $D=1$, $m=2$; (b) $D=2$, $m=1$; (c) $D=3$, $m=2$.

Replacing the line segment of unit length, divided by m points, by a block of unit volume, divided by m parallel planes, the probability density of the sub-blocks will be exactly the same as shown above.

Let R denote the radius of a spherical bubble of volume V, and R_0 the radius of a bubble of volume V_0, such that the normalized bubble radii and volume are $r=R/R_0$ and $v=V/V_0$, respectively. The density of the distribution of bubble radii is given by:

$$\frac{dP}{d\{\ln(r)\}}=\frac{dP}{dV}\frac{dV}{d\{\ln(r)\}}$$
$$=p(v)3v=m(1-v)^{(m-1)}3v. \quad (4)$$

In the two-dimensional model a cubical block of unit volume, which is split by two sets of perpendicular and independent planes, is considered. The distributions are calculated using numerical integration. In a similar way the model can be extended to three dimensions, with distributions having a much smaller mean. In all three cases the distributions for a large number of planes tend to limiting values of nonzero standard deviation and skewness.

In this paper visual observations of air disc fragmentation lead to studies of bubble entrainment in a waterfall and under a plunging jet in the laboratory. The comparison between the data sets and the theory is discussed in terms of the assumptions inherent in the model, and the relative effectiveness of visual and acoustic techniques for monitoring bubble entrainment.

I. OPTICAL METHODS

A so-called ''two-dimensional'' bubble (an ''air disk'') was formed by injecting an air bubble into degassed water between two plates of polymethylmethacrylate (PMMA) separated by an O-ring. The energy for fragmentation is provided by a repeatable combination of flow and pressure disturbance. The apparatus consists of a U-tube which is partially filled with degased water. It is connected by a hose to the PMMA plates at one end, and is stoppered at the other by a top plate which can be opened (Fig. 2). The bubble, which is initially positioned at the center of the plates using a magnetic slider, expands when the vacuum pump reduces the ambient pressure. The top plate is then rapidly opened, and the resulting liquid pressure change causes bubble oscillation, which is video recorded at 50 f.p.s. (frames per second), 1-ms exposure. The minimum pressure applied to the bubble was 20 mbar (approximately 10 mbar higher than the vapor pressure of water at 20 °C). The pre-expansion and pre-collapse size of the bubble and its daughters were measured using a graticule magnifying glass. The partial vacuum caused deflection of the plates of about 0.2 mm in the center and 0.06 mm at the O-ring when the lowest pressure was applied. This deflection was taken into account when the bubble volumes were calculated. The release of the pressure leads to an oscillation of the plates (measured by accelerometer to be at 30 Hz and of 0.5-s duration) and an oscillating water flux. Bubbles having radii smaller than the distance between the plates were hemispherical, rather than cylindrical (Fig. 2). Since the third dimension may influence the dynamics of such bubbles significantly more than for the larger air disks, the transitional size between the two is indicated on the plots of the results (Fig. 6). In general 10–30 breakups of one bubble size at a particular pressure were recorded to obtain representative histograms of daughter bubbles. The accuracy of the measurement of the bubble volume was $\pm 5\%$. (Reference 34 describes how the same apparatus can be modified to generate sonoluminescence from more energetic bubble collapses.)

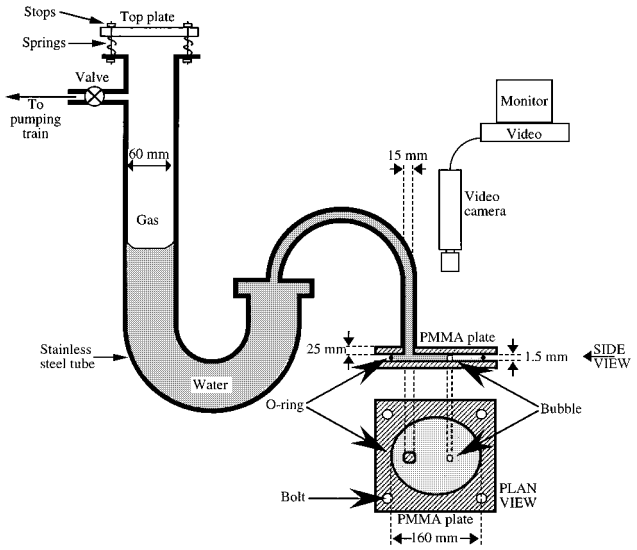


FIG. 2. The apparatus employed to fragment air disks, showing the U-tube equipment; and both plan and side views of the PMMA plates.

II. ACOUSTIC MEASUREMENTS OF ENTRAINMENT THROUGH LIQUID IMPACT

A. Automated bubble signal recognition

Bubble generation was monitored through the use of the passive acoustic emissions. Hydrophone data from waterfalls and brooks had previously been examined,[5] and the time series found to contain distinct exponentially decaying sinusoids which are characteristic of bubble entrainment. These were used to obtain the size distribution of entrained bubbles from the frequencies of the sinusoids. However in noisy environments and with higher entrainment rates (where such ''signatures'' overlap), individual entrainments may not be detected with this technique. Nevertheless a time-frequency representation (TFR) via the Gabor coefficients (see the Appendix) can readily identify the bubble signatures (e.g., from

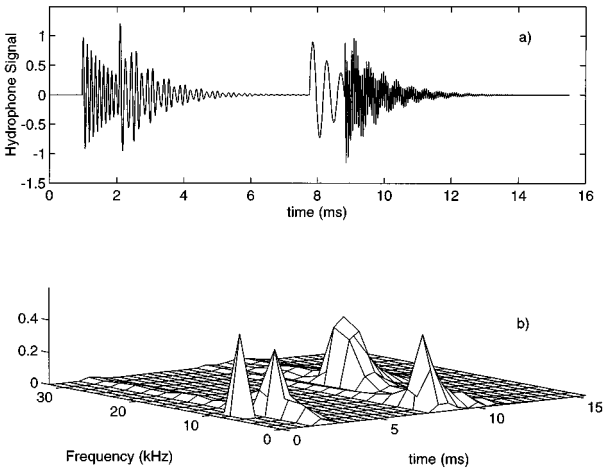


FIG. 3. (a) A noise-free test signal (artificial data) consisting of four exponentially decaying sinusoids. All have unit amplitude, with center frequencies 8, 6, 2.3, and 15.2 kHz, with epochs at 1, 2.12, 8, and 9.1 ms, respectively. All the signals have e-folding decay rates of 0.5 ms^{-1}. (b) Time-frequency representation of the magnitude of the Gabor coefficients associated with the time series of (a) (computed using an FFT size of 32 samples and an assumed decay rate matching that of the simulation). The first transient lies exactly on the sampling lattice of the Gabor expansion; the second lies on the lattice in frequency but not in time; the third lies on the lattice in time but not in frequency; and the fourth lies between lattice points in both frequency and time.

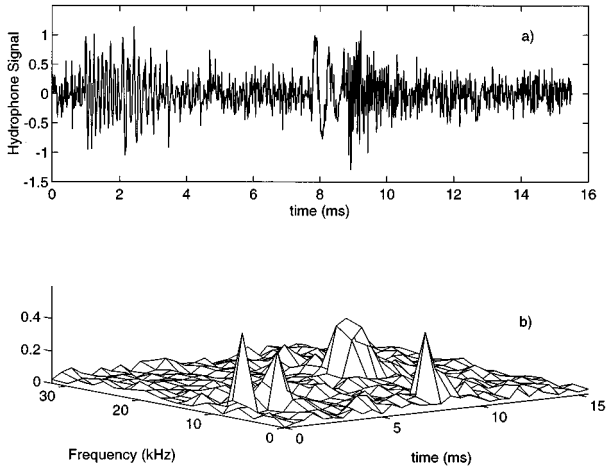


FIG. 4. (a) Same as for Fig. 3(a), but with the addition of Gaussian white noise, the standard deviation of which is 0.2. (b) Time-frequency representation of the magnitude of the Gabor coefficients associated with the time series of (a) (computed using an FFT size of 32 samples and an assumed decay rate matching that of the simulation).

rainfall over the ocean[35]). Resolution in time and frequency is a compromise determined by the size of the window imposed upon the data. The procedure employs a routine which thresholds on the value and gradient of the Gabor coefficients, then automatically counts and sizes the bubbles. For the calculation of the Gabor coefficients, a damping parameter with an e-folding time of 100 s^{-1} was assumed, and for bubble counting the threshold of the Gabor coefficient was set to be three times the mean value of the coefficients, with the gradient set equal to 0.3. These values were robust with respect to bubble counting. Entrainment data were gathered for analysis in this way in a small waterfall, and beneath a water jet which impacts a water surface from above.

Figures 3 and 4 show the efficacy of this technique on artificial data. Figure 3(a) depicts a noise-free test signal consisting of four transient components, each one being an exponentially decaying sinusoid, such that two overlapping pairs occur. In this example all these four components were chosen to have unit amplitude, with center frequencies 8, 6, 2.3, and 15.2 kHz, with epochs at 1, 2.12, 8, and 9.1 ms, respectively. All the signals had decay rates of 0.5 ms^{-1}. The Gabor coefficients are computed using an FFT size of 32 samples and an assumed decay rate matching that of the simulation. This time series was designed so that the first component lay exactly on the sampling lattice of the Gabor expansion, so that it could be represented by a single nonzero coefficient. The second component occurs at a frequency which is on the sampling lattice of the Gabor expansion, but occurs at a time which is between lattice points. The converse is true of the third component (which lies between lattice points in frequency, but on the sampling lattice in time). Finally the fourth component lies at a point which is on neither the time or frequency lattice. The effect of this choice on the Gabor expansion can be seen in Fig. 3(b), which plots the magnitude of the Gabor coefficients as a function of time and frequency. The first component generates a single-point spike. The second signal has a well defined peak in time but is smeared in frequency, whilst for the third the converse is true. The peak corresponding to the fourth component is smeared in both frequency and time. This inconsistency can be mitigated by computing the Gabor expansion at a greater number of points.[36] However, for computational ease, attention here is restricted to the critically sampled distribution,[37] which is adequate for the purposes of this investigation.

Figure 4(a) shows the same signal as Fig. 3(a) but with the addition of Gaussian white noise, the standard deviation of which is 0.2. In the Gabor expansion shown in Fig. 4(b) the peaks are still clearly evident whereas in the time series, Fig. 4(a), identification of signals typical of bubble entrainment is less easy.

B. Entrainment in a waterfall

The acoustic signals of newly entrained bubbles were measured in a brook on the Southampton University Campus on 25.05.93. The hydrophone (Brüel & Kjaer 8104), was mounted on a steel rod at 15 cm depth beneath mean water level in a bubble field created at the base of a small waterfall (height approximately 20 cm, called location A); and at 10 cm depth in a different bubble field approximately 30 cm away (location B). The water was running very smoothly over the step. Two sets of data were obtained at each location. The hydrophone output was amplified (Brüel & Kjaer 2635) and recorded (AIWA HHB 1 PRO DAT recorder, with flat frequency response from 20 Hz to 22 kHz) for 10 min at both locations. Subsequently, the data from the field recordings were transmitted through a low-pass filter (Barr and Stroud EF5/20, roll off approx. 48 dB/oct) into a data acquisition box, where they were acquired and digitized using the MATLAB software package. The time history of the data

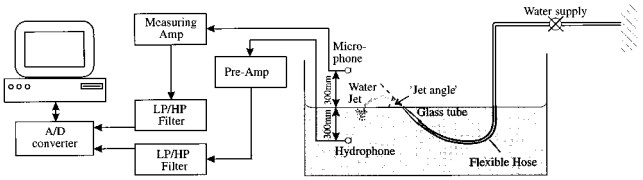


FIG. 5. Apparatus for entrainment by liquid jet.

showed that the most bubbles occurred in the frequency range between 500 Hz and 2 kHz. The signal was acquired with a sampling rate of 20 kHz while the low-pass filter was set to 10 kHz (although this provides no guard band to avoid aliasing, no significant frequency components were present above this frequency).

C. Entrainment by a plunging liquid jet

The experiment was carried out in a 1.8 m $\times$1.2 m$\times$1.2 m deep glass reinforced plastic tank which was filled to 1 m depth with fresh water. The tank was vibration isolated from the floor by Tico pads. A hose, terminating in a glass tube (5 mm inner diameter), was mounted near the water surface at varying angles (Fig. 5). Flow rate and jet angle were varied and recorded (the intention being to investigate the relative effect of these rather than characterize the jet completely[13–15]). The hydrophone (Brüel & Kjaer 8105) was at $\sim$300 mm depth beneath the entrained bubbles, but $\sim$150 mm off-axis. The hydrophone signal was amplified (Brüel & Kjaer 2635) and recorded on a DAT-recorder (AIWA, Type HOB 1 PRO) or fed directly through a low-pass filter (Barr & Stroud EF5/20) into the MATLAB data acquisition toolbox. The sample rate was 44 kHz, and the low-pass filter set with a cutoff frequency of 20 kHz.

III. OPTICAL RESULTS OF AIR DISC FRAGMENTATION

Figure 6 shows the daughter populations produced when a two-dimensional bubble having initial diameter d_0 $=7.5$ mm expands under various pressure reductions to varying values of d_g, the diameter when fully expanded. The applied pressure reduction, and consequently the subsequent pressure step, were increased from the value just required to cause fragmentation. When a small pressure difference between applied pressure and atmospheric pressure was released, giving low excitation energy, two daughter bubbles were produced. With higher excitation energies, more daughters were created. In regard to the dependence on the initial bubble size, the release of a given pressure difference tended to produce an increased number of daughter bubbles, as d_0 increased.

The data in Fig. 6 are compared with the theory. The measured radii of the air discs were converted into radii of spheres R_{sph} with the same volume. Depending on the average number N_{avg} of daughter bubbles per break-up, the best value of m with the appropriate curve was fitted. In all the cases the data is fitted optimally when the one-dimensional theory for fragmentation is applied ($D=1$). In Fig. 6(a) the average number of daughter bubbles is $N_{\mathrm{avg}}=2$. The best fit

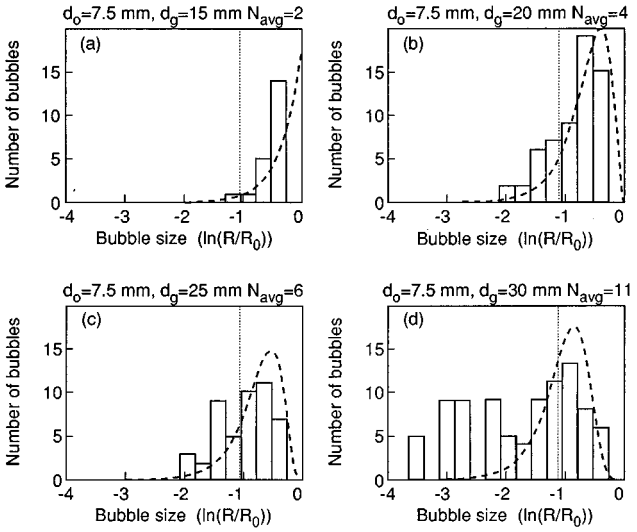


FIG. 6. Histograms of the distribution of daughter bubbles after fragmentation of an air disk of initial diameter $d_0=7.5$ mm for different excitation energies, with the best-fit curve from theory. The dotted vertical line shows the transition diameter, such that measured radii smaller than this are less than the plate spacing.

is given by the one-dimensional theory with $D=1$ and m $=1$, shown as the dashed line in Fig. 6(a). The theoretical curve does not fit the measured data, because no very large daughter bubbles were observed. This suggests that the preferred splitting occurred toward the middle of the bubble. This would give a more narrow distribution than would the random splitting that theory assumes. In Fig. 6(b) the excitation energy was increased, leading to an increased number of daughter bubbles. The mode of the measured bubble size distribution occurs at a smaller radius than does that of the prediction. A theoretical distribution having $D=1$, $m=3$, and $N_{\mathrm{avg}}=4$ [shown as the dashed line in Fig. 6(b)] gives the best fit. Figure 6(c) shows a similar distribution, but with a smaller mean and less sharp peak. Augmenting the applied pressure difference has generated an increased number of daughter bubbles. Again, the theoretical curve for a splitting in one dimension ($D=1$), with an increased number of planes ($m=5$, $N_{\mathrm{avg}}=6$), produces a better fit than is possible with the two- or three-dimensional theories. This optimal fit is shown by the dashed line in Fig. 6(c). Increasing the excitation energy still further leads to a greater number of daughter bubbles, as shown in Fig. 6(d). Most of the bubbles are smaller than the transitional size between disklike and hemispherical daughters, shown by the dotted vertical line. It is therefore not surprising that no single theoretical curve fits the entire data set in Fig. 6(d), since the processes involved in the fragmentation of hemispherical and disklike bubbles would be expected to differ. The figure shows the theoretical curve for $D=1$, $m=10$ tends to cover the disklike bubbles in the distribution.

The $D=1$ best fit to the results suggests that the bubbles fragmented in a process where splitting occurred in one direction only. In addition, if the plane were inserted near the center of the bubble, this would give the more narrow distribution which is observed. These hypotheses are confirmed using high speed video pictures taken from the fragmentation

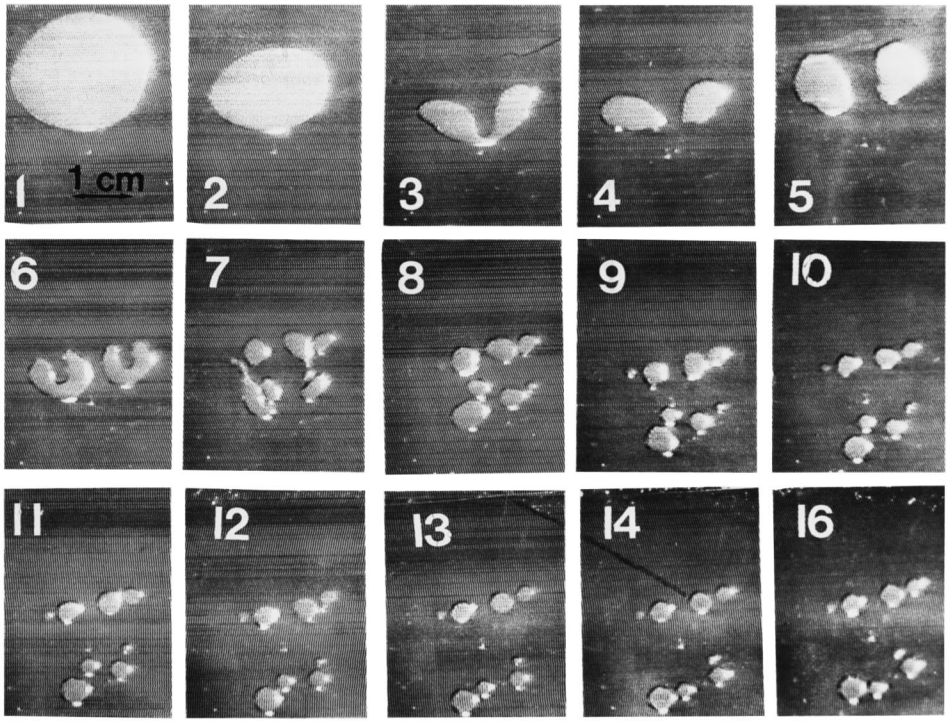


FIG. 7. Successive video images of air disc fragmentation (20-ms interframe time; $d_0 = 7.5$ mm; $d_g = 20$ mm). Between frames 1 and 2 the pressure is released, causing the bubble to lose its initially approximately circular profile.

of bubbles having pre-expansion diameters of $d_0 = 7.5$ mm (for pre-collapse diameters d_g of 20 and 25 mm). In Fig. 7, when the pressure is released the upstream wall involutes and forms a jet which travels across the cavity to impact the downstream wall (frame 3). This leads to the splitting of the bubble (frame 4), forming two daughter bubbles of approximately equal size (frame 5). These smaller bubbles undergo severe distortion from the circular shape and move with the water flux in the direction of the connector. This can be seen by comparing frame 4 and frame 6. It was found that with smaller excitation pressures (for example when $d_0 = 7.5$ mm and $d_g = 15$ mm), the process of bubble fragmentation stopped at this point, such that two ''daughters'' were formed from a single ''mother'' bubble. However with the stronger excitation used for Fig. 7 ($d_0 = 7.5$ mm and $d_g = 20$ mm), further fragmentation occurs. The daughter bubbles formed in frame 4 develop jets in turn, and are further split into smaller bubbles (frame 7). They are influenced by the movement of the plates and the associated water flux, and by the rebound of the pressure wave. The bubbles move as a result of the flux of the water around them (frames 8 and 9) and change their location and their shape for as long as 500 ms (this interval is the same as the time for which the PMMA plates oscillate). All of the daughter bubbles present at the end of the experiment were formed between frame 6 and frame 9. A variety of fragmentation behaviors was seen in other collapses, including the simultaneous formation of two jets (which are parallel in Fig. 8; and perpendicular in Fig. 9). Though the bubbles are subjected to similar excitations in Figs. 7 and 9, the collapses differ, primarily because of slight differences in the initial bubble shape.

IV. ACOUSTIC RESULTS FROM LIQUID IMPACT

A. In brook waterfall

The daughter bubble size distributions obtained at site A are shown in Fig. 10(a), and at site B in Fig. 10(b). Two data sets, (i) and (ii), were acquired at each site, showing that the general form of the histogram at each site was stable. Both sites exhibit two distinct peaks, the first at 400 Hz and the second at 1.2 kHz. At site A [Fig. 10(a)] the hydrophone was near to a bubble field having a frequency distribution from 700 to 1300 Hz, whereas at site B [Fig. 10(b)] the hydrophone record was dominated by the component at around 400 Hz. Visual observation indicated far too few bubbles large enough to account for the generation of this peak from individual entrainment signatures. Additionally, the time series showed that the signatures corresponding to the 400-Hz component were low amplitude and ill-defined, compared to the decaying sinusoids characteristic of the entrainment of smaller bubbles. These two facts suggest that this low-frequency peak arose through collective bubble oscillations,[38,39] and as such it was removed from the data set before comparison with theory. This comparison is shown in Fig. 11 for the same respective sites and times, where $dN/d[\ln(R_0)]$ is plotted as a function of R_0 (in keeping with the analysis of the Introduction with R_0 normalized to 1 mm throughout). The best fit to data is for a one-dimensional event ($D = 1$, $m = \infty$), as illustrated by comparison (through standard deviation and skewness) of the data with models of varying dimensionality and plane number (Fig. 12).

The theoretical curves in Fig. 11 are not so peaked as the experimental data, and show greater standard deviation. Comparing the time history with theory shows that the

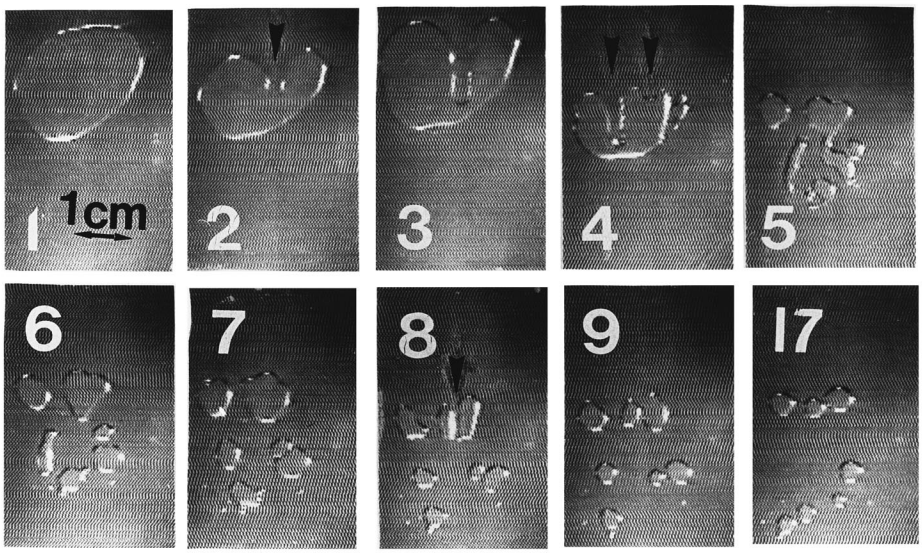


FIG. 8. Successive video images of air disk fragmentation (20-ms interframe time; $d_0 = 7.5$ mm; $d_g = 25$ mm). Between frames 1 and 2 the pressure is released. Two parallel jets (arrowed in frame 4) fragment the bubble (one jet appears before the other, as arrowed in frame 2). In frame 8 another jet, which causes a daughter bubble to fragment, is arrowed.

bubbles are entrained sequentially, rather than simultaneously, so that the data must be considered as summed distributions if comparison with theory is judging like against like.

B. Entrainment by plunging liquid jet

Investigation was made of the bubble size distribution entrained when a water jet strikes the surface of the water. The flow rate and the jet angle (measured between the horizontal and the direction of the water as it leaves the tube) were varied. Visual observation revealed that a minimum jet speed was needed to entrain bubbles, below which no entrainment occurred. The sole exception to this occurred when small surface waves on the water tank disturbed the water jet, resulting in bubble entrainment. Such surface waves could be excited by entrained bubbles which rose to the surface by buoyancy. The maximum jet speed employed was limited to ensure that no subsequent splashes occurred. During the experiment it became clear that, for large angles, the main reason for entrainment was the disturbance of the water surface by bubbles rising to the surface. For smaller angles the water jet carried the entrained bubbles away from the jet and the waves on the tank surface set up when the bubbles burst did not affect the water jet.

Figures 13–15 show histograms of the bubble entrainment distributions. In part (i) of each figure the bubble count (N) is shown as a function of the bubble natural frequency. In part (ii), $dN/d[\ln(R_0)]$ is plotted as a function of R_0 to compare with the theory of the Introduction. Figure 13 shows results from an experiment with a constant water flow

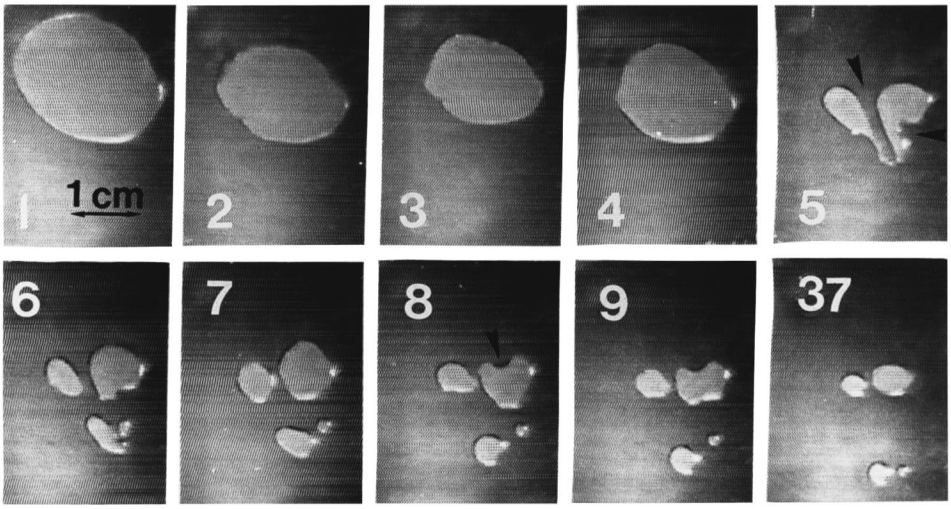


FIG. 9. Successive video images of air disk fragmentation (20-ms interframe time; $d_0 = 7.5$ mm; $d_g = 20$ mm). Between frames 1 and 2 the pressure is released. Two perpendicular jets fragment the bubble (arrowed in frame 5). In frame 8 the involution of a daughter bubble occurs (arrowed), but this is not sufficiently energetic to fragment the bubble.

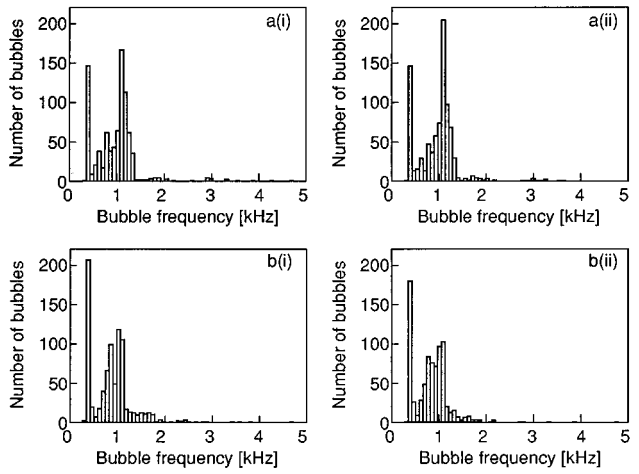


FIG. 10. Bubble resonance distributions (50 s of brook data each): (a): location A; (b) location B. Two data sets, (i) and (ii), are taken at each site. Acquisition: 10 ksamples/s with 5-kHz low-pass filter; damping 100 $[s^{-1}]$; Gabor coefficient for bubble count: threshold=3×mean, gradient=0.3. The number of samples in each data set was 500 000; with 128 sample points in each time interval for the TFR.

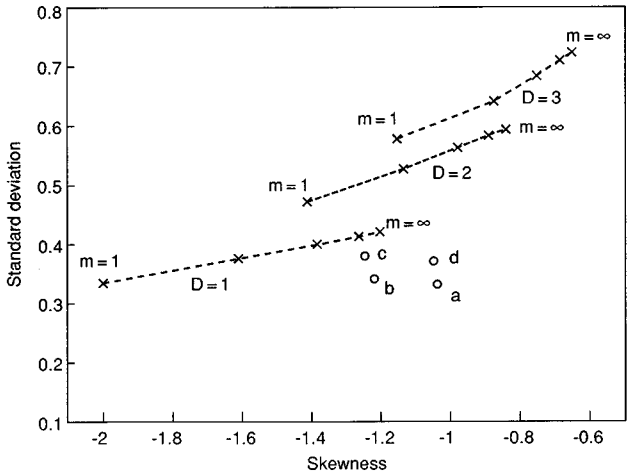


FIG. 12. The data from the waterfalls which were displayed in Fig. 11 are condensed to four points (°) on the standard deviation/skewness plane. The points are labeled (a)–(d) to show the dataset in Fig. 11 to which they refer. This display allows a comparison with the predictions obtained from various (D,m) combinations: For each of $D=1$, 2, and 3, the range of values from $m=1$ to $m=\infty$ is shown.

(2.5 l/min) and jet angles of 80°–30°. As expected from the visual observations outlined above, the number of bubbles entrained per time interval first decreased with a decreasing angle and then increased again. All the distributions showed a maximum at ~2 kHz. In addition they had a peak at ~700 Hz. At 80° the distribution is roughly monomodal, and the best fit to theory is one dimensional, both indicating a simple entrainment process. As the angle reduces to 60° the multimodal distribution is measured. Although the ''best'' fit is for $D=3$, the fit is poor, and suggests that processes not incorporated in the theory (such as coalescence, and the simultaneous occurrence of several types of entrainment and comminution) are involved. The entrainment rate decreases sharply (75°) and then gradually increases again. High frequencies are generated. The rate remains high from 55° to 30°, for which $D=2$ is the best fit, and which have distributions that are less multimodal, but still not as simple as for 80°.

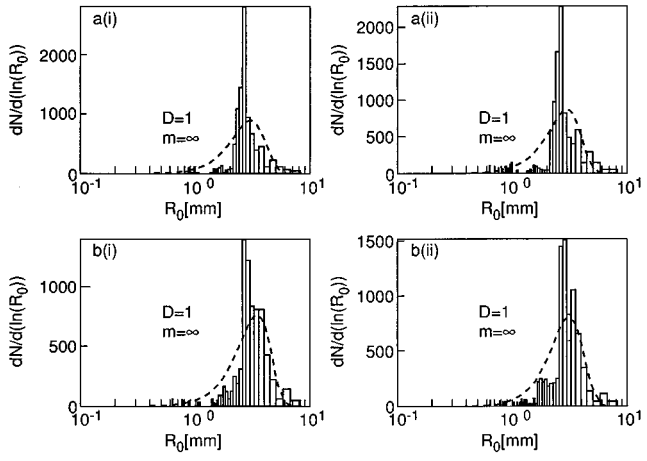


FIG. 11. Comparison of measured bubble density distributions with theoretical curves for $D=1$ and $m=\infty$ are shown. R_0 is normalized to 1 mm.

Figures 14 and 16 record the results for varying flow rates, and constant angles (55° in Fig. 14; 70° in Fig. 15). As expected, the numbers of bubbles entrained increases with flow rate in both cases. Multiple peaks are observed [see for example Figs. 14(b) and 15(b)]. Except for the trivially simple distribution presented in Fig. 15(a), the distributions are of the higher dimensions [$D=2$ or 3 throughout Fig. 14; $D=3$ for Fig. 15(b) and (c)].

V. DISCUSSION

This study reports measurements of three experimental systems: the comminution of a disk of air trapped between two solid plates; and the generation of bubble populations beneath a natural waterfall and a plunging liquid jet. The data sets are then compared with a theory which the author[30] sets out as being a simple, ''rough,'' first approach to the problem. As such, close agreement between theory and model would not be expected.

The best agreement is obtained for the most ''artificial'' scenario (in terms of geometry and mode of fragmentation), that is, the comminution of an air disk. In this, a single cavity breaks up, in keeping with a key assumption of the theory. The fit between theory and data is poorer for the two acoustically measured processes. Although attempts to find a best fit lead to the application of a one-dimensional model for the waterfall, and higher dimensions for the jet, even so the agreement is not good, with the occurrence of multiple peaks in the measured distributions (the current theory can only produce unimodal fits).

There are a number of factors which, while being intrinsic components of generation of bubble populations under waterfalls and plunging jets, do not have immediate counterparts in the theory. Both of these processes involve, to a greater or lesser extent, a surface. To a first approximation,

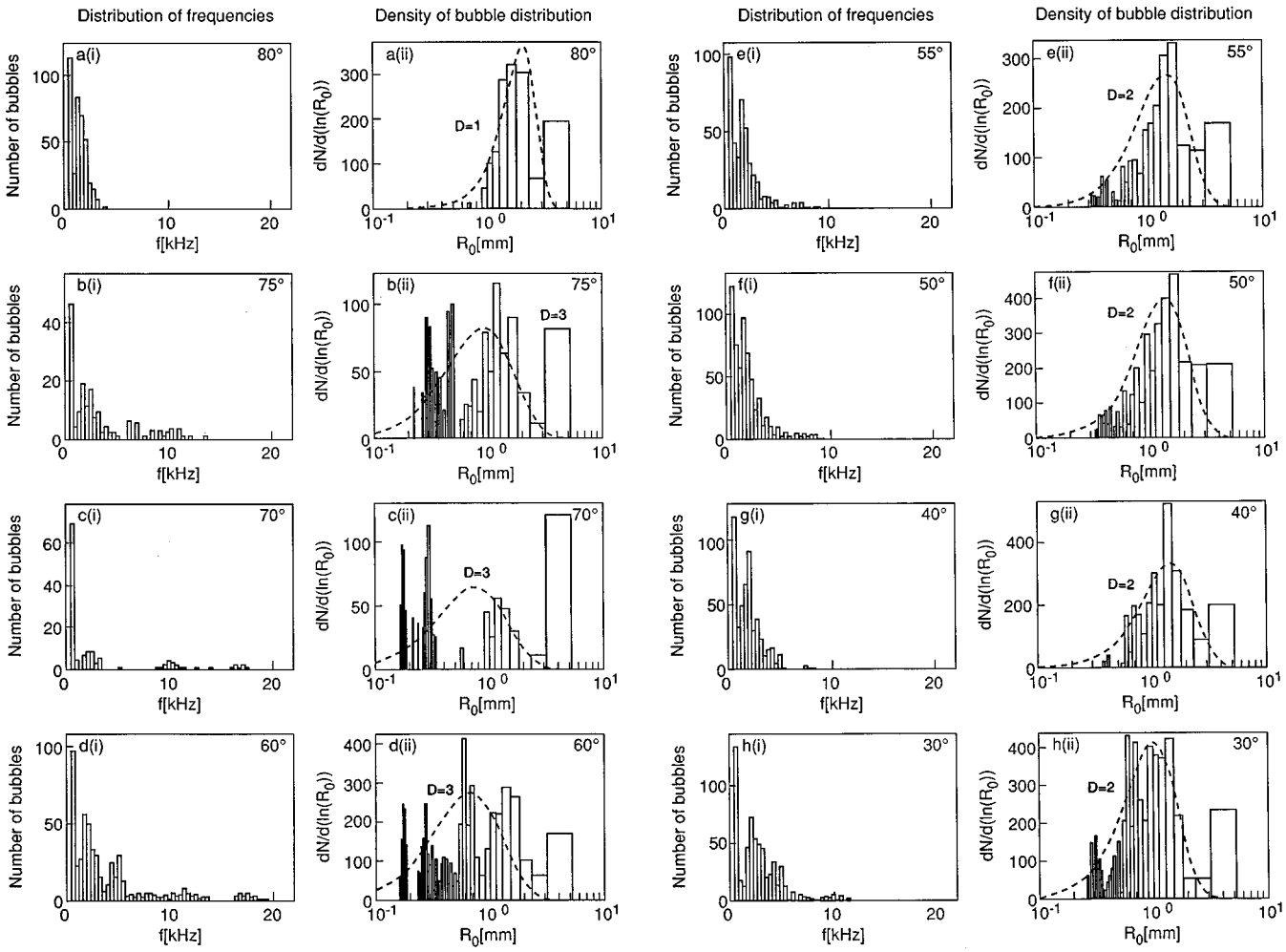


FIG. 13. Histograms of (i) number of bubbles N against bubble natural frequencies f; and (ii) $dN/d[\ln(R_0)]$ as a function of bubble radius R_0. R_0 is normalized to 1 mm. Part (ii) shows the best fit theoretical curves, with the value of D shown (the best-fit value of m is ∞ throughout). Each data set is taken over a period of 5 s. Jet flow rate 2.5 l/min. Surface-to-nozzle angle: (a) 80°; (b) 75°; (c) 70°; (d) 60°; (e) 55°; (f) 50°; (g) 40°; (h) 30°.

this surface is planar and ''infinite,'' and its disturbance causes closure and detachment, and therefore the entrainment of some daughter bubbles. While subsequent break-ups of these bubbles would more closely fit the model, how accurately the free-surface event is described by the theory depends on details of the closure and detachment processes. However, during the period in which these ''subsequent break-ups'' occur, coalescence might also become an issue: A double peak in the population distribution might, for example, be observed if a preferential bubble size is first formed, and, later, coalesces. In addition to these fluid dynamic issues, contributions from collective oscillations in the waterfall do not reflect bubble entrainment in the same way as single-bubble emissions, but are interpreted as such by the measurement system.

There are other factors inherent in bubble entrainment processes which might more readily be discussed in the light of the assumptions of the theory. Figure 16 shows the fragmentation of large bubbles following injection. The lower bubble has the ''hourglass'' form characteristic of a shape oscillation dominated by a second-order ($n=2$) axisymmetric spherical harmonic perturbation, an extreme version of which would divide it into two daughters of roughly equal size (indeed the upper bubble was generated in this way).[40] Such low order shape oscillations may be modeled by the insertion of small numbers of planes at specific sites determined by the mode value, n, and not randomly as the theory[30] assumes. They will generate a number of daughters of similar size, but few bubbles which are either very small or of a size comparable with the mother bubble. As a result one would expect the size distributions obtained experimentally when this occurs to show smaller standard deviations than are predicted by the theory. The assumption of simultaneous insertion will be true for all planes related to minima in a given mode. However fragmentation may occur through a series of sequential shape oscillations, as indicated by the preliminary optical study. Figure 7 shows a ''mother'' bubble undergoing fragmentation into four bubbles through repeated division into halves. The mother is bisected by a jet, the position of which is determined by a second order ($n=2$) oscillation, and the same fate befalls the resulting two daughters. The description entails the insertion of a single plane associated with the $n=2$ oscillation on three occasions (once in the original bubble, and then once in each of its two daughters). Figure 6 shows a bubble being fragmented by a third-order ($n=3$) shape oscillation, corresponding to the

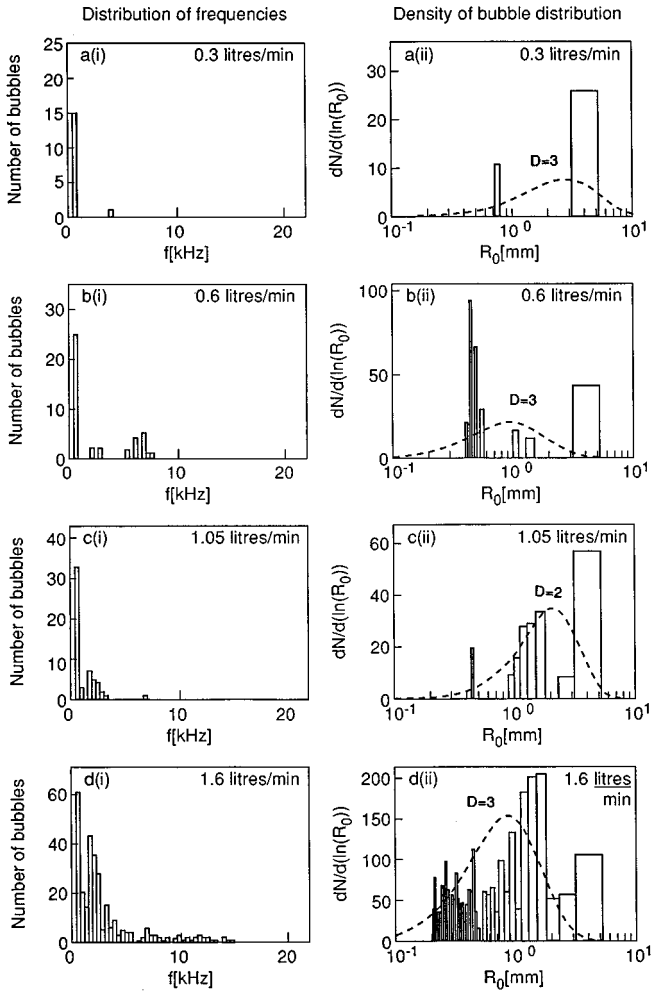


FIG. 14. Histograms of (i) number of bubbles N against bubble natural frequencies f; and (ii) $dN/d[\ln(R_0)]$ as a function of bubble radius R_0. R_0 is normalized to 1 mm. Part (ii) shows the best fit theoretical curves, with the value of D shown (the best-fit value of m is ∞ throughout). Each data set is taken over a period of 4 s. Surface-to-nozzle angle is 55°. Jet flow rate (a) 0.30; (b) 0.60; (c) 1.05; (d) 1.60 l/min.

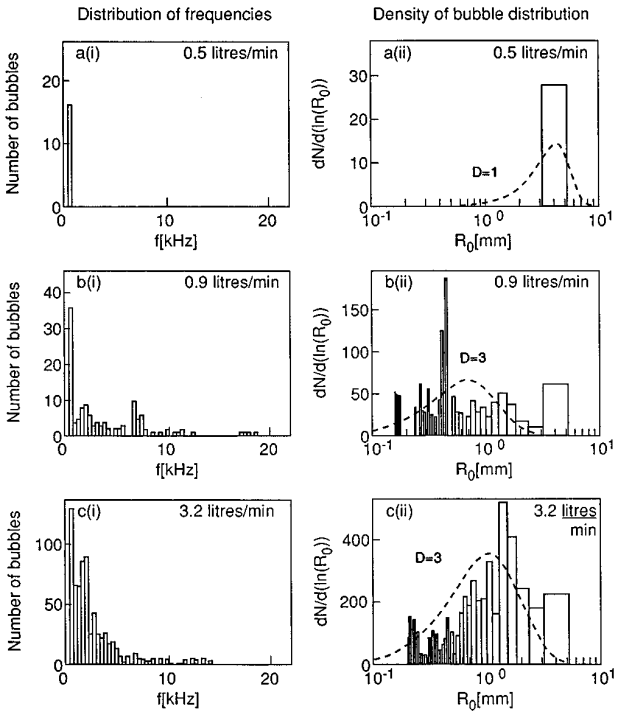


FIG. 15. Histograms of (i) number of bubbles N against bubble natural frequencies f; and (ii) $dN/d[\ln(R_0)]$ as a function of bubble radius R_0. R_0 is normalized to 1 mm. Part (ii) shows the best fit theoretical curves, with the value of D shown (the best-fit value of m is ∞ throughout). Each data set is taken over a period of 4 s. Surface-to-nozzle angle is 70°. Jet flow rate (a) 0.50; (b) 0.90; (c) 1.05; (d) 3.20 l/min.

simultaneous insertion of the two parallel planes associated with the $n=3$ mode, to produce three daughter bubbles.

However Fig. 16 also shows another method of producing daughter bubbles: The small bubbles in the picture were ''pinched off'' from the peaks of capillary waves visible at the narrowest region of the lower bubble, and over much of the upper bubble. Such capillary wave action can leave the mother relatively intact,[41] equivalent to the simultaneous insertion of planes at positions close to the bubble wall; or may disintegrate the mother entirely.[42]

In a complicated entrainment event a great number of processes may occur, involving (as seen in Fig. 16) both shape oscillations (where plane insertion is biased towards the bubble center) and capillary waves (where the bias is toward the walls). Being based upon a random positioning of planes, it is clear that the theory of Longuet Higgins[30] takes account of neither shape oscillations nor capillary waves on any individual bubble. However it is not clear to what extent the superposition of many shape and surface oscillations oc-

FIG. 16. Photograph of the bubble population produced following injection through a cm radius nozzle.

curring in a population may be modeled by a random positioning of planes.

VI. CONCLUSIONS

Data are obtained on the number and size of bubbles produced by bubble entrainment and comminution, both optically (for the comminution of air disk) and acoustically (using an automated signal recognition system for measurements under both waterfalls and plunging liquid jets). Comparison is made with a theory which is based on the breakup of a cubical air pocket. Assumptions regarding the simultaneous insertions of planes at random positions through a single gas body do not take account of the specific ways in which a self-contained air pocket may fragment. Entrainment may often be a population event. It may be the simplicity of the model which makes it versatile as a diagnostic tool for the dimension of bubble breakup under different conditions.

ACKNOWLEDGMENTS

The authors wish to acknowledge the funding support of the Natural Environment Research Council (GR3/9992), and the three referees for their careful and constructive analyses of this manuscript.

APPENDIX: GABOR EXPANSIONS

Computing the Gabor expansion of a signal is one form of time-frequency analysis. The signal, say $x(t)$, is represented as a sum of scaled time and frequency shifted versions of a prototype signal, $g(t)$, specifically:

$$x(t)=\sum_{n=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}C_{n,k}g(t-n\Delta t)e^{2\pi ik\Delta ft}$$

$$=\sum_{n=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}C_{n,k}g_{n,k}(t), \tag{A1}$$

where $g_{n,k}(t)=g(t-n\Delta t)e^{2\pi ik\Delta f}$, $C_{n,k}$ are the coefficients of the expansion, and Δt and Δf are the intervals in the time and frequency domains respectively. It should be noted that the representation (A1) maps the continuous time function $x(t)$ to the doubly infinite discrete set $C_{n,k}$. The existence of the representation (A1) is by no means assured. For a suitable function $g(t)$, the sampling grid has to be sufficiently fine,[43] specifically $\Delta t\Delta f\leqslant 1$.

Representation (A1) allows one to consider $x(t)$ as being made up of a sum of functions of the form of $g_{n,k}(t)$. Many choices of the prototype functions can be made. Gabor[44] considered a Gaussian function of the basic form e^{-t^2}. This choice is problematic since the computation of the expansion coefficients is difficult. Friedlander and co-workers[36,37] present a decomposition based on single sided exponentials, i.e., $g(t)=e^{-\lambda t}$ for $t>0$. The decaying exponential form closely mimics the sound of a decaying bubble and it is this form of Gabor transform which is employed in this paper. By matching the functions, $g_{n,k}(t)$, to the expected data set, the signal representation is efficient and is robust to the adverse affects of noise.

While Eq. (A1) is a representational form, it fails to give information about how to compute the coefficients $C_{n,k}$. This computation is made considerably more simple if the function $g(t)$ and the time-frequency grid are chosen so that $g_{n,k}(t)$ form an orthogonal basis, such that

$$\int g_{n,k}(t)g_{m,1}(t)^*=0 \quad \text{unless } n=m \text{ and } k=1, \tag{A2}$$

and

$$\int g_{n,k}(t)g_{m,1}(t)^*=\xi \quad \text{for } n=k \text{ and } k=1. \tag{A3}$$

Combining this with Eq. (A1) leads directly to

$$C_{n,k}=\frac{1}{\xi}\int x(t)g_{n,k}(t)^*\,dt. \tag{A4}$$

It can be shown[43] that if the basis is to be orthogonal then the grid must be such that $\Delta t\Delta f=1$. However choosing an orthogonal basis is incompatible with the desire for good time and frequency localization of the basis functions.[43]

In cases where $g(t)$ does not generate an orthogonal basis, but is reasonably well behaved, it is possible to define a biorthogonal function,[45] $\gamma(t)$, such that

$$\int g_{n,k}(t)\,\gamma_{m,1}(t)=0 \quad \text{unless } n=m \text{ and } k=1, \tag{A5}$$

and

$$\int g_{n,k}(t)\,\gamma_{m,1}(t)=\xi \quad \text{for } n=k \text{ and } k=1, \tag{A6}$$

where $\gamma_{n,k}(t)=\gamma(t-n\Delta t)e^{2\pi ik\Delta f}$ and hence

$$C_{n,k}=\frac{1}{\xi}\int x(t)\,\gamma_{n,k}(t)dt. \tag{A7}$$

The biorthogonal function for the one-sided exponential has a relatively simple analytic form and consequently the Gabor coefficients based on that expansion are conceptually simple to compute. Further, Friedlander and Porat[37] present a highly efficient algorithm, requiring only 2 FFTs.

[1] M. Minnaert, ''On musical air-bubbles and sounds of running water,'' Philos. Mag. **16**, 235–248 (1933).
[2] C. Devin, ''Survey of thermal, radiation and viscous damping of pulsating air bubbles in water,'' J. Acoust. Soc. Am. **31**, 1654–1667 (1959).
[3] M. Strasberg, ''The pulsation frequency of nonspherical bubbles,'' J. Acoust. Soc. Am. **25**, 536–537 (1953).
[4] M. Strasberg, ''Gas bubbles as sources of sound in water,'' J. Acoust. Soc. Am. **28**, 20–26 (1956).
[5] T. G. Leighton and A. J. Walton, ''An experimental study of the sound emitted from gas bubbles in a liquid,'' Eur. J. Phys. **8**, 98–104 (1987).
[6] M. Nicholas, R. A. Roy, L. A. Crum, H. Oguz, and A. Prosperetti, ''Sound emissions by a laboratory bubble cloud,'' J. Acoust. Soc. Am. **94**, 3171–3182 (1994).
[7] A. R. Kolaini, R. A. Roy, and L. A. Crum, ''An investigation of the acoustic emissions from a bubble plume,'' J. Acoust. Soc. Am. **89**, 2452–2455 (1991).
[8] A. R. Kolaini, R. A. Roy, L. A. Crum, and M. Yi, ''Low-frequency underwater sound generation by impacting transient cylindrical water jets,'' J. Acoust. Soc. Am. **94**, 2809–2820 (1993).
[9] A. R. Kolaini, R. A. Roy, and D. L. Gardner, ''Low-frequency acoustic emissions in fresh and salt-water,'' J. Acoust. Soc. Am. **96**, 1766–1772 (1994).
[10] H. N. Oguz, A. Prosperetti, and A. R. Kolaini, ''Air entrapment by a

falling water mass,'' J. Fluid Mech. **294**, 181–207 (1995).
[11] D. Kusabiraki, H. Niki, K. Yamagiwa, and A. Ohkawa, ''Gas entrainment rate and flow pattern of vertical plunging liquid jets,'' Can. J. Chem. Eng. **68**, 893–903 (1990).
[12] D. Kusabiraki, M. Murota, S. Ohno, K. Yamagiwa, M. Yasuda, and A. Ohkawa, ''Gas entrainment rate and flow pattern in a plunging liquid jet aeration system using inclined nozzles,'' J. Chem. Eng. Jpn. **23**, 704–710 (1990).
[13] P. A. Hwang, Y. K. Yu, and J. Wu, ''Temperature effects on generation and entrainment of bubbles induced by a water jet,'' J. Phys. Oceanogr. **21**, 1602–1605 (1991).
[14] F. Bonetto and R. T. Lahey, ''An experimental study on air carryunder due to a plunging liquid jet,'' Int. J. Multiphase Flow **19**, 281–294 (1993).
[15] A. K. Bin, ''Gas entrainment by plunging liquid jets,'' Chem. Eng. Sci. **48**, 3585–3630 (1993).
[16] A. Prosperetti and H. N. Oguz, ''Air entrainment upon liquid impact,'' Philos. Trans. R. Soc. London **355**, 491–506 (1997).
[17] H. C. Pumphrey and A. J. Walton, ''Experimental study of the sound emitted by water drops impacting on a water surface,'' Eur. J. Phys. **9**, 225–231 (1988).
[18] H. C. Pumphrey, L. A. Crum, and L. Bjørnø, ''Underwater sound produced by individual drop impacts and rainfall,'' J. Acoust. Soc. Am. **85**, 1518–1526 (1989).
[19] H. C. Pumphrey and P. A. Elmore, ''The entrainment of bubbles by drop impacts,'' J. Fluid Mech. **220**, 539–567 (1990).
[20] H. Medwin, A. Kurgan, and J. A. Nystuen, ''Impact and bubble sound from raindrops at normal and oblique incidence,'' J. Acoust. Soc. Am. **88**, 413–418 (1990).
[21] H. Medwin and M. W. Beaky, ''Bubble sources of the Knudsen sea noise spectra,'' J. Acoust. Soc. Am. **86**, 1124–1130 (1989).
[22] G. E. Updegraff and V. C. Anderson, ''Bubble noise and wavelet spills recorded 1 m below the ocean surface,'' J. Acoust. Soc. Am. **89**, 2264–2279 (1991).
[23] M. S. Longuet-Higgins, ''An analytical model of sound production by raindrops,'' J. Fluid Mech. **214**, 395–410 (1990).
[24] H. N. Oguz and A. Prosperetti, ''Bubble entrainment by the impact of drops on liquid surfaces,'' J. Fluid Mech. **219**, 143–179 (1990).
[25] J. A. Nystuen, ''Rainfall measurements using underwater ambient noise,'' J. Acoust. Soc. Am. **79**, 972–982 (1986).
[26] J. A. Nystuen and D. M. Farmer, ''The influence of wind on the underwater sound generated by light rain,'' J. Acoust. Soc. Am. **82**, 270–274 (1987).
[27] H. N. Oguz and A. Prosperetti, ''Numerical calculations of the underwater noise of rain,'' J. Fluid Mech. **228**, 417–442 (1991).
[28] F. Laville, G. D. Abbott, and M. J. Miller, ''Underwater sound generation by rainfall,'' J. Acoust. Soc. Am. **89**, 715–721 (1991).
[29] J. A. Nystuen, ''Quantitative rainfall measurement using the underwater sound field,'' in *Proceedings of the Fourth International Meeting on Natural Physical Processes Related to Sea Surface Sound*, edited by T. G. Leighton (University of Southampton, Southampton, 1997), pp. 73–81.
[30] M. S. Longuet-Higgins, ''The crushing of air cavities in a liquid,'' Proc. R. Soc. London **439**, 611–626 (1992).
[31] Y. Toba, ''Drop production by bursting air bubbles on the sea surface III. Study by use of a wind flume,'' Mem. Coll. Sci. Kyoto A **29**, 313–343 (1961).
[32] G. Updegraff, ''In situ investigation of sea surface noise from a depth of 1 meter,'' Ph.D. University of California at San Diego, 1989.
[33] H. Medwin and A. C. Daniel, ''Acoustical measurements of bubble production by spilling breakers,'' J. Acoust. Soc. Am. **88**, 408–412 (1989).
[34] T. G. Leighton, W. L. Ho, and R. Flaxman, ''Sonoluminescence from the unstable collapse of a conical bubble,'' Ultrasonics **35**, 399–405 (1997).
[35] T. G. Leighton, M. F. Schneider, and P. R. White, ''Study of dimensions of bubble fragmentation using optical and acoustic techniques,'' in *Proceedings of the Third International Meeting on Natural Physical Processes Related to Sea Surface Sound*, edited by M. J. Buckingham and J. Potter (World Scientific, Singapore, 1995), pp. 414–428.
[36] B. Friedlander and A. Zeira, ''Oversampled Gabor representation for transient signals,'' IEEE Trans. Signal Process. **43**, 2088–2094 (1995).
[37] B. Friedlander and B. Porat, ''Detection of transient signals by the Gabor representation,'' IEEE Trans. Acoust. Speech Signal Process. **37**, 169–180 (1989).
[38] S. W. Yoon, L. A. Crum, A. Prosperetti, and N. Q. Lu, ''An investigation of the collective oscillations of a bubble cloud,'' J. Acoust. Soc. Am. **89**, 700–706 (1991).
[39] W. Carey and D. G. Browning, ''Low frequency ocean surface noise: Measurement and theory,'' in *Sea Surface Sound*, edited by B. R. Kerman (Kluwer, Boston, 1988), pp. 361–376.
[40] T. G. Leighton, K. J. Fagan, and J. E. Field, ''Acoustic and photographic studies of injected bubbles,'' Eur. J. Phys. **12**, 77–85 (1991).
[41] T. G. Leighton, *The Acoustic Bubble* (Academic, London, 1994), Fig. 4.31.
[42] G. W. Willard, ''Ultrasonically induced cavitation in water: a step-by-step process,'' J. Acoust. Soc. Am. **25**, 669 (1953).
[43] I. Daubechies, ''The wavelet transform: A method for time frequency localisation,'' in *Advanced Spectral Analysis and Array Processing*, Vol. 1, edited by S. Haykin (Prentice-Hall, Englewood Cliffs, NJ, 1991), pp. 366–417.
[44] D. Gabor, ''Theory of communication,'' J. Inst. Electr. Eng. **93**, 429–457 (1946).
[45] M. J. Bastiaans, ''A sampling theorem for the complex spectrogram, and Gabor's expansion of a signal in Gaussian elementary signals,'' Opt. Eng. (Bellingham) **20**, 594–598 (1981).

Enhancement of the total acoustic field due to the coupling effects from a rough sea surface and a bubble layer

Guy V. Norton
Naval Research Laboratory, Stennis Space Center, Mississippi 39529-5004

Jorge C. Novarini
Planning Systems, Inc., 21294 Johnson Road, Long Beach, Mississippi 39560-9702

(Received 14 March 1997; accepted for publication 19 December 1997)

At wind speeds higher than a few meters per second, when breaking waves are present, the sea surface roughness is accompanied by an assemblage of microbubbles forming different structures or clouds, varying in range and in time. For simplicity, in terms of an average bubble population, the time and space varying assemblage of microbubbles is often assumed to be uniform and referred to as the subsurface bubble layer. Of fundamental importance is the role that this subsurface bubble layer may play in connection with scattering from a rough air/sea interface. The purpose of this paper is to determine the effect that the rough surface in the presence of this subsurface bubble layer has on the total field at low frequencies. Numerical calculations indicate that at low frequencies the bubble layer shifts the incident angle on the surface to steeper angles. In addition it was found that the enhancement of the total field is a consequence of scattering at the rough surface in the presence of the upper refracting bubble layer. An enhancement of approximately 40 dB with respect to the bubble-free medium was obtained very near the surface for a frequency of 400 Hz, with a nominal grazing angle of 20 deg, and a void fraction at the surface of 3.2×10^{-5}. The enhancement decreases to about 10–15 dB at 10 m below the surface, but is still significant at depth exceeding the bubbly region. For a void fraction at the surface of 3.1×10^{-6} the enhancement is approximately 5 dB.
 [S0001-4966(98)01104-7]

PACS numbers: 43.30.Hw, 43.30.Re, 43.30.Es, 43.30.Dr [DLB]

INTRODUCTION

At wind speeds higher than a few meters per second, when breaking waves are present, the sea surface roughness is accompanied by an assemblage of microbubbles forming different structures or clouds, varying in range and in time. For simplicity, in terms of an average bubble population, the time and space varying assemblage of microbubbles is often assumed to be uniform (Crowther, 1980; McDaniel and Gorman, 1981; Hall, 1989; Farmer and Vagle, 1989; Gilbert, 1993; Ye, 1997) and is referred to as the subsurface bubble layer. Although the bubble layer has been the subject of many studies in the last 20 years, mainly in relation to high-frequency reverberation, its impact on sound propagation has received relatively little attention. Examples of articles on acoustic propagation where the bubble layer has been taken into account are Novarini and Bruno (1982), Weston (1989), and Hall (1989). At low frequencies and at the bubble concentrations encountered in the ocean, the presence of bubbles will alter the sound speed close to the surface, inducing a refractive effect that will affect the surface ensonification. Lamarre and Melville (1994) have measured, near the surface, reductions in the sound speed of 400–800 m/s at 5 kHz, in the open ocean, at a moderate wind speed of 8 m/s. The bubble layer also brings additional attenuation, which increases with frequency, but is small at frequencies below a few kilohertz. The true range-dependent nature of the bubble population may present itself in a low-frequency backscattering experiment, since the clouds may act as individual scatterers. The experiment by Farmer and Vagle (1989) on propagation of ambient noise in the layer induced by the presence of bubbles, and the modeling by Hall (1989) show that the results of propagating through the range-dependent layer are consistent with results obtained using an average bubble spectrum. Thus indicating that in the case of forward propagation, the use of a uniform bubble layer with an average spectrum is an acceptable first step in modeling the medium.

In spite of the strong effect that bubbles have on the sound speed near the surface, researchers have remained reluctant to include the refractive effect when dealing with sound interaction with the sea surface, skeptical about the importance of the effect. Except for a few attempts (Weston, 1989; Schneider, 1987), the bubble layer is usually not included when dealing with propagation under conditions where multiple interaction with the surface are expected to occur (surface duct, shallow water propagation). An exact solution for the problem of surface scattering in the presence of an exponential decaying velocity gradient in the underlying medium, as introduced by the bubble layer, is not available. Approximate theories and heuristic approaches are not suitable for elucidating the problem.

In the presence of an exponentially decaying sound speed the effect of scattering (from the rough surface) and the effect of propagation of both, the scattered (surface interacting) and nonscattered (nonsurface interacting) energy are coupled to one another and cannot be separated. Once the conformal mapping algorithm, which allows for the inclusion of rough surface effects, is coupled to the propagation algorithm, extraction or isolation of the scattered field from

the total field is not possible. Therefore, if propagation effects manifest themselves in the course of the experiment, as in this case due to the near-surface bubble layer, the total field must be examined.

The purpose of this paper, therefore, is to determine the effect that the rough surface in the presence of the subsurface bubble layer has on the total field at low frequencies, through highly accurate numerical calculations. To that end we have adopted a propagation algorithm based on the parabolic equation method (Collins, 1993) with a conformal mapping algorithm (Dozier, 1984; Norton *et al.*, 1995) which locally flattens a rough surface segment. The resulting method, which shares the flexibility of PE to handle arbitrary environments, has been shown to produce benchmark quality results for the sea surface scattering problem (Norton and Novarini, 1996). Besides fulfilling the main objective of the paper, the accuracy of the method provides a set of results that could serve as a reference solution against which approximate theories can be compared. A time-averaged, range-independent subsurface bubble layer is brought into the model through the complex index of refraction of the resulting water/bubble mixture. The effective index of refraction was calculated assuming a tenuous medium (Clay and Medwin, 1977) and a modified Johnson and Cooke bubble spectrum.

I. MODELING PROPAGATION IN AN OCEAN WAVEGUIDE WITH A ROUGH SEA SURFACE

For details of the conformal mapping (CM) technique the reader is referred to Dozier (1984), and to Norton *et al.* (1995) concerning its current implementation in the split-step Pade solution for the parabolic equation method (EFEPE) (Collins, 1993). The technique for propagating in an ocean waveguide with sea surface roughness has been validated by Norton *et al.* (1995) through comparisons with a full wave scattering model for the case of periodic surfaces, as well as for single realizations of randomly rough surfaces having a power spectrum characteristic of a sea surface. The method has also been applied to study the effect of sea surface roughness on low-frequency (400-Hz, 800-Hz) propagation, for a point source in a realistic shallow water environment (Norton and Novarini, 1996).

Although the paper focuses on the combined effect of surface roughness and bubble layer, it should be made clear that in this simulation the numerical surfaces do not actually include breaking waves, since they are created through a linear filtering technique. The lack of breaking waves should not affect the forward scattered field, which is mainly controlled by the rms surface height and by the long surface waves, which are adequately included in the surface spectrum. The rough surfaces were generated by standard linear filtering technique, where a set of random numbers (1-D white surface) is passed through a filter whose transfer function (in the wave number domain) is given by the square root of the desired power density spectrum of the output surface. In this case we chose the Pierson–Moskowitz spectrum for a fully developed wind driven sea surface for a 15-m/s wind at 10-m elevation. The spectrum, which is a frequency spectrum, was converted to the 1-D wave number domain, which implicitly assumes all the energy goes in the windward direction. The range step adopted for generating the surface realization was one tenth an acoustic wavelength. The total length of the surfaces was based on the incident angle. For a 15-m/s wind speed (at 10-m above the surface) the rms height of the surface realization is 1.41 m.

II. THE BUBBLY ENVIRONMENT

A. The bubble spectrum

Models that describe a horizontally averaged subsurface bubble layer have been available to acousticians for over a decade. They provide a dependence of the bubble size spectrum (i.e., the number of bubbles per unit volume that have radii within a unit increment in radius) on wind speed and depth. Hall (1989), utilizing published ocean measurements and previous parameterizations (Wu, 1981; Novarini and Bruno, 1982), set forth a comprehensive bubble spectrum for a horizontally averaged bubble layer, suitable for use in acoustic modeling. Hall's spectrum is largely based on Johnson and Cooke's data, which show the bubble distribution peaking at about 50–60 μm. Later experiments (Su *et al.*, 1988; Farmer and Vagle, 1989; Vagle and Farmer, 1992) have shown that ocean bubble distributions peak at smaller radius (around 20–25 μm). The peak around 50–60 μm is believed to be a shortcoming of the photographic technique used by Johnson and Cooke, which misses bubbles at small radii (Breitz and Medwin, 1989; Vagle and Farmer, 1992). More specifically, Breitz and Medwin state that for small bubbles, a -4 slope can extend down to 10 μm, thus placing the peak at that radius. Based on these data, to avoid the peak at 60 μm, we have modified Hall's spectrum by choosing as the peak (and cutoff) the lower limit estimated by Breitz and Medwin (10 μm). For the spectral slope, we kept that obtained by Hall fitting Johnson and Cooke data, and for small bubbles we adopted the -4 slope measured by Medwin and Breitz (1989). For more details see Novarini and Norton (1994). It should be pointed out that, although the modified spectrum better reflects the latest findings, the location of the peak is not critical. The refraction at low frequencies is mainly driven by the total void fraction and small bubbles do not contribute significantly.

The generic functional form for the bubble density spectrum of the bubble layer can be written as (Hall, 1989)

$$N(a,z,u)=N_0 G(a|z)Z(z|u)U(u). \tag{1}$$

The function $G(a|z)$ provides the dependence on the bubble radius at a given depth z. Hall's form, with the modification for small bubbles described above, takes the form

$$G(a,z)=\begin{cases} 0, & a<10\ \ \mu\text{m}, \\ (a_{\text{ref}}/a)^{4.0}, & 10\ \ \mu\text{m}\leqslant a\leqslant a_{\text{ref}}, \\ (a_{\text{ref}}/a)^{p(z)}, & a>a_{\text{ref}}. \end{cases} \tag{2}$$

The reference radius (a_{ref}) and the spectral slope $p(z)$ are obtained from Hall (1989),

$$a_{\text{ref}}=54.4+(1.98)z, \tag{2a}$$

$$p(z)=4.37+(z/2.55)^2, \tag{2b}$$

where z is the depth in meters and the radius is in μm. In this work, a_{ref} is the radius at which the spectral slope changes from -4 to $-p(z)$ (identified as radius a_2 in Hall's paper). As explained, the spectrum truncates at 10 μm because of the uncertainty of its functional form for smaller radii. The maximum radius is set to 1000 μm.

The term $Z(z|u)$ has the form (Hall, 1989)

$$Z(z|u)=\exp(-z/d), \tag{3}$$

with

$$d=\begin{cases}0.4, & u<7.5 \text{ m/s},\\ 0.4+0.115(u-7.5), & u\geqslant 7.5 \text{ m/s}.\end{cases}$$

The term $U(u)$ provides the dependence on wind speed, and has the form (Hall, 1989)

$$U(u)=(u/13)^3. \tag{4}$$

The constant N_0 is the value of $n(a,z,u)$ at the surface, the reference radius and the reference wind speed (13 m/s). In Hall's fitting of Johnson and Cooke's data the constant has the value $N_0=1.6\times10^4$ m^{-3} μm^{-1}.

As mentioned elsewhere, the range-independent bubble layer is a convenient simplification. In reality the bubble population changes not only in depth but also in range and time, and can achieve high bubble densities at localized patches. Hence, even under same wind conditions, N_0 can differ substantially from that assigned to the average layer. On the other hand, at low frequencies the sound-speed perturbation is mainly controlled by the total fraction of air entrained in bubble form, i.e., the void fraction. Hence, instead of using different N_0 to represent different bubble structures (plumes, background, etc.) in the present work the value of N_0 will be adjusted so as to produce void fractions within values encountered at sea. The results will then be parameterized by the total void fraction V_0 at the surface, defined as

$$V_0=\int_0^{\infty} n(a,z=0,u)\,\frac{4\pi}{3}\,a^3\,da. \tag{5}$$

B. Sound speed and attenuation in the bubbly medium

The effective sound speed in a bubbly medium can be characterized by a complex number c_{eff}, the real part of which provides the phase velocity and the imaginary part leads to the attenuation. For simplicity, hereafter we will refer to the phase velocity that occurs in the presence of the bubbles as the perturbed sound speed c_b. At low frequencies the bubbles produce a reduction in the phase velocity, which is mainly controlled by the total void fraction. In this work the resonant characteristics of the bubbles are taken into account. We have adopted the expressions given by Clay and Medwin (1977) to calculate the real and imaginary part of the effective sound speed in a bubbly medium. The calculation of c_{eff} includes the contribution from bubbles of all radii, with due regard to viscous and thermal effects. Multiple scattering between bubbles is ignored because, at the void fractions involved in the simulation (less than 10^{-4}), it has been shown that they have no significant effect on the complex sound speed (Feuillade, 1995). The resulting perturbed sound speed as given by Clay and Medwin (1977) is

$$c_b(z)=c_0\left\{1-\frac{3}{2}\int_{a_{\min}}^{a_{\max}}\frac{u(a)Y^2(Y^2-1)}{a^2k_r^2[(Y^2-1)^2+\delta^2]}\,da\right\} \tag{6}$$

with

$$u(a)=n(a)\tfrac{4}{3}\pi a^3;\quad Y=f_r/f,$$

where f_r is the resonance frequency for bubbles of radius a, k_r is the corresponding wave number, and δ is the total damping constant for the bubble, expressed as a sum of thermal damping (δ_t), reradiation damping (δ_R), and viscous damping (δ_v)

$$\delta=\delta_t+\delta_R+\delta_v. \tag{7}$$

The damping constant is a function of frequency. The expressions for the damping constant and resonance frequency, with due regard to viscous and thermal effects, can be found in Clay and Medwin (1977). The attenuation coefficient (in dB/m) is given by

$$\alpha_{\mathbf{b}}=4.34\int_{a_{\min}}^{a_{\max}}\frac{4\pi a^2(\delta/\delta_{\mathbf{R}})n(a)}{[(f_r/f)^2-1]^2+\delta^2}\,da. \tag{8}$$

In the numerical calculation the maximum bubble radius at the surface is set to 1000 μm. We have assumed that the maximum bubble radius decreases exponentially with depth (Wu, 1981)

$$a_{\max}(z)=a_{\max}(0)\exp\left(\frac{-z}{4.1}\right). \tag{9}$$

It should be pointed out here that, although for the sake of accuracy c_{eff} has been calculated as outlined above, at the low frequencies used in this work the attenuation is negligible and the resulting phase velocity c_b is very close to that obtained by the Wood's equation at the corresponding void fraction.

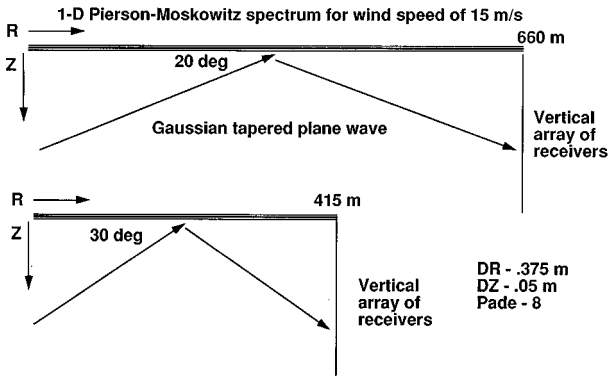


FIG. 1. Geometry used for the numerical calculations.

III. NUMERICAL CALCULATIONS

A. Geometry

The geometry chosen for the numerical calculations is depicted in Fig. 1. The simulated acoustic experiment consists of ensonifying the rough surface with a tapered plane

TABLE I. Parameters for the numerical calculations.

Frequency (Hz)	Incident angle (Deg)	Surface length (m)
400	20	660
400	30	415
800	20	330

wave from below and collecting the total field on a vertical array set at the right edge of the rough surface. The EFEPE-CM model has been shown to have achieved benchmark quality solutions for the forward scattering problem as compared to the solution from an integral equation method (Norton and Novarini, 1996). In the present case we include a bubble layer underneath the rough surface, and examine its influence on the total acoustic field. Since the bubble layer is introduced by modifying the index of refraction, EFEPE-CM will accurately operate within this environment. The bubble layer is modeled as attached to the surface profile and not to an imaginary average flat plane.

The problem is restricted to a 1-D rough surface, and therefore the field is only a function of range and depth. The rough surface profile is specified at range intervals of one-tenth of a wavelength. The field incident on the rough surface is given as a function of depth at the left endpoint of the surface (range 0 m) in the form of a tapered plane wave. The angle of incidence is measured from the horizontal (grazing angle). The total field as a function of depth at the right endpoint of the surface was determined for various bubbly environments as well as when no bubbles were present. An ensemble of 20 surfaces were used with three different bubble environments (different void fractions). In addition, two different acoustic frequencies and two different incident angles were used. The average intensity versus depth was determined for each frequency, incident angle, and bubble environment. This was compared to the corresponding average intensity versus depth for the no bubble environment.

The two frequencies are 400 and 800 Hz. The two incident grazing angles for the bubble-free environment (nominal angles) are 20 and 30 deg. The starting fields were initiated such that the peak value would be 32 wavelengths below the surface and would ensonify the center of the surface. Thus for 20 deg incident angle, the surfaces were of different lengths for the two frequencies. However, the total surface length expressed in number of wavelengths were the same. Table I gives the pertinent information. Three different bubble environments (different void fractions at the surface) were used: 3.2×10^{-5}, 1.6×10^{-5}, and 3.1×10^{-6}. For a wind speed of 15 m/s, these values correspond to N_0 equal to 1.6×10^{5} m^{-3} μm^{-1}; 8.0×10^{4} m^{-3} μm^{-1}; and 1.6×10^{4} m^{-3} μm^{-1}, respectively. The perturbed sound-speed profiles for the three void fractions are shown in Fig. 2. For clarity, only the profiles for 400 Hz are shown. Profiles for 800 Hz are about 3 m/s weaker than the corresponding profile for 400 Hz at the surface, and essentially identical at 10-m depth. At 10-m depth, even for the largest void fraction, the sound speed perturbation due to the bubble layer has dropped to less than 1 m/s. To verify that the numerical results were free of spurious reflections, (which can occur when the field

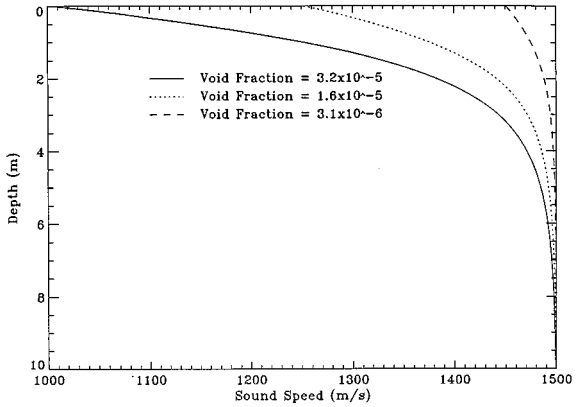


FIG. 2. Sound speed versus depth for the three different void fractions.

is not terminated correctly while marching out in range), the output of EFEPE-CM for a single surface realization for the different incident angles and frequencies were compared to an exact solution obtained with a computer code based on an integral equation method, the method of moments as implemented by Thorsos (1988), modified to allow for finite receiver range. The solution provides the total field (scattered plus direct) at the receiver array. For these comparisons, a homogeneous half-space is assumed below the surface, having a sound speed of 1500 m/s. Figure 3 shows a comparison of the field expressed in dB versus depth for a frequency of 400 Hz and two incident angles, 20 and 30 deg. The method of moments is the dashed line while the EFEPE-CM result is the solid line. Excellent agreement exists throughout depth for both incident angles. Thus indicating that no spurious reflections occurred.

The requirement that the source be located at the same relative distance below the surface for the different cases ensured that variability in the results between different frequencies would not be the result of source placement. In addition, the source was placed far enough below the surface to ensure that the plane wave ensonified the surface from below, traveling through the entire length of the bubble layer and does not start within the bubble layer. The surfaces and hence range coordinate were discretized in range (Δr) equal

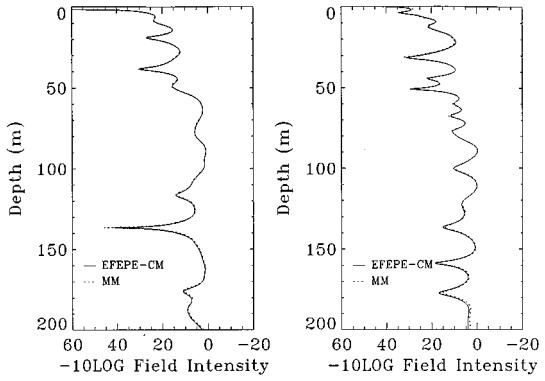


FIG. 3. A comparison of the forward scattered field versus depth expressed in dB of the EFEPE-CM result (solid line) and the method of moment (dotted line) for a frequency of 400 Hz. (a) Incident angle is 20 deg. (b) Incident angle is 30 deg.

to one tenth of a wavelength, the depth coordinate (Δz), at 0.05 m.

B. Result of numerical calculations

In all cases the ''average intensity of the total field expressed in dB'' (hereafter the ''total field'') versus depth will be calculated. Additionally, the difference between the total field when the bubble layer is and is not present will be determined versus depth. There are 20 surface realizations making up the ensemble. The number of surfaces used is the result of monitoring the scintillation index (normalized covariance of the intensity of the acoustic field), which should converge once an adequate number of surfaces are included in the ensemble. For the type of surfaces and frequencies involved, convergence was achieved after including approximately 15 surfaces.

The numerical calculation of the total field versus depth, using the different void fractions are now examined. The analysis of the numerical calculations are discussed in the next section. The total field versus depth, for each frequency, and incident angle are depicted in the next three figures.

Figures 4 and 5 depicts the total field versus depth for a frequency of 400 Hz, the three environments, and a nominal incident angle of 20 and 30 deg, respectively. The total field versus depth for the two largest void fraction cases (solid and dotted line) show a lobe that indicates a trapping of energy near the surface [Figs. 4(a) and 5(a)] within the duct created by the bubble-induced sound speed gradient. The depth and average intensity of the lobe is greatest for the intermediate void fraction case (dotted line) with an incident angle of 20 deg [Fig. 4(a)], while the peak value for the 30-deg incident angle corresponds to the largest void fraction case [solid line Fig. 5(a)]. The total field for the smallest void fraction case (dashed line) shows the least amount of change from the total field when the bubble layer is not present [Figs. 4(a) and 5(a)], independent of incident angle. There is a definite change in slope of the curve at approximately 5–7 m depth for the 20-deg incident case [Fig. 4(a)], which is absent from the 30-deg incident case [Fig. 5(a)].

In order to better quantify the effect that the bubble layer has on the total field, the ratio of the intensities between the total field when a bubble layer is and is not present for a nominal incident angle of 20 and 30 deg is depicted in Figs. 4(b) and 5(b), respectively. The total field is enhanced between 30 and 40 dB near the surface when the two largest void fraction cases (solid and dotted line) were used [Figs. 4(b) and 5(b)]. While the enhancement is approximately 5 dB when the smallest void fraction is used (dashed line). For all cases the enhancement quickly diminishes with depth. At a depth of 10 m the enhancement is approximately 10–15 dB for the two largest void fraction cases (solid and dotted line). This enhancement is nonexistent below a depth of 50 m. The enhancement for the smallest void fraction case (dashed line) is nonexistent below 20 m.

The numerical calculations for a frequency of 800 Hz are now examined. Figure 6(a) depicts the total field versus depth for the three environments, and a nominal angle of 20 deg. The results for the two largest void fraction cases (solid and dotted line) do not show the pronounced lobe that was observed at 400 Hz; instead the intensity is nearly constant in depth. The average intensity of the total field for the smallest void fraction case (dashed line) is the weakest of the three results. As observed in Fig. 4(a), there is a definite change in slope of the curve at approximately 5–7 m.

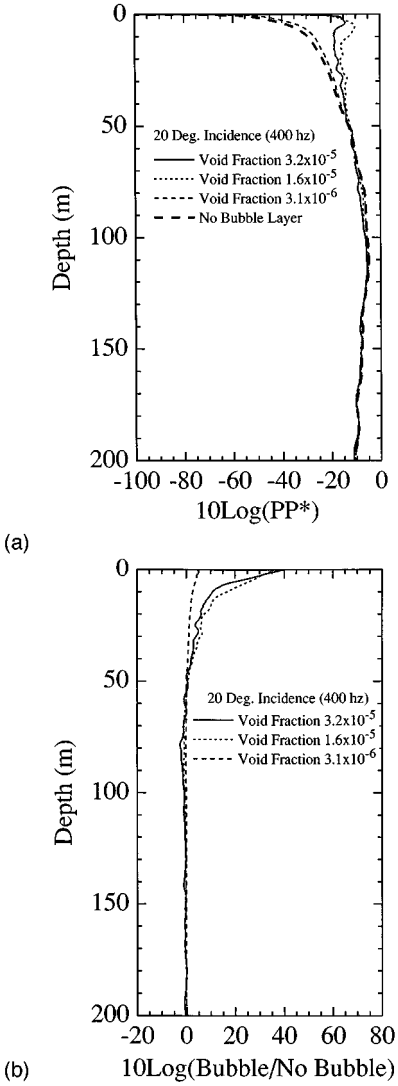
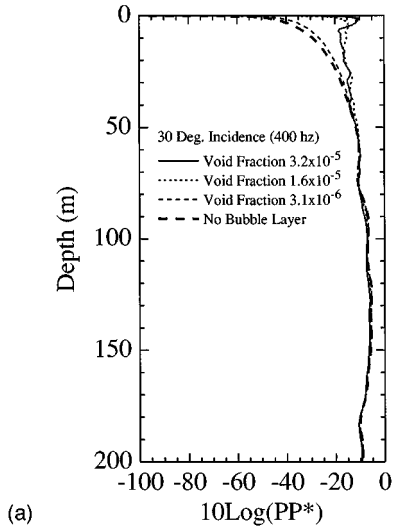


FIG. 4. Results of the numerical calculation for a frequency of 400 Hz, an incident angle of 20 deg, and three void fractions, 3.2×10^{-5} (solid line), 1.6×10^{-5} (dotted line), and 3.1×10^{-6} (dashed line). (a) The average intensity of the total field versus depth. (b) Difference in dB between bubble and no bubble result.

Figure 6(b) depicts the ratio of the bubble layer intensity to the no bubble layer intensity. The total field is enhanced between 30 and 35 dB near the surface for the two largest void fraction cases (solid and dotted line). The enhancement is approximately 15 dB for the smallest void fraction case (dashed line). For all cases the enhancement quickly diminishes with depth. At a depth of 10 m it is approximately 5–15 dB while it is nonexistent below a depth of approximately 15 m.

IV. ANALYSIS

To aid in understanding the results of the numerical calculations just presented, the total field versus range for one



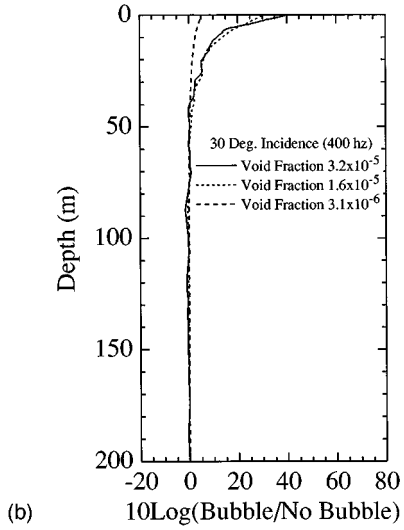
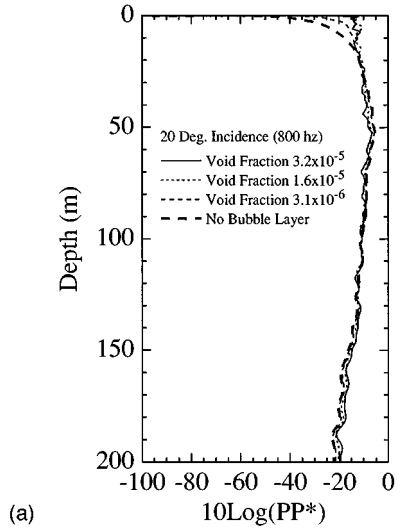


FIG. 5. Results of the numerical calculation for a frequency of 400 Hz, an incident angle of 30 deg, and three void fractions, 3.2×10^{-5} (solid line), 1.6×10^{-5} (dotted line), and 3.1×10^{-6} (dashed line). (a) The average intensity of the total field versus depth. (b) Difference in dB between bubble and no bubble result.



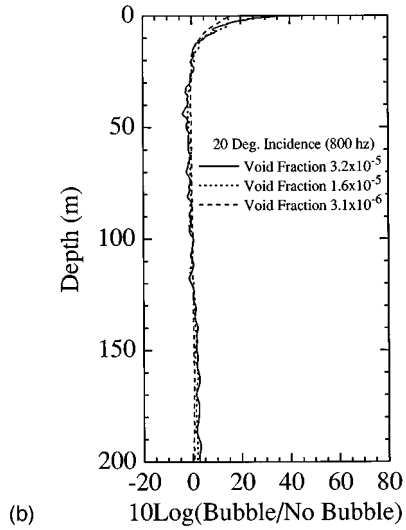


FIG. 6. Results of the numerical calculation for a frequency of 800 Hz, an incident angle of 20 deg, and three void fractions, 3.2×10^{-5} (solid line), 1.6×10^{-5} (dotted line), and 3.1×10^{-6} (dashed line). (a) The average intensity of the total field versus depth. (b) Difference in dB between bubble and no bubble result.

realization of the rough surface at 400 Hz will be spectrally decomposed into horizontal wave number (angle) space. This will be repeated for 10 depths each separated by 1 m. Each horizontal field when spectrally decomposed indicates the (grazing) angle at which the acoustic energy is traveling. Since the sound speed is varying with depth, the angle will likewise vary with depth. Thus by looking at several depths, one can observe the effect that the depth dependent index of refraction has on the angle of propagation and the energy distribution of the total field.

The technique used is based on the work by Thompson (1993). It is a PE-based spectral method, based on an exact relationship between the solution (p) of the wave equation and the solution (ψ) of the standard parabolic equation. This relationship between fields takes the form of a nonlocal integral transform. In contrast, the horizontal wave number spectra of these fields are related locally. This unique spectral relationship forms the basis of an *a posteriori* method of mapping ψ fields into p fields. It is also possible to determine the modal amplitudes of p by spectral analysis of a certain ψ field correlation function. Although this connection between p and ψ fields is restricted to propagation in range-independent environments, the spectral approach can be used to determine the modal properties of fields propagating in range-dependent environments.

The application of the PE-based modal decomposition method to propagation in range-dependent environments is straightforward. Once ψ has been determined for a range-dependent environment, at a specific range and for all mesh depths, this complex field $[\psi(R,z)]$ is used to start a range independent PE calculation for an environment characterized by the ''local'' waveguide properties at range R. This range-independent calculation is marched out to a range determined by the required spectral resolution. A horizontal array of the complex field is collected. This horizontal array is then analyzed spectrally (FFT'd) to determine the mode properties of the initial data $[\psi(R,z)]$.

The results using this technique are shown in Fig. 7 for

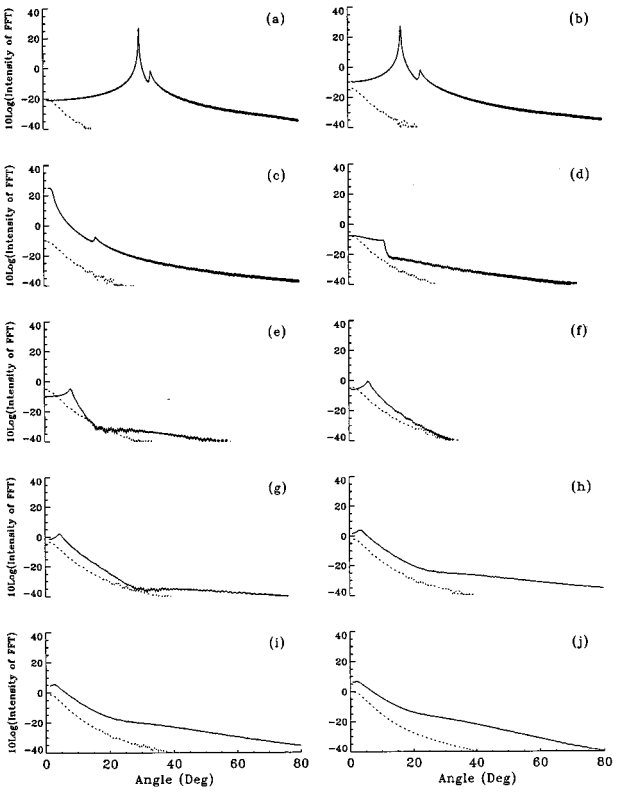


FIG. 7. Comparison of the spectrally decomposed total field versus angle for ten depths. Frequency is 400 Hz, range is 415 m, and the incident angle is 30 deg. Rough surface with bubble layer calculations (solid line) are compared to rough surface with no bubble layer calculations (dotted line). (a) 1-m depth. (b) 2-m depth. (c) 3-m depth. (d) 4-m depth. (e) 5-m depth. (f) 6-m depth. (g) 7-m depth. (h) 8-m depth. (i) 9-m depth. (j) 10-m depth.

various depths, using a frequency of 400 Hz and an incident angle of 30 deg. The complex field (ψ) at a range of 415 m and a Δz of 0.05 m (right edge of the rough surface) was used to start a range-independent (flat surface) PE calculation. Each horizontal array consisted of 8192 points with a Δr of 0.375 m. A Hamming filter was applied to the array and then a FFT consisting of 2^{15} points was performed. The output was then converted from horizontal wave number space to angle space. The horizontal axis is angle (grazing) expressed in degrees and the vertical axis is $10\times\log_{10}$ (intensity of FFT of Ψ). Two curves are plotted, the solid line is when the bubble layer is present and the dotted line is when the bubble layer is absent. The results when the bubble layer is present show energy at predominately two angles. These angles change (toward grazing) with increasing depth showing the refractive effects of the bubble layer, in contrast to the results with no bubble layer, which show maximum energy contribution near grazing at all depths. In addition the bubble layer result is much stronger than the no bubble layer result.

The energy observed for the bubble layer case (solid line in Fig. 7) cannot be isolated into energy going up toward the surface, or down away from the surface. However any energy that is ensonifying the surface is doing so at angles (grazing) steeper than that of the no bubble layer result.

The enhancement of the total field observed near the surface in Figs. 4(b)–6(b) is a result of energy that has been redistributed upon scattering at the rough surface and trapped in the bubble-induced duct. This trapping of energy in the bubble-induced duct can be analyzed in terms of waveguide propagation. Farmer and Vagle (1989) derived an expression for the cutoff frequency in a half-space with an exponential sound-speed profile and a flat surface. Although the perturbed sound-speed profiles are not exactly exponential, and the surface is not flat, the theory provides a first approximation to the problem. The cutoff frequency for the nth mode in the layer is given by

$$f_n^{\text{cutoff}}=\frac{\xi_n^{\text{cutoff}}}{4\pi}\left(\frac{c_1}{d}\right)\left(\frac{c_1}{2\Delta c}\right)^{1/2}, \tag{10}$$

where in this case c_1 is the sound speed in the bubble free medium; Δc is the sound-speed defect at the surface, and d is the e-folding depth of the sound-speed perturbation. The change in sound speed, Δc is 500 m/s, 250 m/s, and 50 m/s for the high, medium, and low void fractions, respectively. For the three cases, the constant d is equal to 1.25, as predicted by Eq. (3) for a 15-m/s wind speed. The parameter ξ_n^{cutoff} corresponding to the cutoff frequency are taken from Farmer and Vagle. They are; 2.40, 5.52, and 8.86 for modes 1, 2, and 3, respectively. The cutoff frequencies for the first three modes are listed in Table II for the three void fractions.

TABLE II. Cutoff frequency (Hz) for the first three modes.

n	$V=3.2\times10^{-5}$	$V=1.6\times10^{-5}$	$V=3.1\times10^{-6}$
1	278	640	1000
2	394	906	1420
3	880	2025	3174

As indicated for the largest void fraction case, modes 1 and 2 can both be trapped for a frequency of 800 Hz, while for 400 Hz mode 1 is trapped and mode 2 is questionable. Mode 1 can also be trapped at the intermediate void fraction case for 800 Hz, but the presence of the rough surface may destroy the effect. Based on this approximate theory, one would expect an enhancement of the total field near the surface for both frequencies with the largest void fraction and for 800 Hz with the intermediate void fraction. For the smallest void fraction, no trapping of energy is predicted by this approximate theory for the frequencies considered here and hence, no enhancement is expected. As stated earlier the trapping occurs because, upon scattering at the rough surface, part of the scattered energy is redirected toward small grazing angles, and remains in the bubble-induced duct. Figures 4(b)–6(b) indicate that there is an enhancement to the total field near the surface for all void fractions, all frequencies, and all incident angles. The fact that the approximate theory does not adequately predict the condition for enhancement at the lowest void fraction is because it cannot take into account the energy redistribution due to the rough surface and it does not adequately model the bubble-induced duct.

The enhancement at depths exceeding the bubble layer, as observed in Figs. 4(b)–6(b), is again due to the combined effect of scattering at the rough surface and the presence of the bubble layer. However, for this case, unlike the near-surface enhancement case, there is no trapping of energy.

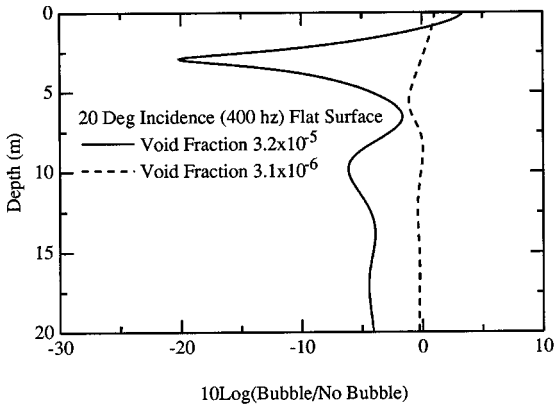


FIG. 8. Difference in dB between bubble layer and no bubble layer result when the surface is flat, for a frequency of 400 Hz, an incident angle of 20 deg, and two void fractions, 3.2×10^{-5} (solid line) and 3.1×10^{-6} (dashed line).

The refractive effect of the layer causes the rough surface to be ensonified at a steeper angle. The energy is then redistributed such that more energy goes into steeper scattered angles than when the bubble layer is not present. To illustrate that this enhancement mechanism is not solely due to the refractive effect from the bubble layer, the ratio of intensities between the bubble and no bubble calculation of the total field when the surface is flat is depicted in Fig. 8. The calculations were made for 400 Hz, the largest and smallest void fractions and for an incident angle of 20 deg. For the largest void fraction, there is a slight enhancement (approximately 3 dB) to the field only near the surface. The smallest void fraction case shows even a smaller enhancement (approximately 1 dB) to the field again near the surface. Note for both cases there is no enhancement of the field at depths below the bubble layer (below approximately 5-m depth). Contrast these results to those of Fig. 4(b) which show an enhancement of approximately 40 dB for the largest and 5 dB for the smallest void fraction cases near the surface. Below the bubble layer, the enhancement of the total field extends down to approximately 50-m depth for the largest void fraction, while the enhancement for the lowest void fraction extends down to approximately 10-m depth. Similar results were obtained for 800 Hz (not shown). It should be emphasized that significant enhancement occurs only when both the rough surface and bubbles are present, and that the enhancement of the total field extends well beyond the bubbly region.

V. CONCLUSIONS

Through numerical calculations, it has been shown that the combination of a range independent subsurface bubble layer and a rough sea surface has a significant effect on the total (forward propagated) field. The effects are the largest near the surface and diminish with depth. It is shown that, at the frequencies considered in this work, the subsurface bubble layer causes a shift of the angle of incidence at the surface toward steeper angles, causing an enhancement of the field, at depths below the bubble layer. At the angles and frequencies considered in this work, and for the wind speed and void fractions involved, the total field close to the surface may be enhanced as much as 40 dB for the largest void fraction considered ($V_0=3.2\times10^{-5}$) and about 5 dB for the smallest ($V_0=3.1\times10^{-6}$). The enhancement of the total field extends well below the bubble region when the two largest void fractions were used.

ACKNOWLEDGMENTS

This work has been supported by the Office of Naval Research. This work has also been supported by a grant of HPC time from the DoD HPC Shared Resource Center, CEWES Cray YMP, and NAVO Cray C90. The authors would like to thank R. Keiffer for his helpful discussions concerning this work.

Breitz, N. D., and Medwin, H. (**1989**). ''Ambient and transient-bubble spectral densities in quiescent seas and under spilling breakers,'' J. Geophys. Res. **94**, C9, 12751–12759.

Clay, C. S., and Medwin, H. (**1977**). *Acoustical Oceanography* (Wiley, New York).

Collins, M. D. (**1993**). ''A split-step Pade solution for the parabolic equation method,'' J. Acoust. Soc. Am. **93**, 1736–1742.

Crowther, P. A. (**1980**). ''Acoustic scattering from near-surface bubble layers,'' in *Cavitation and Inhomogeneities in Underwater Acoustics*, edited by W. Luterborn (Springer-Verlag, Berlin).

Dozier, L. B. (**1984**). ''PERUSE: A numerical treatment of rough surface scattering for the parabolic wave equation,'' J. Acoust. Soc. Am. **75**, 1415–1432.

Farmer, D. M., and Vagle, S. (**1989**). ''Waveguide propagation of ambient noise in the ocean surface bubble layer,'' J. Acoust. Soc. Am. **86**, 1897–1908.

Feuillade, C. (**1995**). ''The collective acoustic properties of water containing resonating air bubbles, and its effect on the propagation of sound,'' J. Acoust. Soc. Am. **98**, 2937(A).

Hall, M. V. (**1989**). ''A comprehensive model of wind generated bubbles in the ocean and predictions of the effects on sound propagation at frequencies up to 40 kHz,'' J. Acoust. Soc. Am. **86**, 1103–1116.

Gilbert, K. E. (**1993**). ''A stochastic model for scattering from the near-surface oceanic layer,'' J. Acoust. Soc. Am. **94**, 3325–3334.

Lamarre, E., and Melville, W. K. (**1994**). ''Sound speed measurements near the ocean surface,'' J. Acoust. Soc. Am. **96**, 3605–3616.

McDaniel, S. T., and Gorman, A. D. (**1982**). ''Acoustic and radar sea surface backscatter,'' J. Geophys. Res. **87**, C6, 4127–4136.

Medwin, H., and Breitz, N. D. (**1989**). ''Instrumentation for in situ acoustical measurement of bubble spectra under breaking waves,'' J. Acoust. Soc. Am. **86**, 739–743.

Novarini, J. C., and Bruno, D. R. (**1982**). ''Effects of the subsurface bubble layer on sound propagation,'' J. Acoust. Soc. Am. **72**, 510–514.

Novarini, J. C., and Norton, G. V. (**1994**). ''Acoustic index of refraction in the background bubble layer of the ocean: An up-dated bubble spectrum and computer program CBUBBLY,'' NRL Tech. Rept. Ref. NRL/FR/7181–93-9432.

Norton, G. V., Novarini, J. C., and Keiffer, R. S. (**1995**). ''Coupling scattering from the sea surface to a one-way marching propagation model via conformal mapping: Validation,'' J. Acoust. Soc. Am. **97**, 2173–2180.

Norton, G. V., and Novarini, J. C. (**1996**). ''The effect of sea surface roughness on waveguide propagation: A coherent approach,'' J. Acoust. Soc. Am. **99**, 2013–2021.

Schneider, H. G. (**1987**). ''Modeling wind dependent acoustic transmission loss due to bubbles in shallow water,'' in *Progress in Underwater Acoustics*, edited by H. Merklinger (Plenum, New York), pp. 509–516.

Su, M. Y., Ling, S. C., and Cartmill, J. (**1988**). ''Optical microbubble measurements in the North Sea,'' in *Sea Surface Sound*, edited by B. D. Kerman (Kluwer Academic, New York).

Thompson, D. J. (**1993**). ''PE-based spectral decomposition,'' PE Workshop II: Proceedings of the second parabolic equation workshop, edited by S. A. Chin-Bing, D. B. King, J. A. Davis, and R. B. Evans, Naval Research Laboratory NRL/BB/7181-93-0001 (U.S. Government Printing Office, Washington, DC, 1993), pp. 296–329.

Thorsos, E. I. (**1988**). ''The validity of the Kirchhoff approximation for rough surface scattering using a Gaussian roughness spectrum,'' J. Acoust. Soc. Am. **83**, 78–92.

Vagle, S., and Farmer, D. M. (**1992**). ''The measurement of bubble-size distributions by acoustical backscatter,'' J. Atmos. Ocean. Technol. **9**, 630–644.

Weston, D. E. (**1989**). ''On the losses due to storm bubbles in oceanic sound transmission,'' J. Acoust. Soc. Am. **86**, 1546–1553.

Wu, J. (**1981**). ''Bubble population and spectra in near-surface ocean: Summary and review,'' J. Geophys. Res. **86**, 453–463.

Ye, Z. (**1997**). ''Further consideration of the waveguide propagation of ambient sound in the ocean bubble layer,'' J. Acoust. Soc. Am. **102**, 788–797.

Leaky waves and the elastic wave resonance reflection on a crystal-thin solid layer interface. II. Leaky waves given rise to by exceptional bulk waves

A. N. Darinskii
Institute of Crystallography, Academy of Sciences of Russia, Leninskii pr. 59, Moscow 117333, Russia

(Received 25 April 1997; accepted for publication 25 November 1997)

A study is made of the elastic wave resonance reflection in a crystal-thin solid layer structure near the velocity of a leaky wave which originates from an exceptional bulk wave. The latter is a homogeneous mode with group velocity parallel to the boundary of a medium which satisfies the condition of the mechanically free surface. On the surface coated with a thin layer, the exceptional bulk wave can transform into the leaky wave and then there arises a narrow interval of incident angles where the coefficients of mode conversion suffer abrupt and strong variations in phase and magnitude. A criterion of occurence of the leaky wave is found. Approximate expressions describing the behavior of the coefficients of mode conversion within the resonance interval of incident angles are derived and analyzed.
[S0001-4966(98)03103-8]

PACS numbers: 43.35.Cg [HEB]

INTRODUCTION

In the preceding paper,[1] see also Ref. 2, it was shown that a thin solid layer coated on the surface of a crystal can cause the transformation of a Rayleigh wave into a leaky one provided this Rayleigh wave is ''supersonic,'' i.e., its phase velocity is larger than that of a bulk wave for certain directions of propagation in the sagittal plane. Owing to the mode mixing through the boundary conditions a possibility of the resonance excitation of the leaky wave by an incident bulk wave then arises. This resonance interaction results in a strong dependence of the phases of the coefficients of mode conversion on the incident angle and in the excitation of surface oscillations with amplitudes far exceeding the amplitude of the incident wave.

Similar resonance phenomena also appear in the case when an exceptional bulk wave exists on the mechanically free surface of a crystal. The investigation of specific properties of leaky waves given rise to by exceptional bulk waves in consequence of the presence of a coating and the analysis of the resonance behavior of the coefficients of mode conversion will be performed in the present work.

An exceptional bulk wave is a bulk wave with group velocity parallel to the boundary of the medium which meets the condition of the mechanically free surface. A bulk mode with group velocity parallel to the surface but not leaving it traction free is customarily termed a limiting wave. It is apparent that there are at least three limiting waves for any orientation of two mutually orthogonal unit vectors **n** and **m** specifying the normal to the surface and the direction of propagation in it. But for a limiting wave to become exceptional the orientation of the pair **m** and **n** must be specially selected. The simplest example is the situation when the plane (**m**, **n**) is a plane of crystallographic symmetry. The limiting wave which corresponds to the branch of shear waves polarized perpendicular to the plane of symmetry then meets the condition of free surface. However, exceptional bulk waves have been shown to exist in crystals of unrestricted symmetry and for any of the three branches of uniform modes.[3,4] The set of orientations of **m**, **n** allowing the existence of exceptional bulk waves occupies continuous one-dimensional subspaces (lines) in the three-dimensional space of orientation angles which define the position of the pair **m**, **n**.

It should be underlined that a leaky wave branch and accordingly the resonance reflection cannot occur in the case when the substrate is isotropic. Indeed, a necessary requirement is that an exceptional wave should not be the slowest bulk wave for a given direction **m**. But in isotropic media, on the one hand, the shear wave, which satisfies the condition of free surface, is the slowest bulk mode. On the other hand, the limiting longitudinal wave does not meet the condition of free surface.

As in Ref. 1, we shall assume that both the substrate and the layer are of arbitrary anisotropy. In the next section we shall discuss briefly a number of important relations to be frequently used in the following. Afterwards, we shall derive approximate expressions for the speed of the leaky wave and for the coefficients of mode conversion. Finally, we shall analyze the behavior of the coefficients of mode conversion under conditions of the resonance excitation of the leaky wave.

I. GENERAL REMARKS

Elastic wave propagation in anisotropic media is governed by the equation

$$\rho \frac{\partial^2 u_i}{\partial t^2} = C_{ijkl} \frac{\partial^2 u_k}{\partial x_j \partial x_l}, \quad i,j,k,l = 1,2,3, \tag{1}$$

where ρ is the density, C_{ijkl} are the elastic moduli, $\mathbf{u}(\mathbf{r},t)$ is the displacement vector, and $\mathbf{r}$ is the position vector. Let us consider a displacement field

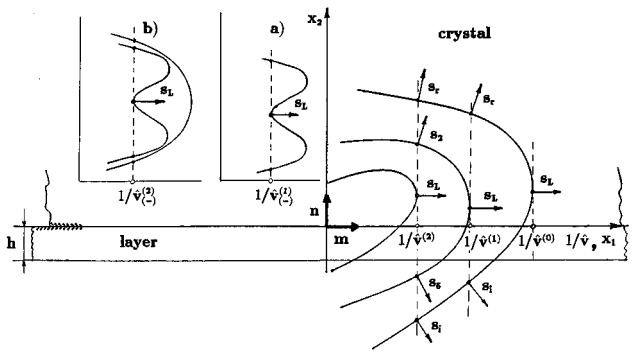


FIG. 1. Geometry of the problem. Dashed lines are vertical tangents to the slowness curves of the bulk modes in the crystal and the tangency points represent limiting waves. Insets **a)** and **b)** show a concave part of a slowness curve. $\mathbf{s}_\alpha$, $\alpha=i,r,2,5$, are the projections of the group velocities of incident and reflected waves onto the plane $(\mathbf{m},\mathbf{n})$; $\mathbf{s}_L$ is the projection of the group velocity of a limiting wave onto the plane $(\mathbf{m},\mathbf{n})$.

$$\mathbf{u}(\mathbf{r},t)=\mathbf{a}(x_2)e^{ik[x_1-vt]}, \quad x_1=\mathbf{mr}, \quad x_2=\mathbf{nr}, \tag{2}$$

moving along the direction $\mathbf{m}$ (Fig. 1) with trace speed $v=\omega/k$, where ω is the frequency and k is the tangential projection of the wave vector, the vector function $\mathbf{a}(x_2)$ describes the displacement distribution depending on the coordinate x_2. The wave (2) produces on unit area of the surface $x_2=\text{const}$ a traction with amplitude $\mathbf{f}(x_2)=-ik\mathbf{l}(x_2)$, where

$$\mathbf{l}(x_2)=-\left[(nm)\mathbf{a}+\frac{1}{ik}(nn)\frac{\mathrm{d}\mathbf{a}}{\mathrm{d}x_2}\right], \tag{3}$$

the symbols (nm), (nn) denote 3×3 matrices with elements

$$(nm)_{ij}=C_{kijl}n_km_l, \quad (nn)_{ij}=C_{kijl}n_kn_l. \tag{4}$$

In combination with (3), Eq. (1) yields the equation[5,6]

$$\hat{\mathbf{N}}\boldsymbol{\mu}=\frac{1}{ik}\frac{\mathrm{d}\boldsymbol{\mu}}{\mathrm{d}x_2} \tag{5}$$

upon the six-component vector column $\boldsymbol{\mu}(x_2)=(\mathbf{a}(x_2),\mathbf{l}(x_2))^t$; the symbol t means the transposition. $\hat{\mathbf{N}}$ is a 6×6 real matrix,

$$\hat{\mathbf{N}}=-\begin{Bmatrix}(nn)^{-1}(nm) & (nn)^{-1}\\ (mn)(nn)^{-1}(nm)-(mm)+\rho v^2\hat{\mathbf{I}} & (mn)(nn)^{-1}\end{Bmatrix}, \tag{6}$$

$(nn)^{-1}$ is the inverse of the matrix (nn), (mn) is the transpose of the matrix (nm), and $\hat{\mathbf{I}}$ is the 3×3 unit matrix. Note that the matrix $\hat{\mathbf{N}}$ obeys the relation

$$(\hat{\mathbf{T}}\hat{\mathbf{N}})^t=\hat{\mathbf{T}}\hat{\mathbf{N}}, \quad \hat{\mathbf{T}}=\begin{pmatrix}\hat{\mathbf{O}} & \hat{\mathbf{I}}\\ \hat{\mathbf{I}} & \hat{\mathbf{O}}\end{pmatrix}, \tag{7}$$

and $\hat{\mathbf{O}}$ is the 3×3 zero matrix.

Generally matrix (6) is not degenerate so that Eq. (5) has then six independent solutions of the form $\boldsymbol{\mu}_\alpha(x_2)=\boldsymbol{\xi}_\alpha\exp(ikp_\alpha x_2)$. The vectors $\boldsymbol{\xi}_\alpha$ and the parameters p_α are found from the eigenvalue problem

$$\hat{\mathbf{N}}\boldsymbol{\xi}_\alpha=p_\alpha\boldsymbol{\xi}_\alpha, \quad p_\alpha\neq p_\beta \quad \text{if} \quad \alpha\neq\beta, \tag{8}$$

$$\boldsymbol{\xi}_\alpha=\begin{pmatrix}\mathbf{A}_\alpha\\ \mathbf{L}_\alpha\end{pmatrix}, \tag{9}$$

where $\mathbf{A}_\alpha$ is the polarization vector of the mode α, $\mathbf{L}_\alpha$ is the amplitude of vector (3), $\mathbf{L}_\alpha=-[(nm)+p_\alpha(nn)]\mathbf{A}_\alpha$. Thus for a given v Eq. (1) generally admits six partial solutions

$$\mathbf{u}_\alpha(\mathbf{r},t)=\mathbf{A}_\alpha e^{ik[(x_1+p_\alpha x_2-vt]}, \tag{10}$$

the six parameters p_α being distinct and occurring both pairwise real and in pairs of complex conjugate.

The trace speed at which the limiting bulk wave appears is referred to as the limiting speed. In Fig. 1 these are $\hat{v}^{(0)}$, $\hat{v}^{(1)}$, $\hat{v}^{(2)}$, $\hat{v}^{(1)}_{(-)}$, and $\hat{v}^{(2)}_{(-)}$. We shall also use a general notation $\hat{v}$ for a limiting speed. The situation when $v=\hat{v}$ is termed a transonic state. The concept of transonic states has been introduced in papers devoted to the general theory of Rayleigh waves in anisotropic media and detailed analyses of properties of transonic states in relation to the surface wave theory are given in Refs. 6–10. In the present paper we confine ourselves to considering so-called nonzero curvature type 1 transonic states (Fig. 1) which are characterized by that only one limiting bulk wave exists at $\hat{v}$, and the curvature of the slowness curve at the tangency point is not equal to zero, and, besides, the tangency point is not a point of contact of slowness curves. This is the most often met situation in practice, e.g., transonic states associated with the shear waves, except for certain orientations of $\mathbf{m}$, $\mathbf{n}$ in the plane of crystallographic symmetry, refer to type 1.

Of our concern are the positive curvature transonic states $\hat{v}^{(1)}$, $\hat{v}^{(2)}$ and the negative curvature transonic states $\hat{v}^{(1)}_{(-)}$, $\hat{v}^{(2)}_{(-)}$. It is seen from Fig. 1 that there are two and four bulk waves with group velocity not parallel to the surface at $\hat{v}^{(1)}$, $\hat{v}^{(1)}_{(-)}$ and $\hat{v}^{(2)}$, $\hat{v}^{(2)}_{(-)}$, respectively. We denote a pair of bulk waves by the subscripts $\alpha=i$ (incident) and $\alpha=r$ (reflected) and later we shall investigate the reflection problem for the mode $\alpha=i$. The subscripts $\alpha=2,5$ will be ascribed to the pair of partial modes which have the complex conjugate p_α's at $\hat{v}^{(1)}$, $\hat{v}^{(1)}_{(-)}$ and to the second pair of bulk waves at $\hat{v}^{(2)}$, $\hat{v}^{(2)}_{(-)}$. It will be born in mind that

$$\begin{aligned}&\mathrm{Im}(p_2)>0 \quad \text{at} \quad \hat{v}^{(1)}, \ \hat{v}^{(1)}_{(-)};\\ &\qquad \alpha=2 \quad \text{is the reflected wave}\\ &\qquad \text{and} \quad \alpha=5 \quad \text{is the incident wave at} \quad \hat{v}^{(2)}, \ \hat{v}^{(2)}_{(-)}.\end{aligned} \tag{11}$$

Finally, there are two modes (10) which have the complex conjugate p_α's from one side of a transonic state ($v<\hat{v}^{(1)}$, $v<\hat{v}^{(2)}$ and $v>\hat{v}^{(1)}_{(-)}$, $v>\hat{v}^{(2)}_{(-)}$) and coalesce to the limiting wave at $\hat{v}$. In the adjacent speed range ($v>\hat{v}^{(1)}$, $v>\hat{v}^{(2)}$, and $v<\hat{v}^{(1)}_{(-)}$, $v<\hat{v}^{(2)}_{(-)}$) they become bulk waves with distinct real p_α's. We shall identify these modes by the subscripts $\alpha=3,6$ and choose the ordering

$$\begin{aligned}&\mathrm{Im}(p_3)>0 \quad \text{if} \quad p_3, \ p_6 \quad \text{are complex;}\\ &\qquad \alpha=3 \quad \text{is the reflected wave} \quad \text{and}\\ &\qquad \alpha=6 \quad \text{is the incident wave} \quad \text{if} \quad p_3, \ p_6 \quad \text{are real.}\end{aligned} \tag{12}$$

From the point of view of the properties of the matrix $\hat{\mathbf{N}}$, a nonzero curvature type 1 transonic state is a speed at which the matrix $\hat{\mathbf{N}}(v)$ (6) becomes non-semisimple degenerate because of its two eigenvalues coalescing to one real eigenvalue, $p_3(\hat{v})=p_6(\hat{v})=p_{d3}$ in our notations.[6] The eigenvalue problem for matrix (6) associated with $v=\hat{v}$ reads

$$\begin{aligned}&\hat{\mathbf{N}}\boldsymbol{\xi}_\alpha=p_\alpha\boldsymbol{\xi}_\alpha, \qquad \alpha=i,r,2,5,\\ &\hat{\mathbf{N}}\boldsymbol{\xi}_{d3}=p_{d3}\boldsymbol{\xi}_{d3}, \quad \hat{\mathbf{N}}\boldsymbol{\xi}_{d6}=p_{d3}\boldsymbol{\xi}_{d6}-\boldsymbol{\xi}_{d3},\end{aligned} \tag{13}$$

where $\boldsymbol{\xi}_{d3}$ is the degenerate eigenvector and $\boldsymbol{\xi}_{d6}$ is the generalized eigenvector, so that Eq. (5) has only five solutions of the form $\boldsymbol{\mu}_\alpha(x_2)=\boldsymbol{\xi}_\alpha\exp(ikp_\alpha x_2)$, $\alpha=i,r,2,5,d3$. The solution $\boldsymbol{\mu}_{d3}(x_2)$ corresponds to the limiting wave, and the sixth solution of Eq. (5) is $\boldsymbol{\mu}_{d6}(x_2)=(\boldsymbol{\xi}_{d6}-ikx_2\boldsymbol{\xi}_{d3})\exp(ikp_{d3}x_2)$.

Using the perturbation technique one can show that $\boldsymbol{\xi}_\alpha$, p_α, $\alpha=i,r,2,5$, are differentiable functions at $v=\hat{v}$ and, hence, near $\hat{v}$

$$\begin{aligned}&\boldsymbol{\xi}_\alpha(v)\approx\boldsymbol{\xi}_\alpha+\frac{\partial\boldsymbol{\xi}_\alpha}{\partial v}(v-\hat{v}), \quad p_\alpha(v)\approx p_\alpha+\frac{\partial p_\alpha}{\partial v}(v-\hat{v}),\\ &\qquad \alpha=i,r,2,5,\end{aligned} \tag{14}$$

while $\boldsymbol{\xi}_\alpha$ and p_α, $\alpha=3,6$, behave in the vicinity of $\hat{v}$ as

$$\begin{aligned}&\boldsymbol{\xi}_3(v)\approx\boldsymbol{\xi}_{d3}-\Delta p\boldsymbol{\xi}_{d6}+(\Delta p)^2\sum_{\alpha\neq d6}a_\alpha^{(2)}\boldsymbol{\xi}_\alpha,\\ &\qquad p_3(v)\approx p_{d3}+\Delta p,\\ &\boldsymbol{\xi}_6(v)\approx\boldsymbol{\xi}_{d3}+\Delta p\boldsymbol{\xi}_{d6}+(\Delta p)^2\sum_{\alpha\neq d6}a_\alpha^{(2)}\boldsymbol{\xi}_\alpha,\\ &\qquad p_3(v)\approx p_{d3}-\Delta p,\end{aligned} \tag{15}$$

where $\boldsymbol{\xi}_\alpha$, $\alpha=i,r,2,5$ are referred to $v=\hat{v}$,

$$\Delta p=\sqrt{\rho}|\mathbf{A}_{d3}|f(v), \quad \text{at } \hat{v}=\hat{v}^{(1)},\ \hat{v}^{(2)}, \tag{16}$$

$$\Delta p=i\sqrt{\rho}|\mathbf{A}_{d3}|f(v), \quad \text{at } \hat{v}=\hat{v}_{(-)}^{(1)},\ \hat{v}_{(-)}^{(2)}, \tag{17}$$

$$f(v)=\begin{cases}\sqrt{v^2-\hat{v}^2}, & v>\hat{v}\\ i\sqrt{\hat{v}^2-v^2}, & v<\hat{v},\end{cases} \tag{18}$$

and $a_\alpha^{(2)}$, $\alpha=i,r,2,5,d3$, are certain coefficients. The ''unusual'' form of the perturbation series (15) is related to the matrix $\hat{\mathbf{N}}$ being non-semisimple degenerate at $\hat{v}$. In more detail, properties of the eigenvectors of the matrix $\hat{\mathbf{N}}$ as well as of the vectors $\mathbf{L}_\alpha$, $\alpha=i,r,2,5,d3$, and $\mathbf{L}_{d6}$, where $\mathbf{L}_{d6}$ is the ''traction'' part (the last three components) of the vector $\boldsymbol{\xi}_{d6}$, are discussed in Appendix. Note that the problem of deriving perturbation series for the non-semisimple degenerate matrix $\hat{\mathbf{N}}$ also arises in the theory of static elasticity, specifically in the theory of dislocations.[11]

II. EXCEPTIONAL BULK AND LEAKY WAVES

We consider that a solid layer of thickness h is superimposed on the surface of a semi-infinite medium and perfect rigid contact between them is realized (Fig. 1). The exterior face of the layer is mechanically free. The substrate and the layer are taken to have arbitrary anisotropy. Our final objective is to study specific features of the behavior of the coefficients of mode conversion near a transonic state at which the limiting wave is exceptional.

Accounting for the notations (11) and (12) we seek for the solution of the reflection problem in terms of a linear combination of modes $\alpha=i,r,2,3$ (crystal) and in terms of a linear combination of all six partial modes $\alpha=1,\ldots,6$ obeying Eq. (1) at a given v (layer),

$$\mathbf{u}(\mathbf{r},t)=\sum_{\alpha=i,r,2,3}b_\alpha\mathbf{A}_\alpha e^{ik[x_1+p_\alpha x_2-vt]}, \quad x_2>0, \tag{19}$$

$$\mathbf{u}^{(f)}(\mathbf{r},t)=\sum_{\alpha=1}^{6}b_\alpha^{(f)}\mathbf{A}_\alpha^{(f)}e^{ik[x_1+p_\alpha^{(f)}x_2-vt]}, \quad -h<x_2<0. \tag{20}$$

The coefficients b_α, $b_\alpha^{(f)}$ must be found to satisfy the boundary conditions

$$\sum_{\alpha=1}^{6}b_\alpha^{(f)}\boldsymbol{\xi}_\alpha^{(f)}=\sum_{\alpha=i,r,2,3}b_\alpha\boldsymbol{\xi}_\alpha, \quad x_2=0, \tag{21}$$

$$\sum_{\alpha=1}^{6}b_\alpha^{(f)}\mathbf{L}_\alpha^{(f)}e^{-ip_\alpha^{(f)}H}=0, \quad x_2=-h, \tag{22}$$

where $H=kh$, the superscript (f) is assigned to the parameters of partial modes in the layer.

For our purposes it is sufficient to envisage only the coefficients $b_\alpha(v,H)$, $\alpha=i,r,2,3$. Assuming the matrix (6) for the layer, $\hat{\mathbf{N}}^{(f)}$, not to be non-semisimple degenerate at $v=\hat{v}$, one can take advantage of the relation $\hat{\mathbf{N}}^{(f)}\boldsymbol{\xi}_\beta^{(f)}=p_\beta^{(f)}\boldsymbol{\xi}_\beta^{(f)}$, $\beta=1,\ldots,6$, to reduce Eqs. (21) and (22) to

$$\sum_{\alpha=i,r,2,3}b_\alpha\mathbf{L}'_\alpha=0, \tag{23}$$

where $\mathbf{L}'_\alpha$ is the ''traction'' part of the vector $\boldsymbol{\xi}'_\alpha=\exp(-iH\hat{\mathbf{N}}^{(f)})\boldsymbol{\xi}_\alpha$,

$$\begin{aligned}&\boldsymbol{\xi}'_\alpha=\boldsymbol{\xi}_\alpha-iH\boldsymbol{\xi}_\alpha^{(1)}-0.5H^2\boldsymbol{\xi}_\alpha^{(2)}+\cdots,\\ &\boldsymbol{\xi}_\alpha^{(n)}=\{\hat{\mathbf{N}}^{(f)}\}^n\boldsymbol{\xi}_\alpha, \quad n=1,2,...,\end{aligned} \tag{24}$$

the symbol $\{\hat{\mathbf{N}}^{(f)}\}^n$ stands for the matrix $\hat{\mathbf{N}}^{(f)}$ in power n. Thus we have

$$\begin{aligned}&b_i(v,H)=[\mathbf{L}'_r\mathbf{L}'_2\mathbf{L}'_3], \quad b_r(v,H)=[\mathbf{L}'_i\mathbf{L}'_3\mathbf{L}'_2],\\ &b_2(v,H)=[\mathbf{L}'_r\mathbf{L}'_3\mathbf{L}'_i], \quad b_3(v,H)=[\mathbf{L}'_r\mathbf{L}'_i\mathbf{L}'_2],\end{aligned} \tag{25}$$

where $[\cdots]$ designates the scalar triple product.

In what follows, the quantity H will be viewed as a small parameter, $H\ll 1$. It occurs then that in the vicinity of a transonic state $\hat{v}$ the most significant peculiarities in the behavior of the coefficients of mode conversion appear provided the exceptional wave transforms into the leaky wave. In the crystal the wave field of the latter involves the modes $\alpha=r,2,3$ and the leaky wave speed v_l is a complex root of the equation $b_i(v,H)=0$,

$$v_l = v_l' - iv_l'', \quad v_l'' > 0, \tag{26}$$

the value of v_l' slightly differing from $\hat{v}$ and v_l'' being far smaller than $\hat{v}$. The imaginary component characterizes the attenuation due to the outflow of the energy from the surface, this radiation being provided by the ''reflected'' constituents of the leaky wave field (the modes $\alpha = r$ near $\hat{v}^{(1)}$, $\hat{v}_{(-)}^{(1)}$ and $\alpha = r,2$ near $\hat{v}^{(2)}$, $\hat{v}_{(-)}^{(2)}$).

Let us estimate v_l and the amplitudes of modes incorporated into the leaky wave. One can limit oneself to considering the neighborhood of positive curvature transonic states because in view of (16)–(18) and (A16) and (A17) the results to be obtained for $\hat{v}^{(1)}$ and $\hat{v}^{(2)}$ admit a straightforward extension to $\hat{v}_{(-)}^{(1)}$ and $\hat{v}_{(-)}^{(2)}$, respectively.

We begin with representing $b_i(v,H)$ in the form $b_i(v,H) \propto \Delta p + \Psi$, where Ψ does not depend on v and contains the terms $O(H)$ and $O(H^2)$. Owing to (24), (15) the vectors $\mathbf{L}_\alpha$, $\alpha = i,r,2,3$, can be written down as

$$\begin{aligned}
&\mathbf{L}_\alpha' \approx \mathbf{L}_\alpha - iH\mathbf{L}_\alpha^{(1)}, \quad \alpha = i,r,2, \\
&\mathbf{L}_3' \approx \mathbf{L}_{d3}' - \Delta p \mathbf{L}_{d6}' + (\Delta p)^2 \sum_{\alpha \neq d6} a_\alpha^{(2)} \mathbf{L}_\alpha' \\
&\quad \approx -\Delta p \mathbf{L}_{d6} - iH(\mathbf{L}_{d3}^{(1)} - \Delta p \mathbf{L}_{d6}^{(1)}) - 0.5H^2 \mathbf{L}_{d3}^{(2)} \\
&\qquad + (\Delta p)^2 \sum_{\alpha \neq d3,d6} a_\alpha^{(2)} \mathbf{L}_\alpha,
\end{aligned} \tag{27}$$

where $\mathbf{L}_\alpha^{(n)}$ are the ''traction'' parts of the vectors $\boldsymbol{\xi}_\alpha^{(n)}$. We have also taken into account that the limiting wave is exceptional, i.e., $\mathbf{L}_{d3} = 0$.

After substitution of (27) for $\mathbf{L}_\alpha'$ in $b_i(v,H)$ (25) one obtains

$$\begin{aligned}
b_i(v,H) &\approx -\Delta p\{[\mathbf{L}_r\mathbf{L}_2\mathbf{L}_{d6}] - iHX_1\} - iH[\mathbf{L}_r\mathbf{L}_2\mathbf{L}_{d3}^{(1)}] \\
&\quad - H^2 X_2 + (\Delta p)^2 \sum_{\alpha \neq d3,d6} a_\alpha^{(2)} [\mathbf{L}_r\mathbf{L}_2\mathbf{L}_\alpha] \\
&\approx -[\mathbf{L}_r\mathbf{L}_2\mathbf{L}_{d6}] \left\{ \Delta p + i \frac{[\mathbf{L}_r\mathbf{L}_2\mathbf{L}_{d3}^{(1)}]}{[\mathbf{L}_r\mathbf{L}_2\mathbf{L}_{d6}]} H + \Xi_i H^2 \right\}
\end{aligned} \tag{28}$$

with

$$X_1 = [\mathbf{L}_r^{(1)}\mathbf{L}_2\mathbf{L}_{d6}] + [\mathbf{L}_r\mathbf{L}_2^{(1)}\mathbf{L}_{d6}] + [\mathbf{L}_r\mathbf{L}_2\mathbf{L}_{d6}^{(1)}], \tag{29}$$

$$X_2 = [\mathbf{L}_r^{(1)}\mathbf{L}_2\mathbf{L}_{d3}^{(1)}] + [\mathbf{L}_r\mathbf{L}_2^{(1)}\mathbf{L}_{d3}^{(1)}] + 0.5[\mathbf{L}_r\mathbf{L}_2\mathbf{L}_{d3}^{(2)}], \tag{30}$$

$$\Xi_i = \frac{1}{[\mathbf{L}_r\mathbf{L}_2\mathbf{L}_{d6}]} \left\{ X_2 - X_1 \frac{[\mathbf{L}_r\mathbf{L}_2\mathbf{L}_{d3}^{(1)}]}{[\mathbf{L}_r\mathbf{L}_2\mathbf{L}_{d6}]} \right\}. \tag{31}$$

Since the velocity range to be of our concern is $|v - \hat{v}| \propto \hat{v}H^2$, having in view (14), we have set $\mathbf{L}_\alpha(v) \approx \mathbf{L}_\alpha(\hat{v})$, $\mathbf{L}_\alpha^{(n)}(v) \approx \mathbf{L}_\alpha^{(n)}(\hat{v})$, and $H \approx \omega h/\hat{v}$. By virtue of (A23) all determinants at $(\Delta p)^2$ turn then into zero. In the final expression (28) we have discarded the term $-iHX_1$ which is small in comparison with $[\mathbf{L}_r\mathbf{L}_2\mathbf{L}_{d6}]$.

Using (A1)–(A3) allows one to represent the vectors $\boldsymbol{\xi}_\alpha^{(n)}(\hat{v})$ as

$$\boldsymbol{\xi}_\alpha^{(n)} = \sum_{\beta = i,r,2,5} (\{\hat{N}^{(f)}\}^n)_{\beta\alpha} \boldsymbol{\xi}_\beta + (\{\hat{N}^{(f)}\}^n)_{6\alpha} \boldsymbol{\xi}_{d3} + (\{\hat{N}^{(f)}\}^n)_{3\alpha} \boldsymbol{\xi}_{d6} \tag{32}$$

and then, accounting for $\mathbf{L}_{d3} = 0$, in (28)–(31)

$$\mathbf{L}_\alpha^{(n)} = \sum_{\beta = i,r,2,5} (\{\hat{N}^{(f)}\}^n)_{\beta\alpha} \mathbf{L}_\beta + (\{\hat{N}^{(f)}\}^n)_{3\alpha} \mathbf{L}_{d6}, \tag{33}$$

where $(\{\hat{N}^{(f)}\}^n)_{\alpha\beta}$ stands for the contraction of the matrix $\hat{\mathbf{T}}\{\hat{\mathbf{N}}^{(f)}\}^n$ with the vectors $\boldsymbol{\xi}_\beta$, $\boldsymbol{\xi}_\alpha$, $(\{\hat{N}^{(f)}\}^n)_{\alpha\beta} = \boldsymbol{\xi}_\alpha \cdot \hat{\mathbf{T}}\{\hat{\mathbf{N}}^{(f)}\}^n \boldsymbol{\xi}_\beta$, the indexes $\alpha = 3$ or $\alpha = 6$ meaning that $\boldsymbol{\xi}_{d3}$ or $\boldsymbol{\xi}_{d6}$, respectively, are involved in the contraction. Replacing $\mathbf{L}_{d3}^{(1)}$ in (28) by (33) and accounting for (A23) leads to that in the lowest approximation the equation $b_i(v,H) = 0$ becomes

$$\Delta p + iH(\hat{N}^{(f)})_{33} = 0, \tag{34}$$

because of (A16) the value of $(\hat{N}^{(f)})_{33}$ being real. The examination of Eq. (34) shows that due to (16) and (18) it has a solution, and consequently $b_i(v,H) = 0$ will have a solution in the vicinity of $\hat{v}$, provided

$$(\hat{N}^{(f)})_{33} < 0. \tag{35}$$

Inequality (35) does not put severe constraints upon material parameters of the substrate and the layer. For example let the substrate be a cubic medium and $\mathbf{n} = (\bar{1}10)$, $\mathbf{m} = (110)$. The exceptional wave is the SH-mode propagating along $\mathbf{m}$ with the speed $v_{SH} = (c_{44}/\rho)^{1/2}$. Let the layer have cubic symmetry too, $\mathbf{n}$ being directed along the crystallographic direction (010) and $\mathbf{m}$ making an angle φ with the direction (100) in the layer; this nonsymmetric orientation of the coating is required in order to furnish the ''mixing'' of modes in the substrate [see the discussion after Eq. (46)]. We obtain

$$\begin{aligned}
(\hat{N}^{(f)})_{33} &= -\frac{c_{44}^{(f)}}{c_{44}} \left\{ \frac{\rho^{(f)}}{\rho} \frac{c_{44}}{c_{44}^{(f)}} - 1 - \sin^2(2\varphi) \left[\frac{1}{\eta^{(f)}} - 1 \right] \right\} \\
&< 0,
\end{aligned} \tag{35'}$$

where $c_{ij}^{(f)}$ and $\rho^{(f)}$ are the elastic moduli and the density of the layer, respectively, $\eta^{(f)} = 2c_{44}^{(f)}/(c_{11}^{(f)} - c_{12}^{(f)})$ is the factor of anisotropy.

If (35) holds true, then

$$v_l \approx \hat{v}(1 - a_1 H^2), \quad a_1 = \frac{(\hat{N}^{(f)})_{33}^2}{2\rho\hat{v}^2 \mathbf{A}_{d3}^2} > 0. \tag{36}$$

Thus the first correction is real and it has the order of $\hat{v}H^2$ (as $a_1 \propto 1$). Note that due to $\mathbf{L}_{d3} = 0$ the contraction $(\hat{N}^{(f)})_{33}$ reduces to $(\hat{N}^{(f)})_{33} = \mathbf{A}_{d3} \cdot \hat{\mathbf{X}}\mathbf{A}_{d3}$, where $\hat{\mathbf{X}}$ is the 3×3 left-hand side bottom block of $\hat{\mathbf{N}}^{(f)}$.

Let us evaluate the parameter Ξ_i (31). Accounting for the relations

$$(\hat{N}^{(f)})_{\alpha\beta} = (\hat{N}^{(f)})_{\beta\alpha}, \tag{37}$$

$$(\{\hat{N}^{(f)}\})_{33} = \sum_{\alpha = i,r,2,5} (\hat{N}^{(f)})_{3\alpha}^2 + 2(\hat{N}^{(f)})_{33}(\hat{N}^{(f)})_{36}, \tag{38}$$

which follow from (7), (A4), and using (A20)–(A23), one can recast Ξ_i into the form

$$\Xi_i=\frac{1}{2[\mathbf{L}_i\mathbf{L}_5\mathbf{L}_{d6}]}\left\{\frac{(\hat{N}^{(f)})_{3P}(\hat{N}^{(f)})_{3Q}}{[\mathbf{L}_i\mathbf{L}_2\mathbf{L}_{d6}]}+\frac{(\hat{N}^{(f)})_{3J}(\hat{N}^{(f)})_{3S}}{[\mathbf{L}_r\mathbf{L}_5\mathbf{L}_{d6}]}\right\}, \tag{39}$$

where $(\hat{N}^{(f)})_{3P}$, etc. signify the contractions of the matrix $\hat{\mathbf{T}}\hat{\mathbf{N}}^{(f)}$ with the vector $\boldsymbol{\xi}_{d3}$ and the vectors

$$\mathbf{P}=\sum_{\alpha=i,r,2}[\mathbf{L}_\alpha\mathbf{L}_5\mathbf{L}_{d6}]\boldsymbol{\xi}_\alpha,\quad \mathbf{Q}=\sum_{\alpha=i,r,5}[\mathbf{L}_\alpha\mathbf{L}_2\mathbf{L}_{d6}]\boldsymbol{\xi}_\alpha,$$
$$\mathbf{S}=\sum_{\alpha=i,2,5}[\mathbf{L}_\alpha\mathbf{L}_r\mathbf{L}_{d6}]\boldsymbol{\xi}_\alpha,\quad \mathbf{J}=\sum_{\alpha=r,2,5}[\mathbf{L}_\alpha\mathbf{L}_i\mathbf{L}_{d6}]\boldsymbol{\xi}_\alpha \tag{40}$$

[e.g., $(\hat{N}^{(f)})_{3P}=\boldsymbol{\xi}_{d3}\cdot\hat{\mathbf{T}}\hat{\mathbf{N}}^{(f)}\mathbf{P}$]. Due to (A23) these vectors are equal to

$$\mathbf{P}=\begin{pmatrix}\mathbf{A}_P\\ \mathbf{O}\end{pmatrix},\ \mathbf{Q}=\begin{pmatrix}\mathbf{A}_Q\\ \mathbf{O}\end{pmatrix},\ \mathbf{S}=\begin{pmatrix}\mathbf{A}_S\\ \mathbf{O}\end{pmatrix},\ \mathbf{J}=\begin{pmatrix}\mathbf{A}_J\\ \mathbf{O}\end{pmatrix}, \tag{41}$$

their ''polarization'' components being found by multiplying (A19) from the left by the relevant vector product $\mathbf{L}_\alpha\times\mathbf{L}_{d6}$,

$$\mathbf{A}_P=\mathbf{L}_5\times\mathbf{L}_{d6},\quad \mathbf{A}_Q=\mathbf{L}_2\times\mathbf{L}_{d6},\quad \mathbf{A}_S=\mathbf{L}_r\times\mathbf{L}_{d6},$$
$$\mathbf{A}_J=\mathbf{L}_i\times\mathbf{L}_{d6}, \tag{42}$$

so that $(\hat{N}^{(f)})_{3P}$, etc. reduce to the contractions of the aforementioned matrix $\hat{\mathbf{X}}$ with $\mathbf{A}_{d3}$ and $\mathbf{A}_P$, etc. The vectors **P** and **Q** are complex conjugate at $\hat{v}^{(1)}$ but at $\hat{v}^{(2)}$ they are real and imaginary, respectively. The vectors **J** and **S** are real and imaginary, respectively, both at $\hat{v}^{(1)}$ and $\hat{v}^{(2)}$.

With the aid of (A6)–(A8), (A16), and (A20) one can show that

$$\mathrm{Re}(\Xi_i)=\frac{|(\hat{N}^{(f)})_{3Q}|^2}{2|[\mathbf{L}_i\mathbf{L}_2\mathbf{L}_{d6}]|^2},$$
$$\mathrm{Im}(\Xi_i)=\frac{(\hat{N}^{(f)})_{3J}(\hat{N}^{(f)})_{3S}}{2i[\mathbf{L}_i\mathbf{L}_5\mathbf{L}_{d6}][\mathbf{L}_r\mathbf{L}_5\mathbf{L}_{d6}]}\quad\text{at } \hat{v}^{(1)}, \tag{43}$$

while at the transonic state $\hat{v}^{(2)}$ the parameter Ξ_i becomes real,

$$\Xi_i=\mathrm{Re}(\Xi_i)=\frac{1}{2[\mathbf{L}_i\mathbf{L}_5\mathbf{L}_{d6}]^2}\{(\hat{N}^{(f)})_{3P}^2+(\hat{N}^{(f)})_{3J}^2\}$$
$$>0\quad\text{at } \hat{v}^{(2)}. \tag{44}$$

Here we used the fact that due to (A23) any of the vectors (41) can be expressed in the form of a linear combination of any two of the other three vectors.

To summarize,

$$b_i(v,H)\approx-[\mathbf{L}_r\mathbf{L}_2\mathbf{L}_{d6}]\{\Delta p+i(\hat{N}^{(f)})_{33}H+\Xi_iH^2\} \tag{45}$$

and, equating (45) to zero, we find v_l' and v_l'' in (26) to within H^3:

$$v_l'\approx\hat{v}\left\{1-a_1H^2+2a_1\frac{\mathrm{Im}(\Xi_i)}{|(\hat{N}^{(f)})_{33}|}H^3\right\},$$
$$v_l''\approx 2a_1\hat{v}\frac{\mathrm{Re}(\Xi_i)}{|(\hat{N}^{(f)})_{33}|}H^3>0. \tag{46}$$

Thus the leaky wave appears provided the material constants of the crystal and the layer fulfill inequality (35). The imaginary part of the leaky wave speed v_l is proportional to H^3 and the real part of v_l is smaller than the limiting speed by a quantity of the order $\hat{v}H^2$. Note that in regards to the aforementioned features this leaky wave is similar to the leaky solution which arises in a medium when a foreign thin layer is immersed into it.[12]

It will be recalled that on the surface of a crystal covered by a thin film leaky waves can also originate from a ''supersonic'' Rayleigh wave.[1,2] But in this situation the imaginary component of the leaky wave speed is proportional to the square of the thickness-to-wavelength ratio and the real correction to the Rayleigh wave speed is linear in this parameter. Moreover, the occurrence of such leaky waves is not conditioned by any relation of type (35) for the layer should only ''mix'' surface and bulk modes. In the case under discussion the coating must also cause the localization of the limiting wave and this takes place provided inequality (35) holds. The contractions $(\hat{N}^{(f)})_{3P}$, etc. in (39) characterize the ''mixing'' of the exceptional wave with other modes in the crystal which occurs because of the boundary conditions being modified. As in Ref. 1, at generic orientations of the plane $(\mathbf{m},\mathbf{n})$ any coating, isotropic included, generally leads to the ''mode mixing'' needed. If $(\mathbf{m},\mathbf{n})$ is a plane of symmetry of elastic properties of the substrate, then the layer must be anisotropic and its plane of symmetry, if any, cannot coincide with the plane $(\mathbf{m},\mathbf{n})$ since in this case the layer is to ''mix'' out-of-plane ($\alpha=3$) and in-plane ($\alpha=i,r,2,5$) modes.

Let us turn to $b_\alpha(v,H)$, $\alpha=i,2,3$. By analogy with $b_i(v,H)$ we derive

$$b_r(v,H)\approx[\mathbf{L}_i\mathbf{L}_2\mathbf{L}_{d6}]\{\Delta p+i(\hat{N}^{(f)})_{33}H+\Xi_rH^2\}, \tag{47}$$

$$b_2(v,H)\approx[\mathbf{L}_r\mathbf{L}_i\mathbf{L}_{d6}]\{\Delta p+i(\hat{N}^{(f)})_{33}H+\Xi_2H^2\}, \tag{48}$$

where

$$\Xi_r=\frac{1}{2[\mathbf{L}_i\mathbf{L}_5\mathbf{L}_{d6}]}\left\{\frac{(\hat{N}^{(f)})_{3J}(\hat{N}^{(f)})_{3S}}{[\mathbf{L}_r\mathbf{L}_5\mathbf{L}_{d6}]}-\frac{(\hat{N}^{(f)})_{3Q}(\hat{N}^{(f)})_{3P}}{[\mathbf{L}_i\mathbf{L}_2\mathbf{L}_{d6}]}\right\}, \tag{49}$$

$$\Xi_2=\frac{1}{2[\mathbf{L}_r\mathbf{L}_i\mathbf{L}_{d6}]}\left\{\frac{(\hat{N}^{(f)})_{3P}(\hat{N}^{(f)})_{3S}}{[\mathbf{L}_i\mathbf{L}_5\mathbf{L}_{d6}]}-\frac{(\hat{N}^{(f)})_{3Q}(\hat{N}^{(f)})_{3J}}{[\mathbf{L}_r\mathbf{L}_2\mathbf{L}_{d6}]}\right\}. \tag{50}$$

The quantity Ξ_r is complex at $\hat{v}^{(1)}$, comparison of (49) and (39) showing that $\Xi_r=-\Xi_i^*$, while Ξ_r is real at $\hat{v}^{(2)}$,

$$\Xi_r = \mathrm{Re}(\Xi_r) = \frac{1}{2[\mathbf{L}_i\mathbf{L}_2\mathbf{L}_{d6}]^2}\{|(\hat{N}^{(f)})_{3Q}|^2 - (\hat{N}^{(f)})_{3J}^2\}$$

$$\text{at } \hat{v}^{(2)}. \tag{51}$$

Similarly to Ξ_r, Ξ_2 is complex at $\hat{v}^{(1)}$, but it is not related to Ξ_i as straightforwardly as Ξ_r, and Ξ_2 becomes real at $\hat{v}^{(2)}$,

$$\Xi_2 = \mathrm{Re}(\Xi_2) = \frac{1}{2[\mathbf{L}_r\mathbf{L}_i\mathbf{L}_{d6}]^2}\{|(\hat{N}^{(f)})_{3S}|^2 - (\hat{N}^{(f)})_{3J}^2\}$$

$$\text{at } \hat{v}^{(2)}. \tag{52}$$

In distinction from $b_\alpha(v,H)$, $\alpha = i,r,2$, the coefficient $b_3(v,H)$ (25) does not involve the term linear in Δp and, expanding $b_3(v,H)$ with respect to small parameters, it is sufficient to allow for terms linear in H. Owing to (A20), (A23), and (40) $b_3(v,H)$ becomes

$$b_3(v,H) \approx i\frac{[\mathbf{L}_r\mathbf{L}_2\mathbf{L}_{d6}]}{[\mathbf{L}_i\mathbf{L}_5\mathbf{L}_{d6}]}(\hat{N}^{(f)})_{3P}H. \tag{53}$$

Substituting now $-i(\hat{N}^{(f)})_{33}H - \Xi_i H^2$ for Δp into (47) and (48) and canceling the common multiplier which then appears in Eqs. (47), (48), and (53) we obtain

$$b_r(v_l,H) \approx i\frac{(\hat{N}^{(f)})_{3Q}H}{[\mathbf{L}_r\mathbf{L}_2\mathbf{L}_{d6}]},$$

$$b_2(v_l,H) \approx -i\frac{(\hat{N}^{(f)})_{3S}H}{[\mathbf{L}_r\mathbf{L}_2\mathbf{L}_{d6}]}, \quad b_3(v_l,H) \approx 1, \tag{54}$$

so that the modes $\alpha = r,2$ are involved in the leaky wave with amplitudes as small as H in comparison with the amplitude of the mode $\alpha = 3$. The latter becomes weakly localized since due to (15) and (34) $p_3(v_l) \approx p_{d3} + i|(\hat{N}^{(f)})_{33}|H$, i.e., its penetration depth will be of the order $\lambda/H \gg \lambda$, where λ is the wavelength.

III. RESONANCE FEATURES OF THE COEFFICIENTS OF MODE CONVERSION

Before attacking the analysis of the behavior of the coefficients $R_{i\to\alpha}(v,H) = b_\alpha(v,H)/b_i(v,H)$ in a crystal-layer structure, it is advisable to compute these coefficients for the case when the surface is mechanically free. Letting $H = 0$ in (25) we obtain, through Eqs. (15), (A9), and (A23), that near a transonic state $\hat{v}$ at which the limiting wave is exceptional

$$R_{i\to r}(v,0) = \frac{[\mathbf{L}_2\mathbf{L}_i\mathbf{L}_{d6}]}{[\mathbf{L}_r\mathbf{L}_2\mathbf{L}_{d6}]} + O(|\Delta p|),$$

$$R_{i\to 2}(v,0) = \frac{[\mathbf{L}_i\mathbf{L}_r\mathbf{L}_{d6}]}{[\mathbf{L}_r\mathbf{L}_2\mathbf{L}_{d6}]} + O(|\Delta p|), \tag{55}$$

$$R_{i\to 3}(v,0) = O(|\Delta p|),$$

where the vectors $\mathbf{L}_\alpha$ are referred to $v = \hat{v}$, so that for $v \neq \hat{v}$ the coefficients differ from their values $R_{i\to\alpha}^{(0)} = R_{i\to\alpha}(\hat{v},0)$ but in the measure of the smallness of $|\Delta p|$. Note that in particular cases $R_{i\to 3}(v,0)$ can vanish identically.

Let us pass on to studying reflection from a crystal-thin layer interface near $\hat{v}$. Introducing for the sake of convenience the functions

$$F_\alpha(v) = -i\Delta p + (\hat{N}^{(f)})_{33}H + \mathrm{Im}(\Xi_\alpha)H^2, \quad \alpha = i,r,2, \tag{56}$$

and having in view Eqs. (45), (47), (48), and (53) we find

$$R_{i\to\alpha}(v,H) \approx R_{i\to\alpha}^{(0)}\frac{F_\alpha(v) - i\ \mathrm{Re}(\Xi_\alpha)H^2}{F_i(v) - i\ \mathrm{Re}(\Xi_i)H^2}, \quad \alpha = r,2 \tag{57}$$

$$R_{i\to 3}(v,H) \approx -\frac{1}{[\mathbf{L}_i\mathbf{L}_5\mathbf{L}_{d6}]}\frac{(\hat{N}^{(f)})_{3P}H}{F_i(v) - i\ \mathrm{Re}(\Xi_i)H^2}. \tag{58}$$

It should be said that due to Eqs. (A5) and (A3), (A6)–(A8), and (A10) the law of energy conservation results in the following relations: near $\hat{v}^{(1)}$ $|R_{i\to r}| = 1$ and $|R_{i\to r}|^2 + 2\Delta p|R_{i\to 3}|^2 = 1$ for $v < \hat{v}^{(1)}$ and $v > \hat{v}^{(1)}$, respectively; near $\hat{v}^{(2)}$ $|R_{i\to r}|^2 + |R_{i\to 2}|^2 = 1$ and $|R_{i\to r}|^2 + |R_{i\to 2}|^2 + 2\Delta p|R_{i\to 3}|^2 = 1$ for $v < \hat{v}^{(2)}$ and $v > \hat{v}^{(2)}$, respectively. On account of (A20)–(A23) the coefficients (57) and (58) can be shown to fulfill these relations.

The most marked features in the behavior of coefficients (57) and (58) emerge in the range

$$|v - v_l'| \propto v_l'' \tag{59}$$

providing inequality (35) holds true, i.e., in the case when the exceptional wave gives rise to the leaky wave, and, from the physical viewpoint, the resonance excitation of the latter accounts for the peculiarities to be discussed. In considering the resonance reflection we do not allow for dissipative attenuation of the leaky wave. This attenuation can be described by an additional imaginary component $-iv_d''$ in v_l, the magnitude of v_d'' being determined mainly by the damping of the exceptional wave. Neglect of dissipation is justified until $v_l'' \gg v_d''$ so that the range of admissible values of H is bounded from below by the condition $H^3 \gg v_d''/\hat{v}$.

Consider the transonic state $\hat{v}^{(1)}$. With regard for $\Xi_r = -\Xi_i^*$ the reflection coefficient reads

$$R_{i\to r}(v,H) \approx R_{i\to r}^{(0)}\frac{F_i(v) + i\ \mathrm{Re}(\Xi_i)H^2}{F_i(v) - i\ \mathrm{Re}(\Xi_i)H^2}, \tag{60}$$

by (16), (18), and (56) $F_i(v) \equiv F_r(v)$ being real for $v < \hat{v}^{(1)}$. The magnitude of $R_{i\to r}(v,H)$ is identically equal to unity but its phase

$$\psi_r(v) = 2\ \mathrm{arccot}\left\{\frac{F_i(v)}{\mathrm{Re}(\Xi_i)H^2}\right\} \tag{61}$$

suffers an abrupt change in the neighborhood of the speed v_l' at which to our accuracy $F_i(v)$ vanishes, e.g., $\psi_r(v_l' - v_l'') \approx \pi/2$ and $\psi_r(v_l' + v_l'') \approx 3\pi/2$.

The coefficient $R_{i\to 2}(v,H)$, which near $\hat{v}^{(1)}$ has the significance of the coefficient of excitation of the inhomogeneous mode $\alpha = 2$, undergoes abrupt changes both in phase and magnitude. The latter is approximately equal to $|R_{i\to 2}^{(0)}|$ outside the interval (59) whereas

$$|R_{i\to 2}(v_l')|\approx|R_{i\to 2}^{(0)}|\frac{|\mathrm{Im}(\Xi_2-\Xi_i)-i\ \mathrm{Re}(\Xi_2)|}{\mathrm{Re}(\Xi_i)}. \tag{62}$$

Clearly $|R_{i\to 2}(v_l')|$ can appreciably differ from $|R_{i\to 2}^{(0)}|$ but in any case $|R_{i\to 2}(v_l')|$ remains of the order of unity.

In distinction from $R_{i\to 2}$, the magnitude of the coefficient $R_{i\to 3}$ increases significantly in the region (59) and at $v=v_l'$ it reaches its maximum

$$|R_{i\to 3}(v_l')|\approx\frac{1}{|[\mathbf{L}_i\mathbf{L}_5\mathbf{L}_{d6}]|}\frac{|(\hat{N}^{(f)})_{3P}|}{\mathrm{Re}(\Xi_i)H}\gg 1. \tag{63}$$

Changes in magnitude are attended by changes in phase of $R_{i\to 3}(v,H)$, $\psi_3(v)=0.5\psi_r(v)$, as well.

Let us turn to the transonic state $\hat{v}^{(2)}$. The behavior of the coefficient $R_{i\to 3}(v,H)$ does not exhibit a modification. A different situation takes place regarding $R_{i\to\alpha}(v,H)$, $\alpha=r,2$. The quantities Ξ_α, $\alpha=i,r,2$, become real so that $F_i(v)\equiv F_2(v)$ and the reflection coefficients $R_{i\to\alpha}(v,H)$, $\alpha=r,2$, ($\alpha=2$ is now the second reflected wave) are of the form

$$R_{i\to\alpha}(v,H)\approx R_{i\to\alpha}^{(0)}\frac{F_i(v)-i\Xi_\alpha H^2}{F_i(v)-i\Xi_i H^2},\quad \alpha=r,2. \tag{64}$$

Inside the interval (59) the magnitudes of both the coefficients undergo dramatic changes and at $v=v_l'$ they take their extreme values

$$|R_{i\to r}(v_l')|\approx|R_{i\to r}^{(0)}|\frac{|\Xi_r|}{\Xi_i}$$

$$=|R_{i\to r}^{(0)}|^{-1}\frac{||(\hat{N}^{(f)})_{3Q}|^2-(\hat{N}^{(f)})_{3J}^2|}{(\hat{N}^{(f)})_{3P}^2+(\hat{N}^{(f)})_{3J}^2}, \tag{65}$$

$$|R_{i\to 2}(v_l')|\approx|R_{i\to 2}^{(0)}|\frac{|\Xi_2|}{\Xi_i}$$

$$=|R_{i\to 2}^{(0)}|^{-1}\frac{||(\hat{N}^{(f)})_{3S}|^2-(\hat{N}^{(f)})_{3J}^2|}{(\hat{N}^{(f)})_{3P}^2+(\hat{N}^{(f)})_{3J}^2}. \tag{66}$$

Outside (59) the coefficients are close to $R_{i\to\alpha}^{(0)}$, $\alpha=r,2$. The phases of coefficients (64),

$$\psi_\alpha(v)=\mathrm{arccot}\left\{\frac{F_i(v)}{\Xi_i H^2}\right\}-\mathrm{arccot}\left\{\frac{F_i(v)}{\Xi_\alpha H^2}\right\},\quad \alpha=r,2, \tag{67}$$

also manifest a ''jump'' in the vicinity of v_l'.

It is of interest to discuss the situation when, say, $|(\hat{N}^{(f)})_{3Q}|=\pm(\hat{N}^{(f)})_{3J}$. One should distinguish between two cases: (a) $\mathbf{L}_i$ and $\mathbf{L}_2$ are not collinear, (b) $\mathbf{L}_i$ and $\mathbf{L}_2$ are collinear. In case (a) the above equality can hold due to special relations among material constants of the crystal and the layer. By virtue of (51) we arrive at

$$R_{i\to r}(v,H)\approx R_{i\to r}^{(0)}\frac{F_i(v)}{F_i(v)-i\ \mathrm{Re}(\Xi_i)H^2}, \tag{68}$$

$$R_{i\to 2}(v,H)\approx R_{i\to 2}^{(0)}\frac{F_i(v)-i\Xi_2 H^2}{F_i(v)-i\Xi_i H^2}, \tag{69}$$

so that $R_{i\to r}(v_l')\approx 0$ while $|R_{i\to 2}(v_l')|\approx|R_{i\to 2}^{(0)}\Xi_2|/\Xi_i=1$. In case (b) the collinearity of $\mathbf{L}_i$ and $\mathbf{L}_2$ entails the collinearity of $\mathbf{L}_r$ and $\mathbf{L}_5$, see, e.g., (A23). In view of (A1) and (A3) $\mathbf{L}_2=B_i\mathbf{L}_i$, $\mathbf{L}_r=B_5\mathbf{L}_5$ where $B_i^2=B_5^2=-1$. Hence, due to (42) $\mathbf{Q}=B_i\mathbf{J}$, $\mathbf{S}=B_5\mathbf{P}$ and then independently of the material constants of the layer $(\hat{N}^{(f)})_{3Q}=B_i(\hat{N}^{(f)})_{3J}$, $(\hat{N}^{(f)})_{3S}=B_5(\hat{N}^{(f)})_{3P}$. One cannot use Eq. (51) since, in deriving it, $\mathbf{L}_i$, $\mathbf{L}_2$ were assumed not to be collinear. We have to go back to Eq. (49) and, employing (A20)–(A23), after some evaluations one obtains

$$R_{i\to r}(v,H)\approx\frac{1}{B_i[\mathbf{L}_i\mathbf{L}_5\mathbf{L}_{d6}]^2}\frac{(\hat{N}^{(f)})_{3P}(\hat{N}^{(f)})_{3J}H^2}{F_i(v)-i\ \mathrm{Re}(\Xi_i)H^2}, \tag{70}$$

$$R_{i\to 2}(v,H)\approx B_i\frac{F_i(v)-i\Xi_2 H^2}{F_i(v)-i\Xi_i H^2}. \tag{71}$$

Owing to (56) $|R_{i\to r}|\propto H$ outside the region (59) but with v approaching v_l' $|R_{i\to r}|$ increases and $|R_{i\to 2}|$ diminishes, and at v_l' they are equal to

$$|R_{i\to r}(v_l')|\approx\frac{2(\hat{N}^{(f)})_{3P}(\hat{N}^{(f)})_{3J}}{(\hat{N}^{(f)})_{3P}^2+(\hat{N}^{(f)})_{3J}^2}, \tag{72}$$

$$|R_{i\to 2}(v_l')|\approx\frac{(\hat{N}^{(f)})_{3J}^2-(\hat{N}^{(f)})_{3P}^2}{(\hat{N}^{(f)})_{3P}^2+(\hat{N}^{(f)})_{3J}^2}. \tag{73}$$

If in addition $(\hat{N}^{(f)})_{3J}=\pm(\hat{N}^{(f)})_{3P}$, then $R_{i\to 2}(v_l')=0$, $|R_{i\to r}(v_l')|=1$ so that under resonance conditions the incident wave will totally convert into the mode $\alpha=r$ whereas outside the region (59) as well as in the absence of the layer it almost totally converts into the mode $\alpha=2$.

The case when $(\hat{N}^{(f)})_{33}>0$ is now discussed, i.e., the leaky wave does not appear. $F_i(v)$ (56) does not vanish in the neighborhood of $\hat{v}$ and its values have the order of the smallness not higher than H. Hence by (57) the coefficients $R_{i\to\alpha}$, $\alpha=i,2$, will slightly differ from $R_{i\to\alpha}^{(0)}$. Neither their moduli nor phases will undergo strong changes. On the other hand, the magnitude of the coefficient (58),

$$R_{i\to 3}(v,H)\approx-\frac{1}{[\mathbf{L}_i\mathbf{L}_5\mathbf{L}_{d6}]}\frac{(\hat{N}^{(f)})_{3P}H}{-i\Delta p+(\hat{N}^{(f)})_{33}H}, \tag{74}$$

increases as v approaches the limiting speed and at $v=\hat{v}$ it becomes

$$|R_{i\to 3}(\hat{v})|\approx\frac{|(\hat{N}^{(f)})_{3P}|}{|[\mathbf{L}_i\mathbf{L}_5\mathbf{L}_{d6}]|(\hat{N}^{(f)})_{33}}\propto 1 \tag{75}$$

while on the mechanically free surface $R_{i\to 3}(\hat{v})=0$, see (55) [we need not account for the term $O(H^2)$ in the denominator of (74) when $(\hat{N}^{(f)})_{33}>0$]. The alternating part of the phase of $R_{i\to 3}$ in the range $v>\hat{v}$, $\psi_3(v)\approx\arctan(\Delta p/(\hat{N}^{(f)})_{33}H)$, tends quickly to zero and for $v<\hat{v}$ it remains close to zero. Such a difference in the behavior of $R_{i\to 3}$ at $H\neq 0$ and $H=0$ can be interpreted as follows. On the mechanically free surface the limiting wave, the mode $\alpha=3$ at $v=\hat{v}$, is exceptional, i.e., it is a characteristic solution of the boundary-value problem. This wave ceases to meet the boundary con-

ditions in the presence of the layer. However, a thin layer does not strongly violate the existence conditions of the exceptional wave and therefore there appears a possibility of exciting the mode $\alpha=3$ with fairly large amplitude at $v=\hat{v}$ because of the ''mode mixing'' brought in by the coating. Note that Eq. (74) describes correctly the behavior of $R_{i\to 3}(v,H)$ until $H^2 \gg v_d''/\hat{v}$, i.e., until the correction to the exceptional wave speed due to dissipation, v_d'', is far smaller than the width $\hat{v}H^2$ of the speed range within which the mode $\alpha=3$ can be efficiently excited.

IV. CONCLUSION

We have developed the theory of elastic wave reflection under conditions of resonance excitation of a leaky wave originating from an exceptional bulk wave in the case when the surface of an anisotropic medium is covered by a thin solid foreign layer. The coating provides the localization of the limiting wave and it also has the effect of mixing modes in the substrate which are not coupled through the boundary conditions on the mechanically free surface. The leaky wave appears providing that the material constants of the crystal and the layer obey certain relations of a type of inequality and in the crystal it is constructed from a weakly localized limiting mode to which two more modes are added with small amplitudes, between these adjacent modes at least one being ''reflected.'' The parameters of the leaky wave and of the coefficients of mode conversion are expressed straightforward in terms of the elastic moduli and density of the film.

The resonance reflection occurs at angles of incidence at which the trace speed of the incident wave is equal or close to the real part of the leaky wave speed. It is characterized first by excitation of the main component of the leaky wave with amplitude much exceeding the amplitude of the incident wave. The analogous effect takes place in the situation discussed in Refs. 1 and 2 (a leaky wave given rise to by a Rayleigh wave). At the same time, the width of the resonance interval of incident angles in the case being considered in the present paper is ''narrower'' than in the cases of Refs. 1 and 2. It is proportional not to the square of the thickness-to-wavelength ratio but to the third power of this small parameter. Besides, the reflection coefficients near the speed of the leaky wave arising from an exceptional wave exhibit strong changes not only in phase but also in magnitude. The latter takes place on the condition that the incident wave converts into two reflected modes in the vicinity of a transonic state. It can even happen that outside the resonance interval of incident angles or in the absence of the layer the incident wave almost totally converts into the first reflected mode while under resonance conditions the incident wave almost totally converts into the second reflected mode.

ACKNOWLEDGMENTS

A. N. Darinskii is grateful to V. I. Alshits, J. Lothe, V. N. Lyubimov, and A. L. Shuvalov for fruitful discussions and to the reviewer for pointing out a number of misprints in the initial text of the paper. The author thanks the Russian Foundation for Basic Investigations (Grant No. 97-02-16338) for partial financial support.

APPENDIX

Due to the matrix $\hat{\mathbf{T}}\hat{\mathbf{N}}$ being symmetric, Eq. (7), it follows from (13) that

$$\boldsymbol{\xi}_{d3}\cdot\hat{\mathbf{T}}\boldsymbol{\xi}_{\alpha}=\boldsymbol{\xi}_{d6}\cdot\hat{\mathbf{T}}\boldsymbol{\xi}_{\alpha}=\boldsymbol{\xi}_{\beta}\cdot\hat{\mathbf{T}}\boldsymbol{\xi}_{\alpha}=\boldsymbol{\xi}_{d3}\cdot\hat{\mathbf{T}}\boldsymbol{\xi}_{d3}=0, \quad \alpha,\beta=i,r,2,5, \quad \alpha\neq\beta. \tag{A1}$$

The generalized eigenvector $\boldsymbol{\xi}_{d6}$ can be defined such as

$$\boldsymbol{\xi}_{d6}\cdot\hat{\mathbf{T}}\boldsymbol{\xi}_{d6}=0, \tag{A2}$$

since if $\boldsymbol{\xi}_{d6}\cdot\hat{\mathbf{T}}\boldsymbol{\xi}_{d6}\neq 0$ for a chosen $\boldsymbol{\xi}_{d6}$, then the latter is replaced by the vector $\boldsymbol{\xi}_{d6}-(\boldsymbol{\xi}_{d6}\cdot\hat{\mathbf{T}}\boldsymbol{\xi}_{d6})\boldsymbol{\xi}_{d3}/2(\boldsymbol{\xi}_{d3}\cdot\hat{\mathbf{T}}\boldsymbol{\xi}_{d6})$. This change does not violate Eq. (13). In view of (A1) neither $\boldsymbol{\xi}_{\alpha}\cdot\hat{\mathbf{T}}\boldsymbol{\xi}_{\alpha}\neq 0$, $\alpha=i,r,2,5$ nor $\boldsymbol{\xi}_{d3}\cdot\hat{\mathbf{T}}\boldsymbol{\xi}_{d6}\neq 0$ because $\boldsymbol{\xi}_{\alpha}$, $\alpha=i,r,2,5,d3,d6$ are linearly independent and provided

$$\boldsymbol{\xi}_{\alpha}\cdot\hat{\mathbf{T}}\boldsymbol{\xi}_{\alpha}=1, \quad \alpha=i,r,2,5, \quad \boldsymbol{\xi}_{d3}\cdot\hat{\mathbf{T}}\boldsymbol{\xi}_{d6}=1, \tag{A3}$$

the following equality is valid

$$\sum_{\alpha=i,r,2,5}\boldsymbol{\xi}_{\alpha}\otimes\hat{\mathbf{T}}\boldsymbol{\xi}_{\alpha}+\boldsymbol{\xi}_{d3}\otimes\hat{\mathbf{T}}\boldsymbol{\xi}_{d6}+\boldsymbol{\xi}_{d6}\otimes\hat{\mathbf{T}}\boldsymbol{\xi}_{d3}=\begin{pmatrix}\hat{\mathbf{I}} & \hat{\mathbf{O}}\\ \hat{\mathbf{O}} & \hat{\mathbf{I}}\end{pmatrix}, \tag{A4}$$

where the symbol ''$\otimes$'' designates dyadic multiplication.

The projection $E_{n\alpha}$ onto the inner normal $\mathbf{n}$ of the energy flux of a bulk mode α having unit amplitude is found in terms of the vector $\boldsymbol{\xi}_{\alpha}$ as

$$E_{n\alpha}=-\frac{\omega k}{4}\boldsymbol{\xi}_{\alpha}\cdot\hat{\mathbf{T}}\boldsymbol{\xi}_{\alpha}^{*}. \tag{A5}$$

Accounting for (A3) and (A5) we conclude that the vectors $\boldsymbol{\xi}_{\alpha}$ associated with the incident and reflected waves are real and imaginary, respectively,

$$\boldsymbol{\xi}_{i}=\boldsymbol{\xi}_{i}^{*}, \quad \boldsymbol{\xi}_{r}=-\boldsymbol{\xi}_{r}^{*}, \tag{A6}$$

$$\boldsymbol{\xi}_{5}=\boldsymbol{\xi}_{5}^{*}, \quad \boldsymbol{\xi}_{2}=-\boldsymbol{\xi}_{2}^{*} \quad \text{at } \hat{v}^{(2)},\ \hat{v}_{(-)}^{(2)}, \tag{A7}$$

the asterisk means complex conjugation. At the transonic states $\hat{v}^{(1)}$, $\hat{v}_{(-)}^{(1)}$ the modes $\alpha=2,5$ are inhomogeneous and therefore, having in view the relation $p_2=p_5^{*}$, we set

$$\boldsymbol{\xi}_{2}=\boldsymbol{\xi}_{5}^{*} \quad \text{at } \hat{v}^{(1)},\hat{v}_{(-)}^{(1)}. \tag{A8}$$

As to $\boldsymbol{\xi}_{d3}$ and $\boldsymbol{\xi}_{d6}$, it follows from Eqs. (13) and (A3) and from the reality of the matrix $\hat{\mathbf{N}}$ and p_{d3} that these vectors can occur either both real or both imaginary. A more exact conclusion will be made later.

In the neighborhood of $\hat{v}$ the matrix $\hat{\mathbf{N}}$ becomes nondegenerate. Hence by virtue of (7) all six eigenvectors satisfy the relation $\boldsymbol{\xi}_{\alpha}\cdot\hat{\mathbf{T}}\boldsymbol{\xi}_{\beta}=0$, $\alpha\neq\beta$. We shall assume that the vectors $\boldsymbol{\xi}_{\alpha}$, $\alpha=i,r,2,5$, obey Eq. (A3) at $v\neq\hat{v}$ as well. Note that Eqs. (A1)–(A3) allow one to rewrite (14) as

$$\boldsymbol{\xi}_\alpha(v)\approx\boldsymbol{\xi}_\alpha+\sum_\beta\left\{\left(\boldsymbol{\xi}_\beta\cdot\hat{\mathbf{T}}\frac{\partial\boldsymbol{\xi}_\alpha}{\partial v}\right)\boldsymbol{\xi}_\beta+\left(\boldsymbol{\xi}_{d6}\cdot\hat{\mathbf{T}}\frac{\partial\boldsymbol{\xi}_\alpha}{\partial v}\right)\boldsymbol{\xi}_{d3}+\left(\boldsymbol{\xi}_{d3}\cdot\hat{\mathbf{T}}\frac{\partial\boldsymbol{\xi}_\alpha}{\partial v}\right)\boldsymbol{\xi}_{d6}\right\}(v-\hat{v}),\tag{A9}$$

where $\beta\neq\alpha,d3,d6$. On the other hand, it is expedient to set

$$\boldsymbol{\xi}_3\cdot\hat{\mathbf{T}}\boldsymbol{\xi}_3=-2\Delta p,\quad \boldsymbol{\xi}_6\cdot\hat{\mathbf{T}}\boldsymbol{\xi}_6=2\Delta p,\tag{A10}$$

where the parameter Δp is introduced through the equation

$$(\Delta p)^2=\rho(v^2-\hat{v}^2)\mathbf{A}_{d3}^2.\tag{A11}$$

Taking into account that (a) $\boldsymbol{\xi}_{3,6}$ to within phase factors must be real when $p_{3,6}$ are real, (b) if p_3 and p_6 are complex, then there must hold $p_3=p_6^*$ and to within a phase factor $\boldsymbol{\xi}_3=\boldsymbol{\xi}_6^*$, (c) Δp can have either real or imaginary values, we look for $\boldsymbol{\xi}_{3,6}$, $p_{3,6}$ in the form

$$\begin{aligned}&\boldsymbol{\xi}_3=\boldsymbol{\xi}_{d3}+\sum_{m=1}^{\infty}(-\Delta p)^m\boldsymbol{\psi}_m,\quad p_3=p_{d3}+\sum_{m=1}^{\infty}n_m(\Delta p)^m,\\&\boldsymbol{\xi}_6=\boldsymbol{\xi}_{d3}+\sum_{m=1}^{\infty}(\Delta p)^m\boldsymbol{\psi}_m,\quad p_6=p_{d3}+\sum_{m=1}^{\infty}n_m(-\Delta p)^m,\end{aligned}\tag{A12}$$

where the vectors $\boldsymbol{\psi}_m$, $m=1,2,\ldots$ are linear combinations of the vectors $\boldsymbol{\xi}_\alpha$, $\alpha=1,2,4,5,d3,d6$, $\boldsymbol{\psi}_m=\Sigma_\alpha a_\alpha^{(m)}\boldsymbol{\xi}_\alpha$, and they are either real or imaginary depending on whether $\boldsymbol{\xi}_{d3}$ is real or imaginary, respectively, and all n_m, $m=1,2,\ldots$ are real.

We confine ourselves to computations allowing for terms $O(|\Delta p|^2)$. Introducing (A12), e.g., into $\hat{\mathbf{N}}(v)\boldsymbol{\xi}_3=p_3\boldsymbol{\xi}_3$, where $\hat{\mathbf{N}}(v)$ is taken in the form

$$\hat{\mathbf{N}}(v)=\hat{\mathbf{N}}(\hat{v})-\rho(v^2-\hat{v}^2)\hat{\mathbf{K}},\quad \hat{\mathbf{K}}=\begin{pmatrix}\hat{\mathbf{O}}&\hat{\mathbf{O}}\\ \hat{\mathbf{I}}&\hat{\mathbf{O}}\end{pmatrix},\tag{A13}$$

and employing Eqs. (13) and (A1)–(A3) we derive the expressions for the terms $O(|\Delta p|)$ and $O(|\Delta p|^2)$. Equating them to zero provides two equations,

$$\sum_{\alpha=1,2,4,5}(p_{d3}-p_\alpha)a_\alpha^{(1)}\boldsymbol{\xi}_\alpha+(n_1-a_{d6}^{(1)})\boldsymbol{\xi}_{d3}=0,\tag{A14}$$

$$\frac{1}{\mathbf{A}_{d3}^2}\hat{\mathbf{K}}\boldsymbol{\xi}_{d3}+\sum_{\alpha=1,2,4,5}(p_{d3}-p_\alpha)a_\alpha^{(2)}\boldsymbol{\xi}_\alpha+(a_{d6}^{(2)}-n_1a_{d3}^{(1)}+n_2)\boldsymbol{\xi}_{d3}-n_1a_{d6}^{(1)}\boldsymbol{\xi}_{d6}=0,\tag{A15}$$

respectively, wherefrom, taking advantage of (A1)–(A3) we obtain $a_\alpha^{(1)}=0$, $\alpha=i,r,2,5$, $n_1=a_{d6}^{(1)}=1$, and also the coefficients $a_\alpha^{(2)}$, $\alpha=i,r,2,5$. Accounting for (A10) and for the relation $\boldsymbol{\xi}_3\cdot\hat{\mathbf{T}}\boldsymbol{\xi}_6=0$ yields $a_{d3}^{(1)}=a_{d6}^{(2)}=0$ and afterwards from (A15) one can get n_2. To compute $a_{d3}^{(2)}$, one should allow for terms of higher order than $O(|\Delta p|^2)$. However, as a matter of fact for our purposes it is essential only that $a_{d6}^{(2)}=0$ and, thus, we arrive at (15).

Inspection of Fig. 1 shows that in the vicinity of positive curvature transonic states p_3 and p_6 are real when $v>\hat{v}$ and complex when $v<\hat{v}$. Hence in these intervals $(\Delta p)^2$ must be positive and negative, respectively. Equation (12) implies that $\mathbf{A}_{d3}$ and correspondingly $\boldsymbol{\xi}_{d3}$, $\boldsymbol{\xi}_{d6}$ will be real vectors,

$$\boldsymbol{\xi}_{d3}=\boldsymbol{\xi}_{d3}^*,\quad \boldsymbol{\xi}_{d6}=\boldsymbol{\xi}_{d6}^*,\quad \text{at } \hat{v}^{(1)},\quad \hat{v}^{(2)}.\tag{A16}$$

Conversely, when the curvature is negative, p_3, p_6 are real for $v<\hat{v}$ and complex in the range $v>\hat{v}$ so that $\boldsymbol{\xi}_{d3}$, $\boldsymbol{\xi}_{d6}$ occur to be purely imaginary vectors,

$$\boldsymbol{\xi}_{d3}=-\boldsymbol{\xi}_{d3}^*,\quad \boldsymbol{\xi}_{d6}=-\boldsymbol{\xi}_{d6}^*,\quad \text{at } \hat{v}_{(-)}^{(1)}\ \hat{v}_{(-)}^{(2)}.\tag{A17}$$

With the aid of Eqs. (A5), (A10), (A16), (A17), and (16)–(18) one can check that the subscripts $\alpha=3,6$ are ascribed to p_α, $\boldsymbol{\xi}_\alpha$ in (A11) and (15) in conformity with the rule (12). Besides, it follows from (A1), (A16), and (A17) that E_{nd3} (A5) vanishes, i.e., the degenerate eigensolution $(p_{d3},\ \boldsymbol{\xi}_{d3})$ does correspond to the limiting wave.

Consider now algebraic properties of the vectors $\mathbf{L}_\alpha$, $\alpha=i,r,2,5,d3,d6$. First, $\mathbf{L}_{d3}$ and $\mathbf{L}_{d6}$ cannot vanish simultaneously, since otherwise there would be $\boldsymbol{\xi}_{d3}\cdot\hat{\mathbf{T}}\boldsymbol{\xi}_{d6}=0$. Furthermore, due to Eq. (A1) none of the vectors $\mathbf{L}_\alpha$, $\alpha=i,r,2,5$, can be collinear either to $\mathbf{L}_{d3}$ or to $\mathbf{L}_{d6}$.

Let the limiting wave be exceptional, $\mathbf{L}_{d3}=0$. Equation (A4) then yields

$$\sum_{\alpha=i,r,2,5}\mathbf{L}_\alpha\otimes\mathbf{L}_\alpha=\hat{\mathbf{O}},\tag{A18}$$

$$\sum_{\alpha=i,r,2,5}\mathbf{L}_\alpha\otimes\mathbf{A}_\alpha+\mathbf{L}_{d6}\otimes\mathbf{A}_{d3}=\hat{\mathbf{I}}.\tag{A19}$$

Multiplying (A18) by the vector products $\mathbf{L}_\beta\times\mathbf{L}_{d6}$ and $\mathbf{L}_\gamma\times\mathbf{L}_{d6}$ leads to

$$[\mathbf{L}_\alpha\mathbf{L}_\beta\mathbf{L}_{d6}][\mathbf{L}_\alpha\mathbf{L}_\gamma\mathbf{L}_{d6}]+[\mathbf{L}_\mu\mathbf{L}_\beta\mathbf{L}_{d6}][\mathbf{L}_\mu\mathbf{L}_\gamma\mathbf{L}_{d6}]=0,\quad \gamma\neq\beta,\tag{A20}$$

where $[\cdots]$ denotes the scalar triple product. We can combine a pair of equalities (A20) to obtain

$$[\mathbf{L}_\alpha\mathbf{L}_\beta\mathbf{L}_{d6}]^2=[\mathbf{L}_\mu\mathbf{L}_\gamma\mathbf{L}_{d6}]^2,\quad \alpha\neq\beta\neq\gamma\neq\mu.\tag{A21}$$

Yet another equality can be derived by multiplying (A18) twice by the same vector $\mathbf{L}_\beta\times\mathbf{L}_{d6}$,

$$[\mathbf{L}_\alpha\mathbf{L}_\beta\mathbf{L}_{d6}]^2+[\mathbf{L}_\gamma\mathbf{L}_\beta\mathbf{L}_{d6}]^2+[\mathbf{L}_\mu\mathbf{L}_\beta\mathbf{L}_{d6}]^2=0.\tag{A22}$$

Successive multiplication of (A18) by the vectors $\mathbf{L}_\alpha\times\mathbf{L}_{d6}$, $\alpha=5,2,r,i$, yields the relations

$$\begin{aligned}&\sum_{\alpha=i,r,2}[\mathbf{L}_\alpha\mathbf{L}_5\mathbf{L}_{d6}]\mathbf{L}_\alpha=\mathbf{0},\quad \sum_{\alpha=i,r,5}[\mathbf{L}_\alpha\mathbf{L}_2\mathbf{L}_{d6}]\mathbf{L}_\alpha=\mathbf{0},\\&\sum_{\alpha=i,2,5}[\mathbf{L}_\alpha\mathbf{L}_r\mathbf{L}_{d6}]\mathbf{L}_\alpha=\mathbf{0},\quad \sum_{\alpha=r,2,5}[\mathbf{L}_\alpha\mathbf{L}_i\mathbf{L}_{d6}]\mathbf{L}_\alpha=\mathbf{0},\end{aligned}\tag{A23}$$

i.e., any three of the four vectors $\mathbf{L}_\alpha$, $\alpha=i,r,2,5$, are linearly dependent. In principle, these vectors can split into collinear pairs but pairwise collinearity at a transonic state is not related straightforwardly to the condition $\mathbf{L}_{d3}=0$. We shall assume, unless otherwise mentioned, that the vectors $\mathbf{L}_\alpha$, $\alpha=i,r,2,5$, are not pairwise collinear. In this connection we note that certain pairs of the $\mathbf{L}_\alpha$'s just cannot fall into collinearity, e.g., $\mathbf{L}_r$ never happens to be parallel to $\mathbf{L}_2$ which is proved with the help of Eqs. (A1), (A3), and (A6)–(A8).

[1] A. N. Darinskii, ''Leaky waves and the elastic wave resonance reflection on a crystal-thin solid layer interface,'' J. Acoust. Soc. Am. **102**, 283–291 (1997).
[2] A. N. Darinskii, ''Resonance phenomena in the reflection of an elastic wave at the boundary between a hexagonal crystal and an anisotropic film,'' JETP **80**, 317–323 (1995) [Zh. Eksp. Teor. Fiz. **107**, 585–598 (1995)].
[3] V. I. Alshits and J. Lothe, ''Elastic waves in triclinic crystals. III. The problem of existence and some general properties of exceptional surface waves,'' Sov. Phys. Crystallogr. **24**, 1122–1130 (1979) [Kristallografia **24**, 1122–1130 (1979)].
[4] V. I. Alshits, V. N. Lyubimov, N. F. Naumenko, N. V. Perelomova, and A. L. Shuvalov, ''Exceptional elastic body waves in crystals of various symmetries,'' Sov. Phys. Crystallogr. **30**, 123–126 (1985) [Kristallografia **30**(2), 213–219 (1985)].
[5] A. N. Stroh, ''Steady state problems in anisotropic elasticity,'' J. Math. Phys. **41**, 77–103 (1962).
[6] P. Chadwick and G. D. Smith, ''Foundation of the theory of surface waves in anisotropic elastic media,'' Adv. Appl. Mech. **17**, 303–376 (1977).
[7] P. Chadwick, ''A general analysis of transonic states in an anisotropic elastic body,'' Proc. R. Soc. London, Ser. A **401**, 203–223 (1985).
[8] V. I. Alshits and J. Lothe, ''Comments on the relation between surface wave theory and the theory of reflection,'' Wave Motion **3**, 297–310 (1981).
[9] D. M. Barnett and J. Lothe, ''Free surface (Rayleigh) waves in anisotropic half-spaces: the surface impedance method,'' Proc. R. Soc. London, Ser. A **402**, 135–152 (1985).
[10] D. M. Barnett, J. Lothe, and S. A. Gundersen, ''Zero curvature transonic states and free surface waves in anisotropic elastic media,'' Wave Motion **12**, 341–360 (1990).
[11] K. Nishioka and J. Lothe, ''Isotropic limiting behaviour of the six-dimensional formalism of anisotropic dislocation theory and anisotropic Green's function theory. II. Perturbation theory on the isotropic $\hat{\mathbf{N}}$-matrix,'' Phys. Status Solidi B **52**, 45–54 (1972).
[12] A. N. Darinskii and G. A. Maugin, ''The elastic wave resonance reflection from a thin solid layer in a crystal,'' Wave Motion **23**, 362–385 (1996).

Signal response of elastically coated plates

Richard F. Keltie
Center for Sound and Vibration, Department of Mechanical and Aerospace Engineering, North Carolina State University, Raleigh, North Carolina 27695-7910

(Received 1 July 1997; revised 27 October 1997; accepted 17 January 1998)

An analytical model of a compliant elastic coating attached to a submerged thin plate has been developed. A plane acoustic wave, representing the signal, is incident from the water on to the elastic coating. The normal and tangential velocity components induced in the coating by the acoustic wave are calculated. The effects of incidence angle, frequency, location throughout the coating, and coating properties on the signal response velocity components are evaluated. Many details of the coating velocity components are explained based on simple uncoupled waveguide characteristics of the coating. In general, the normal velocity components are several tens of decibels higher in amplitude than the tangential velocity components. It was shown, however, that the tangential velocity component could be increased by an average amount of about 20 dB, without serious degradation of the normal component, by altering the coating properties. In particular, combinations of high longitudinal wave speed and low shear wave speed were found to be advantageous. [S0001-4966(98)04704-3]

PACS numbers: 43.40.At, 43.40.Qi [CBB]

INTRODUCTION

Recent interest[1,2] has been focused on the use of kinematic sensors, such as accelerometers or velocity sensors, instead of pressure hydrophones in underwater acoustic sensing applications. These sensors are typically placed in or on a relatively soft coating that is backed by a stiff structural element. The soft coating leads, ideally, to a pressure release boundary condition at the water/coating interface, which produces an enhanced velocity response. It is this response which makes the concept attractive for use as a signal receiver. In these applications the incoming acoustic pressure wave is considered to be the signal, and the goal of the sensing system is to recognize the signal in the presence of noise. By the nature of these applications, it is reasonable for a first-order model to assume the incoming acoustic wave to be plane, and the receiving system to be of infinite linear extent.

One of the novel aspects of measuring a kinematic response variable is their vector nature. It has been shown that vector velocity measurement at a point is equivalent to the pressure measurement within a finite volume.[3] Thus, there is inherently much more information contained in a vector velocity measurement than in a scalar hydrophone output.

The work reported here involved an analytical study of the signal response of an elastic coating attached to a submerged plate. The signal response was induced by an incident acoustic plane wave. The main points of interest were the normal and tangential velocity components of the coating response as functions of frequency, signal arrival angle, location through the coating thickness, and physical parameters of the coating. Primary attention was focused on developing physical understanding of the dynamic processes affecting coating response, and the various mechanisms by which this response could be tailored to achieve desired results. This work builds on an earlier study by Keltie and Crank[4,5] in which a similar system was analyzed to determine the transmission characteristics of the coating to structural noise sources. That work, coupled with the present work, allows for a complete signal-to-noise analysis of the acoustic sensing system.

I. MODEL FORMULATION

A sketch of the model used in the study is shown in Fig. 1. The acoustic pressure in the half-space $z>0$ consists of an incoming plane wave, p_i, and a reflected plane wave, p_r. Taking harmonic time dependence to be of the form $e^{-j\omega t}$, the expressions for the incident and reflected pressures are given by

$$p_i(x,z)=P_i e^{jk_o(x\ \sin\ \theta_i - z\ \cos\ \theta_i)}, \tag{1a}$$

$$p_r(x,z)=P_r e^{jk_o(x\ \sin\ \theta_r + z\ \cos\ \theta_r)}, \tag{1b}$$

where P_i is the complex amplitude of the incident wave, P_r is the complex amplitude of the reflected wave, θ_i is the angle of incidence, θ_r is the angle of reflection, and $k_0 = \omega/c_0$ is the acoustic wave number in the water.

The coating is modeled as a 2-D elastic solid. It is well known that the general response can be written in the form of two potential functions, the longitudinal potential $\phi(x,z)$, and the shear potential $\psi(x,z)$. Each potential satisfies a 2-D wave equation, so that the general solutions are of the form[6,7]

$$\phi(x,z)=\Phi_1 e^{jk_\phi(x\ \sin\ \theta_1 - z\ \cos\ \theta_1)} + \Phi_2 e^{jk_\phi(x\ \sin\ \theta_2 + z\ \cos\ \theta_2)}, \tag{2a}$$

$$\psi(x,z)=\Psi_1 e^{jk_\psi(x\ \sin\ \gamma_1 - z\ \cos\ \gamma_1)} + \Psi_2 e^{jk_\psi(x\ \sin\ \gamma_2 + z\ \cos\ \gamma_2)}, \tag{2b}$$

where Φ_1 is the complex amplitude of the transmitted longitudinal wave, θ_1 is the angle of the transmitted longitudinal wave (measured from the vertical), Φ_2 is the complex amplitude of the reflected longitudinal wave, θ_2 is the angle of the reflected longitudinal wave, Ψ_1 is the complex amplitude

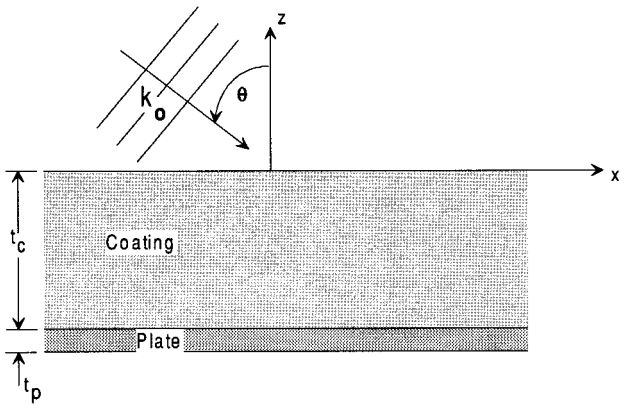


FIG. 1. Geometry of the coated plate model.

of the transmitted shear wave, γ_1 is the angle of the transmitted shear wave, Ψ_2 is the complex amplitude of the reflected shear wave, γ_2 is the angle of the reflected shear wave, k_ϕ is the longitudinal wave number ($k_\phi=\omega/c_\phi$), and k_ψ is the shear wave number ($k_\psi=\omega/c_\psi$). The coating displacement components, u_x and u_z are obtained from the potential functions

$$u_x=\frac{\partial\phi}{\partial x}-\frac{\partial\psi}{\partial z}, \tag{3a}$$

$$u_z=\frac{\partial\phi}{\partial z}+\frac{\partial\psi}{\partial x}. \tag{3b}$$

The displacement of the elastic plate, w_p, is described by the thin plate equation under the combined action of normal stress on the upper surface due to the attached elastic layer and the acoustic pressure on the lower surface due to the transmission of the incident acoustic wave:

$$D_p\frac{d^4w_p}{dx^4}-\omega^2\rho_p t_p w_p=\sigma_{zz}(x,z=-t_c) - p_{tr}(x,z=-(t_c+t_p)), \tag{4}$$

where $D_p=Et_p^3/12(1-\nu^2)$ is the complex plate flexural stiffness, E is the Young's modulus, ν is the Poisson's ratio, ρ_p is the plate material density, t_p is the thickness of the plate, t_c is the thickness of the coating, $\sigma_{zz}(x,z=-t_c)$ is the normal stress or pressure at the interface between the plate and the coating, and $p_{tr}(x,z=-(t_c+t_p))$ is the acoustic pressure acting on the lower surface of the plate. The region below the plate is taken to be air-filled.

The normal stress, σ_{zz}, is known in terms of derivatives of the potential functions, In particular,

$$\sigma_{zz}=\lambda_c\left(\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial z^2}\right)+2\mu_c\left(\frac{\partial^2\phi}{\partial z^2}+\frac{\partial^2\psi}{\partial x\,\partial z}\right), \tag{5}$$

where λ_c and μ_c are the Lame constants. Assumed forms for the plate displacement and the transmitted acoustic wave below the plate are given by

$$w_p(x)=W_p e^{jk'x}, \tag{6}$$

$$p_{tr}(x,z)=P_t e^{jk_a(x\,\sin\,\theta_t-z\,\cos\,\theta_t)}, \tag{7}$$

where W_p is the complex amplitude of the plate displacement, k' is the wave number of the wave on the plate's surface, P_t is the complex amplitude of the acoustic wave transmitted below the plate, k_a is the acoustic wave number in the air-filled space, and θ_t is the angle of the transmitted acoustic wave.

The unknowns consist of the various wave amplitudes, the angles of transmission and reflection, and the horizontal wave numbers. These may be determined from the six coupling and boundary conditions:

(i) continuity of velocity at $z=0$:

$$\frac{\partial}{\partial z}(p_i+p_r)|_{z=0}=\omega^2\rho_o u_z|_{z=0}; \tag{8a}$$

(ii) continuity of pressure at $z=0$:

$$(p_i+p_r)|_{z=0}=-\sigma_{zz}|_{z=0}; \tag{8b}$$

(iii) continuity of normal velocity at $z=-t_c$:

$$u_z|_{z=-t_c}=w_p; \tag{8c}$$

(iv) continuity of tangential velocity at $z=-t_c$:

$$\frac{t_p}{2}\frac{\partial w_p}{\partial x}=-u_x; \tag{8d}$$

(v) zero shear stress at $z=0$:

$$\left(2\frac{\partial^2\phi}{\partial x\,\partial z}+\frac{\partial^2\psi}{\partial x^2}-\frac{\partial^2\psi}{\partial z^2}\right)\Bigg|_{z=0}=0; \tag{8e}$$

(vi) continuity of normal velocity at $z=-(t_c+t_p)$:

$$\frac{\partial p_t}{\partial z}\Bigg|_{z=-(t_c+t_p)}=\omega^2\rho_a w_p. \tag{8f}$$

Application of these conditions yields the following relations between the angle of incidence, the angles of transmission and reflection, and wave numbers:

$$\theta_i=\theta_r; \quad \theta_1=\theta_2; \quad \gamma_1=\gamma_2, \tag{9a}$$

$$k_0\sin\,\theta_i=k_\phi\,\sin\,\theta_1=k_\psi\,\sin\,\gamma_1=k_a\,\sin\,\theta_t=k'. \tag{9b}$$

In addition, a system of seven simultaneous algebraic equations remains to determine the seven amplitudes, W_p, P_r, P_t, Φ_1, Φ_2, Ψ_1, and Ψ_2. These equations are given by

TABLE I. Model parameters.

Parameter	Value
Plate material	Steel
Plate thickness	0.04127 m (1.625 in.)
Coating thickness	0.09081 m (3.575 in.)
Coating dilatational wave speed	190*(1–0.3j) m/s
Coating shear wave speed	75*(1–0.3j) m/s

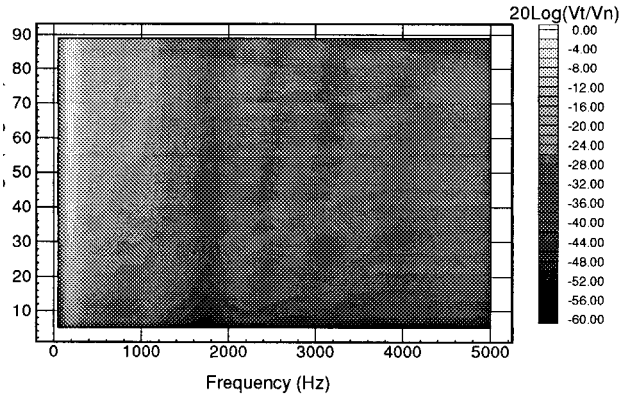


FIG. 2. Ratio of tangential velocity component to normal velocity component for nominal layer.

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \alpha_{14} & \alpha_{15} & 0 \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \alpha_{24} & 1 & 0 \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \alpha_{34} & 0 & -1 \\ \alpha_{41} & \alpha_{42} & \alpha_{43} & \alpha_{44} & 0 & \alpha_{46} \\ \alpha_{51} & \alpha_{52} & \alpha_{53} & \alpha_{54} & 0 & 0 \\ \alpha_{61} & \alpha_{62} & \alpha_{63} & \alpha_{64} & 0 & \alpha_{66} \end{bmatrix} \begin{Bmatrix} \Phi_1 \\ \Phi_2 \\ \Psi_1 \\ \Psi_2 \\ P_r \\ W_p \end{Bmatrix} = P_i \begin{Bmatrix} \beta_1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix}. \tag{10}$$

The various coefficients shown in Eq. (10) are defined in the Appendix. This set of equations may be solved numerically using a Gaussian elimination algorithm for any value of frequency and signal arrival angle desired.

The value of material damping used in the model was found to have a significant effect on the predicted results. Although it is recognized that the coating material is inherently dispersive, in order to strike a balance between accuracy and simplicity, it was decided to use non-dispersive material properties in the analysis. Furthermore, a ''loss factor'' was used, and the loss factors were applied to the two wave speeds of the coating. The nominal values of these complex wave speeds, along with other model parameters used, are shown in Table I.

II. RESULTS

A. Signal velocity components

The coating velocity components at the coating/water interface generated by a plane acoustic wave impinging on the coating at a variety of incidence angles, measured from the vertical, were computed and normalized by the acoustic particle velocity amplitude. Figure 2 shows the ratio of the tangential velocity component to the normal velocity component as a function of incidence angle and frequency for the nominal layer. It is clear that the tangential component is comparable to the normal component only in a rather narrow, low frequency range for nearly grazing angles of incidence. Outside this range, the tangential velocity component is significantly smaller than the normal velocity component.

It is of interest to examine the individual velocity components in greater detail to determine their dependence on arrival angle, frequency, and location through the coating. To focus on the effects of signal arrival angle, the individual velocity components were computed at a fixed frequency of 1000 Hz over a range of incidence angles from broadside (0°) to grazing (90°). The results are shown in Fig. 3 for the tangential velocity component and in Fig. 4 for the normal velocity component. In these figures, the individual velocity components have been normalized by the amplitude of the particle velocity of the incoming acoustic plane wave, $\rho_o c_o / |P_i|$.

The tangential component shown in Fig. 3 is seen to peak when the arrival angle is nearly 45°. Given the nature of the model, this is quite reasonable since the coating response is excited only through normal stress acting at the coating/water interface. The tangential velocity decreases significantly near broadside incidence, and also as the signal approaches grazing incidence. It is noteworthy that, at this

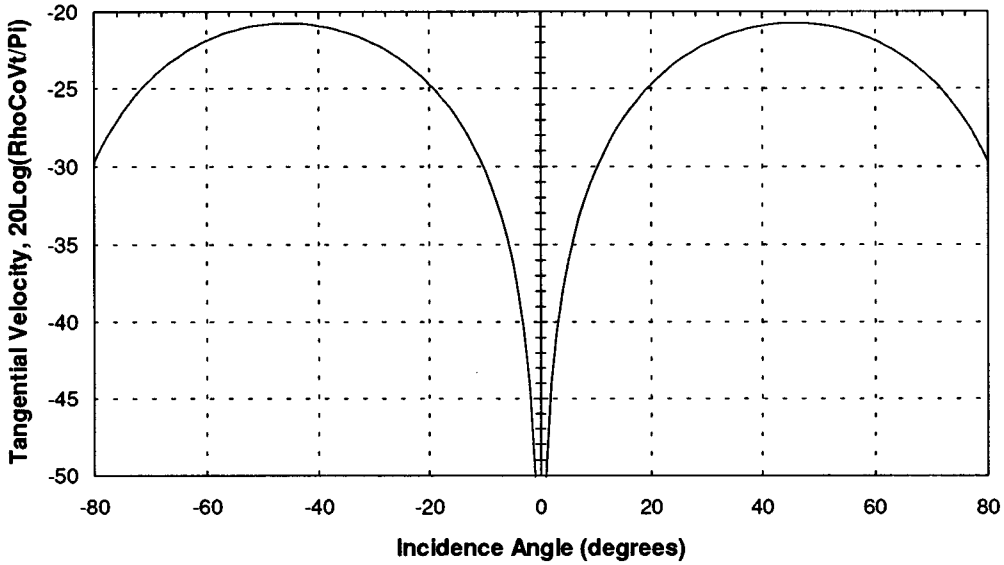


FIG. 3. Normalized tangential velocity as a function of incidence angle at 1000 Hz.

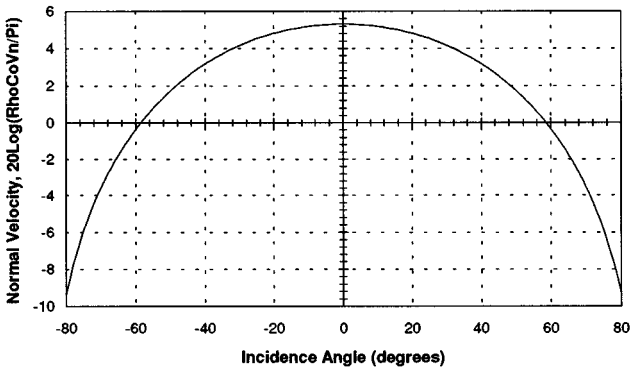


FIG. 4. Normalized normal velocity as a function of incidence angle at 1000 Hz.

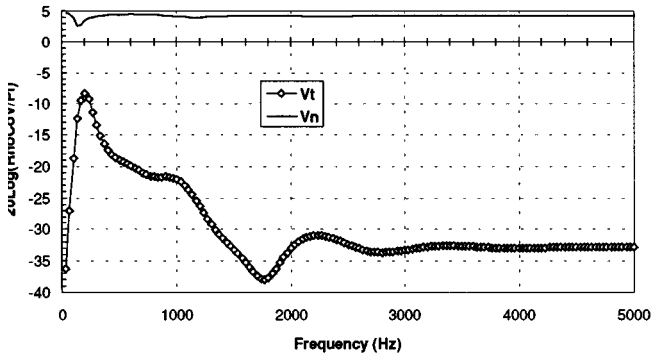


FIG. 5. Normalized coating (loss factors 0.3) velocity components due to 30°—incident plane wave.

frequency of 1000 Hz, the maximum tangential velocity is over 20 dB smaller than the incident wave particle velocity.

The variation of the normal velocity component with incidence angle shown in Fig. 4 exhibits the expected cosine-dependence with the arrival angle. That is, the normal velocity of the incident wave is proportional to the cosine of the incidence angle. It is shown that for arrival angles near broadside, the coating velocity is approximately 5 dB above the particle velocity of the incident wave. Velocity amplification persists to arrival angles of approximately $\pm 60°$. As mentioned in the Introduction, it is this enhancement of the normal velocity over the incident particle velocity that makes this system attractive as an acoustic sensor.

In order to examine the individual velocity components more closely, the normalized tangential and normal components generated by a plane wave incident at a 30° incidence angle are shown as a function of frequency in Fig. 5. Except for a narrow low frequency region, the normal velocity component demonstrates a nearly uniform $+4$ dB signal gain over the frequency range examined. The tangential, or in-plane, component varies considerably with frequency, and is generally several tens of decibels below the normal velocity component.

The observed velocity components change somewhat if a different value of coating damping is used. Figure 6 shows the velocity components when the coating loss factors are changed to 0.1. The normal velocity response is nearly unchanged, except for a more pronounced minimum at low frequency, and the appearance of another dip near 1 kHz. The tangential velocity response, however, demonstrates considerably more structure and variation than in Fig. 5. In addition, the peak tangential levels, while still rather small compared to the normal component, and higher than those observed in Fig. 5.

The reasons for the structure observed in the magnitude of the velocity components as a function of frequency and loss factor may be determined by making use of the waveguide behavior of the elastic coating presented in Ref. 4. Briefly, it was found that the coating response in the coupled plate/coating/water system could be well approximated by a combination of waveguide or mode-like processes. These consisted of a longitudinal waveguide model and a shear waveguide model. The longitudinal waveguide model comprised a rigid–rigid acoustic waveguide whose height was equal to the coating thickness. The physical basis for this model was the relatively small normal velocity of the coating at the plate interface and also, at least for certain frequencies, at the water interface. The shear waveguide behavior was correlated with a rigid-free acoustic waveguide. These boundary conditions are justified by the unrestrained tangen-

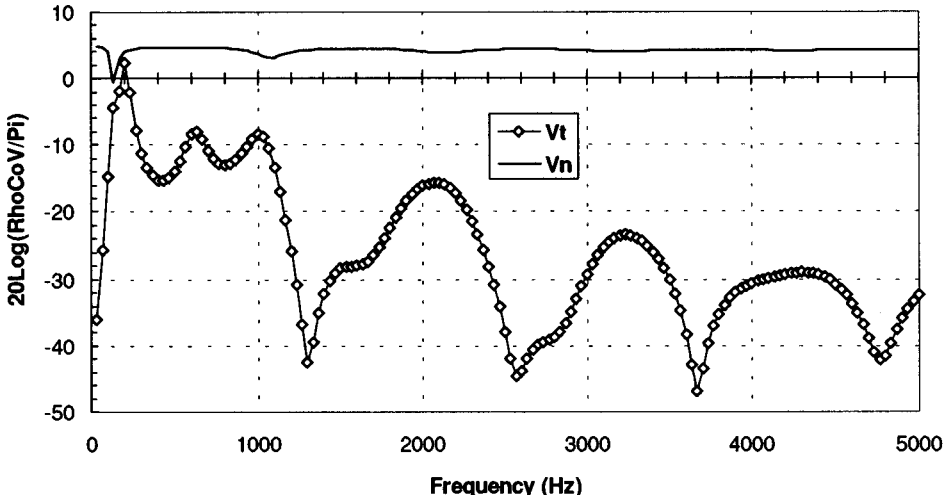


FIG. 6. Normalized coating (loss factors 0.1) velocity components due to 30°—incident plane wave.

TABLE II. Waveguide resonant frequencies of the elastic coating.

Mode number	Shear mode (Hz)	Longitudinal mode (Hz)
0	207	⋯
1	620	1048
2	1033	2096
3	1446	3145
4	1859	4193
5	2272	5241
6	2685	6289
7	3098	7338
8	3511	8386
9	3924	9434
10	4337	10 482

tial velocity at the water interface, coupled with the nearly rigid boundary condition at the plate interface. For the nominal coating properties, the resonant frequencies associated with these modal responses are given in Table II.

Figure 7 shows the normalized coating velocity components (30° incidence) for the case in which the shear loss factor has been set to 0.1 and the longitudinal loss factor has been set to zero. These loss factors accentuate the effects of the longitudinal resonances. Notice, for example, the pronounced minima in the normal velocity at the longitudinal resonances and the corresponding maxima in the tangential velocities. In addition, there are two low frequency peaks in the tangential response that correspond to the first two shear resonances from Table II.

These observations make sense in comparison to the waveguide behavior. The longitudinal resonances correspond to ideally zero normal velocity at the plate and at the water interface. This agrees with the observed behavior in Fig. 7. In addition, shear resonances correspond to large tangential velocity at the plate/water interface. Again, this agrees with the observed behavior in Fig. 7. It may be observed in Fig. 7 that only the first two shear resonances are apparent in the tangential velocity. This may be attributed to the fact that a significant shear loss factor is present in the analysis. On the other hand, one observes a large tangential velocity response at the longitudinal resonances. This behavior is due to the fact that the longitudinal waveguide model does not impose constraints on the tangential velocity component. Thus, as is true in any non-plane wave, hard-walled waveguide mode, the tangential velocity is large. Furthermore, the presence of longitudinal wave damping dramatically eliminates the nulls in the normal velocity response, and has a significant reduction and broadening effect on the peak tangential velocities.

There is another low frequency minimum in the normal velocity data shown in Fig. 7, at approximately 150 Hz, that is not correlated with the waveguide behavior. This minimum may be explained through a simple lumped mass model. Consider the 2-mass model shown in Fig. 8. This model consists of a lumped mass/unit length for the plate, a lumped mass/unit length for the water, and a lumped stiffness for the coating.

Assuming that the height of the water column is one-half an acoustic wavelength, the equations for the system parameters are as follows:

$$\text{water mass/unit area:} \quad m_w = \frac{\pi \rho_o c_o}{\omega}, \tag{11a}$$

$$\text{plate mass/unit area:} \quad m_p = \rho_p t_p, \tag{11b}$$

$$\text{coating stiffness/area:} \quad k_c = \frac{E_c}{h}. \tag{11c}$$

The expression for the natural frequency of a 2-DOF system such as that shown in Fig. 8 is well known,[8] and given by:

$$\omega_n = \frac{E_c}{h} \frac{\pi \rho_o c_o + \rho_p t_p \omega}{\pi \rho_o c_o \rho_p t_p}. \tag{12}$$

Using the values appropriate for the coated plate model, the resonant frequency is calculated to be approximately 160 Hz. Furthermore, given the relative sizes of the masses involved, the mode is characterized by a nearly stationary coating/water interface and a rather mobile coating/plate interface. This behavior correlates very well with the observations from Fig. 7.

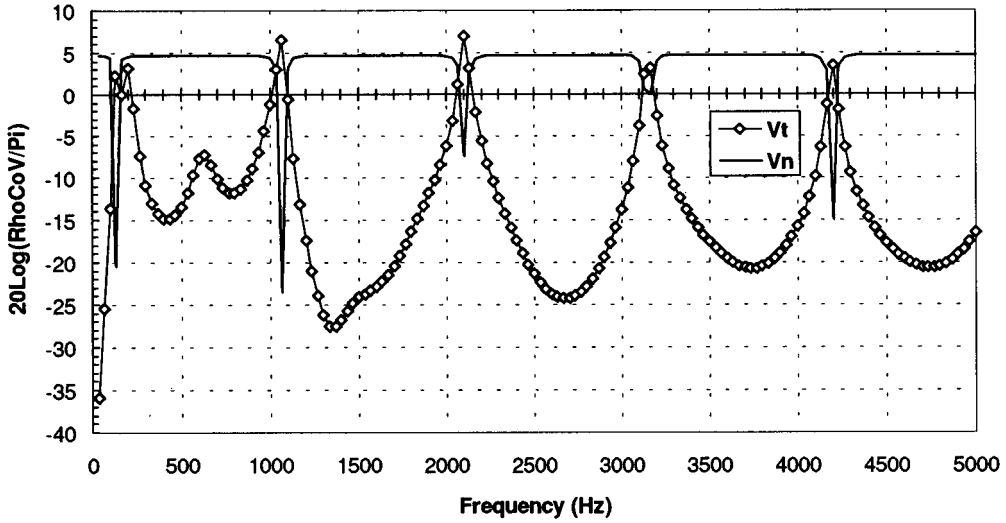


FIG. 7. Normalized coating velocity components due to 30°—incident plane wave (shear loss factor 0.1, longitudinal loss factor 0.0).

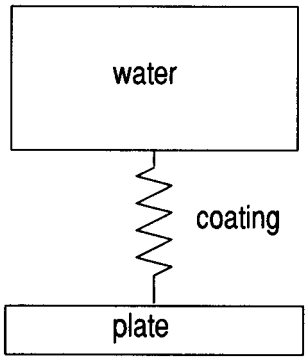


FIG. 8. Lumped mass model of the coated plate system.

B. Signal velocity response through the coating depth

The above discussion addressed the coating velocity components at the coating/water interface. Since it is conceptually possible to locate velocity transducers at any depth within the coating thickness, a series of calculations were performed to determine the signal velocities at various depths. These calculations were done over a range of frequencies and angles of incidence. The velocity ratio, V_t/V_n, was calculated and expressed in decibel form. These results are shown in Fig. 9 for an incidence angle of 5°.

The results shown in Fig. 9 are qualitatively identical to those that were observed for other values of arrival angle as well. Although the amplitudes of the velocity ratio depend on signal arrival angle, with the ratios increasing as the signal moves from broadside to grazing, the relative pattern through the coating depth remains similar. From Fig. 9 it is clear that there are two regions of increased velocity ratio. The first of these occurs at a frequency near the lumped mass resonance described above, and persists throughout most of the coating's depth. The second region occurs over a frequency band centered near 700 Hz, and is localized near to the surface of the plate. There is also a region of low velocity ratio that is apparent in the figure. This region is seen to decrease in depth through the coating as frequency increases. At the present time there is no clear explanation for this feature.

The individual velocity components that were used to generate Fig. 9 are shown in Fig. 10 and Fig. 11 for the tangential and normal velocity components, respectively. One observes a gradual degradation in each velocity response component with increasing frequency and with increasing depth in the elastic coating.

C. Effects of changing coating material properties

It was of interest to determine how the signal velocity components at the coating/water interface would be affected by changes in the coating elastic properties. In particular, the prospect of significantly increasing the tangential or in-plane velocity component was addressed. This was examined by independently varying the amplitudes of the dilatational wave speed and the shear wave speed. The values of the respective loss factors were not changed. The dilatational wave speed was varied from 50 to 900 m/s, and the shear wave speed was varied from 20 to 400 m/s. Furthermore, it was necessary to perform some type of averaging over frequency in order to reduce the volume of data. Five frequency bands were defined: 0–250 Hz; 250–500 Hz; 500–1000 Hz; 1000–2000 Hz; and 2000–4000 Hz. Within each frequency band, the scaled velocity components, in response to a plane wave incident at a 30° angle, were determined for each specific combination of dilatational and shear wave speed. These amplitudes were averaged over all the frequencies within the given frequency band, and the resulting average number saved. This calculation was performed for both the normal and the tangential velocity components, and the results examined.[9]

Based on such calculations over an extensive range of wave speeds, a general conclusion was obtained that reasonably high dilatational speeds (greater than about 300 m/s), together with shear speeds that are less than approximately one-half the dilatational speed, produced desirable results. Such combinations tend to provide the desired normal velocity amplification, and also yield significant tangential velocities (although these will generally be less than the normal component).

Results were calculated for the case of rather large dila-

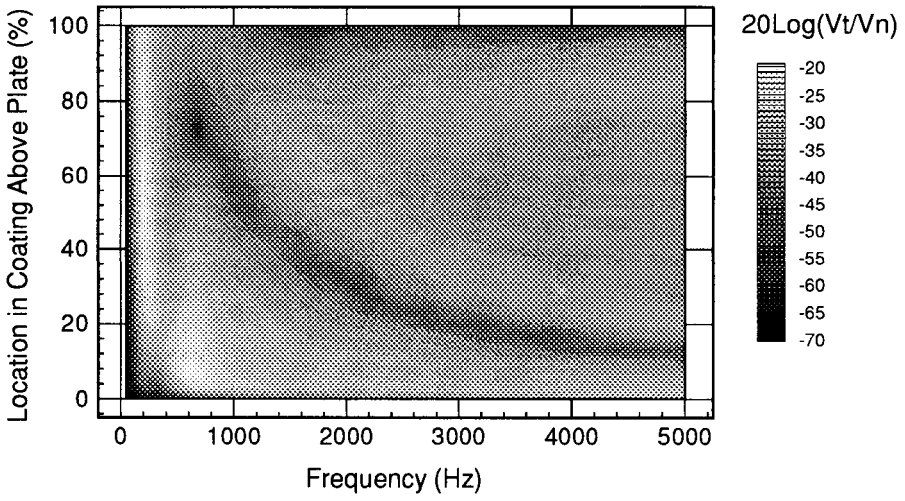


FIG. 9. Velocity ratio as a function of frequency and location through coating at incidence angle of 5°.

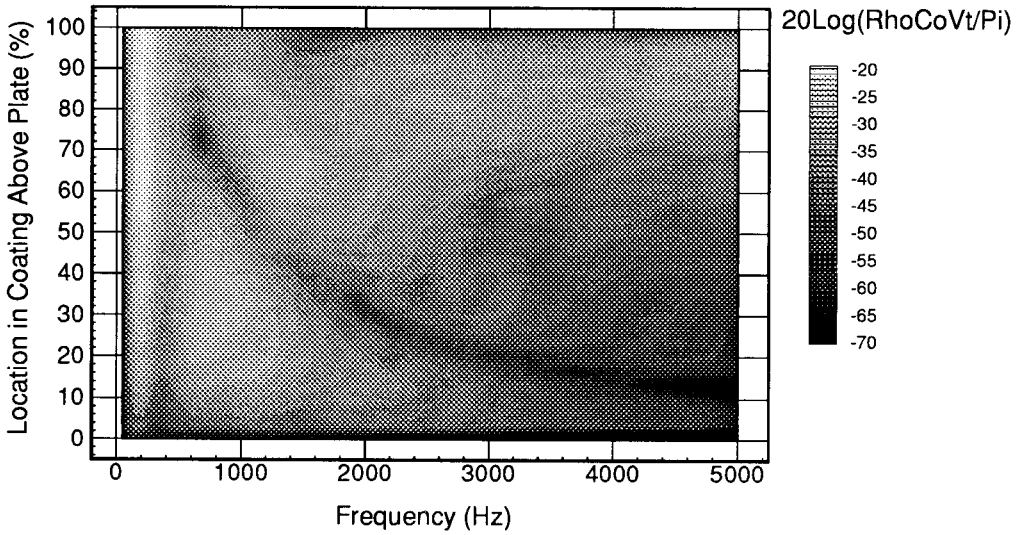


FIG. 10. Tangential velocity component as a function of frequency and location through coating at incidence angle of 5°.

tational speed (900 m/s) and low shear speed (100 m/s). The normalized velocity components were computed for a 30° arrival angle and nominal values of the coating loss factors. The individual velocity components are shown in Fig. 12. It is perhaps easier to observe the differences induced by the changes in coating wave speeds by comparing these results with those of the nominal layer shown in Fig. 5. The decibel differences between the alter coating response and the nominal coating response are shown in Fig. 13. These results show that there has been some degradation in the normal velocity response. There is an -8 dB loss of response at approximately 700 Hz. At higher frequency the degradation is on the order of -2 dB. An average value of loss is about -3 dB. In comparison, there has been a significant increase in the tangential velocity component. While there is still considerable variation with frequency in the tangential component, the average response level has been increased by approximately $+20$ dB. The peak value of enhanced tangential response of $+31$ dB occurred near 1700 Hz.

III. SUMMARY

The response of an elastic coating, attached to a submerged plate, due to an incident acoustic plane wave has been formulated. The plate was considered to be thin, and the coating was modeled as a 2-D elastic solid. The coating could support a combination of shear and dilatational wave motions. The acoustic wave was incident from a water-filled half-space adjoining the coating, propagated through the coating and the plate, and exited into an air-filled half-space below the plate. The system model was developed analytically, utilizing appropriate coupling and compatibility conditions at the various interfaces. The model for the response of the coating/plate system consisted of a set of seven coupled algebraic equations.

Primary interest was focused on examination of the normal and tangential velocity components in the elastic coating produced by the incident acoustic plane wave. The frequency and the angle of incidence of the plane wave were both var-

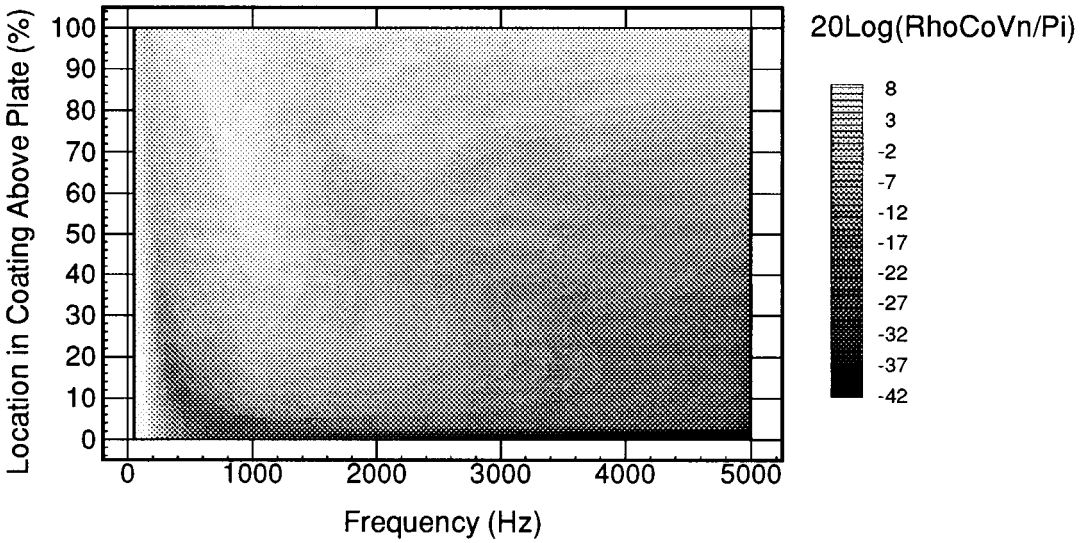


FIG. 11. Normal velocity component as a function of frequency and location through coating at incidence angle of 5°.

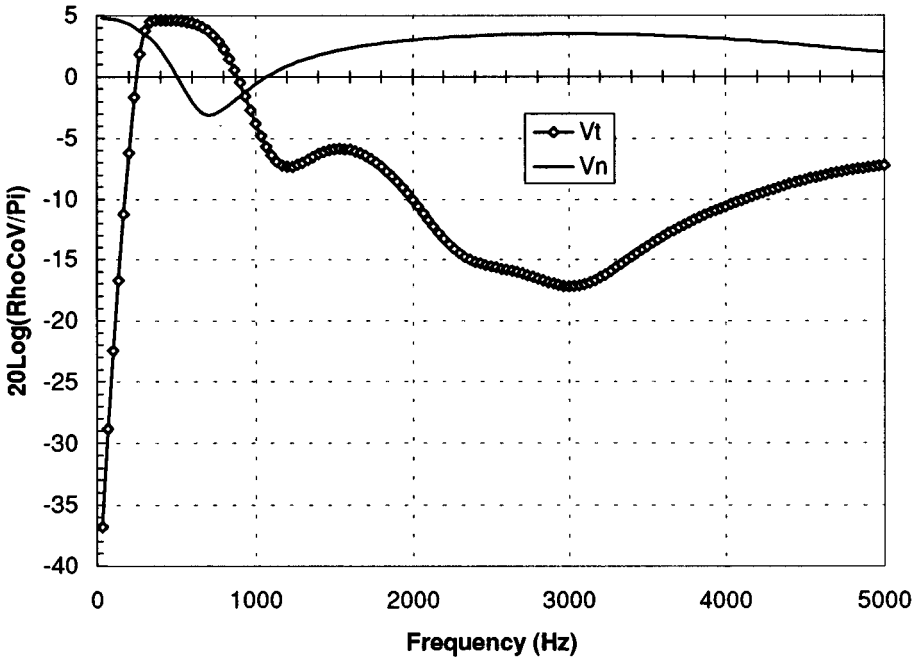


FIG. 12. Normalized velocity components for a coating with longitudinal wave speed of 900 m/s and a shear wave speed of 100 m/s due to 30°—incident plane wave.

ied, as was the location in the depth of the coating at which the velocities were computed.

The magnitudes of the velocity components were found to be sensitive to the amount of dissipation in the coating material. For low values of damping, the waveguide responses of the layer produced dramatic peaks and valleys associated with the shear and dilatational resonances. In general, minima in the normal velocity component correlated with dilatational resonances, and maxima in the tangential velocity component correlated with shear resonances.

When the value of damping was set at physically realistic levels, most of the excursions in the velocity response were eliminated. The tangential velocity component retained significant variations with frequency, although the mean levels were tens of decibels below the normal velocity component. The normal component became nearly constant (at approximately +5 dB above the incident particle velocity), except for a null at low frequency corresponding to a lumped

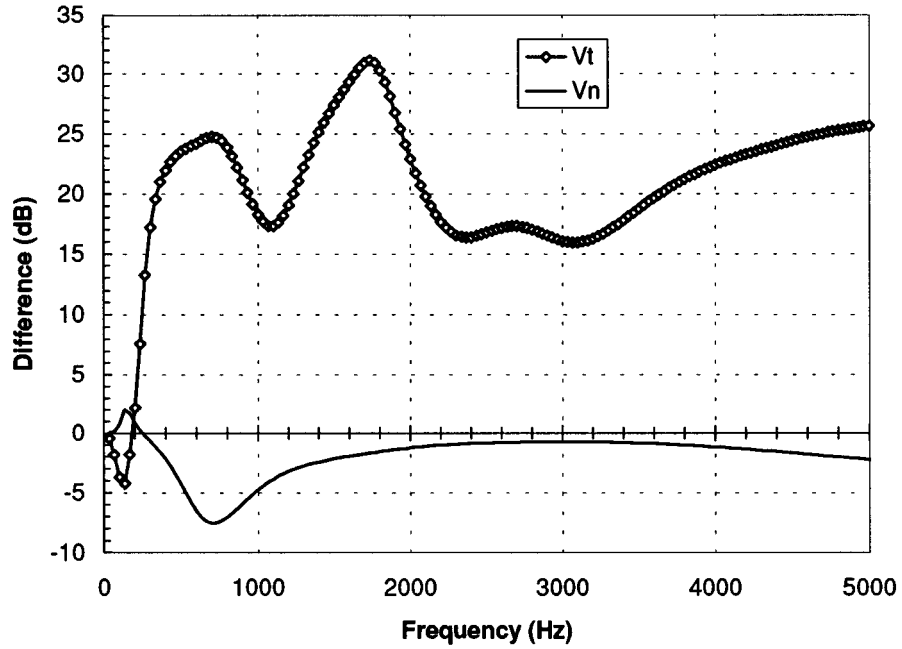


FIG. 13. Difference in velocity components between nominal coating and coating with longitudinal wave speed of 900 m/s and a shear wave speed of 100 m/s due to 30°—incident plane wave.

mass resonance. At this frequency of approximately 150 Hz, the system is described by an equivalent mass of water connected to an equivalent mass of the plate by an equivalent stiffness of the coating. The normal motion of the water/coating interface is much less than the normal motion of the coating/plate interface.

The effects of coating wave speeds on the velocity response components were also examined. The general conclusion from this study showed that the tangential velocity component could be increased, without having a deleterious effect on the normal velocity component, by achieving a rather high dilatational speed (greater than about 300 m/s) in combination with a shear speed that is less than about one-half the dilatational speed. In any case, it remained true that the tangential response would be significantly lower than the normal response.

ACKNOWLEDGMENTS

The author gratefully acknowledges the support of the Naval Undersea Warfare Center, New London Detachment, Code 2133, for their sponsorship of the IPA activity under which this work was performed. Special thanks are extended to Dr. Benjamin A. Cray, who served as technical liaison for this activity.

APPENDIX

The coefficients of the terms in Eq. (10) are given by:

$$\alpha_{11}=\omega^2\rho_0 k_\phi \cos\theta_1, \tag{A1}$$

$$\alpha_{12}=-\omega^2\rho_0 k_\phi \cos\theta_2, \tag{A2}$$

$$\alpha_{13}=-\omega^2\rho_0 k_\psi \sin\gamma_1, \tag{A3}$$

$$\alpha_{14}=-\omega^2\rho_0 k_\psi \sin\gamma_2, \tag{A4}$$

$$\alpha_{15}=k_0 \cos\theta_r, \tag{A5}$$

$$\alpha_{21}=-k_\phi^2(\lambda_c+2\mu_c \cos^2\theta_1), \tag{A6}$$

$$\alpha_{22}=-k_\phi^2(\lambda_c+2\mu_c \cos^2\theta_2), \tag{A7}$$

$$\alpha_{23}=2\mu_c k_\psi^2 \sin\gamma_1 \cos\gamma_1, \tag{A8}$$

$$\alpha_{24}=-2\mu_c k_\psi^2 \sin\gamma_2 \cos\gamma_2, \tag{A9}$$

$$\alpha_{31}=-jk_\phi \cos\theta_1 \exp(jk_\phi t_c \cos\theta_1), \tag{A10}$$

$$\alpha_{32}=jk_\phi \cos\theta_2 \exp(-jk_\phi t_c \cos\theta_2), \tag{A11}$$

$$\alpha_{33}=jk_\psi \sin\gamma_1 \exp(jk_\psi t_c \cos\gamma_1), \tag{A12}$$

$$\alpha_{34}=jk_\psi \sin\gamma_2 \exp(-jk_\psi t_c \cos\gamma_2), \tag{A13}$$

$$\alpha_{41}=k_\phi \sin\theta_1 \exp(jk_\phi t_c \cos\theta_1), \tag{A14}$$

$$\alpha_{42}=k_\phi \sin\theta_2 \exp(-jk_\phi t_c \cos\theta_2), \tag{A15}$$

$$\alpha_{43}=k_\psi \cos\gamma_1 \exp(jk_\psi t_c \cos\gamma_1), \tag{A16}$$

$$\alpha_{44}=-k_\psi \cos\gamma_2 \exp(-jk_\psi t_c \cos\gamma_2), \tag{A17}$$

$$\alpha_{46}=t_p k'/2, \tag{A18}$$

$$\alpha_{51}=2\mu_c k_\phi^2 \sin\theta_1 \cos\theta_1, \tag{A19}$$

$$\alpha_{52}=-2\mu_c k_\phi^2 \sin\theta_2 \cos\theta_2, \tag{A20}$$

$$\alpha_{53}=\mu_c k_\psi^2(\cos^2\gamma_1-\sin^2\gamma_1), \tag{A21}$$

$$\alpha_{54}=\mu_c k_\psi^2(\cos^2\gamma_2-\sin^2\gamma_2), \tag{A22}$$

$$\alpha_{61}=k_\phi^2(\lambda_c+2\mu_c \cos^2\theta_1)\exp(jk_\phi t_c \cos\theta_1), \tag{A23}$$

$$\alpha_{62}=k_\phi^2(\lambda_c+2\mu_c \cos^2\theta_2)\exp(-jk_\phi t_c \cos\theta_2), \tag{A24}$$

$$\alpha_{63}=-2\mu_c k_\psi^2 \sin\gamma_1 \cos\gamma_1 \exp(jk_\psi t_c \cos\gamma_1), \tag{A25}$$

$$\alpha_{63}=2\mu_c k_\psi^2 \sin\gamma_2 \cos\gamma_2 \exp(-jk_\psi t_c \cos\gamma_2), \tag{A26}$$

$$\alpha_{66}=D_p k'^4-\rho_p t_p \omega^2, \tag{A27}$$

$$\beta_1=k_0 \cos\theta_i. \tag{A28}$$

[1] S. H. Ko, ''Signal reception by a velocity sensor placed on the surface of a microvoided elastomer layer backed by a hull plate,'' NUWC-NPT Technical Report 10,642, Newport, Rhode Island, 6 September 1996, pp. 1–2.

[2] G. L. D'Spain *et al.*, ''Initial analysis of the data from the vertical DIFAR array,'' *Proceedings of Mastering the Oceans Through Technology (OCEANS 92)*, pp. 346–351, Newport, Rhode Island, 26–29 October 1992.

[3] B. A. Cray and A. H. Nuttall, ''A comparison of vector-sensing and scalar-sensing linear arrays,'' NUWC-NPT Technical Report 10,632, Newport, Rhode Island, 12 January 1997, pp. 2–4.

[4] R. F. Keltie and W. D. Crank, ''Structural acoustic response of elastic coatings attached to vibrating plates,'' IPA Activity Report to NUWC-New London, New London, Connecticut, 21 October 1995, pp. 2–10.

[5] W. D. Crank, ''Response of an elastic layer which separates a vibrating plate from an acoustic half-space,'' Ph.D. dissertation, North Carolina State University, December 1997.

[6] L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1980), pp. 28–32.

[7] Karl F. Graff, *Wave Motion in Elastic Solids* (Ohio State U.P., Columbus, OH, 1975), pp. 273–281.

[8] A. Dimarogonas, *Vibration for Engineers*, 2nd ed. (Prentice-Hall, Englewood Cliffs, NJ, 1996), pp. 277–280.

[9] R. F. Keltie, ''Response of elastically-coated plates to incident acoustic plane waves,'' IPA Activity Report to NUWC-New London, New London, Connecticut, 1 November 1996, pp. 13–19.

Rib resonances present in the scattering response of a ribbed cylindrical shell

Martin H. Marcus and Angie Sarkissian
Naval Research Laboratory, Washington, DC 20375-5350

(Received 2 May 1997; revised 2 October 1997; accepted 19 December 1997)

The presence of ribs in an evacuated, finite cylindrical shell, placed in a fluid medium, produces highlights in its scattering response at certain resonant frequencies. These resonances are identified from the end-incident bistatic response of a ribbed cylindrical shell and are shown to be produced by the flexural vibrations of individual ribs satisfying the boundary conditions of an annular plate. [S0001-4966(98)00804-2]

PACS numbers: 43.40.Dx [CBB]

INTRODUCTION

The scattering response of an evacuated ribbed cylindrical shell in an unbounded fluid medium contains, among many features, highlights produced by the presence of individual ribs. Rib resonances[1,2] are different from Bloch wave resonances,[1,3] which may also be produced by ribs if the ribs have regular spacing. Rib resonances do not require regular spacing and may be produced by the vibrations of a single rib present in the structure.

Using a finite elements/infinite elements computation for the scattering response of a ribbed cylindrical shell, rib resonances are identified and shown to be produced by flexural vibrations of individual ribs satisfying boundary conditions of an annular plate.

I. END-INCIDENT BISTATIC RESPONSE OF A RIBBED CYLINDER

Figure 1 shows the end-incident bistatic response of a ribbed finite cylindrical shell with hemispherical end caps. The cylinder, shown in Fig. 2, contains 85 ribs, all of which are of equal size. The computations were performed using program Sara-2d[4] which uses finite elements to model the cylindrical structure and finite and infinite elements to model the unbounded fluid medium external to the structure. For additional details on the finite elements and infinite elements approach, the reader is referred to Ref. 5 or Ref. 6.

Figure 3 shows enlarged version of the ribs. The cylinder used for the computations has radius a, a total length $L=13.7a$ and shell thickness $t=0.0074a$. The ribs have length $l=0.074a$, thickness $h=0.0065a$, and rib spacing of $d=0.14a$. The elastic parameters used are that of nickel with modulus of elasticity $E=2.1\times10^{11}$ Pa, Poisson's ratio $\nu=0.30$, density $\rho=8800$ kg/m^3, and a loss factor $\eta=0.005$. The external fluid is water with sound speed of $c=1500$ m/s and density of $\rho_1=1000$ kg/m^3.

The bistatic response is shown for a frequency range of $2.5\leqslant ka\leqslant 10$, where $k=\omega/c$. Angles range from 0° representing end-incident forward scattering to 180° representing end-incident backscattering. The target strength plot is normalized to reach a maximum value of 0 dB. The highlight observed near $ka=8$ and higher is due to Bloch waves produced by the periodicity of the rib spacing. A horizontal highlight is observed near $ka=3$. For this end-incident condition, the resonant behavior at this frequency seems independent of the scattered angle. This resonance may be identified most easily in the forward direction where it appears as a dip in the target scattering response.

Figure 4 shows plots of the end-incident forward scattered target strength as a function of ka for three different rib lengths. Again, each curve is normalized to reach a maximum value of 0 dB. In each case a resonance may be identified as a dip followed by a peak in the scattering response as a function of frequency. The solid line in Fig. 4 represents a line plot of Fig. 1 at an angle of 0° where the resonance condition near $ka=3$ appears as a dip in the forward scattering response. It can be seen from the three curves in Fig. 4 that the forward target strength values drop by several dBs at the rib resonances. We also observe from Fig. 1 that rib resonances occur at all receiver angles.

II. RIB RESONANCES

To determine the frequencies where rib resonances occur we use a thin plate model which is applicable as long as the plate thickness does not exceed $\lambda_s/20$, where λ_s is the shear wavelength.[7] The results shown in this article are all below $ka=10$, where $h/\lambda_s=0.0051$, well within the limit where thin plate theory is applicable. The lateral displacement field w on the surface of a rib, as shown in Fig. 3, satisfies the plate equation[8]

$$D\nabla^4 w+\rho h\,\frac{\partial^2 w}{\partial t^2}=0, \tag{1}$$

where

$$D=\frac{Eh^3}{12(1-\nu^2)}.$$

Choosing the origin of the coordinate system to be at the center of a rib and oriented such that the rib lies in the x-y plane, the solution to the above equation may be written as

$$w(r)=\sum_{n=0}^{\infty}[a_nJ_n(k_fr)+b_nY_n(k_fr)+c_nK_n(k_fr)+d_nI_n(k_fr)]e^{in\theta}, \tag{2}$$

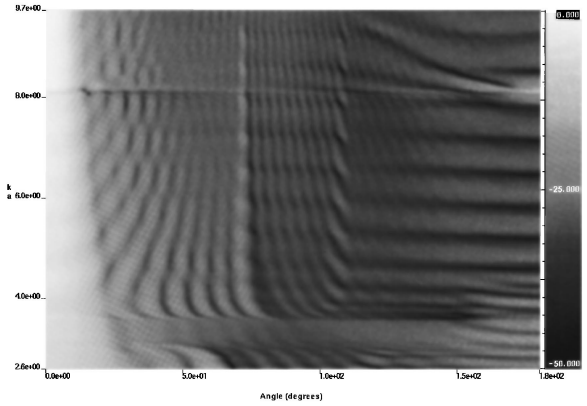


FIG. 1. End-incident bistatic scattering response of the ribbed cylindrical shell.

where J_n, Y_n, K_n, and I_n are the cylindrical Bessel functions, $r=\sqrt{x^2+y^2}$ and k_f is the flexural wave number,

$$k_f=\left(\frac{\rho h\omega^2}{D}\right)^{1/4}.$$

The $e^{-i\omega t}$ time dependence has been suppressed. The axisymmetric case is considered here where the field has no θ dependence in which case only the $n=0$ component of the sum in Eq. (2) contributes to the displacement,

$$w(r)=a_0J_0(k_fr)+b_0Y_0(k_fr)+c_0K_0(k_fr)+d_0I_0(k_fr). \tag{3}$$

The boundary conditions require the vanishing of the bending moment M and the shear Q at the free end,

$$M(b)=-D\left(\frac{\partial^2 w}{\partial r^2}+\frac{\nu}{r}\frac{\partial w}{\partial r}\right)\bigg|_{r=b}=0, \tag{4}$$

where $b=a-l$,

$$Q(b)=-D\,\frac{\partial}{\partial r}\,\nabla^2 w\bigg|_{r=b}=0, \tag{5}$$

while at the attached end, we approximate the plate to be attached to fixed supports where the rotation is related to the bending moment and shear by[9]

$$\frac{\partial w}{\partial r}\bigg|_{r=a}=\frac{16.67M}{Eh^2}+\frac{(1-\nu)Q}{Eh}\bigg|_{r=a}, \tag{6}$$

and the displacement vanishes,

$$w(a)=0. \tag{7}$$

Satisfying the requirements of the boundary conditions results in the matrix equations

$$A_{i1}a_0+A_{i2}b_0+A_{i3}c_0+A_{i4}d_0=0, \quad \text{for } i=1,2,3,4, \tag{8}$$

where

FIG. 2. The ribbed cylinder used for the numerical simulations.

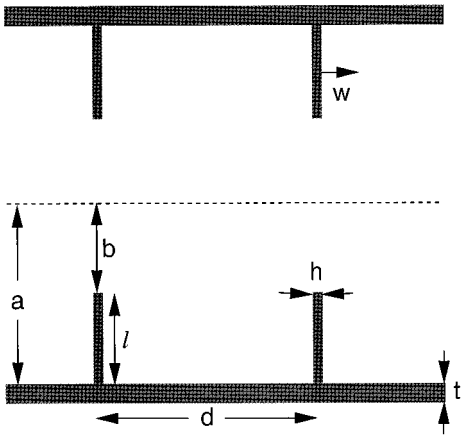


FIG. 3. Enlarged version of the ribs shown attached to the cylinder.

$$A_{11}=J_0(k_fb)+\frac{\nu-1}{k_fb}\,J_1(k_fb),$$

$$A_{12}=Y_0(k_fb)+\frac{\nu-1}{k_fb}\,Y_1(k_fb),$$

$$A_{13}=-K_0(k_fb)+\frac{\nu-1}{k_fb}\,K_1(k_fb),$$

$$A_{14}=-I_0(k_fb)-\frac{\nu-1}{k_fb}\,I_1(k_fb),$$

$$A_{21}=J_1(k_fb), \quad A_{22}=Y_1(k_fb),$$

$$A_{23}=-K_1(k_fb), \quad A_{24}=I_1(k_fb)$$

$$A_{31}=-\left[1-\frac{k_f^2h^2}{12(1+\nu)}-\frac{16.67h}{12(1+\nu)r}\right]J_1(k_fa)-\frac{16.67k_fh}{12(1-\nu^2)}\,J_0(k_fa),$$

$$A_{32}=-\left[1-\frac{k_f^2h^2}{12(1+\nu)}-\frac{16.67h}{12(1+\nu)r}\right]Y_1(k_fa)-\frac{16.67k_fh}{12(1-\nu^2)}\,Y_0(k_fa),$$

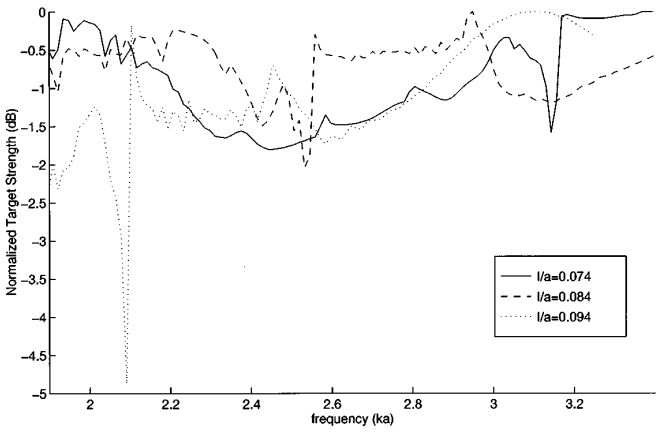


FIG. 4. Forward normalized target strength as a function of ka for three different rib lengths.

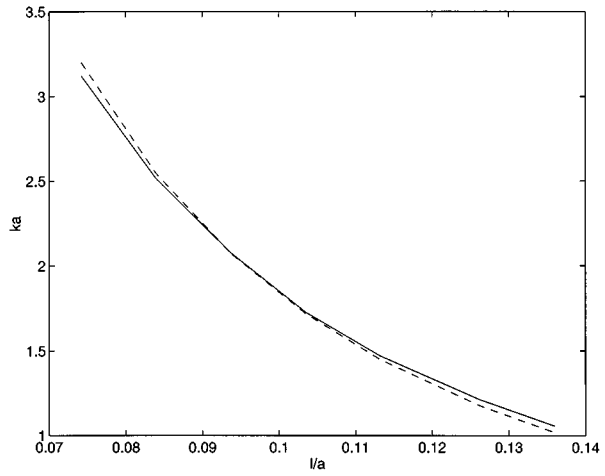


FIG. 5. Resonant frequencies computed numerically using finite element analysis, shown by the solid line, compared to values determined analytically, shown by the dashed line.

$$A_{33} = -\left[1 + \frac{k_f^2 h^2}{12(1+\nu)} - \frac{16.67h}{12(1+\nu)r}\right] K_1(k_f a) + \frac{16.67 k_f h}{12(1-\nu^2)} K_0(k_f a),$$

$$A_{34} = \left[1 + \frac{k_f^2 h^2}{12(1+\nu)} - \frac{16.67h}{12(1+\nu)r}\right] I_1(k_f a) + \frac{16.67 k_f h}{12(1-\nu^2)} I_0(k_f a),$$

$$A_{41} = J_0(k_f a), \quad A_{42} = Y_0(k_f a),$$

$$A_{43} = K_0(k_f a), \quad A_{44} = I_0(k_f a).$$

The resonant frequencies are obtained by requiring the determinant of matrix **A** to vanish. The dashed line in Fig. 5 shows the resulting lowest order resonant frequencies as a function of rib length using dimensionless quantities ka and l/a. The solid line displays resonances numerically computed using finite elements analysis. These resonances were obtained by examining a plot of the forward target strength as a function of frequency for each rib length, from which the resonant frequency was determined by eye. The two curves in Fig. 5 agree very closely.

Although rib resonances occur for all incident angles of the acoustic field, the end-incident case is examined here where the incident field is axisymmetric. Since the rib boundary conditions at the attached end are chosen to be attached to fixed supports, no angular dependence appears in rib resonances. This is apparent in Fig. 1 where the rib resonance occurs near $ka = 3$ for all receiver angles.

III. CONCLUSIONS

Using finite element computations of the scattering response of a ribbed cylindrical shell, rib resonances are identified. The resonances are produced by flexural vibrations on the ribs which satisfy boundary conditions of an annular plate having one free end and one end welded to the structure. A specific algorithm is presented to compute the axisymmetric resonances produced by the ribs. The results of the algorithm are shown to agree very closely to rib resonances determined by finite element computations.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research Code 334.

[1] C. H. Hodges, J. Power, and J. Woodhouse, ''The low frequency vibration of a ribbed cylinder, part 1: Theory,'' J. Sound Vib. **101**, 219–235 (1985).
[2] T. Charnley, R. Perrin, V. Mohanan, and H. Banu, ''Vibrations of thin rings of rectangular cross-section,'' J. Sound Vib. **134**, 455–488 (1989).
[3] D. M. Photiadis, J. A. Bucaro, and B. H. Houston, ''Scattering from flexural waves on a ribbed cylindrical shell,'' J. Acoust. Soc. Am. **96**, 2785–2790 (1994).
[4] H. Allik, R. Dees, S. Moore, and D. Pan, ''Sara-2*d* User's Manual,'' version 95-3, BBN Systems and Technologies, New London, CT, 1995.
[5] O. C. Zienkiewicz, C. Emson, and P. Bettess, ''A novel boundary infinite element,'' Int. J. Numer. Methods Eng. **19**, 393–404 (1983).
[6] O. C. Zienkiewicz, K. Bando, P. Bettess, C. Emson, and T. C. Chiam, ''Mapped infinite elements for exterior wave problems,'' Int. J. Numer. Methods Eng. **21**, 1229–1251 (1985).
[7] M. C. Junger and D. Feit, *Sound, Structures, and Their Interaction* (Acoustical Society of America, Woodbury, NY, 1993), p. 215.
[8] S. Timoshenko and S. Woinowsky-Krieger, *Theory of Plates and Shells* (McGraw-Hill, New York, 1959), 2nd ed., p. 283.
[9] R. J. Roark and W. C. Young, *Formulas for Stress and Strain* (McGraw-Hill, New York, 1982), 5th ed., p. 146.

Local admittance model for acoustic scattering from a cylindrical shell with many internal oscillators

J. A. Bucaro, A. J. Romano,[a] A. Sarkissian, D. M. Photiadis, and B. H. Houston
Naval Research Laboratory, Washington, DC 20375

(Received 5 April 1997; accepted for publication 10 December 1997)

This paper considers whether local admittance fluctuations can explain the acoustic scattering from a submerged shell with many internal oscillators observed by Bucaro *et al.* [Proc. ASME, Noise Control and Acoustics Div., NCA **22**, 87–92 (1996)] and Photiadis *et al.* [J. Acoust. Soc. Am. **101**, 895–899 (1997)]. The scattering is computed using a Helmholtz-based approach for a locally reacting surface whose admittance is affected by the random placement of internal oscillators. The predicted scattering patterns versus frequency and monostatic scattering angle are shown to agree with what is observed in the measured results. © *1998 Acoustical Society of America.* [S0001-4966(98)00104-0]

PACS numbers: 43.40.At, 43.20.Fn, 43.40.Qi, 43.30.Jx [CBB]

INTRODUCTION

Recently, measurements related to the effect of internal oscillators on the acoustic scattering from submerged shells have demonstrated that the presence of such oscillators can have a profound effect on the angle/frequency monostatic scattering patterns.[1,2] Figure 1, taken from Refs. 1 and 2, wherein is compared the monostatic angle/frequency scattering patterns for a framed cylindrical shell both with and without internal oscillators, clearly shows this dramatic change. As can be seen, the effect of the internal oscillators has been to increase the scattering in the ''off-beam'' angular aspects and to do so all the more as frequency is lowered. This increased character of the angular scattering, together with the somewhat ''speckled'' appearance, forms what might be called a ''christmas tree'' in frequency-angle space. This lies in stark contrast to the sharp, discrete ''chalicelike'' [3] features found in the empty shell result. For aspects away from beam (90°) and ends (0°), angularly spread scattering from smooth cylindrical shells in this frequency regime is typically connected with elastic shell-wave propagation effects. For example, note in Fig. 1 the clear angular features from the empty shell associated with helical membrane waves[3] (the ''chalice'') and the Bloch or Floquet flexural waves[4] (the horizontal feature extending across all angles). Accordingly, it is natural to consider modified elastic wave effects in order to interpret the ''christmas tree'' pattern seen in the oscillator shell. Indeed, Bucaro *et al.*[1] and Photiadis *et al.*[2] have discussed oscillator modification to the flexural waves (from circumferential mode coupling) and to the membrane waves (from frequency-dependent mass coupling) as mechanisms potentially playing a role in producing the new scattering characteristics.

In this paper, we seek to demonstrate that there exists another mechanism, one not necessarily connected directly to shell elastic wave phenomena, which can lead to the angular scattering patterns that are seen in the case of the oscillator shell. In particular, we consider scattering from a shell having random spatial fluctuations in its local admittance, Y. In the case of the oscillator shell of Refs. 1 and 2, these fluctuations would presumably arise from the random azimuthal placement of the 80 sets of 11 oscillators, each of the 11 oscillators having a unique resonance band.

Our approach is to consider the special case of a locally reacting surface without any contributions from elastic wave propagation. We intend to determine what the scattered pressure spectrum would look like from such a mechanism and to what extent the experimental measurements from the oscillator shell have the same essential features. We do not intend to determine whether elastic wave scattering effects are entirely missing but rather whether scattering from an oscillator-caused spatially fluctuating admittance is the dominant source of the unusual ''christmas tree'' scattering patterns observed in angle-frequency space.

In the following section (Sec. I), we formulate the theoretical approach used in this study. In Sec. II, we present calculated scattering patterns for several cases of interest and discuss the implications of these results regarding the particular experimental case studied by Bucaro *et al.*[1] and Photiadis *et al.*[2]

I. THEORETICAL FORMULATION

The geometry for the shell scattering calculation is shown in Fig. 2. We chose to use a Helmholtz integral-based approach since it reduces to surface integrals involving the spatially dependent admittance, $Y(\phi,z)$. In addition, we are specifically interested in local mechanisms (rather than, for example, wave effects), and for this case, the Helmholtz integral approach leads fairly simply to expressions for the scattered field in terms of $Y(\phi,z)$.

The total pressure, p, can be written in terms of the following integral:[5]

$$p=\int GJ\,dV+\int [p\hat{n}\cdot\nabla G-i\omega\rho Gv]dS, \tag{1}$$

where G is any Green's function of the Helmholtz equation, J is the spatial function for a distant acoustical source, ρ the

[a]Also at SFA, Inc., Landover, MD 20785.

 0001-4966/98/103(4)/1867/7/$10.00

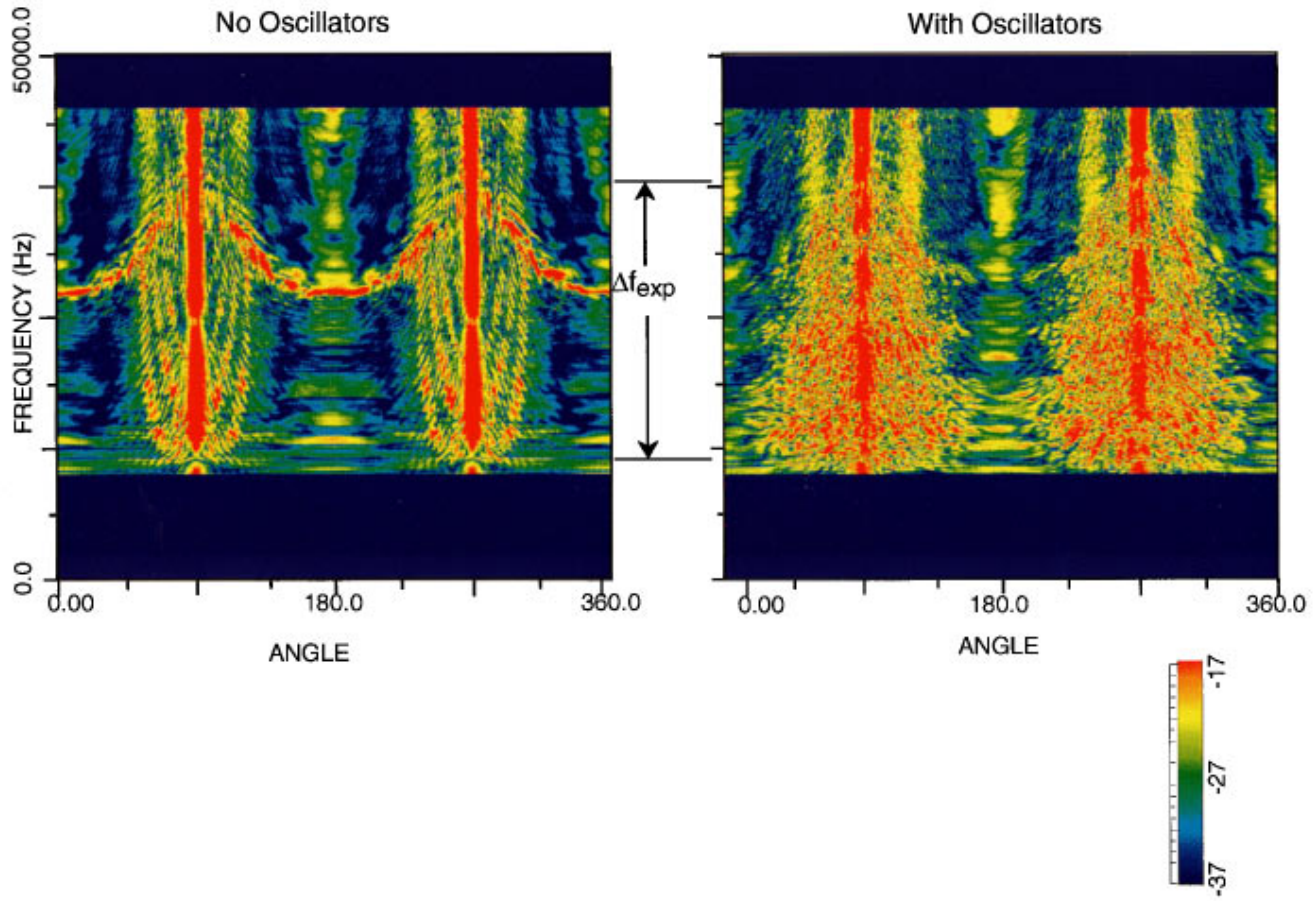


FIG. 1. Monostatic scattering versus frequency and angle for an empty, framed cylindrical shell (left) and for the same shell with 880 dumbbell mechanical oscillators mounted inside (right). Here the color scale gives the magnitude of the backscattered pressure in decibels. These data are from Refs. 1 and 2.

density, ω the angular frequency, and v the surface normal velocity.

We assume a locally reacting surface such that $v=-Yp$ and define the admittance, Y, in terms of a uniform part and fluctuating part:

$$Y=Y_o+Y_a\,. \tag{2}$$

If the Neumann Green's function, G_{N}, is chosen, Eq. (1) can then be used to express the scattered pressure, p_s, as

$$p_s=p_{os}+p_{as}\,, \tag{3}$$

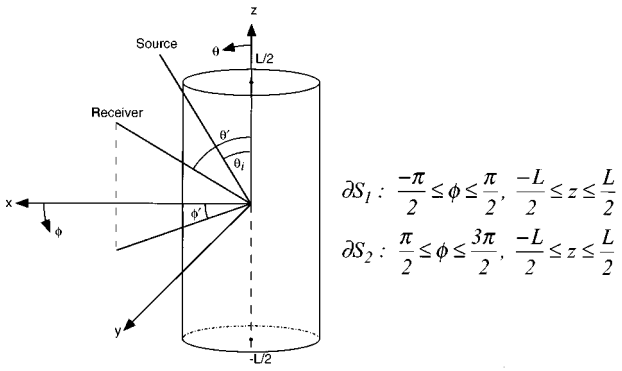


FIG. 2. The geometry used to consider general scattering from a finite cylindrical shell of radius a and length L.

with

$$p_{as}=i\omega\rho\int G_{\mathrm{N}}Y_a p\; dS$$

and

$$p_{os}=p_{os}(\text{rigid})+i\omega\rho\int G_{\mathrm{N}}Y_o p\; dS. \tag{4}$$

Here, p_{as} and p_{os} represent the scattering (at a distant receiver) due to the fluctuating and uniform admittances, respectively. The term labeled rigid for p_{os} refers to scattering from an identical uniform shell with rigid walls and comes from the first term on the right-hand side of Eq. (1) when the Neumann Green's function is employed.

In order to obtain a simple approximation to the scattered pressure, we make a Born[6] approximation by taking the surface pressure in Eq. (4) to be $p\simeq p_i+p_{os}$, i.e., the sum of incident and scattered pressures. Next, a Kirchhoff approximation for p_{os} is made by assuming p_{os} and p_i to be related (on the illuminated side of the cylinder) by a simple plane reflection coefficient, C_τ.[7] Also, p_i and p_{os} (when evaluated on the surface) are assumed to be zero on the shadow side. With these approximations, Eqs. (4) become

$$p_{as}=i\omega\rho\int (1+C_r(\theta_i,\phi))p_i G_{\mathrm{N}}Y_a\; dS$$

and (5)

$$p_{os}=p_{os}(\text{rigid})+i\omega\rho\int (1+C_r(\theta_i,\phi))p_i G_N Y_o\, dS,$$

with

$$C_r(\theta_i,\phi)=\left.\frac{p_{os}(\theta_i,\phi)}{p_i(\theta_i,\phi)}\right|_{r=a}. \tag{6}$$

Taking the Neumann Green's function to be simply twice the free space Green's function (given in terms of $R=|\mathbf{r}-\mathbf{r}'|$, where the prime indicates the field point and unprimed the target surface point) gives

$$G_N=\frac{e^{ikR}}{2\pi R}, \tag{7}$$

where k is the acoustic wave number in the fluid. For $r'\gg r$ we have

$$R\sim r'-z\cos\theta'-\rho\sin\theta'\cos(\phi-\phi'). \tag{8}$$

The first of Eqs. (5) then becomes

$$p_{as}(r',\theta',\phi')$$
$$=i\omega\rho\frac{e^{ikr'}}{2\pi r'}\int e^{-ik[z\cos\theta'+a\sin\theta'\cos(\phi-\phi')]}$$
$$\times(1+C_r(\theta_i,\phi))p_i Y_a\, dS. \tag{9}$$

Since our primary interest is the effect of the attached oscillators (all of which are attached to frames on the shell side wall), we will ignore the integration over the shell end caps and perform the integration in Eq. (9) over the straight section of the shell with length L. Noting that $p_i=p_{io}e^{-ik(z\cos\theta_i+x\sin\theta_i)}$ and $x=a\cos\phi$ on the shell surface ∂S gives

$$p_{as}(r',\theta',\phi')$$
$$=i\omega\rho a\frac{e^{ikr'}}{2\pi r'}p_{io}\int_{-L/2}^{L/2}\int_{-\pi/2}^{\pi/2}[e^{-ikz(\cos\theta'+\cos\theta_i)}$$
$$\times(1+C_r(\theta_i,\phi))$$
$$\times e^{-ika[\sin\theta'\cos(\phi-\phi')+\cos\phi\sin\theta_i]}Y_a(\phi,z)]dz\, d\phi. \tag{10}$$

We are interested in monostatic scattering ($\theta'=\theta_i$) contained in the $\phi'=0$ plane. For this case Eq. (10) becomes

$$p_{as}(r',\theta_i)=i\omega\rho\frac{ae^{ikr'}}{2\pi r'}p_{io}\int_{-\pi/2}^{\pi/2}\int_{-L/2}^{L/2}(1+C_r(\theta_i,\phi))$$
$$\times e^{-2ikz\cos\theta_i}e^{-2ika\sin\theta_i\cos\phi}Y_a(\phi,z)d\phi\, dz. \tag{11}$$

In this same manner the reflected pressure from the ''average'' structure, p_{os}, reduces in the backscattered direction (for high ka) to the familiar geometrical acoustics formula[8] modified by a simple phase shift:

$$p_{os}=C_r(\theta_i)p_{io}\frac{e^{i(kr'-\pi/4-2ka\sin\theta_i)}}{r'}\left(\frac{ka\sin\theta_i}{4\pi}\right)^{1/2}L$$
$$\times j_0(kL\cos\theta_i), \tag{12}$$

where j_0 is the zero-order spherical Bessel function.

Equations (11) and (12) can now be used to predict the monostatic scattering features as a function of angle, θ_i, and frequency, $kc/2\pi$, for various local admittance functions, Y_o+Y_a.

II. SCATTERING PATTERNS

Our interest here is focused on the observed changes in scattering from the shell structure reported in Refs. 1 and 2 with and without the internal oscillators present, and since the oscillators have been attached, to a large degree, randomly, we consider here cases in which Y_a is a pseudorandom function. Later in this section we will obtain a numerical solution to Eq. (11) appropriate to the specifics of the oscillator shell. Before doing this, we believe it is instructive to consider several simple analytic cases. This will demonstrate how the spatial correlations of $Y_a(z)$ might be related to the angular scattering patterns. Specific cases for which Eq. (11) can be solved fairly straightforwardly include the following.

A. Fluctuations along a narrow axial strip

Consider $Y_a(\phi,z)=Y_a(z)$ over a narrow angular strip at ϕ_0 of width $\Delta\phi$. This case would apply had the oscillators been attached only at *one* fixed azimuthal position, rather than all around each frame. Inserting this form of Y_a into Eq. (11) yields

$$p_{as}(r',\theta_i)\simeq i\omega\rho\frac{a\Delta\phi e^{ikr'}(1+C_r(\theta_i,\phi_0))p_{io}}{2\pi r'}$$
$$\times e^{-2ika\sin\theta_i\cos\phi_0}\int_{-L/2}^{L/2}dz\; Y_a(z)e^{-2ikz\cos\theta_i}. \tag{13}$$

Then

$$|p_{as}^2|\simeq\left(\frac{\omega\rho a\Delta\phi(1+C_r(\theta_i,\phi_0))p_{io}}{2\pi r'}\right)^2$$
$$\times\left[\int_{-L/2}^{L/2}dz\; Y_a(z)e^{-2ikz\cos\theta_i}\right]\text{c.c.}, \tag{14}$$

where c.c. refers to the complex conjugate. If $L/L_c\to\infty$ [where L_c is a measure of the correlation length for $Y_a(z)$], the Wiener–Khinchin theorem[9] can be used to express the integral in Eq. (14) as the Fourier transform of the autocorrelation function of Y_a, $f(\tau)$. Then

$$|p_{as}^2|=\left(\frac{\omega\rho a\Delta\phi(1+C_r(\theta_i,\phi_0))p_{io}}{2\pi r'}\right)^2$$
$$\times\langle(Y_a)^2\rangle\int_{-\infty}^{\infty}d\tau\; e^{ik'\tau}f(\tau), \tag{15}$$

where $k'=2k\cos\theta_i$.

Generally speaking, Eq. (15) predicts that the angular spread will be inversely related to the correlation length, L_c. For example, consider the case where $f(\tau)$ is exponential. The integral in Eq. (15) then yields the Lorentzian function

$$\int_{-\infty}^{\infty} d\tau\, e^{ik'\tau} f(\tau) = \int_{-\infty}^{\infty} d\tau\, e^{ik'\tau} e^{-\tau/L_c} = \frac{(1/\pi L_c)^2}{(1/L_c)^2 + (2k\,\cos\,\theta_i)^2}. \tag{16}$$

The half-width of this function, $(\theta_i)_{1/2}$, is $(\theta_i)_{1/2} = \arccos 1/2kL_c$. For example, if L_c is 1 cm (the frame spacing of the shell in Ref. 1), $(\theta_i)_{1/2}$ would be 67° and 14° at the lowest (13-kHz) and highest (50-kHz) frequencies, respectively. This would form a *smooth* ''christmas-tree''-like scattering pattern in frequency angle space of the general shape as that observed in the data of Refs. 1 and 2.

B. Fluctuations over entire cylinder

The result in Sec. II A can be heuristically extended to a case closer to the one of interest namely fluctuations over the entire cylinder surface. Recall that for fluctuations on a narrow strip, Eq. (14) could be expressed in terms of the Fourier transform of the autocorrelation of $Y_a(z)$ if $L/L_c \to \infty$. If we now think of covering the shell with a large number of such strips, we can invoke a form of the Ergodic hypothesis,[10] i.e., that N random realizations over aperture L is equivalent to a random realization over aperture LN. Under the condition that cross correlation between strips can be ignored compared to autocorrelation of each strip, Eq. (15) ought to apply even more stringently, albeit not exactly [due to the $\cos\phi$ term in the exponential and the dependence of C_r on ϕ in Eq. (11)]. Thus for fluctuations over the entire cylinder surface, we can expect to find an angle-frequency pattern related approximately to the Fourier transform of the spatial correlation function of Y_a.

C. Axisymmetric fluctuations

Consider next the case in which Y_a is a function only of z having a spatial correlation length of L_c. In a sense, one could consider this to be the multi-strip case above, but with the opposite condition regarding correlation between strips. In particular, at each z position there is now perfect correlation as one moves around ϕ.

For this case, Eq. (11) can be solved using the stationary phase method[11] to give

$$p_{as}(r',\theta_i) \simeq \frac{i\omega\rho a e^{ikr'}}{2\pi r'} \times \left\{ -ie^{-2ika\,\sin\,\theta_i} \sqrt{\frac{\pi}{2ka\,\sin\,\theta_i}}\,(1 + C_r(\theta_i)) \right\} \times \int_{-L/2}^{L/2} Y_a(z) e^{-2ikz\,\cos\,\theta_i} dz. \tag{17}$$

Comparison with Eq. (13) shows that the scattering pattern will be that for the fluctuating narrow strip, but with additional angle/frequency modulation from the terms in the curly brackets of Eq. (17).

D. Scattering from the oscillator shell

In order to examine scattering patterns for spatial distributions more representative of that found in the oscillator shell of Ref. 1, we shift to numerical integration of Eq. (11). We will consider two cases: one in which the spatial distribution is independent of frequency and one where it is not. The motivation for the latter lies in two facts: (1) our intuition is that the spatial correlation length for the local admittance would be determined by the bending near field[12] and therefore might be directly proportional to the flexural wavelength, which has a $\omega^{-1/2}$ dependence and (2) the oscillators used in Ref. 1 have resonance bandwidths which depend on frequency.

The internal oscillator system used in the shell of Ref. 1 consisted of 880 ''dumbell'' mechanical oscillators, eleven equidistant around each of the 80 frames. Each frame had a random permutation of the 11 different oscillator types, each of the latter having a particular resonance band determined by the dumbell beam thickness (see Fig. 1 in Ref. 1). The band of each of the frame-mounted oscillator resonances is fairly broad with a bandwidth $\Delta\omega/\omega$ varying from 0.34 for the lowest-frequency oscillator to 0.019 for the highest. As pointed out in Ref. 1, these resonance bands (indicated in Fig. 1 by Δf_{exp} for the complete set of 11 oscillators) are associated with the *coupled* frame-oscillator system and are quite different from those of the isolated oscillators. Accordingly, they have been determined as described in Ref. 1 by measurements made with a laser Doppler vibrometer.

At any particular frequency, we assume that Y_a in Eq. (2) is zero except near an oscillator which is resonant at that frequency. In this case, the admittance would look something like that shown in Fig. 3 which applies to the particular case of frame 1.

1. Spatial distribution independent of frequency

We numerically solved Eq. (11) for the case in which the shell admittance is Y_o except at one oscillator position per frame, where it is $Y_o + Y_a$. Specifically, we employed a Gaussian integration routine where we refined the number of points per structural wavelength well beyond the point where the result became independent of resolution size. We took a rectangular patch about each of the oscillators given by 3 cm (three frame spacings) and $2\pi/11$ in the z and ϕ directions, respectively. The position in ϕ was chosen by a random number generator. For any ϕ, this leads to an autocorrelation function which begins to fall linearly with a correlation length $L_c = 1.5$ cm. However, it eventually decays more slowly, even oscillating somewhat, due to the nonzero probability of encountering another identical oscillator down the shell. In addition, a random phase factor (between $-\pi/4$ and $\pi/4$ rad s) was included in each Y_a amplitude, a reflection of the fact that one would not expect the motion of the various dumbell/frame oscillators to be necessarily in phase. The quantity $20\log(p_{as}/p_{io})$ from the Y_a component of Y is shown in Fig. 4. As can be seen, this pattern has some (but

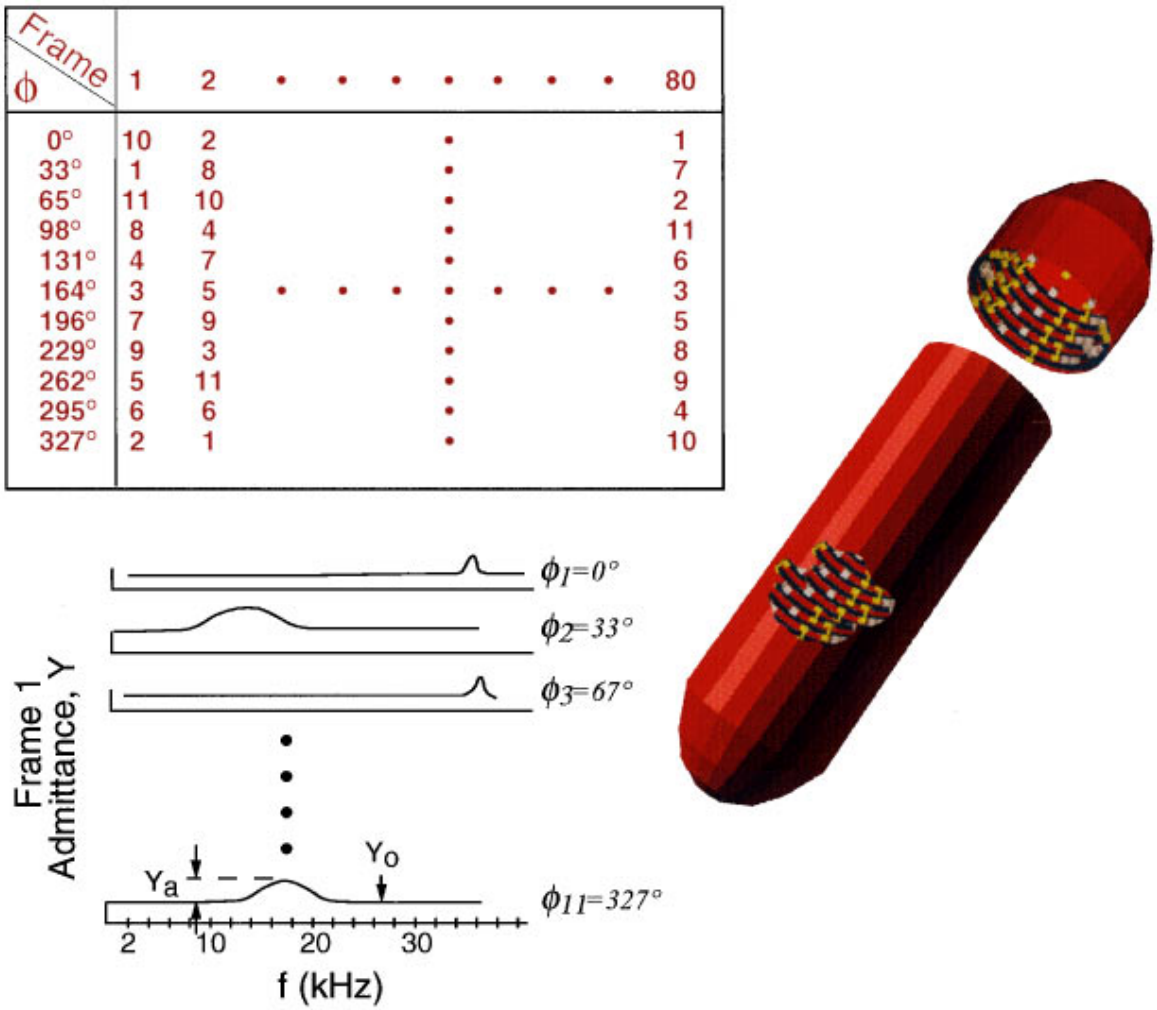


FIG. 3. Partial listing of the positions of the various mechanical oscillators (numbered 1–11) inside the shell of Refs. 1 and 2. The lower left depicts what the admittance, Y, might look like versus frequency for the eleven different angular oscillator positions for the first frame. A rough 3-D drawing of the oscillator shell is shown on the right.

not all) of the essential features we are looking for to explain the ''christmas-tree-like'' scattering patterns seen from the oscillator shell (Fig. 1).

2. Spatial distribution dependent on frequency

As mentioned previously, the oscillators used in Ref. 1 have coupled dumbell/frame resonance frequencies and bandwidths which depend on frequency. This leads to a changing Y_a morphology as the frequency sweeps through the 11 different oscillator bands. To account for this, we used the actual spatial positions of the ''in-band'' oscillators. We assumed a Lorentzian line shape with frequency, f, of the form $(\Delta f_i)^2/[(f-f_i)^2+(\Delta f_i)^2]$ with f_i and Δf_i the resonance frequency and half-width, respectively, of the ith dumbell/frame resonance. At the center of each band, Y_a was assumed to be $4/\rho c$. This value for Y_a was chosen by trial and error in order to produce the best overall visual correspondence between the data and the numerical model. We will discuss the reasonableness of this value below. Again, we included a random phase factor (now between -2.2 and 2.2 rads) in each Y_a amplitude. However, at each oscillator location centered at (z_i,ϕ_i), we implemented a patch shape function consistent with the structural near field[12] of the form $A(z_i,\phi_i)\cos^2[\kappa(z_i-z)]\cos^2[b\kappa a(\phi_i-\phi)]$, where A is unity inside (and zero outside) the patch of sides π/κ and $\pi/b\kappa$. The quantity $20\log[(p_{as}+p_{os})/p_{io}]$ is shown in Fig. 5 for the case where κ is one-fifth the flexural wave number and $b=1.6$. [Here p_{os} was computed from Eq. (12).] For direct comparison to Fig. 1 we display this for θ_i between 0° and 360°. As can be seen, there exists a fair amount of correspondence between the displays in these figures. We point out that the broad vertical lines around 60° and 120° in the experimental data (Fig. 1) are due to scattering from the conical end cap designs used in the actual shell, a structural detail not included in our model. Regarding our use of nonequal z and ϕ patch side dimensions, we point out that only small details were changed using different values for b including $b=1$.

In Fig. 6 we show the normalized squared magnitude $|p_{os}+p_{as}|^2$ averaged over the band 22–25 kHz. For comparison, we also show the measured result provided by the authors of Ref. 1 averaged over the same band. As can be seen, there is reasonable agreement regarding both the relative levels of p_{os} and p_{as} as well as the general shapes of

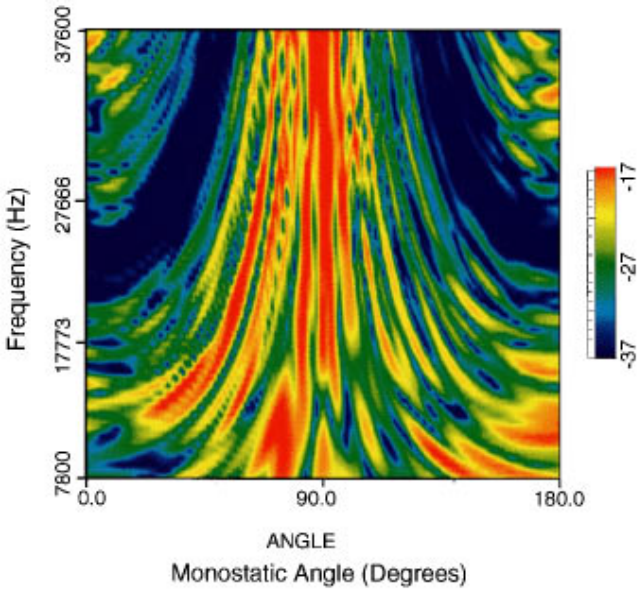


FIG. 4. Monostatic scattering versus frequency and angle for the Y_a component of Y obtained from numerical integration of Eq. (11). Y_a is assumed to be constant except for a random phase factor at each oscillator site over a patch 3 cm along the z axis and $2\pi/11$ degrees azimuthally. Here the color scale gives the magnitude in decibels of the backscattered pressure divided by p_{io}.

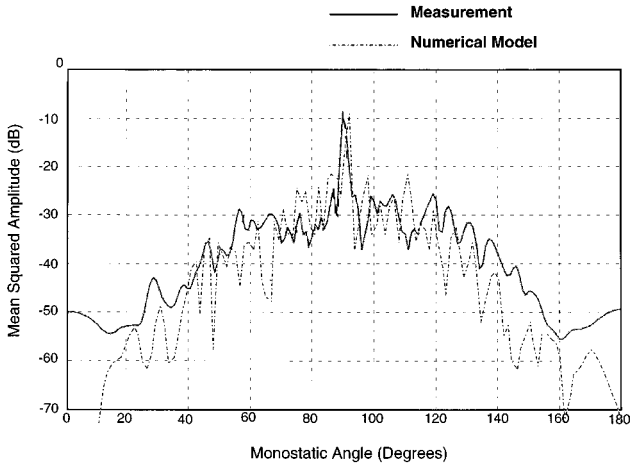


FIG. 6. Monostatic scattering versus angle averaged over the band 22–25 kHz. Dashed: $10\log\langle|p_{os}+p_{as}/p_{io}|2\rangle$; solid: measured results provided by the authors of Ref. 1 for the oscillator shell.

their respective angular patterns, especially for off-beam aspects.

We now discuss the reasonableness of our choice of $4/\rho c$ for Y_a. First, we point out that the major effect of changing the magnitude of Y_a is to alter the relative magni-

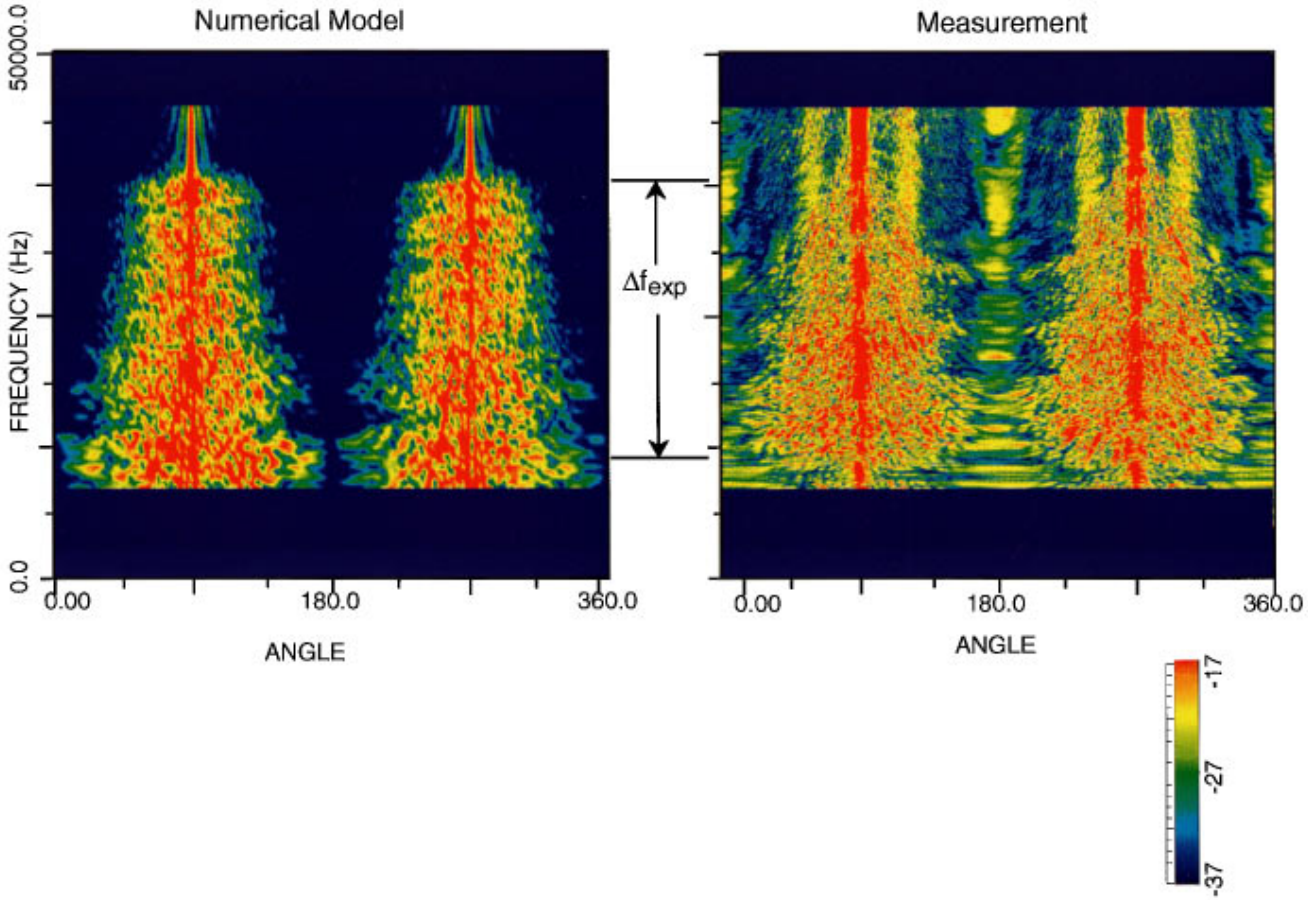


FIG. 5. Monostatic scattering versus frequency and angle for $Y=Y_o+Y_a$ obtained from numerical integration of Eq. (11) (left). In arriving at the spatial distribution for Y_a, the actual positions of the ''in-band'' oscillators of Refs. 1 and 2 were used for each of the eleven frequency bands. Here the color scale gives the magnitude in decibels of the quantity $20\log[(p_{as}+p_{os})/p_{io}]$. Also shown (right) is the experimental monostatic scattering versus the frequency for the oscillator shell.

tude of the narrow scattering peak associated with the uniform admittance (Y_o) and that of the broader, diffuse scattering (see Fig. 6). The uniform shell admittance, Y_o, and the reflection coefficient, C_r, were calculated using the development of Burroughs,[13] with slight modification. In particular, while the Epstein-Kennard[14] shell equations were still used to model the cylindrical section, the effect of the internal ribs was included through averaging their spatially periodic impedances over the surface area of the shell. The displacements and scattered pressure due to an incident plane wave were then calculated for each axial wave number [$k_z = k\cos(\theta_i)$] and 32 circumferential harmonics in a Fourier series, and Y_o and C_r were obtained by dividing the surface velocity and pressure, respectively, by the local value of the incident plane wave. Over the frequency band of interest, the magnitude of Y_o, although fluctuating, ranges from about $3/\rho c$ at the low frequencies to about $1/\rho c$ at the high end. Thus the value of $4/\rho c$ used for Y_a implies that, at the center of each oscillator band, the acoustically induced velocity levels (due to the resonance of the oscillator) would range from a factor of about 1.3–4 times larger than the average shell velocity at the lowest and highest frequencies, respectively. These do not appear to be unreasonable levels.

III. DISCUSSION

The preceeding demonstrates that the surprising oscillator shell scattering data reported by Bucaro *et al.*[1] and Photiadis *et al.*[2] is in fact consistent with scattering from a cylindrical shell whose local admittance has been altered by the spatially pseudorandom placement of internal oscillators. If this is indeed the case, this scattering mechanism would be seen to dominate the off-beam monostatic scattering in this shell over the frequency bands for which the oscillators are resonant. Further, this would imply that the off-beam scattering caused by elastic shell-wave propagation is relatively unimportant compared to the simple local effect treated here. This may suggest that in the ka region of interest here the body of work on scattering from idealized shells with no internal structure—particularly that associated with propagating elastic shell waves—might not be relevant to the complex submarinelike structures which motivated many of those studies.

That the off-beam scattering from submerged shells with complex internal structure might be simply related to pseudorandom spatial admittance fluctuations and their spatial correlations has other interesting implications. Specifically, the spatial correlations can be thought of as spatial apertures having simply predicted angular scattering patterns. For example, long apertures along z (perhaps internal structure attached to many frames) would lead to less off-beam scattering than short apertures (perhaps attachment to only several adjacent frames). On the other hand, azimuthal apertures would tend to control the fine-structure of these off-beam frequency-angle patterns. In particular, short azimuthal apertures (attachment at a single point of a frame) tend to produce little such fine structure whereas large azimuthal apertures (attachment all around a frame) result in significant fine-structure modulation of the off-beam scattering patterns [refer to Eq. (17)]. These, of course, are simple rules which would have to, in general, be applied to systems with not only many, but also overlapping, such apertures. Nonetheless, they would provide principles by which to organize one's thinking about such effects.

IV. CONCLUDING REMARKS

We have shown that considering the spatial admittance fluctuations presented by the presence of internal oscillators leads to a scattering pattern in frequency-angle space which matches the characteristics reported in the experimental measurements of Ref. 1. Put simply, we have interpreted the ''christmas-tree-like'' scattering pattern in terms of a simple aperture function where the aperture is directly related to the spatial autocorrelation of the fluctuating admittance.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research Code 334. We thank Leslie Chaplin for her creative work in generating the figures for this paper and Dr. Earl Williams for his illuminating discussions regarding the structural near field.

[1] J. A. Bucaro, D. M. Photiadis, and B. H. Houston, ''Acoustic scattering from a submerged shell with many internal oscillators,'' Proceedings of the ASME, Noise Control and Acoustics Division, NCA **22**, 87–92 (1996).
[2] D. M. Photiadis, J. A. Bucaro, and B. H. Houston, ''The effect of internal oscillators on the acoustic response of a submerged shell,'' J. Acoust. Soc. Am. **101**, 895–899 (1997).
[3] M. L. Rumerman, ''Contribution of membrane wave reradiation to scattering from finite cylindrical steel shells in water,'' J. Acoust. Soc. Am. **93**, 55–65 (1993).
[4] B. H. Houston, J. A. Bucaro, and D. M. Photiadis, ''Broadband acoustic scattering from a ribbed shell,'' J. Acoust. Soc. Am. **98**, 2851–2853 (1995).
[5] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton U. P., Princeton, NJ, 1986), p. 321.
[6] See, for example, E. Merzbacher, *Quantum Mechanic* (Wiley, New York, 1961), pp. 226–228.
[7] P. M. Morse and K. U. Ingard, Ref. 5, pp. 259–263.
[8] M. J. Skolnik, *Introduction to Radar Systems* (McGraw-Hill, New York, 1962), p. 43.
[9] R. C. Houts, *Signal Analysis in Linear Systems* (Holt, Rinehart and Winston, Philadelphia, 1991), p. 66.
[10] Y. W. Lee, *Statistical Theory of Communication* (Wiley, New York, 1960), p. 208.
[11] See, for example, E. Yamashita, *Analysis Methods for Electromagnetic Wave Problems* (Artech, Boston, 1990), pp. 308–310.
[12] See, for example, L. Cremer and M. Heckl, *Structure-Borne Sound* (Springer-Verlag, New York, 1985), p. 289.
[13] C. B. Burroughs, ''Acoustic radiation from fluid-loaded infinite circular cylinders with doubly periodic ring support,'' J. Acoust. Soc. Am. **75**, 715–722 (1984).
[14] E. H. Kennard, ''The new approach to shell theory: Circular cylinders,'' J. Appl. Mech. **20**, 33–40 (1983).

Input mobilities and power flows for edge-excited, semi-infinite plates

Christiaan Kauffmann
Delft University of Technology, Faculty of Technical Mathematics and Informatics, P.O. Box 5031, 2600 GA Delft, The Netherlands

(Received 8 April 1991; accepted for publication 22 December 1997)

The basis of this paper is a paper by E. Eichler [J. Acoust. Soc. Am. **36**, 344–348 (1964)], which deals with the response of a semi-infinite plate, driven by a uniform, time-harmonic load applied at a finite part of its free edge. Eichler used Fourier transforms in the edge direction and obtained closed form integral expressions for the plate response. The present paper provides full details of the analysis (some of them not previously published) and gives corrections to some of Eichler's results. New, closed form integral expressions for the power flows are derived. Numerical results for $0<ka<10$ are presented for the case of uniform excitation over an excitation area of length $2a$ (k is the structural wave number). A simple physical classification facilitates the interpretation of the results: (i) $ka \ll 1$: point excitation; (ii) $ka \to \infty$: line excitation (beam mode); and (iii) $ka = \mathcal{O}(1)$: modulation due to interference effects at intermediate frequencies. Additionally, the case of N point loads (forces, tangential moments, and normal moments) with nonuniform spacing and arbitrary amplitude and phase distributions is discussed. Numerical results for the power flow include the coupling term between two point loads at separation s and are presented for $0<ks<10$. Depending on the type of excitation interference shows up as strong, intermediate or weak modulation. [S0001-4966(98)00204-5]

PACS numbers: 43.40.Dx [PJR]

INTRODUCTION

A great variety of engineering structures are built up by elements like beams, plates, and shells, connected to each other by structural joints. Typically, these connections are located at or close to the edges of the structural elements. Therefore, the study of edge-excited plates is of importance for the quantification of vibration transmission through connected structures. In particular, the type of excitation that is dealt with in this paper, i.e., forces and moments, is representative for those situations where there is a large impedance mismatch between the adjacent structural elements.

In Ref. 1, the harmonic response of an edge-excited, semi-infinite plate was analyzed using Fourier transforms in the direction parallel to the edge. This approach leads to explicit expressions for the plate response, be it in the form of wave number integrals. The numerical evaluation of the integrals for arbitrary points of observation should be performed in the complex plane. The larger part of the paper presents numerical results for the input power generated by various kinds of excitation (forces, tangential moments, and normal moments) having uniform and nonuniform spatial distributions.

Section I extends Eichler's work by deriving new, closed form integral expressions for the input power delivered by arbitrary force and moment distributions along the edge. Meanwhile, analytical formulas for the poles of the Fourier transforms are given (not included in Ref. 1).

Section II deals with excitation by a uniformly distributed load, i.e., the plate is driven by a constant force and a constant moment per unit length applied at a finite part of the edge of length $2a$. Numerical results for the various power flows are presented for $0<ka<10$, where k is the flexural wave number. A physical interpretation is sought that can improve our qualitative understanding of this problem. First, for $ka \to \infty$ the problem is essentially one dimensional and the plate response is expected to be similar to that of a semi-infinite beam to corresponding loads. Second, for $ka \ll 1$ the loads are point loads and the response is compared to the well known results for infinite plates. Third, for $ka = \mathcal{O}(1)$ a transition regime between these limit cases is expected. Note that Ref. 1 only gives results for the response at the center of the excitation area ($y=0$) for $0<ka<1$.

Dealing with the same excitation Sec. III gives the plate response at $y=0$, the center of the excitation area. In particular, numerical results for the point mobilities ($ka=0$) are given and are compared to similar results for infinite plates. Meanwhile, Eichler's results on the point mobilities are corrected and some remarkable new properties are discussed.

Section IV treats excitation by N point loads of various kinds (forces, tangential moments, and normal moments) with nonuniform spacing and arbitrary distributions of amplitude and phase. The power input is given in terms of the mutual interaction term between two point loads at distance s. Numerical results are presented for $0<ks<10$ and include the coupling term as well as the total power injected by two point loads for various values of the phase difference.

I. GOVERNING EQUATIONS AND ANALYSIS

A. Governing equations

Consider a semi-infinite, thin elastic plate of constant thickness h, made of linear, isotropic elastic material, E is Young's modulus, ν is Poisson's ratio, and m'' is density of mass per unit area. It is assumed that the plate's flexural

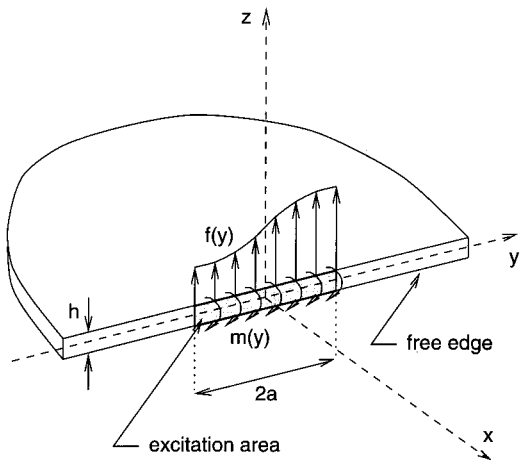


FIG. 1. Semi-infinite plate configuration.

vibrations are described by classical plate theory. The time-harmonic behavior [the complex time factor $\exp(i\omega t)$ is omitted for brevity] is governed by the equation of motion for the normal displacement $w(x,y)$ of the plate:

$$D\Delta\Delta w - m''\omega^2 w(x,y) = 0, \tag{1}$$

where $D = Eh^3/12(1-\nu^2)$ is the bending stiffness. The plate has a free edge at $x=0$ except for the area $|y|<a$ where the plate is excited by an arbitrary load; see Fig. 1. The transverse force and the tangential moments at the edge are related to the plate displacement by the boundary conditions

$$-D\frac{\partial}{\partial x}\left[\frac{\partial^2 w}{\partial x^2}(0,y)+(2-\nu)\frac{\partial^2 w}{\partial y^2}(0,y)\right] = \begin{cases} f(y), & |y|<a, \\ 0, & |y|>a, \end{cases} \tag{2}$$

$$-D\left[\frac{\partial^2 w}{\partial x^2}(0,y)+\nu\frac{\partial^2 w}{\partial y^2}(0,y)\right] = \begin{cases} m(y), & |y|<a, \\ 0, & |y|>a, \end{cases} \tag{3}$$

where $f(y)$ is the force per unit length and $m(y)$ is the prescribed tangential moment per unit length.

B. Solution by Fourier transforms

Following Ref. 1, Fourier transform pairs in the edge direction are introduced by

$$w(x,y) = \int_{-\infty}^{\infty} \tilde{w}(x,q)\exp(-iqy)dq, \tag{4}$$

$$\tilde{w}(x,q) = \frac{1}{2\pi}\int_{-\infty}^{\infty} w(x,y)\exp(iqy)dy. \tag{5}$$

Using that $w(x,y)$ as well as its first, second, and third order partial derivatives with respect to y vanish at infinity, the Fourier transform of Eq. (1) reads

$$\left[\frac{d^4}{dx^4}+2(-iq)^2\frac{d^2}{dx^2}+(-iq)^4\right]\tilde{w}(x,q)-k^4\tilde{w}(x,q)=0, \tag{6}$$

where k denotes the flexural wave number defined by $k^4 = m''\omega^2/D$. Looking for the free modes of flexural vibration $\exp(\lambda x)$, the characteristic equation leads to

$$\lambda_1 = (q^2+k^2)^{1/2} > 0 \quad \text{for all } q, \tag{7}$$

$$\lambda_2 = -\lambda_1, \tag{8}$$

$$\lambda_3 = (q^2-k^2)^{1/2} = \begin{cases} |q^2-k^2|^{1/2}, & |q|>k, \\ i|k^2-q^2|^{1/2}, & -k<q<k, \end{cases} \tag{9}$$

$$\lambda_4 = -\lambda_3. \tag{10}$$

Taking into account the requirements for real physical modes, i.e.,

(a) $\mathrm{Im}[\lambda]=0 \Rightarrow \mathrm{Re}[\lambda]>0$, which excludes the nonpropagating, exponentially increasing field components at infinity ($x\to-\infty$), and
(b) $\mathrm{Re}[\lambda]=0 \Rightarrow \mathrm{Im}[\lambda]>0$, which excludes modes propagating from infinity towards the edge $x=0$,

it is concluded that λ_2 and λ_4 are not admissible. Consequently,

$$\tilde{w}(x,q) = A(q)\exp(\lambda_1 x) + B(q)\exp(\lambda_3 x), \tag{11}$$

where $A(q)$ and $B(q)$ are constants, but still functions of q. Transforming the boundary conditions (2) and (3) leads to

$$-D\left[\frac{d^3\tilde{w}}{dx^3}(0,q)+(2-\nu)(-iq)^2\frac{d\tilde{w}}{dx}(0,q)\right] = \tilde{f}(q), \tag{12}$$

$$-D\left[\frac{d^2\tilde{w}}{dx^2}(0,q)+\nu(-iq)^2\tilde{w}(0,q)\right] = \tilde{m}(q), \tag{13}$$

where $\tilde{f}(q)$ and $\tilde{m}(q)$ are the Fourier transforms of $f(y)$ and $m(y)$ given by

$$\tilde{f}(q) = \frac{1}{2\pi}\int_{-a}^{a} f(y)\exp(iqy)dy, \tag{14}$$

$$\tilde{m}(q) = \frac{1}{2\pi}\int_{-a}^{a} m(y)\exp(iqy)dy. \tag{15}$$

Inserting expression (11) into (12) and (13), and solving for $A(q)$ and $B(q)$ yields

$$\begin{pmatrix} A(q) \\ B(q) \end{pmatrix} = \frac{1}{\det(p)}\begin{pmatrix} M_-(p) & -\mu_-(p)M_+(p) \\ M_+(p) & -\mu_+(p)M_-(p) \end{pmatrix}\cdot\begin{pmatrix} \dfrac{\tilde{f}(q)}{Dk^3} \\ \dfrac{\tilde{m}(q)}{Dk^2} \end{pmatrix}, \tag{16}$$

where $q=kp$ and p is a nondimensional wave number. $\mu_\pm(p)$, $M_\pm(p)$, and $\det(p)$ are defined by

$$\mu_\pm(p) = (p^2\pm 1)^{1/2}, \quad M_\pm(p) = 1\pm(1-\nu)p^2, \tag{17}$$

$$\det(p) = \mu_-(p)[M_+(p)]^2 - \mu_+(p)[M_-(p)]^2. \tag{18}$$

Finally, the integral representation for the plate displacement is

$$w(x,y)=\int_{-\infty}^{\infty}\exp(-ikpy)\{A(q)\exp[kx\mu_+(p)]+B(q)\exp[kx\mu_-(p)]\}k\ dp, \tag{19}$$

where $A(q)$ and $B(q)$ are given by Eq. (16) and the values of the square root functions $\mu_\pm(p)$ for real values of p corresponds to those of λ_1 and λ_3 as given in Eqs. (7) and (9). Other quantities of interest such as the normal and tangential rotation at the edge can be found directly from Eq. (19):

$$-\frac{\partial w}{\partial x}(0,y)=\int_{-\infty}^{\infty}\exp(-ikpy)[\mu_+(p)A(q)+\mu_-(p)B(q)]k^2\ dp, \tag{20}$$

$$\frac{\partial w}{\partial y}(0,y)=\int_{-\infty}^{\infty}\exp(-ikpy)[A(q)+B(q)]ik^2p\ dp. \tag{21}$$

C. Excitation by a distributed normal moment

A special kind of excitation is a normal moment, which, apart from transverse forces and tangential moments corresponds to a distinct type of mechanical drive. A normal point moment can be constructed mathematically by two point forces of equal strength and opposite phases. These two forces produce a normal point moment if $s\to 0$, while $Fs=T$ is kept constant, where s is the distance between the forces and F their amplitude. The construction of a normal point moment that excites the plate at $y=0$ starts by taking

$$f(y)=-F\delta(y)+F\delta(y-s),\quad m(y)=0,$$

which gives the Fourier transforms

$$\tilde{f}(q)=\frac{F}{2\pi}[\exp(iqs)-1],\quad \tilde{m}(q)=0.$$

Because $\tilde{m}(q)=0$ relation (16) simplifies to

$$A(q)=\frac{M_-(p)}{\det(p)}\frac{\tilde{f}(q)}{Dk^3},\quad B(q)=\frac{M_+(p)}{\det(p)}\frac{\tilde{f}(q)}{Dk^3}.$$

Inserting these expressions in Eq. (19) gives the plate response. By changing the order of the integration and the limit process and noting that

$$\lim_{s\to 0}\tilde{f}(q)=\frac{T}{2\pi}iq,\quad \text{while}\quad Fs=T=\text{constant},$$

the plate response to a normal point moment at $y=0$ takes the form

$$w(x,y)=\int_{-\infty}^{\infty}\exp(-ikpy)\{M_-(p)\exp[kx\mu_+(p)]+M_+(p)\exp[kx\mu_-(p)]\}\frac{T/2\pi}{Dk}\frac{ip\,dp}{\det(p)}. \tag{22}$$

Using the principle of superposition, the response to a continuously distributed normal moment is obtained straightforwardly: take $t(\eta)d\eta$ instead of T and replace y by $y-\eta$ (within the argument of the exponential function) and integrate the resulting expression over the area of excitation, i.e., $-a<\eta<a$. Changing the order of the integrations gives the plate displacement due to a distributed normal moment of strength $t(y)$ per unit length:

$$w(x,y)=\int_{-\infty}^{\infty}\exp(-ikpy)\{M_-(p)\exp[kx\mu_+(p)]+M_+(p)\exp[kx\mu_-(p)]\}\frac{\tilde{t}(q)}{Dk}\frac{ip\ dp}{\det(p)}, \tag{23}$$

where $\tilde{t}(q)$ is the Fourier transform of $t(y)$.

D. Input power flows

By definition an expression for the power flow is:

$$\Pi_{\text{in}}=\frac{1}{2}\,\text{Re}\left[\int_{-a}^{a}f(y)^*[i\omega w(0,y)]dy\right]+\frac{1}{2}\,\text{Re}\left[\int_{-a}^{a}m(y)^*\left[i\omega.-\frac{\partial w}{\partial x}(0,y)\right]dy\right]. \tag{24}$$

Inserting for $w(0,y)$ and $-(\partial/\partial x)w(0,y)$ from Eqs. (19) and (20) and changing the order of the integrations yields:

$$\Pi_{\text{in}}=\frac{1}{2}\,\text{Re}\left[2\pi\int_{-\infty}^{\infty}\tilde{f}(q)^*\{A(q)+B(q)\}i\omega k\ dp\right]+\frac{1}{2}\,\text{Re}\left[-2\pi\int_{-\infty}^{\infty}\tilde{m}(q)^*\{\mu_+(p)A(q)+\mu_-(p)B(q)\}i\omega k^2\ dp\right].$$

Finally, by substituting for $A(q)$ and $B(q)$ the expressions (16), the input power splits up into four additive contributions:

$$\Pi_{\text{in}}=\text{Re}\left[2\pi\int_{-\infty}^{\infty}\frac{\tilde{f}(q)^*\tilde{f}(q)}{Dk^3}\frac{i\omega k\ dp}{\det(p)}\right]+\text{Re}\left[-\pi\int_{-\infty}^{\infty}\frac{\tilde{f}(q)^*\tilde{m}(q)}{Dk^2}\{\mu_-(p)M_+(p)+\mu_+(p)M_-(p)\}\frac{i\omega k\ dp}{\det(p)}\right]$$
$$+\text{Re}\left[-\pi\int_{-\infty}^{\infty}\frac{\tilde{m}(q)^*\tilde{f}(q)}{Dk^3}\{\mu_+(p)M_-(p)+\mu_-(p)M_+(p)\}\frac{i\omega k^2\ dp}{\det(p)}\right]$$
$$+\text{Re}\left[2\pi\int_{-\infty}^{\infty}\frac{\tilde{m}(q)^*\tilde{m}(q)}{Dk^2}\mu_+(p)\mu_-(p)\frac{i\omega k^2\ dp}{\det(p)}\right]. \tag{25}$$

Note that $M_-(p)+M_+(p)=2$ is used. It is clear that the coupling terms are equal if $\tilde{f}(q)^*\tilde{m}(q)\equiv\tilde{m}(q)^*\tilde{f}(q)$. This condition is fulfilled if $f(y)$ and $m(y)$ have the same spatial distribution (both for amplitude and phase), i.e., $f(y)/f_0=m(y)/m_0$. As a special case, excitation by a uniform distribution is considered in Sec. II. The four terms are explicitly distinguished by writing

$$\Pi_{\text{in}}=\sum_{i,j=1}^{2}\Pi_{i,j},$$

where $\Pi_{i,j}$ is the power supplied to the plate by load i due to the response induced by load j. Here, the index 1 refers to force excitation and 2 refers to moment excitation.

The input power for the distributed normal moment is

$$\Pi_{\text{in}}=\frac{1}{2}\,\text{Re}\left[\int_{-a}^{a}t(y)^*\left[i\omega\,\frac{\partial w}{\partial y}\,(0,y)\right]dy\right]. \tag{26}$$

Inserting the integral representation for $(\partial/\partial y)w(0,y)$ from Eq. (21) and changing the order of the integrations yields

$$\Pi_{\text{in}}=\text{Re}\left[2\pi\int_{-\infty}^{\infty}\tilde{t}(q)^*\tilde{t}(q)\,\frac{\omega}{D}\,\frac{ip^2\,dp}{\det(p)}\right]. \tag{27}$$

For notational reasons only, expression (27) will be referred to as $\Pi_{3,3}$.

E. Location of poles

The problem that remains is the evaluation of the integral representations, i.e., the inverse Fourier transforms, for w and the other quantities of interest related to it. All integrands in Eqs. (19)–(27) have a common denominator, $\det(p)$. The poles of the integrands satisfy:

$$\det(p)=0. \tag{28}$$

This equation may be rationalized upon multiplying by

$$\text{adjdet}(p)=\mu_-(p)[M_+(p)]^2+\mu_+(p)[M_-(p)]^2. \tag{29}$$

As a result, a quadratic equation in p^4 remains with the solutions

$$p^4=\tau_+,\ \text{ or }\ p^4=\tau_-, \tag{30}$$

where $\tau_\pm$ are given by

$$\tau_\pm=\frac{1-3\nu\pm2\sqrt{\nu^2+(1-\nu)^2}}{(1-\nu)^2(3+\nu)}. \tag{31}$$

It is easy to show that $1<\tau_+<\frac{4}{7}(2\sqrt{2}-1)$ and $-\frac{4}{7}(1+2\sqrt{2})<\tau_-<-\frac{1}{3}$ for $0<\nu<\frac{1}{2}$. Consequently, the roots of $\det(p)\cdot\text{adjdet}(p)=0$ are

$$p_1=\tau_+^{1/4},\ p_2=i\tau_+^{1/4},\ p_3=-p_1,\ p_4=-p_2, \tag{32}$$

$$p_5=|\tau_-|^{1/4}e^{i\pi/4},\ p_6=|\tau_-|^{1/4}e^{3i\pi/4},\quad p_7=-p_5,\ p_8=-p_6. \tag{33}$$

Tedious, but straightforward analysis reveals that these roots split up into two sets according to:

$$\det(p_{1,2,3,4})=0\ \text{ and }\ \text{adjdet}(p_{5,6,7,8})=0. \tag{34}$$

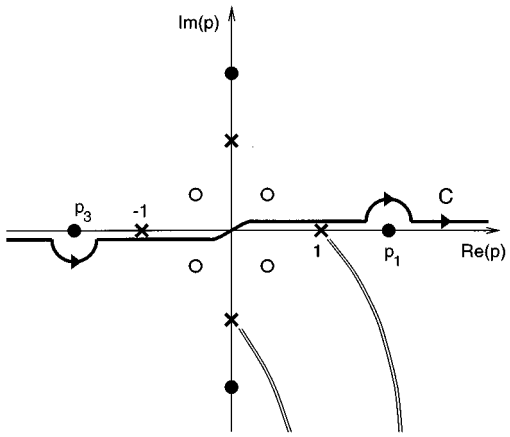


FIG. 2. Contour of integration C for evaluation of inverse Fourier transforms in the complex p plane. × branch points of $\mu_\pm(p)$; ● poles of $1/\det(p)$; ○ poles of $1/\text{adjdet}(p)$; = branch cuts of $\mu_\pm(p)$.

F. Contour of integration

The presence of the real-valued, simple poles p_1 and p_3 on the contour of integration in Eqs. (19)–(23), (25), and (27) is accounted for by taking the Cauchy principal value for the integral and adding a residue contribution times $\pm\frac{1}{2}$ associated with a semi-circular path (of vanishing radius) around each pole. The residue carries a positive algebraic sign if the integration along the semi-circle runs in anticlockwise direction. Otherwise, the sign is negative.

Consider the inverse Fourier transform for $w(x,y)$ in Eq. (19). The contour of integration along the real p axis is closed by a semi-circle at infinity in the lower half-plane if $y>0$ and in the upper half-plane if $y<0$. The integral along this semi-circle vanishes if its radius approaches infinity. The presence of the branch cuts for $\mu_\pm(p)$ is accounted for by integrals along the branch cuts; since branch cuts may not be crossed. Application of the residue theorem at poles $p_j=p_j'+ip_j''$ yields contributions of the form $C(p_j)\times\exp(-ikp_jy)=C(p_j)\exp(-ikp_j'y+kp_j''y)$. Radiation and boundedness conditions require this wave to decay exponentially for $y\to\pm\infty$ or to represent an outgoing wave. The boundedness condition is met by $p_j''<0$ if $y>0$ and $p_j''>0$ if $y<0$. For real-valued poles ($p_j''=0$) the wave propagates unattenuated to $y\to+\infty$ if $p_j'>0$ and to $y\to-\infty$ if $p_j'<0$. The latter condition is fulfilled if the semi-circle around p_1 is extended into the upper half plane and that around p_3 into the lower half plane. This specifies completely the contour of integration and the sign of the residues; see Fig. 2. The same contour applies to $y>0$ and $y<0$ and can be used in Eqs. (19)–(23). However, Eqs. (25) and (27) will be calculated directly and not using the residue theorem, because the integral along a large semi-circle does not vanish if its radius goes to infinity. Nevertheless, the contour along the real p axis in Eq. (25) and (27) is the same as in (19), since this contour is appropriate for any value of y, while Eq. (25) and (27) are derived from (19) to (23) by integration with respect to y from $-\infty$ to $+\infty$. Similarly, for $y=0$ Eqs. (19)–(23) are evaluated directly. Eichler[1] did so to calculate the input mobilities.

II. EXCITATION BY A UNIFORMLY DISTRIBUTED LOAD

A. Expressions for power flows

We consider a simple example. Let the load be uniformly distributed over the excitation area. The force and tangential moment per unit length are

$$f(y)=f_0=\frac{F}{2a},\quad m(y)=m_0=\frac{M}{2a}, \tag{35}$$

where F and M are total force and moment, while f_0 and m_0 are force and moment per unit length. The Fourier transforms of $f(y)$ and $m(y)$ follow from Eqs. (14) and (15)

$$\tilde{f}(q)=\frac{f_0}{2\pi}2a\,\frac{\sin(kap)}{kap},\quad \tilde{m}(q)=\frac{m_0}{2\pi}2a\,\frac{\sin(kap)}{kap}. \tag{36}$$

Inserting these expressions into Eq. (25) yields

$$\Pi'_{1,1}=\frac{f_0^2\omega}{Dk^3}\,ka\left\{\frac{2}{\pi}\int_0^1\left(\frac{\sin(kap)}{kap}\right)^2\frac{-i\mu_-(p)[M_+(p)]^2}{\triangle(p)}\,dp+2\left(\frac{\sin(kap_1)}{kap_1}\right)^2\left[\frac{d}{dp}\det(p_1)\right]^{-1}\right\}, \tag{37}$$

$$\Pi'_{1,2}=-\frac{f_0m_0\omega}{Dk^2}\,ka\left\{\frac{2}{\pi}\int_0^1\left(\frac{\sin(kap)}{kap}\right)^2\times\frac{-i\mu_-(p)\mu_+(p)M_+(p)M_-(p)}{\triangle(p)}\,dp+\left(\frac{\sin(kap_1)}{kap_1}\right)^2[\mu_-(p_1)M_+(p_1)+\mu_+(p_1)M_-(p_1)]\left[\frac{d}{dp}\det(p_1)\right]^{-1}\right\}, \tag{38}$$

$$\Pi'_{2,1}=\Pi'_{1,2}, \tag{39}$$

$$\Pi'_{2,2}=\frac{m_0^2\omega}{Dk}\,ka\left\{\frac{2}{\pi}\int_0^1\left(\frac{\sin(kap)}{kap}\right)^2\times\frac{-i\mu_-(p)[\mu_+(p)]^2[M_-(p)]^2}{\triangle(p)}\,dp+2\left(\frac{\sin(kap_1)}{kap_1}\right)^2\mu_-(p_1)\mu_+(p_1)\left[\frac{d}{dp}\det(p_1)\right]^{-1}\right\}, \tag{40}$$

where

$$\triangle(p)=-\det(p)\cdot\mathrm{adjdet}(p)=2[-(3+\nu)(1-\nu)^3p^8+2(1-3\nu)(1-\nu)p^4+1] \tag{41}$$

and

$$\frac{d}{dp}\det(p_1)=p_1\left[\frac{[M_+(p_1)]^2}{\mu_-(p_1)}-\frac{[M_-(p_1)]^2}{\mu_+(p_1)}\right]+[\mu_-(p_1)M_+(p_1)+\mu_+(p_1)M_-(p_1)]\times(1-\nu)4p_1. \tag{42}$$

In Eqs. (37)–(40) primes indicate that the input power is taken per unit length in edge direction, i.e., $\Pi'_{i,j}=\Pi_{i,j}/2a$. The residue contributions are associated with the semicircular paths around the poles p_1 and p_3. The principal value integrals from 1 up to ∞ do not contribute to the power flow, since they are purely imaginary. For the same reason only the real parts of the integrals over $0<p<1$ are retained. The minus sign in the coupling terms (38) and (39) indicates the absorption of power by the force and moment, respectively. If one of these loads was defined as a positive quantity in opposite direction these coupling terms would be positive and contribute to production of power. This might appear as an artificial effect, but it is noted that within this model it is assumed that the excitation is not affected by the plate response.

For a uniformly distributed normal moment, i.e.,

$$t(y)=t_0=\frac{T}{2a}, \tag{43}$$

where T is the total normal moment and t_0 is the normal moment per unit length along the edge, the Fourier transform is

$$\tilde{t}(q)=\frac{t_0}{2\pi}2a\,\frac{\sin(kap)}{kap}. \tag{44}$$

Inserting $\tilde{t}(q)$ into expression (27) yields the power input

$$\Pi'_{3,3}=\frac{t_0^2\omega}{Dk}\,ka\left\{\frac{2}{\pi}\int_0^1\left(\frac{\sin(kap)}{ka}\right)^2\frac{-i\mu_-(p)[M_+(p)]^2}{\triangle(p)}\,dp+2\left(\frac{\sin(kap_1)}{ka}\right)^2\left[\frac{d}{dp}\det(p_1)\right]^{-1}\right\}, \tag{45}$$

where $\Pi'_{3,3}=\Pi_{3,3}/2a$ is the power input per unit length.

B. Numerical results and discussion

All results are obtained for $\nu=0.3$. Hence, $p_1=1.000\,950\,194\,184\,853$. The values of the various power flows have been evaluated using expressions (37), (38), (40), and (45) for $0<ka<10$ and are presented in Fig. 3 after normalization by $\Psi'_{i,j}$, which is the input power (per unit width) for a semi-infinite beam excited by a force and a moment at the free edge. These quantities are related to the real part of the corresponding admittances and can be found easily[2]:

$$\Psi'_{1,1}=\frac{1}{2}\frac{f_0^2\omega}{Dk^3},\quad \Psi'_{1,2}=-\frac{1}{2}\frac{f_0m_0\omega}{Dk^2},$$

$$\Psi'_{2,1}=\Psi'_{1,2},\quad \Psi'_{2,2}=\frac{1}{2}\frac{m_0^2\omega}{Dk}.$$

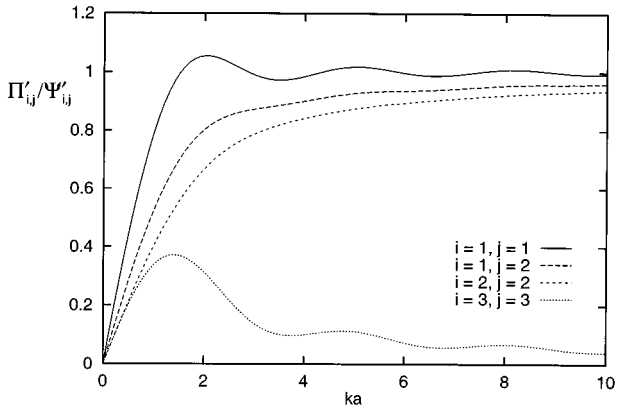


FIG. 3. Normalized power flow $\Pi'_{i,j}/\Psi'_{i,j}$ per unit length vs ka for uniformly distributed load. Force excitation: $i=1$, $j=1$; cross coupling term: $i=1$, $j=2$; tangential moment excitation: $i=2$; $j=2$; normal moment excitation: $i=3$, $j=3$. $\nu=0.3$.

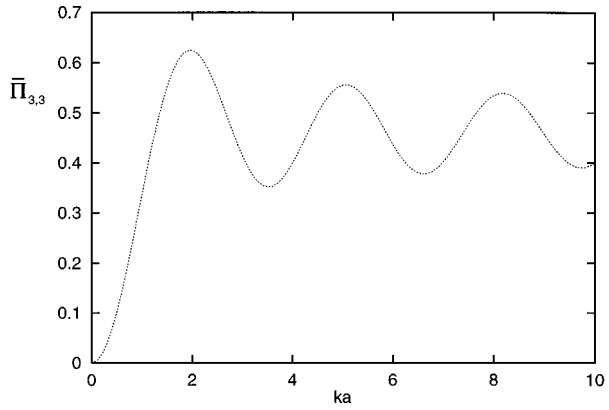


FIG. 4. Normalized total power flow $\bar{\Pi}_{3,3}$ vs ka for a uniformly distributed normal moment over an excitation area of length $2a$; $\nu=0.3$.

Although no beam admittance corresponds to the case of normal moment excitation, it is convenient to normalize $\Pi'_{3,3}$ by $\Psi'_{3,3}=\frac{1}{2}(t_0^2\omega/Dk)$.

For line excitation, i.e., $ka\to\infty$, the power flow approaches the result for the beam, i.e., $\Pi'_{i,j}/\Psi'_{i,j}\to 1$, except for the case of normal moment excitation ($i=3$, $j=3$). Note that $\Pi'_{3,3}/\Psi'_{3,3}\to 0$ for $ka\to\infty$.

For point excitation ($ka\ll 1$), the power flows $\Pi'_{1,1}$, $\Pi'_{2,2}$, and $\Pi'_{3,3}$ are proportional to ka. The constant of proportionality (the slope of the curve at $ka=0$) corresponds to the point admittance of an edge-excited, semi-infinite plate. If the power had been normalized by the power flow from a point load on an infinite plate, the limit $ka\to 0$ is expected to be about 2, indicating that the semi-infinite plate and the infinite plate should behave similar, apart from a geometric factor. However, due to the free edge conditions this number is larger. Actually, the free edge factor is about 1.85 for a point force and 1.73 for a tangential point moment; see Sec. III.

The comparison with the infinite plate does not make sense for the coupling term $\Pi'_{1,2}$ since a point force on an infinite plate induces no rotational velocity at its drive point and conversely, a point moment cannot induce a drive point transverse velocity. Nevertheless, the coupling term $\Pi'_{1,2}$ is proportional to ka too and the corresponding point mobility is given in Sec. III.

Modulation typifies the results at intermediate values of ka. This is caused by interference of waves emitted from different edge locations. Modulation depth is large in the case of force excitation ($i=1$, $j=1$), while it is virtually absent in the case of tangential moment excitation ($i=2$, $j=2$). For the coupling term ($i=1$, $j=2$) modulation is a minor effect. The absence of modulation in the case of tangential moment excitation is due to the highly directional response in the direction perpendicular to the edge. Consequently, virtually no waves are emitted parallel to the edge that may interfere with each other.

For a uniformly distributed normal moment the normalized, total power $\bar{\Pi}_{3,3}=\Pi_{3,3}/[t_0^2/(m''D)^{1/2}]$ is given in Fig. 4, while Fig. 3 shows the power flow per unit length. First, the results in Fig. 4 show deep modulation. The modulation depth is of the same order of magnitude as the power flow itself. This holds not only for intermediate, but also for large values of ka. The strong modulation is due to the highly directional response of the plate in the direction parallel to the edge. Second, the total power flow from a uniformly distributed normal moment is finite, even if the excitation length (and thus ka) is very large. In terms of Fig. 3 this property is $\Pi'_{3,3}/\Psi'_{3,3}\to 0$ for $ka\to\infty$. This feature is a consequence of the antisymmetry of the plate response to a single normal point moment. (Note that a point force and a tangential point moment induce symmetric fields.)

A simplified representation of the uniformly distributed normal moment consists of a large number of uniformly spaced normal point moments of equal strength and phase. Due to the antisymmetry of the plate response to a single normal point moment, the contributions from two adjacent loads cancel completely at their common center and partially at other points on the line segment that connects the drive points. For a large number of normal point moments cancellation dominates everywhere except near the endpoints $y=\pm a$. So, only those parts of the excitation area that are in the vicinity of the endpoints radiate structural energy. The central part does not, because the plate is not excited here due to cancellation. This explains why the power flow per unit length $\Pi'_{3,3}$ vanishes for $ka\to\infty$. Subsequently, interaction between the two radiating regions close to $y=\pm a$ shows up as modulation of the power flow in Fig. 4.

There is a striking resemblance between the results for force excitation and those for a baffled piston radiating sound into a two-dimensional half-space.[3] For $ka\ll 1$ the piston radiates like a monopole, while for $ka\to\infty$ only plane waves are emitted. For $ka=\mathcal{O}(1)$ modulation characterizes the transition region. So the piston problem might well serve as an acoustical analogy. Note that the analogy goes beyond the obvious resemblance: in both cases the boundary conditions at the edge of the semi-infinite domain are of the Neumann type.

III. POINT MOBILITIES

For reference and for comparison with Ref. 1 the expressions for the plate admittance at $y=0$ are listed:

$$i\omega w(0,0)=v_F+v_M, \quad i\omega.-\frac{\partial w}{\partial x}(0,0)=\Omega_F+\Omega_M, \quad (46)$$

where

$$\bar{v}_F\equiv\frac{v_F}{\omega F/Dk^2}$$

$$=\frac{2i}{\pi}\,\mathrm{PV}\int_1^\infty\frac{\sin(kap)}{kap}\frac{1}{\det(p)}\,dp$$
$$-\frac{2i}{\pi}\int_0^1\frac{\sin(kap)}{kap}\frac{\mu_+(p)[M_-(p)]^2}{\triangle(p)}\,dp$$
$$+\frac{2}{\pi}\int_0^1\frac{\sin(kap)}{kap}\frac{-i\mu_-(p)[M_+(p)]^2}{\triangle(p)}\,dp$$
$$+2\frac{\sin(kap_1)}{kap_1}\left[\frac{d}{dp}\det(p_1)\right]^{-1}, \quad (47)$$

$$\bar{v}_M\equiv\frac{v_M}{-\omega M/Dk}=\frac{i}{\pi}\,\mathrm{PV}\int_1^\infty\frac{\sin(kap)}{kap}\frac{\mu_-(p)M_+(p)+\mu_+(p)M_-(p)}{\det(p)}\,dp$$
$$+\frac{i}{\pi}\int_0^1\frac{\sin(kap)}{kap}\frac{[-i\mu_-(p)]^2[M_+(p)]^3-[\mu_+(p)]^2[M_-(p)]^3}{\triangle(p)}\,dp$$
$$+\frac{2}{\pi}\int_0^1\frac{\sin(kap)}{kap}\frac{-i\mu_-(p)\mu_+(p)M_-(p)M_+(p)}{\triangle(p)}\,dp$$
$$+\frac{\sin(kap_1)}{kap_1}[\mu_-(p_1)M_+(p_1)+\mu_+(p_1)M_-(p_1)]\left[\frac{d}{dp}\det(p_1)\right]^{-1}, \quad (48)$$

$$\bar{\Omega}_F\equiv\frac{\Omega_F}{-\omega F/Dk}=\frac{v_M}{-\omega M/Dk}\equiv\bar{v}_M, \quad (49)$$

$$\bar{\Omega}_M\equiv\frac{\Omega_M}{\omega M/D}=\frac{2i}{\pi}\,\mathrm{PV}\int_1^\infty\frac{\sin(kap)}{kap}\frac{\mu_-(p)\mu_+(p)}{\det(p)}\,dp+\frac{2i}{\pi}\int_0^1\frac{\sin(kap)}{kap}\frac{[-i\mu_-(p)]^2\mu_+(p)[M_+(p)]^2}{\triangle(p)}\,dp$$
$$+\frac{2}{\pi}\int_0^1\frac{\sin(kap)}{kap}\frac{-i\mu_-(p)[\mu_+(p)]^2[M_-(p)]^2}{\triangle(p)}\,dp+2\frac{\sin(kap_1)}{kap_1}[\mu_-(p_1)\mu_+(p_1)]\left[\frac{d}{dp}\det(p_1)\right]^{-1}. \quad (50)$$

Two obvious misprints in Ref. 1 are noted: the denominator appearing in the principal value integrals in expressions (21) and (25) of Eichler's paper should read ''det'' instead of ''Det.''

For the case of normal moment excitation the admittance at $y=0$ is

$$i\omega\frac{\partial w}{\partial y}(0,0)=\Omega_T, \quad (51)$$

where

$$\bar{\Omega}_T\equiv\frac{\Omega_T}{\omega T/D}$$
$$=\frac{2i}{\pi}\,\mathrm{PV}\int_1^\infty\frac{\sin(kap)}{kap}\frac{p^2}{\det(p)}\,dp$$
$$-\frac{2i}{\pi}\int_0^1\frac{\sin(kap)}{kap}\frac{\mu_+(p)[M_-(p)]^2}{\triangle(p)}\,p^2\,dp$$
$$+\frac{2}{\pi}\int_0^1\frac{\sin(kap)}{kap}\frac{-i\mu_-(p)[M_+(p)]^2}{\triangle(p)}\,p^2\,dp$$
$$+2\frac{\sin(kap_1)}{kap_1}p_1^2\left[\frac{d}{dp}\det(p_1)\right]^{-1}. \quad (52)$$

Note that expression (52) differs from the corresponding formulas (18) and (19) in Ref. 1, due to a different definition of $t(y)$. However, they coincide for $ka\to 0$.

For point loads, i.e., for $ka\to 0$, expressions (47)–(50) and (52) can be simplified by replacing $[\sin(kap)/kap]$ and $[\sin(kap_1)/kap_1]$ by their limiting value, which is equal to 1. The resulting formulas for the point mobilities have been evaluated numerically and are listed in Table I. It is interesting to compare these results to those for infinite plates excited by corresponding loads. The input mobility for an infinite plate excited by a point force is purely real:

$$v_F^\infty=\frac{1}{8}\frac{\omega F}{Dk^2}.$$

The imaginary part of the tangential moment admittance $\bar{\Omega}_M$

TABLE I. Point mobilities for edge-excited, semi-infinite plates, $\nu=0.3$.

	Real part	Imaginary part
$\overline{v_F}$	0.461 96	-5.97×10^{-7}
$\overline{v_M}\equiv\overline{\Omega_F}$	0.291 48	0.29149
$\overline{\Omega_M}$	0.216 44	∞
$\overline{\Omega_T}$	0.216 44	∞

is singular for $ka\to0$, i.e., the imaginary part becomes infinitely large as ka goes to zero. This is well known[2] for the case of moment excitation of an infinite plate:

$$\Omega_M^{\infty}=\frac{\omega M}{8D}\left[\frac{1}{2}-\frac{2i}{\pi}\left\{\gamma_{\mathrm{E}}-\frac{1}{2}+\ln(ks/4)\right\}\right],\quad \text{for}\quad ks\to0. \tag{53}$$

One should, however, be cautious in comparing the leading order imaginary part in Eq. (53) to the result for the semi-infinite plate, because in Eq. (50) $2a$ is the width of the excitation area, which is parallel to the direction of the moment vector, while in Eq. (53) s is the separation of the moment producing pair of forces and is oriented perpendicular to the direction of the moment vector.

From Table I we compute the ratios

$$\frac{\mathrm{Re}[\overline{v_F}]}{\mathrm{Re}[\overline{v_F^{\infty}}]}\approx3.6957,\quad \frac{\mathrm{Re}[\overline{\Omega_M}]}{\mathrm{Re}[\overline{\Omega_M^{\infty}}]}\approx3.4630.$$

These values contain a geometric factor, which is 2, and a free edge factor which accounts for the boundary conditions. It is interesting to compare these ratios to their counterparts for a beam, which are 4 in both cases.

Apart from $\mathrm{Im}[\overline{v_F}]$ and $\mathrm{Re}[\overline{\Omega_M}]$ the values listed in Table I agree with those given by Eichler within about 10%; this discrepancy is probably caused by an inaccurate approximation of the location of the pole p_1 in Ref. 1. The value for $\mathrm{Re}[\overline{\Omega_M}]$, however, differs considerably from the one given by Eichler. His value was 0.082. This would give the ratio $\mathrm{Re}[\overline{\Omega_M}]/\mathrm{Re}[\overline{\Omega_M^{\infty}}]\approx1.321$, which is even smaller than the geometric factor. So that result is obviously wrong. Anyhow, the order of magnitude of the present result is reasonable.

The numerical value for $\mathrm{Im}[\overline{v_F}]$ in Table I has no significant figures, since the results for the two integrals in Eq. (47) cancel within the relative accuracy. However, the absolute value is certainly smaller than 10^{-6}. Furthermore, the numerical results for both $\mathrm{Re}[\overline{v_M}]$ and $\mathrm{Im}[\overline{v_M}]$ are equal within the relative accuracy. Similarly, the results for $\mathrm{Re}[\overline{\Omega_M}]$ and $\mathrm{Re}[\overline{\Omega_T}]$ are equal. These remarkable observations suggest that these results could be established analytically. So it is conjectured that the corresponding integral expressions in (47)–(50) and (52) are identical for all values of ν, i.e.,

$$\mathrm{Im}[\overline{v_F}]\equiv0,\quad \mathrm{Re}[\overline{v_M}]\equiv\mathrm{Im}[\overline{v_M}],\quad \mathrm{Re}[\overline{\Omega_M}]\equiv\mathrm{Re}[\overline{\Omega_T}].$$

IV. EXCITATION BY *N* POINT LOADS

A. *N* point forces

Let N distinct point forces having strength $F_1,\ldots,F_N$ and phase $\phi_1,\ldots,\phi_N$ respectively, excite the plate at positions $y_1,\ldots,y_N$. All forces are independent and have frequency ω. The excitation function and its Fourier transform reads

$$f(y)=\sum_{j=1}^{N}F_j\exp(i\phi_j)\,\delta(y-y_j),$$

$$\tilde{f}(q)=\sum_{j=1}^{N}\frac{F_j}{2\pi}\exp[i(\phi_j+qy_j)].$$

The power flow follows from the first integral in Eq. (25). Observe that

$$\tilde{f}(q)^*\tilde{f}(q)=\sum_{j=1}^{N}\left\{\frac{F_j^{\,2}}{4\pi^2}+\sum_{k=1}^{j-1}\frac{F_jF_k}{2\pi^2}\cos[\phi_{jk}+qy_{jk}]\right\}. \tag{54}$$

The first term corresponds to the power flow injected into the plate by N point forces disregarding cross coupling. The second term accounts for all mutual interactions of two forces, where $\phi_{jk}=\phi_j-\phi_k$ is their phase difference and $y_{jk}=y_j-y_k$ their separation. The form in which Eq. (54) is cast shows that it suffices to investigate the case $N=2$. In particular, the results for the cross coupling term for a wide range of ks and for all possible values of ϕ, where $s=|y_2-y_1|$ and $\phi=|\phi_2-\phi_1|$, provide a building block for the calculation of the total power flow, which is a linear combination of such terms. This will be the subject of Sec. IV B.

As a special case, consider a periodic distribution of forces of equal strength and linearly varying phase, i.e., let $F_j=F$, $\phi_j=j\psi$ and $y_j=js$, where ψ is the phase difference between two adjacent forces and s is the spacing. The double series in Eq. (54) simplifies to

$$\tilde{f}(q)^*\tilde{f}(q)=\frac{F^2}{4\pi^2}\frac{1-\cos(\alpha N)}{1-\cos(\alpha)},\quad \text{where}\quad \alpha=\psi+qs.$$

The case of a uniformly distributed force is retained if one takes the limit for $N\to\infty$ and $s\to0$, while keeping $Ns=2a$ as well as $F/s=f_0$ constant and choosing $\psi=0$. This yields the square of expression (36).

B. Two point forces

For $N=2$ expression (54) has one cross coupling term. The power flow reads

$$\overline{\Pi}_{\mathrm{in}}=\frac{\Pi_{\mathrm{in}}}{F_1F_2/(m''D)^{1/2}}=\frac{1}{2}\left\{\frac{F_1}{F_2}+\frac{F_2}{F_1}\right\}\Phi_{11,22}+\cos\,\phi\;\Phi_{12,21}, \tag{55}$$

where

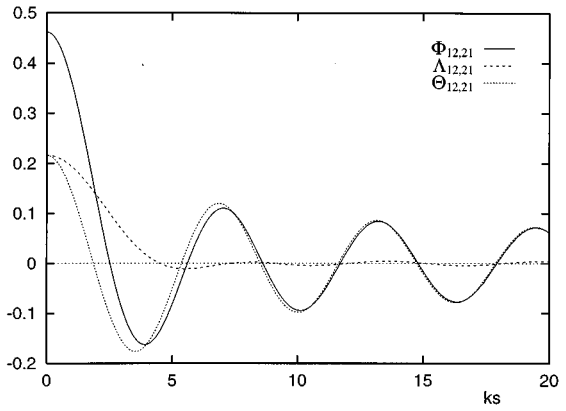


FIG. 5. Normalized-cross coupling power flow $\Phi_{12,21}$, $\Lambda_{12,21}$, and $\Theta_{12,21}$ vs ks for two point loads. $\nu=0.3$.

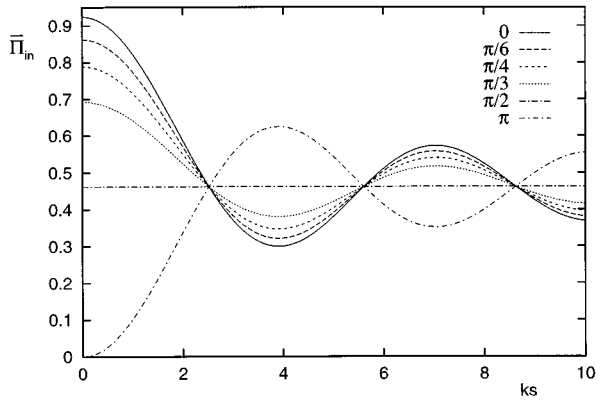


FIG. 6. Normalized total power flow $\overline{\Pi}_{in}$ vs ks for two point forces of equal strength and separation s. The key gives the phase difference ϕ. $\nu=0.3$.

$$\Phi_{11,22}=\frac{2}{\pi}\int_0^1 \frac{-i\mu_-(p)[M_+(p)]^2}{\triangle(p)}\,dp + 2\left[\frac{d}{dp}\det(p_1)\right]^{-1}, \tag{56}$$

$$\Phi_{12,21}=\frac{2}{\pi}\int_0^1 \cos(ksp)\,\frac{-i\mu_-(p)[M_+(p)]^2}{\triangle(p)}\,dp + 2\cos(ksp_1)\left[\frac{d}{dp}\det(p_1)\right]^{-1}. \tag{57}$$

Here, s denotes the separation of the forces and ϕ the phase difference. Note that $\Phi_{11,22}=\mathrm{Re}[\overline{v_F}]$; see Sec. III. Its numerical value for $\nu=0.3$ is given in Table I. Also note that $\Phi_{12,21}=\Phi_{11,22}$ for $s=0$. Finally, comparison with Eq. (45) shows that $\Phi_{12,21}=\Phi_{11,22}-\overline{\Pi}_{3,3}$ if $s=2a$.

The cross coupling term (57) has been calculated for $0<ks<20$ and is presented in Fig. 5 (solid line). The result shows rather strong modulation which may be explained as follows: The plate response due to a single force has a traveling wave part that dominates over the nonpropagating near-field component if ks is sufficiently large. Therefore, the phase of the response varies linearly along the edge. This phase variation causes a constructive or destructive effect on the power produced by the second force depending on the position of its drive point. The same argument applies if the forces are interchanged. The constructive and destructive effects show up as modulation of the cross coupling power flow.

The functional form of $\Phi_{12,21}$ can be used to find the total power flow from N point forces. Just evaluate $\Phi_{12,21}$ for each value of $s=y_{jk}$. The total power flow is then

$$\Pi_{\mathrm{in}}=\frac{1}{(m''D)^{1/2}}\sum_{j=1}^{N}\left\{\frac{1}{2}F_j^2\Phi_{11,22} + \sum_{k=1}^{j-1}F_jF_k\cos(\phi_{jk})\Phi_{12,21}[s=y_{ik}]\right\}. \tag{58}$$

For two forces of equal strength, i.e., $F_1=F_2$, the total power flow was calculated for various values of ϕ. The results are presented in Fig. 6. If $F_1\neq F_2$ the power should be raised by a fixed level, i.e.,

$$\left[\frac{F_1/F_2+F_2/F_1}{2}-1\right]\Phi_{11,22},$$

which is independent of s. The phase difference ϕ determines the sign and the amplitude of the modulation.

C. Two tangential point moments

Proceeding along the same lines the power flow from two tangential point moments is discussed briefly. The results can be used to deal with N tangential point moments. Let M_1 and M_2 be the amplitudes of the moments, s the separation and ϕ their phase difference. The total power flow is

$$\overline{\Pi}_{\mathrm{in}}=\frac{\Pi_{\mathrm{in}}}{M_1M_2\omega/D}=\frac{1}{2}\left\{\frac{M_1}{M_2}+\frac{M_2}{M_1}\right\}\Lambda_{11,22} + \cos\phi\,\Lambda_{12,21}, \tag{59}$$

where

$$\Lambda_{11,22}=\frac{2}{\pi}\int_0^1 \frac{-i\mu_-(p)[\mu_+(p)]^2[M_-(p)]^2}{\triangle(p)}\,dp + 2\mu_-(p_1)\mu_+(p_1)\left[\frac{d}{dp}\det(p_1)\right]^{-1}, \tag{60}$$

$$\Lambda_{12,21}=\frac{2}{\pi}\int_0^1 \cos(ksp)\,\frac{-i\mu_-(p)[\mu_+(p)]^2[M_-(p)]^2}{\triangle(p)}\,dp + 2\mu_-(p_1)\mu_+(p_1)\cos(ksp_1)\left[\frac{d}{dp}\det(p_1)\right]^{-1}. \tag{61}$$

Note that $\Lambda_{11,22}=\mathrm{Re}[\overline{\Omega}_M]$; see Sec. III. Its numerical value for $\nu=0.3$ is given in Table I. Also note that $\Lambda_{12,21}=\Lambda_{11,22}$ for $s=0$.

Numerical values for the cross coupling term (61) have been obtained for $0<ks<20$ and are presented in Fig. 5 (dashed line). In this case modulation is only a small effect, because the plate response is highly directional in the direction perpendicular to the edge. So the response along the

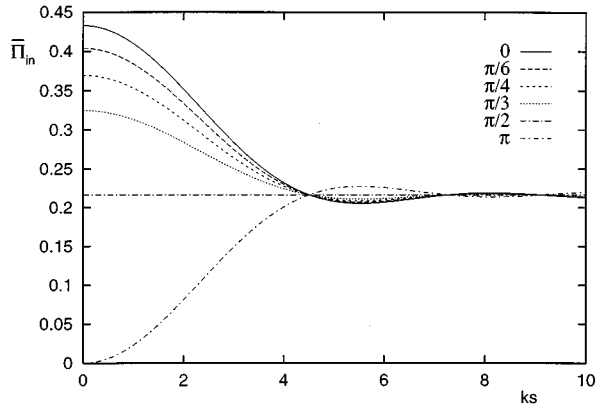


FIG. 7. Normalized total power flow $\overline{\Pi}_{in}$ vs ks for two tangential point moments of equal strength and separation s. The key gives the phase difference ϕ. $\nu=0.3$.

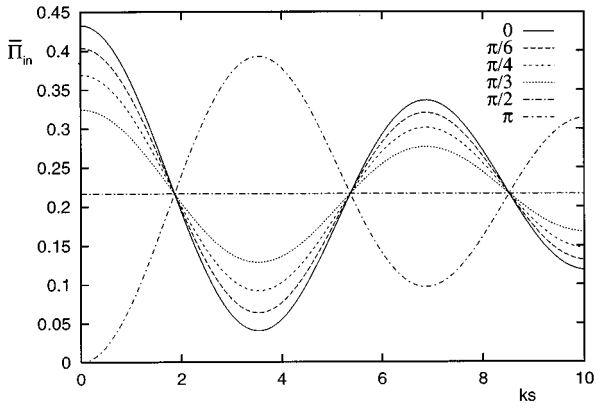


FIG. 8. Normalized total power flow $\overline{\Pi}_{in}$ vs ks for two normal point moments of equal strength and separation s. The key gives the phase difference ϕ. $\nu=0.3$.

edge line due to a tangential point moment is very weak. Consequently, interaction with the other tangential point moment is a minor effect.

Numerical results for the total power flow for two tangential point moments of equal strength are given in Fig. 7 and show that modulation is small. Amplification by near field interaction is important for $ks<\pi$.

D. Two normal point moments

The results are easily extended to the case of two normal point moments. The analysis is similar and can also be used to deal with N normal moments. Let T_1 and T_2 denote the amplitudes of the normal moments, ϕ their phase difference and s their separation. The total power flow is

$$\overline{\Pi}_{\text{in}}=\frac{\Pi_{\text{in}}}{T_1T_2\omega/D}=\frac{1}{2}\left\{\frac{T_1}{T_2}+\frac{T_2}{T_1}\right\}\Theta_{11,22}+\cos\,\phi\,\Theta_{12,21}, \tag{62}$$

where

$$\Theta_{11,22}=\frac{2}{\pi}\int_0^1\frac{-i\mu_-(p)[M_+(p)]^2}{\triangle(p)}\,p^2\,dp + 2p_1^2\left[\frac{d}{dp}\det(p_1)\right]^{-1}, \tag{63}$$

$$\Theta_{12,21}=\frac{2}{\pi}\int_0^1\cos(ksp)\,\frac{-i\mu_-(p)[M_+(p)]^2}{\triangle(p)}\,p^2\,dp + 2p_1^2\cos(ksp_1)\left[\frac{d}{dp}\det(p_1)\right]^{-1}. \tag{64}$$

Note that $\Theta_{11,22}=\text{Re}[\overline{\Omega}_T]$; see Sec. III. Its numerical value for $\nu=0.3$ is given in Table I. Also note that $\Theta_{12,21}=\Theta_{11,22}$ for $s=0$.

Numerical results for the cross coupling term $\Theta_{12,21}$ are presented in Fig. 5 (dotted line). Unlike the previous case the modulation depth is very large, because the plate response to a normal point moment is highly directional along the edge, i.e., the response along the edge line carries most of the energy of the radiated field. Therefore, the coupling with the other normal point moment is almost complete and the modulation is very deep.

Figure 8 presents numerical results for the total power flow for two normal point moments of equal strength. It appears that modulation dominates the result even if the separation of the moments is large.

V. SUMMARY OF RESULTS

Following Eichler,[1] the response of a semi-infinite plate excited along its edge is analyzed using Fourier transforms. New, closed form integral expressions are derived for the power flow into the plate due to arbitrary force and moment distributions along the edge. It is shown that, under certain conditions, the coupling terms (between force and moment) are identical to each other. In addition to Ref. 1, analytical formulas are given for the poles of the integrand in the complex plane. Numerical results are presented for the case of a uniformly distributed load (force, tangential moment, and normal moment) applied at a finite part of the edge. The results are classified as follows: (i) $ka\ll 1$: point excitation. A ''free edge factor'' is introduced to facilitate the comparison with the point mobilities for infinite plates. Meanwhile, Eichler's results[1] for point mobilities are corrected. (ii) $ka\to\infty$: line excitation. The results compare well to those for a semi-infinite beam excited by corresponding loads. (iii) $ka=\mathcal{O}(1)$: the transition region is characterized by modulation which is very pronounced in the case of normal moment excitation, moderate in the case of force excitation and neglectable in the case of tangential moment excitation.

Multiple point excitation with nonuniform spacing and arbitrary distributions of amplitude and phase is shown to decompose into a sum of terms including all mutual interactions of two point loads. Numerical results that show the variation with phase difference and spacing are presented for the basic problem of two point loads including the coupling term. Modulation is very strong for normal moments, moderate for forces and weak for tangential moments.

ACKNOWLEDGMENTS

This work is supported by the TNO-Institute of Applied Physics, Delft, The Netherlands. The author is indebted to Professor J. W. Verheij for drawing his attention to this problem and to Ir. M. H. A. Janssens for his assistance in obtaining numerical results for the case of uniform force excitation (solid curve in Fig. 3). Discussions with Dr. B. A. T. Petersson and Ir. H. Rijnks on the point moment mobility are gratefully acknowledged. Valuable comments on an earlier version of this paper was kindly provided by Ir. M. G. Dittrich, Dr. B. A. T. Petersson, and Dr. D. J. Thompson.

[1] E. Eichler, ''Plate-edge admittances,'' J. Acoust. Soc. Am. **36**, 344–348 (1964).
[2] L. Cremer and M. Heckl, *Structure-Borne Sound* (Springer-Verlag, Berlin, 1973), Chap. IV, §3, pp. 249–281.
[3] O. A. Lindemann, ''Radiation impedance of a long narrow rectangular piston in a plane baffle,'' J. Acoust. Soc. Am. **52**, 1045–1048 (1972).

Analytical/numerical matching for efficient calculation of scattering from cylindrical shells with lengthwise constraints

R. C. Loftman and D. B. Bliss
Department of Mechanical Engineering, Duke University, Durham, North Carolina 27708-0300

(Received 4 April 1997; revised 18 November 1997; accepted 16 December 1997)

Structural discontinuities in highly coupled fluid–structure systems are modeled by an approach called analytical/numerical matching (ANM). This method separates the low-resolution global influence of a discontinuity from the relatively high-resolution local effects. A continuous, smoothed replacement for a fundamental structural discontinuity is constructed so that the system is identically unchanged beyond a small smoothing region. Simultaneously, the precise local effect of smoothing the discontinuity is retained in analytical form. The smoothed problem is solved by numerical techniques, with rapid convergence and reduced computational cost. The original discontinuous character is restored using the analytical expression for the local difference between the smoothed and the original problems. ANM has been successfully applied to two-dimensional cases of acoustic scattering from a thin, infinitely long cylindrical shell, with multiple structural discontinuities. Local solutions for longitudinal line discontinuities with radial, tangential, and rotational constraints have been formulated using ANM. Line constrained scattering problems, as well as line driven problems, are investigated. The ability of analytical/numerical matching to replace a discontinuous physical problem by a well behaved continuous one for numerical evaluation, while ultimately retaining the original geometry and physical behavior, is illustrated. [S0001-4966(98)00904-7]

PACS numbers: 43.40.Rj, 43.20.Tb [CBB]

INTRODUCTION

The vibration of an elastic body immersed in heavy fluid is a complex problem in which the dynamics of the structure and fluid must be solved simultaneously. There exist few instances of fully coupled fluid–structure interaction problems that can be solved in closed form. Usually, the problem must be discretized in order to study the dynamics. Even the few academic problems for which formal expression of analytic solutions are tractable generally require approximation in the form of modal series truncation or numerical integration. A particular complication in the analysis of engineering structures arises from inhomogeneities. There may be, for example, complex joints due to the attachment of internal structures such as bulkheads or reinforcements. Such discontinuities have a profound effect on the dynamics and complicate the analysis considerably. Furthermore, the analysis of structural discontinuities is inherently high resolution in nature. A discrete structural inhomogeneity, such as a rib or stiffener, causes a local region of rapid change, and even discontinuous derivatives across the inhomogeneity. This high-resolution region challenges the accuracy and efficiency of computational methods. This paper is concerned with efficiently modeling such structural inhomogeneities. The model problem presented in this paper is an important step in the development of a method called analytical/numerical matching (ANM), which reduces the computational effort required to calculate fully coupled fluid–structure interaction problems with structural inhomogeneities. The advantage of the method is accomplished by separating the most computationally intensive part of the problem, the local rapid variation associated with the discontinuity, from the overall response and treating it separately and analytically, thus easing the burden on the computational technique.

The solution of planar (2-D) vibrations of a fluid coupled thin elastic circular cylindrical shell is a fundamental problem in structural acoustics, with significant practical relevance. The physics of this problem has been well explored by previous authors. The solution of acoustic scattering from the unconstrained shell was first discovered analytically by Junger in 1952.[1] Structural inhomogeneities were addressed soon afterward.[2] It is possible to arrive at analytical solutions for the cylindrical geometry in the presence of structural discontinuities so long as the nonuniformities lie along coordinate lines.[3] In the case of planar vibrations of the cylindrical shell, this implies longitudinal constraint orientation. The efficient modeling of such discrete structural constraints is the subject of this work. Cuschieri and Feit have treated the scattering from a discretely constrained thin cylinder by formulating a hybrid Green's function that handles mathematical singularities arising in transform space analytically, while the evaluation of the inverse transform in general is performed numerically.[4] The present work, however, is most similar to the mode acceleration method[5,6] with regard to how computational advantage is gained. Both methods improve computational efficiency by capturing most of the high wave number content of a total solution in a separate constituent solution. In the mode acceleration method, the high-resolution content is contained in the solution of the equivalent pseudo-static problem. In the present work, the high-resolution content is contained in a separate analytic solution of local extent, which is associated directly with a

discontinuity. This local solution has the advantages of being derived independently of overall problem geometry and being very efficiently expressed.

Analytical/numerical matching (ANM), is a general analysis method developed by Bliss. The method achieves high accuracy using low-resolution numerics by treating the most rapidly varying parts of a problem analytically, and separately, from the overall computational evaluation of the problem. The method was first applied to problems in vortex dynamics and rotorcraft free wake analysis.[7–11] More recently, ANM has been applied to wing aerodynamics for compressible unsteady flow,[12,13] and to the development of an acoustic boundary element method.[14] The present work is part of a broad research effort to apply ANM to structural-acoustic problems with structural discontinuities.[15–17] The ANM approach leads to an accurate alternative formulation of the basic problem, with a large reduction in computational cost. In addition, the alternative formulation can lead to an improved understanding of the effect of discontinuities.

In the following sections, the ANM method is explored in the context of two-dimensional scattering from a cylindrical shell with discrete structural constraints. An overview explanation of the method is given. The specific problem geometry is presented and the solution is given for a general constraint configuration in terms of constituent solutions in the form of classical modal expansions. The ANM formulation of the discrete constraint is discussed, and compared with the classical treatment. Finally two example constraint configurations are treated, and the efficiency of the ANM method is compared to the classical modal solution. The sensitivity of the constrained scattering problem to accurate treatment of the constraint influence is also discussed.

I. OVERVIEW OF ANM FOR STRUCTURAL ACOUSTICS

The motivation for ANM in the case of structural acoustics is the presence of a small region characterized by rapid spatial variation around any discrete load applied to an elastic structure. In fact, certain properties of the response of an elastic body are discontinuous at the point of application of a discrete load. Resolving the small region about a discontinuity often requires far greater computational cost than is warranted by the system otherwise. In addition, failure to resolve the solution about a discontinuity can lead to significant errors throughout the system. Constrained structural/acoustic scattering problems are an important example of such a situation where failure to converge locally can lead to global errors.

There exists a disparity in the length scales that characterize the local effect of the discontinuity compared to the overall, or global, features of the problem. The global features, which must be resolved computationally, are characterized by the structural and acoustic wavelengths. In contrast, the length scale that characterizes the local effect of the structural discontinuity may be very small or even infinitesimal. ANM introduces an intermediate smoothing length scale with which to separate the problem into two parts. The high-resolution local effect of the discontinuity is isolated in what is called the ANM local solution, leaving only the smooth overall response in what is called the ANM global solution. The local solution is treated analytically, independent of the overall geometry. The global solution that complements this analytic local solution could be thought of as a spatially filtered version of the original problem, but with an important distinction. The high-resolution content of the problem is not merely discarded, it is preserved in the local solution. The result of the ANM process is a composite solution, given as the ANM local solution plus the ANM global solution. The local solution is known analytically, requiring negligible computational effort to represent. The global solution has been efficiently smoothed, requiring significantly less computational effort to solve. The result is a very computationally efficient ANM composite solution.

The local solution is the keystone of the ANM method. However, it is not actually solved in the usual sense. The local solution is an artificially constructed solution. It is carefully prescribed so that it contains the fine scale features associated with the particular discontinuity but very little else. In fact, the displacement of the structure is prescribed to be identically zero outside of the chosen smoothing region. The global solution, which completes the ANM composite solution, follows from the choice of local solution, as will be explained subsequently.

The local solution is constructed subject to the following general requirements. First, the local solution must contain, at the very point of the discontinuity, the correct physical reaction of the system to that discontinuity. For instance, a line acting torque applied to an elastic plate produces a jump in the second derivative of transverse displacement that is proportional to the torque. A local solution that is designed for this system must contain such a jump in second derivative. Second, the local solution must be expressed explicitly, without the need for numerical approximation, for example, as a piecewise polynomial. Expressing the high-resolution effect of a discontinuity without approximation is how numerical efficiency is increased. Third, because the local solution is artificially generated, some basic compatibility conditions must be explicitly addressed. For example, discontinuous displacement and extraneous discrete loads must be eliminated. Fourth, the convergence of the associated global solution can be enhanced by a judicious choice of the local solution. Fifth, due to fluid coupling, the local solution will have some influence globally, however, measures can be taken, such as eliminating the net monopole strength of an acoustically compact local solution, that will effectively limit the range of influence.

The complementary global solution is determined as follows. The local solution is not a solution of the homogeneous governing equation of the structure. However, it is a solution of the inhomogeneous governing equation. The externally applied forcing that is required for the local solution to exist can be found by substituting the known local displacement into the governing equations. This forcing is, by construction of the local solution, the discrete influence of the discontinuity in question, opposed by a certain smooth pressure distribution. Since the discrete influence of the discontinuity is known (it is the starting point for the construction of the local solution), the opposing smooth pressure distribution

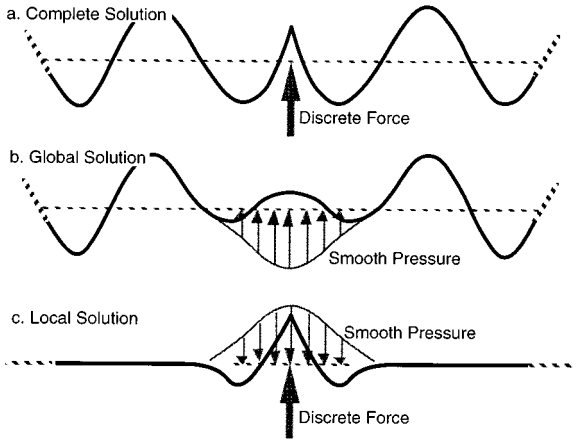


FIG. 1. Pictorial representation of ANM decomposition. (a) Complete solution. (b) Global solution. (c) Local solution.

can be found. This pressure distribution is called the global driving pressure for the following reason. The definition of the problem decomposition requires that the local solution is the difference between the original and the global solutions. It follows from linearity that the forcing required to create the local solution is the difference between the forcings required to create the original and the global solutions. Therefore, the global solution is just the original physical system driven by the global driving pressure described above, rather than the discrete influence of the discontinuity.

The ANM decomposition of a generic discretely driven structure is pictorially represented in Fig. 1. The original discretely driven structure, with a sharp peak and global features such as structural waves, is traded for two constituent solutions. The global solution, which contains the global features, is the smoothed version of the original problem. The local solution, which contains the sharp peak is the difference between the original and the global solutions and is known analytically. Note that the constituent solutions are separated using superposition. There are no specific regions of the physical domain where one or the other of the local and global solutions are valid. The ANM local solution is more an overlay than a subdomain.

II. MATHEMATICAL MODEL

A circular cylindrical shell will be considered under the framework of small disturbance linear theory. Standard r, θ polar coordinates are used. Displacement from mean position is called w and v in the radial and tangential directions, respectively. The problem geometry is illustrated in Fig. 2. The mid-line of the problem is defined as the line that bisects the shell through $\theta=0,\pi$.

The shell is modeled by the same equations that are found in the text by Junger and Feit,[18] which are described there as a ''dynamic counterpart of Donnell's formulation, and are taken from Kraus.'' [19] As a platform for developing ANM these simple equations are deemed acceptable in spite of shortcomings. Using the radius of the cylinder, a, and the reciprocal of frequency, $1/\omega$, as the characteristic length and

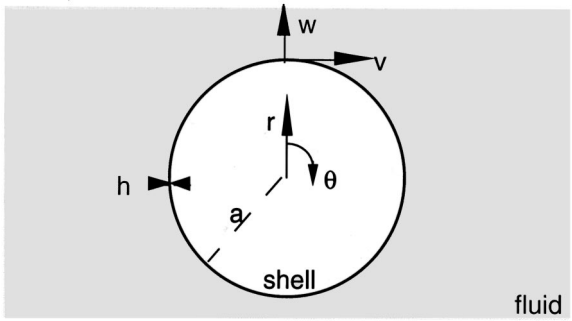


FIG. 2. Problem geometry.

time scales, the dimensionless form of the equations for planar shell motion are

$$\frac{\partial \bar{v}}{\partial \theta}+\bar{w}+\beta^2\frac{\partial^4 \bar{w}}{\partial \theta^4}+K_p^2\frac{\partial^2 \bar{w}}{\partial \bar{t}^2}=\bar{p}_r-\bar{p}_f, \tag{1}$$

$$-\frac{\partial^2 \bar{v}}{\partial \theta^2}-\frac{\partial \bar{w}}{\partial \theta}+K_p^2\frac{\partial^2 \bar{v}}{\partial \bar{t}^2}=\bar{p}_\theta, \tag{2}$$

where, $\bar{w}=w/a$, $\bar{v}=v/a$, and $\bar{t}=\omega t$. Dimensionless frequency is $K_p=k_p a=(\omega/c_p)a$ related to a, the shell radius and c_p, the longitudinal wave speed in a plate. The dimensionless measure of thickness is $\beta^2=h^2/12a^2$. The dimensionless pressure, defined by $\bar{p}=ap/h\rho_s c_p^2=ap(1-\nu^2)/E$ gives applied pressure related to the stiffness of the shell. The pressures applied to the shell and acting in the radial and tangential directions, respectively, are p_r and p_θ. The fluid pressure on the outside surface of the shell is p_f.

In the surrounding fluid, the familiar acoustic wave equation governs the pressure field, p_f, and the acoustic momentum equation relates the pressure solution to its velocity field, $\mathbf{V}_f$:

$$\nabla^2 p_f=\frac{1}{c_f^2}\frac{\partial^2 p_f}{\partial t^2}, \tag{3}$$

$$\rho_f\frac{\partial \mathbf{V}_f}{\partial t}=-\nabla p_f, \tag{4}$$

where c_f is the acoustic wave speed, and ρ_f is the fluid density.

Harmonic time dependence of the form $e^{i\omega t}$ is assumed hereafter.

The formalism for modeling constraint conditions applied to the elastic shell is left to Section III C.

III. SCATTERING FROM THE CONSTRAINED SHELL

The solution for planar excitation of a cylindrical shell immersed in an acoustic medium and described by the previous equations [Eqs. (1) and (2)] is analytically tractable in the form of a modal series.[1] With the general solution of the forced, unconstrained, fluid loaded shell known, scattering from a constrained shell can be formulated as the superposition of the unconstrained scattering problem and the individual force influences of the constraints.

A. The driven shell

The response of the unconstrained shell driven by external forcing, $\overline{p_r}$ and $\overline{p_\theta}$, can be solved by assuming a series representation of the following form:

$$\overline{w}=\sum_{n=0}^{\infty}\left[A_n^{\overline{w}}\cos(n\theta)+B_n^{\overline{w}}\sin(n\theta)\right]e^{i\omega t}, \tag{5}$$

$$\overline{v}=\sum_{n=0}^{\infty}\left[B_n^{\overline{v}}\sin(n\theta)+A_n^{\overline{v}}\cos(n\theta)\right]e^{i\omega t}. \tag{6}$$

Substitution into the governing equation for the shell yields an independent algebraic system of linear equations for each order n. This system is solved in closed form to give the coefficients of the series exactly, provided the coefficients of the external forcing are known. In general, the Fourier coefficients of a known driving force can always be found, however, radiation loading (acoustic pressure on the shell surface due to shell motion) is dependent upon shell response and so is not known *a priori*. In this special geometry, however, each shell mode excites acoustical waves that influence only the same shell mode that excited it. The independence of the algebraic coefficient equations in n is preserved. Therefore, the coupled fluid-structure problem can be solved in closed form.

Performing separation of variables on the wave equation in cylindrical polar coordinates reveals that the fluid solution is naturally expressed as an expansion of Bessel functions in r and trigonometric functions of integral order in θ. Satisfying the Sommerfield radiation condition as $r\to\infty$ for outgoing waves yields the following solution for acoustic radiation due to shell motion:

$$\overline{p_f}=\sum_{n=0}^{\infty}\left[A_n^{\overline{p_f}}\cos(n\theta)+B_n^{\overline{p_f}}\sin(n\theta)\right]H_n^{(2)}(K_f\overline{r})e^{i\omega t}, \tag{7}$$

where $K_f=k_f a=(\omega/c_f)a$, and $H_n^{(2)}$ is the Hankel function of the second kind. Note that the azimuthal dependence is identical to that of the natural modes of shell vibration.

Using the momentum equation, continuity of radial velocity at the shell surface is the final condition which determines the acoustic solution. Acoustic radiation is given by a one-to-one linear relationship between each mode of radial shell response [each individual term in Eq. (5)] and the resulting outgoing acoustic wave mode [each individual term in Eq. (7)]. The equations of shell response coefficients are therefore determined. For example, the coefficients of shell response in the case of symmetry about the midline are

$$nB_n^{\overline{v}}+A_n^{\overline{w}}+\beta^2 n^4 A_n^{\overline{w}}-K_p^2 A_n^{\overline{w}} = -\alpha K_f\frac{H_n^{(2)}(K_f)}{H_n'^{(2)}(K_f)}A_n^{\overline{w}}+A_n^{\overline{p_r}}, \tag{8}$$

$$n^2 B_n^{\overline{v}}+nA_n^{\overline{v}}-K_p^2 B_n^{\overline{v}}=B_n^{\overline{p_\theta}}, \tag{9}$$

where $\alpha=a\rho f c_f^2/h\rho_s c_p^2$ is the fluid loading parameter. The forcing coefficients, $A_n^{p_r}$ and $B_n^{\overline{p_\theta}}$, are given by Fourier decomposition of the forcing, $\overline{p_r}$ and $\overline{p_\theta}$, in θ, as follows:

$$\begin{Bmatrix}A_n^{\overline{p_r}}\\ B_n^{\overline{p_\theta}}\end{Bmatrix}=\frac{\epsilon_n}{2\pi}\int_{-\pi}^{\pi}\begin{Bmatrix}\overline{p_r}\cos(n\theta)\\ \overline{p_\theta}\sin(n\theta)\end{Bmatrix}d\theta. \tag{10}$$

B. Scattering from unconstrained shell

Scattering of an incident acoustical plane wave from the unconstrained elastic cylinder can be formulated as the superposition of two problems. The first problem is the scattering of the plane wave from a geometrically equivalent rigid cylinder. The second problem is the fluid loaded elastic cylinder driven by the surface pressure found in the first problem. The reason for dividing the scattering problem in this way is that it allows the driving influence of the incoming wave, which has complicating fluid displacement, to be treated as a simple externally applied shell load. The formulation of the shell driven by an externally applied load has already been presented.

With the help of a mathematical identity[20] the incoming plane wave train can be written as a series of Bessel's functions in r, and trigonometric functions in θ. Balancing the radial velocity of this incident wave at the surface of the cylinder with outgoing waves yields the following solution for scattering from a rigid cylinder:

$$p=P\sum_{n=0}^{\infty}\left\{\epsilon_n i^n J_n(k_f r)\cos(n\theta) - \epsilon_n i^n J_n'(K_f)\frac{H_n^{(2)}(k_f r)}{H_n'^{(2)}(K_f)}\cos(n\theta)\right\}e^{i\omega t}. \tag{11}$$

The plane wave train, of magnitude P, originates at $\theta=0$.

This solution leads to the following coefficients of radial driving pressure, in dimensionless form:

$$A_n^{\overline{p_r}}=-\overline{P}\epsilon_n i^n\left\{J_n(K_f)-J_n'(K_f)\frac{H_n^{(2)}(K_f)}{H_n'^{(2)}(K_f)}\right\}. \tag{12}$$

Adding the solution of the shell driven by the previous applied pressure [found by Eqs. (8) and (9)] to the rigid scattering solution [Eq. (11)] gives the solution of scattering from the elastic shell.

C. Modeling shell constraints

A general formalism is presented so that any configuration of linear line-acting constraints can be modeled by the superposition of discontinuities belonging to a fundamental set of three discretely driven solutions. The fundamental solutions are those of a shell driven by line-acting radial force, tangential force, and moment.

Any line-acting constraint influences the shell at most by applying a radial force, tangential force, and moment, and can therefore be modeled by such a set of influences. The strength of application of each discrete load is determined by the nature of the constraint which is represented by a set of linear equations built up from the concept of the mechanical impedance. The strengths are linearly dependent upon velocity response of the shell, however, the strength of each load can depend on all three velocity components (radial, tangential, and rotational) of the shell. The loads could also depend

on shell response at a remote location if multiple constraints are coupled to one another directly (aside from coupling through the shell). The equations of constraint can be combined into a single matrix equation representing the potentially complicated dependence of all the component forces on all of the component velocities of the discontinuities present in the shell, as follows:

$$-\{F\}=[Z]\{V\}. \tag{13}$$

The vector $\{F\}$ represents the forcing strengths applied to the shell by each forcing component at each of the discontinuities. The vector $\{V\}$ represents the components of velocity response at each of the discontinuities. The matrix $[Z]$ gives the contribution to each forcing strength component due to unit velocity response components at each of the discontinuities. This matrix defines the physical nature of the discontinuities, it is independent of the shell.

The vector $\{V\}$ is not independent of $\{F\}$ therefore, in order to solve for $\{F\}$, $\{V\}$ must be represented in terms of its constituents, as follows:

$$\{V\}=[VI]\{F\}+\{V_{\text{unconstrained}}\}. \tag{14}$$

The total response, $\{V\}$, is the unconstrained response, $\{V_{\text{unconstrained}}\}$, plus the response due to the discrete loads, which serve to model the constraints. The response due to the discrete loads is given by the product of the matrix $[VI]$, representing the response due to the individual loads of unit strength acting on the unconstrained shell, times $\{F\}$, the forcing strengths.

The forcing magnitudes required to satisfy the configuration of internal complexities is now given as follows, from Eqs. (13) and (14):

$$\{F\}=-[[I]+[Z][VI]]^{-1}[Z]\{V_{\text{unconstrained}}\}. \tag{15}$$

With the proper strengths known, the solution is complete. The constrained problem is constructed by adding the now completely determined constituent solutions.

Examples of the constrained problem will be held until Section VI, after the three line driven solutions are investigated in depth.

IV. ANM FORMULATION OF DISCRETELY DRIVEN SHELL

In order to more efficiently model the rapidly varying regions around driving point discontinuities, the discretely driven problem is separated into two parts using the property of superposition. The overall response, called the global solution, is modeled by classical methods except that the original discretely driven problem is spatially smoothed, and therefore, converges more rapidly. Meanwhile, the difference between the original and the smoothed problems is represented by the local solution, as follows. This procedure is explained in thorough detail in the context of a simpler, abstract, example problem in Ref. 17.

Each of the three components (radial, tangential and rotational) of a discrete longitudinal line acting load, or point load in two dimensions, causes a unique reaction in the shell at the point of application. This local reaction contains rapid variation, requiring more computational effort to resolve than the overall shell response, and therefore hampers convergence. In each case, ANM will remove the high-resolution effect from the computational solution in order to enhance the efficiency of the overall problem.

The radially and tangentially applied discrete line forces are modeled with the Dirac delta function in radial and tangential applied pressures, respectively. The discrete moment is modeled by the first derivative of the Dirac delta function in radial pressure. These loads are simply represented by Fourier series, and therefore are easily incorporated into the previously described solution for the driven shell (Sec. III A). The reaction of the structure at the point of application of each discrete load is found by integrating the inhomogeneous governing equation across this point and taking the limit as the range of integration reduces to include only the point of application. This procedure is demonstrated for example in the case of a radially applied load at zero. Integrating the radial governing equation [Eq. (1)] across the load from $-\epsilon$ to ϵ yields:

$$[\overline{v}]_{-\epsilon}^{\epsilon}+\int_{-\epsilon}^{\epsilon}\overline{w}\,d\theta+\beta^2\left[\frac{\partial^3\overline{w}}{\partial\theta^3}\right]_{-\epsilon}^{\epsilon}+\int_{-\epsilon}^{\epsilon}-K_p^2\overline{w}\,d\theta$$
$$=\int_{-\epsilon}^{\epsilon}(F\overline{\delta}(\theta)-\overline{p_f})d\theta. \tag{16}$$

Taking the limit as $\epsilon\to 0$ under the assumptions that $\overline{v}$ is continuous, and that $\overline{w}$, K_p, and $\overline{p_f}$ are finite yields:

$$\beta^2\left[\frac{\partial^3\overline{w}}{\partial\theta^3}\right]_{0^-}^{0^+}=\frac{F}{h\rho_s c_p^2}. \tag{17}$$

Similar relationships are found for the tangentially applied force and for the moment. These relationships allow the discrete loads to be included in the generation of the ANM local solution as a boundary condition.

A. Local solution

The generation of each ANM composite solution begins with the construction of the local solution. A polynomial form is assumed in the positive half of the smoothing width (to the right of the discontinuity) for the local shell displacements in each coordinate direction, $\overline{w_l}$ and $\overline{v_l}$. These local displacements are completed within the smoothing region by the appropriate even or odd reflection across the discontinuity. Outside the smoothing region, local displacement is prescribed to be zero. For convenience in generating the local solution, the angular coordinate is chosen with the discontinuity of interest at zero and is normalized by the smoothing half-width, $\tilde{\theta}=\theta/\theta_s$. For example, the radial local displacement for a radial discrete load, which is even, is given as follows. A similar expression is assumed for the local tangential displacement, except that it is completed by an odd reflection about $\tilde{\theta}=0$:

$$\overline{w_l}=\begin{cases}\sum_{q=0}^{Q} a_q \overline{\theta^q}, & 0\leqslant\widetilde{\theta}\leqslant 1,\\ \sum_{q=0}^{Q} a_q(-\widetilde{\theta})^q, & -1\leqslant\widetilde{\theta}<0,\\ 0, & 1<|\widetilde{\theta}|\leqslant\pi/\overline{\theta_s}.\end{cases} \tag{18}$$

Once the form of the local solution is assumed, the appropriate constraint relationships are generated, and the necessary polynomial order, Q, is determined. The coefficients are then found by solving the resulting system of linear equations. The constraint relationships are generated as follows.

The effect of the discrete load applied to the system must be included in the local solution. This is accomplished by constraining the value of the appropriate derivative of displacement at zero. Ensuring the proper jump in derivative value accounts for the discrete load in the local solution. For example, a jump in the third derivative of radial displacement given by Eq. (17) ensures the inclusion of a radial discrete load in the local solution.

No other discrete loads are allowed in the solution by ensuring the continuity of the appropriate derivatives of displacement. This condition must be explicitly enforced at $\theta=0$ and at the edge of the smoothing region, $\theta=\theta_s$.

Continuity conditions must be applied in order to ensure a physically compatible local solution. Discontinuous displacement in either coordinate direction, and discontinuous slope of radial displacement must not be allowed for this physical system.

No net area under the curve representing radial displacement of the local solution is allowed in order to eliminate the net monopole strength of the local solution. This condition significantly reduces the radiated pressure due to the local solution when the smoothing is chosen to be acoustically compact. The extent of influence of the local solution is thereby reduced.

Finally, using the governing equations, conditions are imposed on the local solution to guarantee that the global driving pressure, which is a result of the forcing necessary to create the local solution, is smooth. Namely, the value and slope of the applied forcing pressure necessary to create the local solution is chosen to be continuous. This condition enhances the convergence rate of the global solution.

Although the global driving pressure is an important part of the forcing field applied to the local solution, it is more directly relevant to the global solution. Therefore, the determination of the global driving pressure, which is a direct result of the prescribed local solution, will be left to the following section.

Some of the necessary conditions on the local solution will be satisfied automatically at $\theta=0$ by the symmetry or anti-symmetry of the local solution about zero. In general, though, each constraint relationship constitutes a linear equation in terms of the coefficients of the assumed polynomials. The resulting system of linear equations determines the local solution.

The radially driven problem is symmetric about the mid-line of the driven problem, which is defined as the radius line which intersects the point of application of the force ($\theta=0$). Therefore, to generate a local solution for the discrete radial force, radial and tangential local displacements are assumed that are even and odd about zero, respectively. A jump in the third derivative of radial displacement at zero given by Eq. (17) accounts for the point force. Nine other conditions on the radial displacement must be explicitly written in order to ensure a useful local solution. Determining the local tangential displacement is trivial in this case, since none is necessary to satisfy the requirements of the local solution. The lack of tangential local displacement, however, does not imply that there will be no tangential component of applied pressure, due to the coupled nature of the governing equations. The polynomial for local radial displacement is given in the Appendix, and is illustrated for sample values in Fig. 3.

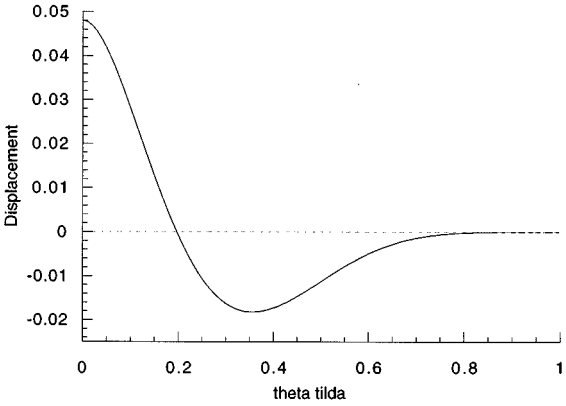


FIG. 3. Local solution for radially applied force.

The tangentially driven problem is anti-symmetric about the mid-line. Therefore, radial and tangential local displacements must be odd and even about zero, respectively. A jump in the first derivative of tangential displacement given by the following equation is required to account for the tangential point force in the local solution:

$$-\left[\frac{\partial \overline{v}}{\partial\theta}\right]_{0^-}^{0^+}=\frac{F}{h\rho_s c_p^2}. \tag{19}$$

Five other conditions are explicitly required of the tangential displacement in order to ensure a useful local solution. Nine constraints are required of the radial local displacement in this case. Coupling with the tangential displacement requires a nonzero radial displacement in order to satisfy the smoothness conditions required of the global driving pressure. The polynomials for local displacement in this case are given in the Appendix.

The moment driven problem is also anti-symmetric about the mid-line, requiring odd and even radial and tangential local displacements respectively. A jump in the second derivative of radial displacement at zero given by the following equation accounts for the point moment:

$$\beta^2\left[\frac{\partial^2 \overline{w}}{\partial\theta^2}\right]_{0^-}^{0^+}=\frac{M/a}{h\rho_s c_p^2}. \tag{20}$$

Eight other conditions are explicitly required of the local radial displacement, and six of the tangential displacement. The polynomials for local displacement are given in the Appendix.

B. Global solution

The global solution is essentially a smoothed version of the original problem. The global solution is calculated by the same conventional methods as the original discretely forced problem, except that the structure is driven by a new smooth pressure distribution, called the global driving pressure. Because the system is driven by a smooth forcing, rather than a discrete one, it converges more rapidly. The global driving pressure is found using the fact that, by the problem decomposition, the local solution is the difference between the original solution and the smooth global solution. Therefore, by linearity, the forcing that is required to create the local solution is the difference between the original discrete forcing, which is known, and the global forcing pressure. Substituting the known local solution into the governing equations of shell motion [Eqs. (8) and (9)] yields the external forcing that is required to create it. Therefore, solving for the global driving pressure is a simple matter of rearranging the terms, as shown by the following equations, where, $\overline{p}_{gf}$ and $\overline{p}_{\theta gf}$ is the global forcing pressure and $\overline{p}_{r\ \mathrm{discrete}}$ and $\overline{p}_{\theta\ \mathrm{discrete}}$ represent the original discrete forcing:

$$\overline{p}_{rgf}=-\frac{\partial \overline{v}_l}{\partial\theta}-\overline{w}_l-\beta^2\frac{\partial^4\overline{w}_l}{\partial\theta^4}-K_p^2\frac{\partial^2\overline{w}_l}{\partial\overline{t^2}}+\overline{p}_{r\ \mathrm{discrete}}-\overline{p}_f, \tag{21}$$

$$\overline{p}_{\theta gf}=\frac{\partial^2\overline{v}_l}{\partial\theta^2}+\frac{\partial\overline{w}_l}{\partial\theta}-K_p^2\frac{\partial^2\overline{v}_l}{\partial\overline{t}^2}+\overline{p}_{\theta\ \mathrm{discrete}}. \tag{22}$$

Recognizing that the applied discrete load, or infinite pressure, is canceled in the governing equation by the correctly prescribed discontinuous derivative of local displacement (which causes a canceling infinity of the proper integrated strength when one more derivative is taken), the component of the smooth driving pressure that arises from structural terms can be simply represented as a polynomial. These polynomials for the three forcing cases are given in the Appendix. The component of the smooth driving pressure that is due to fluid pressure resulting from the local displacement is guaranteed to be smooth, and is generally small compared to the structural component. This fluid pressure due to shell displacement is found by the same method that is used to handle fluid coupling to shell motion in the general solution of the forced system, in this case, a modal approach given by Eq. (7). The Fourier coefficients of the smooth driving pressures are found as in Eqs. (10). With the coefficients of driving pressure known, the global problem can now be solved by determining the coefficients of shell response as in Eqs. (8) and (9).

Taking the radially driven problem as an example, the component of the smooth pressure due to structural terms, which dominates the smooth pressure, is illustrated in Fig. 4. To illustrate the relative smoothness of the pressure distribution that was generated to replace the discrete force in the computational solution, the magnitude of the Fourier coefficients of the smooth forcing is plotted on the same axis as the Fourier coefficients of the discrete load in Fig. 5. This shows the relatively small high-harmonic content of the smooth pressure.

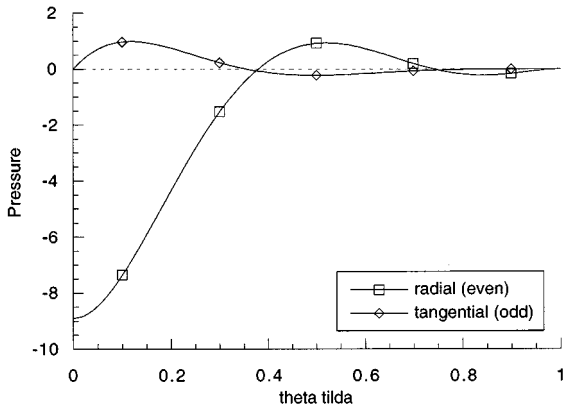


FIG. 4. Smooth driving pressure to replace radial discrete force.

C. ANM composite solution

Assembling the composite solution is a matter of superposition. The local and global solutions are simply added throughout the domain. The local solution is a piecewise polynomial description of shell motion, as well as the associated fluid reaction. The global solution is the result of a conventional solution of the driven system, but in the case of the special global forcing pressure. The series representation of the global solution contains the overall, low-resolution, shell motion, while the polynomial description of the local solution contains high-resolution content in the immediate vicinity of the discrete load. For example, the three solutions together are shown for the radially driven case in Figs. 6 and 7. Figure 6 shows the exact, global, and local radial velocity magnitudes for representative values. Figure 7 shows the Fourier decomposition of each of the three solutions to illustrate the harmonic content of each.

Figure 7 shows the harmonic magnitude of the two ANM constituent solutions, the local and global solution,

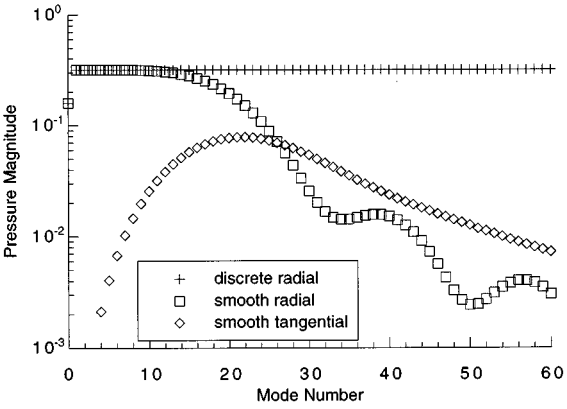


FIG. 5. Fourier decomposition of smooth driving pressure to replace radial discrete force.

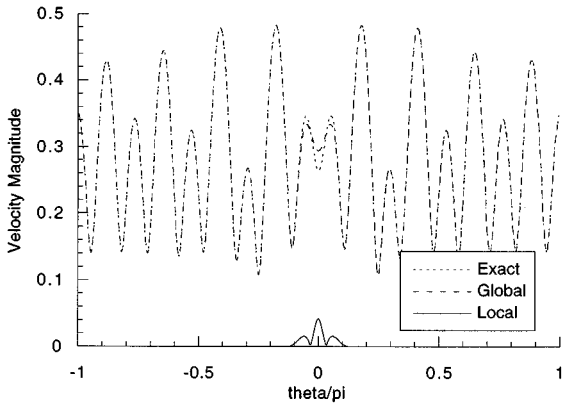


FIG. 6. Radial velocity magnitude, complete, local, and global solutions.

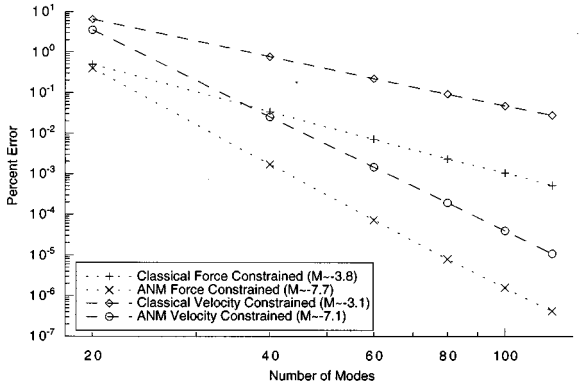


FIG. 8. Convergence of radial velocity, shell driven by discrete radial force.

compared to the original discretely driven solution. This comparison shows that the ANM method separates the shell response into high- and low-resolution parts. The local solution captures the higher-harmonic content, or the high-resolution part of the solution, while the global solution contains the lower-harmonic modes, or low-resolution content. The removal of the dominant high-harmonic content in the global solution allows the higher harmonics of the global solution to fall at a much faster rate, which, in turn, causes the solution to converge more rapidly. Varying the smoothing width changes the point at which the local solution begins to dominate the harmonic content of the composite solution. If the width is decreased, the magnitude of the local solution is decreased and the point of crossover is shifted towards higher harmonic number. The converse is true as well.

The considerations involved in choosing an appropriate smoothing width are as follows. First, if the width is too small, the local solution will not take over until too great a wave number, producing no appreciable advantage over simply solving the method by conventional techniques. Second, if the smoothing width is too large the local solution itself will begin to become global, in that it will begin to influence the lower harmonics of the global solution and alter its overall character, not just smoothing the high-resolution content. Generally, experience indicates that choosing a smoothing width that is about half the length of the dominant structural wave produces good results. It should be emphasized, however, that the ANM composite solution given in the present work is without approximation for any choice of smoothing. The choice of smoothing width, therefore, affects the separation of features, and the rate of convergence, not the converged solution itself.

The prevalence of the local solution in the radially driven case is relatively small compared to the other two cases. The effect of the ANM method is less significant for this forcing configuration. This is due to the stiffness of the shell to bending, which acts as a fairly efficient mechanical filter to the radially applied discrete load. In the moment driven case, the very significant high-harmonic content of the driving force somewhat overcomes the natural filtering effect of the shell stiffness. In addition, the constraint condition on moment is primarily dependent upon slope of radial displacement, which magnifies the prevalence of the local solution. The tangential local solution sees far less stiffness, as the tangential governing equation is only second order. In situations where all three discrete forces are needed to satisfy the constraint conditions, it is the tangential and moment solutions which hinder the convergence of the problem the most. The difficulty these solutions pose to convergence is primarily due to the prominence of high-resolution content at the point of discontinuity. ANM provides the greatest advantage in these sensitive cases.

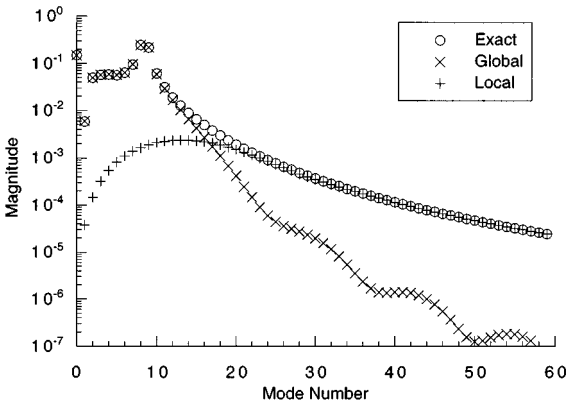


FIG. 7. Harmonic magnitude of radial velocity magnitude, complete, local, and global solutions.

V. COMPARISON OF CLASSICAL AND ANM DRIVEN SOLUTIONS

In order to smooth the influence of a line acting discontinuity, the discrete loads with which the discontinuity is modeled using the formulation in Section III C were each smoothed using the ANM method. In this section, the effectiveness of the ANM treatment of each of the three forcing components of the discontinuity, radial, tangential, and rotational, is examined. The convergence, and therefore the computational efficiency of the method is evaluated by plotting on log–log scales the error in the radial surface velocity averaged in azimuth using an L2 norm, versus the number of modes calculated. The error convergence plots for the three forcings are shown in Figs. 8, 9, and 10, for typical parameter values of interest. The order of convergence is estimated

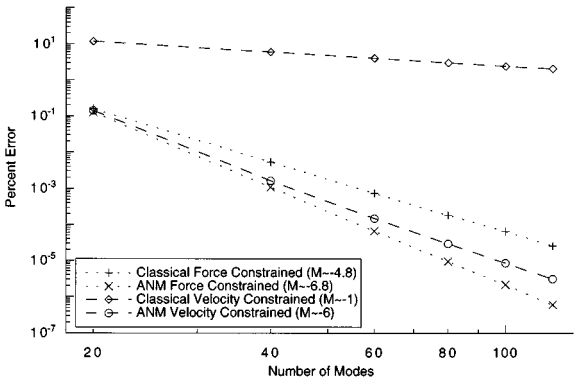


FIG. 9. Convergence of radial velocity, shell driven by discrete tangential force.

by linear regression, the slope of which is indicated on the legend. The radial velocity is chosen because it couples to the fluid, and is therefore of primary interest in scattering problems. Error is measured relative to a solution that is sufficiently converged so that the error of this control solution is far less than the error of the cases plotted. In each of the three forced cases, error convergence is plotted for the ANM composite driven solution and for the classical discretely driven solution for comparison. Two cases of each driven solution are shown as well. In the first case, the strength of the discrete forcing is held constant. In the second case, the response of the structure at the point of application of the discrete forcing is held constant. The constant forcing strength case is perhaps the more obvious case from the standpoint of measuring convergence of the driven solution. However, the constant response case is pertinent with regard to a constrained system because a constraint condition is written in terms of shell response.

The ANM solutions are uniformly more accurate than the classical modal solutions. In addition, the rate of convergence of the ANM solutions is better by several power of modes calculated. The increased rate of convergence is attributed to the smoothed, low-resolution, nature of the ANM global solution, the only part of the ANM solution that must be converged computationally. The increased rate of convergence of the ANM solution is dependent upon the degree of smoothness imposed on the local solution and the resulting global forcing pressure. The tradeoff encountered by increasing this smoothness is that the complexity of the ANM decomposition is also increased. The difference in absolute error between the methods was observed to vary significantly dependent upon the physical parameters of the system.

In general, it was found that the case of constant drive point response is more difficult to converge than the case of constant forcing strength. In the case of constant forcing strength, the first several modes of the solution are evaluated without error regardless of the higher-harmonic content. So, the overall shape of the solution is correct with only a few modes calculated and it is only the local region about the point of application of the discrete load that is slow to converge. In the constant response case, though, the entire solution, even the first several modes, depends directly on the response of the system at the point of application of the discrete load, and therefore, the entire solution is slower to converge. It can be seen that constraining the response of the shell at a point leads to significant error by considering that the constant response case can be constructed by dividing the constant forcing strength case by the response at the point of application of the discrete load. In doing so, a solution that is generally well converged, except for the isolated region of the discontinuous forcing, is traded for the same solution functionally, except that it is scaled so that the poorly converged point is held constant. The result is that the overall solution now displays the higher error that was once restricted to the isolated region of the discontinuity.

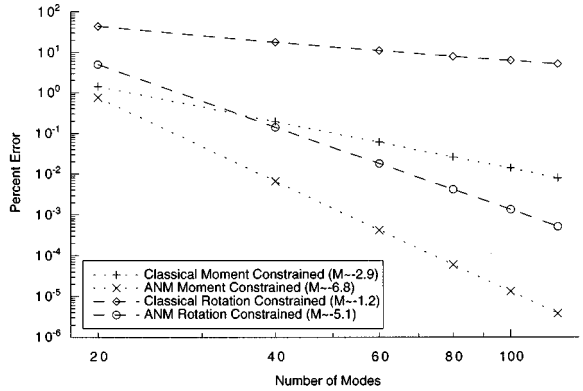


FIG. 10. Convergence of radial velocity, shell driven by discrete moment.

VI. CONSTRAINED SCATTERING SAMPLE PROBLEMS

To illustrate the use of ANM in constrained scattering problems modeled by the formalism presented in Section III C, two sample geometries are presented. The first example is scattering from a thin cylindrical shell that is clamped along four equally spaced lines. This is actually a degenerate case of the modeling formalism presented, but is included because the clamped case is the most stringent constraint condition. The second example is scattering from a shell bisected by a thin elastic plate. This example demonstrates the ability to treat complicated problems in which the components of the constraints, as well as separate constraints themselves, are coupled together through the dynamics of another system. These cases are illustrated pictorially in Fig. 11.

A. Multiply clamped shell

In the first example problem, a plane acoustic wave originating in the $\theta=0$ direction is scattered by a cylindrical shell that is clamped in four places, $\theta=\pm\pi/4$, $\pm 3\pi/4$. All three components of shell response, radial, tangential, and rotational, are required to be zero where clamped. By the formalism presented in Section III C, these constraints are represented by letting the matrix Z be diagonal, with entries that go to infinity. In practice, however, setting the velocity

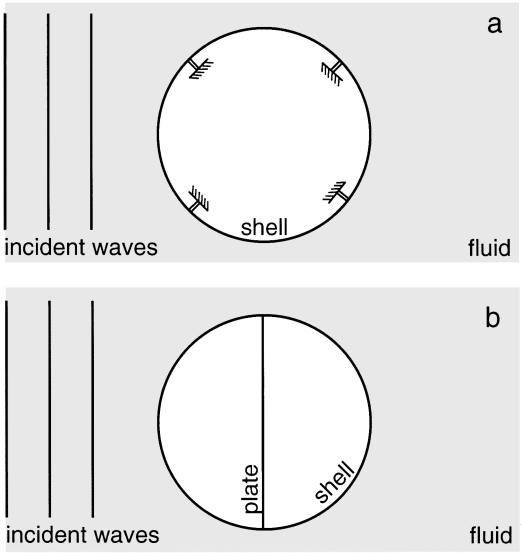


FIG. 11. Pictorial representation of constraint configurations. (a) Clamped at $\pm\pi/4$ and $\pm 3\pi/4$, (b) bisected by elastic plate.

at the constraints, which is represented by the matrix V in Eq. (14), to zero yields an equation that can be simply solved for the forcing magnitudes, F, as follows:

$$\{F\} = -[VI]^{-1}\{V_{\text{unconstrained}}\}. \tag{23}$$

Because the clamped line transmits substantial force to the shell in each of the three discrete driving components, it is expected that this case will be quite difficult to resolve. The convergence of radial surface velocity and the convergence of scattered pressure for this constrained scattering problem is shown in Fig. 12 for typical physical parameters in the mid-frequency range ($Ka=5$) with subsonic flexural waves. The classical modal solution is very slow to converge, revealing significant error even after one hundred modes are calculated. The ANM solution, however, converges rapidly, even for this sensitive problem. The convergence of scattered power is remarkably similar to that of surface velocity, even though only the first several modes are efficient radiators. This is due to the fact that, in this constrained scattering problem, the first few modes are sensitive to the resolution of shell response in the region of the discrete forces. (The

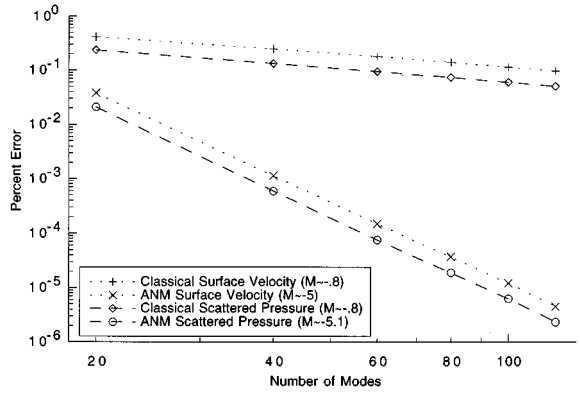


FIG. 12. Convergence of radial velocity and scattered pressure, scattering from shell constrainted at $\pm\pi/4$, $\pm 3\pi/4$.

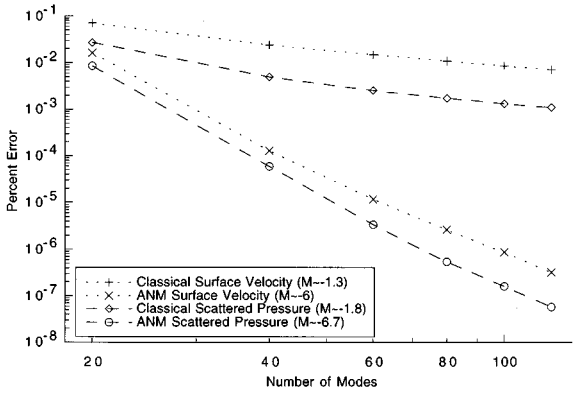


FIG. 13. Convergence of radial velocity and scattered pressure, scattering from shell bisected by a plate connected at $\pm\pi/2$.

strength of the discrete forces themselves are sensitive to the resolution at the point of application of the force.)

B. Bisected shell

The second example problem is the scattering of a plane acoustic wave from a shell that is bisected by a thin elastic plate. The plate is fixed to the shell at $\theta=\pm\pi/2$. The constraint conditions on the shell are determined by solving for the dynamic response of the plate, and assuming compatible displacements and rotation at the plate/shell juncture. The plate, like the shell, is considered to be thin and is modeled by the classical equations. These equations, normalized in the same way as the shell equations, are as follows:

$$\beta^2 \frac{\partial^4 \overline{u}}{\partial \overline{x}^4} + K_p^2 \frac{\partial^2 \overline{u}}{\partial \overline{t}^2} = \overline{p}_y, \tag{24}$$

$$-\frac{\partial^2 \overline{q}}{\partial \overline{x}^2} + K_p^2 \frac{\partial^2 \overline{q}}{\partial \overline{t}^2} = \overline{p}_x. \tag{25}$$

The mid-line of the undisplaced plate lies in the x plane. $\overline{u}$ and $\overline{q}$ are the normalized displacements in the transverse, or y direction, and the longitudinal, or x direction, respectively. The plate is chosen of the same material and thickness as the shell, which leads to identical parameters in the normalized equations. The plate equations have exact solutions in terms of trigonometric, and hyperbolic trigonometric functions.

The entries of the Z matrix are found by allowing only one boundary harmonic motion of unit magnitude (a transverse or longitudinal displacement or a rotation at one end), and solving for the reactions at the boundaries (six quantities). Doing so for each boundary deflection in turn provides all 36 entries of the matrix Z. Special care must be taken in the degenerate case of anti-resonances in the plate. An anti-resonance effectively produces a rigid link between the attachment points on the shell. Also, in the case of a plate resonance no load is applied to the shell. This constraint relationship exhibits coupling not only between radial and tangential influences at a point of attachment, but also between the two points of attachment. The convergence of radial velocity and scattered pressure for this problem is shown in Fig. 13 for typical physical parameters. Again it is found

that the entirety of this constrained scattering problem is sensitive to the modeling resolution at the discrete constraint line.

VII. SUMMARY AND CONCLUSIONS

A novel approach for improving the computational efficiency of fully coupled fluid-structure interaction problems involving structural inhomogeneities has been demonstrated on physically relevant model problems. This application of the analytical/numerical matching methodology is an embodiment of a general approach to problems that require numerical approximation, and contain computationally taxing regions of rapid variation. In engineering problems, complex geometry often requires the use of numerical solution methods. In addition, another common feature of engineering problems is the presence of discontinuities which are very difficult to resolve computationally. Although the overall problem may need to be approached computationally, often the most difficult part of the problem to resolve, that associated with the discontinuity, may be treated separately, and analytically. This situation is the motivation for using analytical numerical matching to improve the efficiency of computational problems that involve discontinuities. If the most computationally taxing part of the problem can be treated separately and analytically, the entire problem can be solved far more efficiently and a clearer understanding of the difficulty associated with the discontinuity can be gained.

Analytical numerical matching for structural acoustics isolates the high-resolution influence of a structural discontinuity by attacking the fundamental forced problems, with which the influence of the discontinuity can be modeled. Each fundamental, discretely forced problem is separated into local and global solutions. The local solution is generated by assuming a general form for the local displacement of the structure, and then satisfying a number of constraints in order to guarantee that the local solution contains the characteristic rapidly varying part of the solution and as little else as is possible. Once the local solution is known, the loading that is required to enforce it is found by direct substitution into the governing equations of the structure, and the smooth forcing that will generate the global solution is determined.

Analytical numerical matching has been successfully applied to the problem of planar vibrations of a circular cylindrical shell immersed in fluid with a general configuration of line constraints and a general forcing. Specifically, example problems of scattering from the cylindrical shell have been worked for several constraint conditions. In each case, modeling the influence of the constraints with analytical numerical matching produced a significant increase in computational efficiency. When the problem is formulated with ANM rather than classical methods, errors are uniformly less by one or more orders of magnitude for a given numerical resolution (given number of modes calculated), and the error diminishes at a significantly faster rate as resolution is increased, by approximately a power of three in mode number.

It is well known in acoustics that the far field radiation from a vibrating structure is influenced almost entirely by the low-resolution content of the solution. Supersonic structural waves, those with wavelengths longer than the acoustic wavelength, are the primary radiators. This would seem to imply that the higher-resolution content of the structural problem is unimportant to the far field. However, in situations where a discontinuity places a constraint on the motion of the structure, it is found that the accuracy with which the response at the point of this discontinuity is modeled affects the accuracy of the entire solution. The reason is that the lower harmonics of such a constrained problem cannot be converged until the effect of the constraint is well converged, which is high resolution in nature. Therefore, even the far field scattering is sensitive to the convergence issues that plague the local region of a structural discontinuity. This fact reiterates the need to model the influence of a discontinuity correctly.

ACKNOWLEDGMENTS

This work was sponsored by the Office of Naval Research. The authors would especially like to thank Dr. L. Couchman, the technical monitor.

APPENDIX

The polynomial description of local displacement in the upper half of the smoothing region for each of the components (radial, tangential, and rotational) of a longitudinal line acting load are given in the following.

The smooth global forcing pressure due to structural terms which complements these local displacements are also given. Note that the fluid pressure feedback due to local radial displacement must be accounted for in the global problem.

1. Radial force

$$\bar{w}_l=\frac{\theta_s^3 F}{\beta^2 h\rho_s c_p^2}\left[\frac{5}{16128}-\frac{9}{448}\,\tilde{\theta}^{\,2}+\frac{1}{12}\,\tilde{\theta}^{\,3}-\frac{15}{128}\,\tilde{\theta}^{\,4}+\frac{35}{192}\,\tilde{\theta}^{\,6}-\frac{3}{14}\,\tilde{\theta}^{\,7}+\frac{27}{256}\,\tilde{\theta}^{\,8}-\frac{5}{252}\,\tilde{\theta}^{\,9}\right], \tag{A1}$$

$$\tilde{v}_l=0, \tag{A2}$$

$$\bar{p}_r=\frac{F}{\theta_s h\rho_s c_p^2}\left[\frac{45}{16}-\frac{525}{8}\,\tilde{\theta}^2+180\tilde{\theta}^3-\frac{2835}{16}\,\tilde{\theta}^4+60\tilde{\theta}^5+\frac{(1-K_p^2)\,\theta_s^4}{\beta^2}\left(-\frac{5}{16128}+\frac{9}{448}\,\tilde{\theta}^2-\frac{1}{12}\,\tilde{\theta}^3+\frac{15}{128}\,\tilde{\theta}^4-\frac{35}{192}\,\tilde{\theta}^6+\frac{3}{14}\,\tilde{\theta}^7-\frac{27}{256}\,\tilde{\theta}^8+\frac{5}{252}\,\tilde{\theta}^9\right)\right], \tag{A3}$$

$$\bar{p}_\theta=\frac{F\theta_s}{\beta^2 h\rho_s c_p^2}\left[-\frac{9}{224}\,\tilde{\theta}+\frac{1}{4}\,\tilde{\theta}^2-\frac{15}{32}\,\tilde{\theta}^3+\frac{35}{32}\,\tilde{\theta}^5-\frac{3}{2}\,\tilde{\theta}^6+\frac{27}{32}\,\tilde{\theta}^7-\frac{5}{28}\,\tilde{\theta}^8\right]. \tag{A4}$$

2. Tangential force

$$\overline{w_l}=\frac{\theta_s^4 F}{\beta^2 h\rho_s c_p^2}\left[\frac{1}{3360}\,\tilde{\theta}-\frac{1}{160}\,\tilde{\theta}^3+\frac{1}{48}\,\tilde{\theta}^4-\frac{1}{32}\,\tilde{\theta}^5+\frac{1}{40}\,\tilde{\theta}^6-\frac{1}{96}\,\tilde{\theta}^7+\frac{1}{560}\,\tilde{\theta}^8\right], \tag{A5}$$

$$\overline{v_l}=\frac{\theta_s F}{h\rho_s c_p^2}\left[\frac{3}{20}-\frac{1}{2}\,\tilde{\theta}+\frac{1}{2}\,\tilde{\theta}^2-\frac{1}{4}\,\tilde{\theta}^4+\frac{1}{10}\,\tilde{\theta}^5+K_p^2\theta_s^2\left(\frac{1}{240}-\frac{1}{24}\,\tilde{\theta}^2+\frac{1}{12}\,\tilde{\theta}^3-\frac{1}{16}\,\tilde{\theta}^4+\frac{1}{60}\,\tilde{\theta}^6\right)\right], \tag{A6}$$

$$\overline{p_r}=\frac{F}{h\rho_s c_p^2}\left[\frac{11}{40}\,\tilde{\theta}-9\tilde{\theta}^2+\frac{39}{4}\,\tilde{\theta}^3-\frac{7}{2}\,\tilde{\theta}^4+\theta_s^2K_p^2\times\left(\frac{1}{12}\,\tilde{\theta}-\frac{1}{4}\,\tilde{\theta}^2+\frac{1}{4}\,\tilde{\theta}^3-\frac{1}{12}\,\tilde{\theta}^4\right)\frac{\theta_s^4}{\beta^2}\,(1-K_p^2)\times\left(-\frac{1}{3360}\,\tilde{\theta}+\frac{1}{160}\,\tilde{\theta}^3-\frac{1}{48}\,\tilde{\theta}^4+\frac{1}{32}\,\tilde{\theta}^5-\frac{1}{40}\,\tilde{\theta}^6+\frac{1}{96}\,\tilde{\theta}^7-\frac{1}{560}\,\tilde{\theta}^8\right)\right], \tag{A7}$$

$$\overline{p}_\theta=\frac{F}{\theta_s h\rho_s c_p^2}\left[1-3\tilde{\theta}^2+2\tilde{\theta}^3+\theta_s^2K_p^2\left(\frac{1}{15}-\frac{1}{4}\,\tilde{\theta}^2+\frac{1}{3}\,\tilde{\theta}^3-\frac{1}{4}\,\tilde{\theta}^4+\frac{1}{10}\,\tilde{\theta}^5\right)\frac{\theta_s^4}{\beta^2}\left(\frac{1}{3360}-\frac{3}{160}\,\tilde{\theta}^2+\frac{1}{12}\,\tilde{\theta}^3-\frac{5}{32}\,\tilde{\theta}^4+\frac{3}{20}\,\tilde{\theta}^5-\frac{7}{96}\,\tilde{\theta}^6+\frac{1}{70}\,\tilde{\theta}^7\right)\theta_s^4K_p^4\times\left(\frac{1}{240}-\frac{1}{24}\,\tilde{\theta}^2+\frac{1}{12}\,\tilde{\theta}^3-\frac{1}{16}\,\tilde{\theta}^4+\frac{1}{60}\,\tilde{\theta}^5\right)\right]. \tag{A8}$$

3. Moment

$$\overline{w_l}=\frac{\theta_s^2 M/a}{\beta^2 h\rho_s c_p^2}\left[-\frac{3}{56}\,\tilde{\theta}+\frac{1}{4}\,\tilde{\theta}^2-\frac{3}{8}\,\tilde{\theta}^3+\frac{5}{8}\,\tilde{\theta}^5-\frac{3}{4}\,\tilde{\theta}^6+\frac{3}{8}\,\tilde{\theta}^7-\frac{1}{14}\,\tilde{\theta}^8\right], \tag{A9}$$

$$\overline{v_l}=\frac{\theta_s^3 M/a}{\beta^2 h\rho_s c_p^2}\left[-\frac{1}{240}+\frac{1}{24}\,\tilde{\theta}^2-\frac{1}{12}\,\tilde{\theta}^3+\frac{1}{16}\,\tilde{\theta}^4-\frac{1}{60}\,\tilde{\theta}^5\right], \tag{A10}$$

$$\overline{p_r}=\frac{M/a}{\theta_s^2 h\rho_s c_p^2}\left[-75\tilde{\theta}+270\tilde{\theta}^2-315\tilde{\theta}^3+120\tilde{\theta}^4+\frac{\theta_s^4}{\beta^2}\left(-\frac{1}{12}\,\tilde{\theta}+\frac{1}{4}\,\tilde{\theta}^2-\frac{1}{4}\,\tilde{\theta}^3+\frac{1}{12}\,\tilde{\theta}^4\right)\times\left(\frac{3}{56}\,\tilde{\theta}-\frac{1}{4}\,\tilde{\theta}^2+\frac{3}{8}\,\tilde{\theta}^3-\frac{5}{8}\,\tilde{\theta}^5+\frac{3}{4}\,\tilde{\theta}^6-\frac{3}{8}\,\tilde{\theta}^7+\frac{1}{14}\,\tilde{\theta}^8\right)\right], \tag{A11}$$

$$\overline{p}_\theta=\frac{\theta_s M/a}{\beta^2 h\rho_s c_p^2}\left[\frac{5}{168}-\frac{3}{8}\,\tilde{\theta}^2-\frac{1}{3}\,\tilde{\theta}^3+\frac{25}{8}\,\tilde{\theta}^4-\frac{9}{2}\,\tilde{\theta}^5+\frac{21}{8}\,\tilde{\theta}^6-\frac{4}{7}\,\tilde{\theta}^7+\theta_s^2K_p^2\left(-\frac{1}{240}+\frac{1}{24}\,\tilde{\theta}^2-\frac{1}{12}\,\tilde{\theta}^3+\frac{1}{16}\,\tilde{\theta}^5-\frac{1}{60}\,\tilde{\theta}^6\right)\right]. \tag{A12}$$

[1] M. C. Junger, ''Sound scattering by thin elastic shells,'' J. Acoust. Soc. Am. **24**, 366–373 (1952).
[2] M. C. Junger, ''The concept of radiation scattering and its application to reinforced cylindrical shells,'' J. Acoust. Soc. Am. **25**, 899–903 (1952).
[3] A. Klauson and J. Metsaveer, ''Sound scattering by a cylindrical shell reinforced by lengthwise ribs and walls,'' J. Acoust. Soc. Am. **91**, 1834–1843 (1992).
[4] J. M. Cushieri and D. Feit, ''A hybrid solution for the response Green's function of a fluid-loaded cylindrical shell,'' J. Acoust. Soc. Am. **96**, 2776–2784 (1994).
[5] J. L. Humar, *Dynamics of Structures* (Prentice-Hall, Engelwood Cliffs, NJ, 1968), pp. 567–572.
[6] R. L. Blisplinghoff and H. Ashley, *Principles of Aeroelasticity* (Dover, New York, 1962), pp. 350, 471.
[7] D. B. Bliss and W. O. Miller, ''Efficient free wake calculations using analytical/numerical matching,'' J. Am. Helicopter Soc. **38**, 870–879 (1993).
[8] W. O. Miller, ''Analytical/numerical matching and periodic inversion: Two advances in free wake analysis,'' Doctoral thesis, Duke University, 1990.
[9] T. R. Quackenbush and D. B. Bliss, ''Free wake calculation of rotor flow fields for interactional aerodynamics,'' Vertica **14**, 313–327 (1990).
[10] T. R. Quackenbush, C-M. G. Lam, and D. B. Bliss, ''Vortex methods for computational analysis of rotor/body interaction,'' J. Am. Helicopter Soc. **39**, 14–24 (1994).
[11] D. B. Bliss and R. J. Epstein, ''Free vortex problems using analytical/numerical matching with solution pyramiding,'' AIAA J. **33**, 894–903 (1995).
[12] D. B. Bliss and R. J. Epstein, ''A novel approach to aerodynamic analysis using analytical/numerical matching,'' AIAA J. **34**, 2225–2232 (1996).
[13] R. J. Epstein and D. B. Bliss, ''An aeroacoustic boundary element method using analytical/numerical matching,'' *Proceedings of the First Joint CEAS/AIAA Aeroacoustics Conference*, Munich, Germany, June, 1995 (German Society for Aeronautics and Astronics, 1995).
[14] R. J. Epstein and D. B. Bliss, ''An acoustic boundary element method using analytical/numerical matching,'' J. Acoust. Soc. Am. **101**, 92–106 (1997).
[15] R. C. Loftman, ''Analytical/numerical matching for structural acoustics,'' Masters thesis, Duke University, 1994.
[16] D. B. Bliss and R. C. Loftman, ''Local and global effects in structural-acoustic scattering using analytical/numerical matching,'' *Proceedings of the 15th International Congress on Acoustics*, Trondheim, Norway, June, 1995 (Acoustical Society of Norway, 1995).
[17] R. C. Loftman and D. B. Bliss, ''The application of analytical numerical matching to structural discontinuities in structural acoustic problems,'' J. Acoust. Soc. Am. **101**, 925–935 (1997).
[18] M. C. Junger and D. Feit, *Sound, Structures, and Their Interaction* (Massachusetts Institute of Technology, Cambridge, MA, 1972).
[19] H. Kraus, *Thin Elastic Shells* (Wiley, New York, 1967).
[20] M. Abromowitz and I. Stegun, *Handbook of Mathematical Functions* (U.S. Bureau of Standards, Washington, DC, 1964).

Use of acoustic basis functions for active control of sound power radiated from a cylindrical shell

Koorosh Naghshineh
Mechanical and Aeronautical Engineering Department, Western Michigan University, Kalamazoo, Michigan 49008

Weicheng Chen and Gary H. Koopmann
Center for Acoustics and Vibration, The Pennsylvania State University, 157 Hammond Building, University Park, Pennsylvania 16802

(Received 23 June 1997; accepted for publication 20 January 1998)

In a previous paper by the authors [J. Acoust. Soc. Am. **92**, 856–870 (1992)], a general strategy was devised for achieving minimum radiation of sound from structures subjected to a harmonic excitation force. In this strategy, a quadratic expression is written for the total sound power radiated from the structure in terms of the primary and actuator (control) forces. Using quadratic optimization of this expression, a single control force vector is found which minimizes the radiated sound power for the particular acoustic model of the structure. The structural response to this control force is referred to as the ''weak radiator'' response and is then used as an objective function for an adaptive algorithm. In previous work, this strategy was used to control the sound radiated from a baffled beam. The work presented in this paper extends the application of this strategy to the case of a cylindrical shell with endcaps. Numerical predictions were verified experimentally. These results show that this strategy is successful in controlling the sound radiated from the cylindrical shell by yielding large reductions in the structural supersonic wave number content. © *1998 Acoustical Society of America.* [S0001-4966(98)04204-0]

PACS numbers: 43.40.Vn [PJR]

INTRODUCTION

Sound attenuation is usually achieved by means of energy absorbing materials placed on the surface of the radiating structure. This technique works quite well for sounds of medium and high frequencies but is very inefficient at low frequencies because the thickness of the absorbing material necessarily increases with decreasing frequency. Active noise control is an alternative to this ''passive'' strategy where destructive interference is used to reduce the sound pressure field. For comprehensive references on active noise control see excellent reviews given in Refs. 1–8.

A more advanced form of this technique is based on altering the vibrations of a noisy structure such that it radiates very little sound. This alteration is achieved by introducing discrete force actuators at select points on the structure. These actuators are driven via a digital controller that receives the noise or vibration signal from a sensor array located on the surface of the structure. This technique, referred to as ''Active Structural Acoustic Control'' (ASAC), has been a major topic of research in recent years. An initial paper documenting application of such an approach on a panel was by Vyalyshev *et al.*[9] More recent examples are given by Naghshineh *et al.*,[10–12] by Clark *et al.*,[13] and many others.

Previous work by the authors[10,11] has focused on the control of total sound power radiated from vibrating structures via ASAC. This was achieved by quadratic optimization of an expression written for total radiated sound power from any structure. A single control force vector was found that, in combination with the primary force vector, produced a structural response that reduced sound. This structural response was referred to as the ''weak radiator'' response. In Ref. 10, quadratic optimization was carried out on the complete power expression, whereas in Ref. 11 the power expression was approximated by a truncated series sum of ''surface velocity basis functions.'' The examples provided in Refs. 10 and 11 were of control of sound power radiated from a baffled beam. The focus of the current paper is to extend the previous work by its application to the case of a cylindrical shell with endcaps in air. Several other authors have also proposed the use of this technique in active control applications; of these, the two most notable are Borgiotti *et al.*[14] and Elliott *et al.*[15]

A brief overview of the theory presented in previous work will be presented in the next section. This will be followed by the methodology used in modeling the radiation of sound from a cylindrical shell. Next, a numerical example of the application of the above strategies along with experimental verifications of the predicted results will be presented.

I. THEORETICAL DEVELOPMENT

Both of the methods introduced previously are based on minimization of a quadratic expression for total sound power radiated from a structure. The power radiated from any vibrating structure can be written in terms of its discretized normal surface velocity vector $\mathbf{v}$. This was done by Cunefare *et al.*[16,17] and Mollo *et al.*[18,19] for three-dimensional radiators. The expression derived by these authors was

$$W = \tfrac{1}{2}\,\mathbf{v}^H \mathfrak{R}\mathbf{v}, \tag{1}$$

 0001-4966/98/103(4)/1897/7/$10.00

where the superscript H refers to Hermitian (transpose conjugate) operation. The matrix operator $\mathfrak{R}$, the acoustic resistance, is positive definite and Hermitian.

A. Minimization of complete power expression

The response of any structure due to harmonic excitation forces can be written in terms of the structural mobility matrix $\mathbf{Y}$ such that

$$\mathbf{v}=\mathbf{Y}\mathbf{f}, \tag{2}$$

where $\mathbf{f}$ is the external point-force vector exciting the structure in the direction normal to the structural surface (includes the primary and the control excitation forces). Although in practice many of the elements of the external force vector $\mathbf{f}$ will be zero, the physical location of these elements will be the same as the corresponding elements of the surface velocity vector $\mathbf{v}$. The mobility matrix of the above equation is written in terms of the structural modal response as

$$\mathbf{Y}\cong\left(\sum_{j=1}^{K}\frac{i\omega\phi_j\phi_j^T}{\omega_j^2-\omega^2+i\eta_j\omega_j^2}\right), \tag{3}$$

where the subscript j refers to the jth mode, ϕ_j is the mass normalized modal displacement vector in the direction perpendicular to the surface, ω_j is the jth natural frequency, η_j is the damping loss factor associated with the jth mode, and $i=\sqrt{-1}$. The harmonic excitation frequency is ω. The mobility matrix $\mathbf{Y}$ is a complex, symmetric matrix. For a complex structure ϕ_j, ω_j, and η_j can be found experimentally via a modal analysis, or numerically by performing a finite element analysis of the structure. Since the experiments presented in this paper were only concerned with active control of total radiated sound power from a cylinder in air, all fluid-loading effects have been ignored. The mass normalized modal displacement vectors, ϕ_j, represent the structural eigenvectors in air. All in-plane displacements have been ignored.

Substituting Eq. (2) in (1) yields the following quadratic expression for the radiated sound power from an arbitrary structure:

$$W=\tfrac{1}{2}\,\mathbf{f}^H\mathbf{D}\mathbf{f}. \tag{4}$$

The matrix $\mathbf{D}$ in the above equation is Hermitian ($\mathbf{D}=\mathbf{Y}^H\mathfrak{R}\mathbf{Y}$) and positive definite. Equation (4) represents an approximation, only valid for light fluid loading. The derivation of the above equation including the fluid-loading effect is beyond the scope of this paper and will not be discussed here. To determine the actuator (control) force vector that yields minimum sound power, the overall force vector, $\mathbf{f}$, is written as the sum of primary ($\mathbf{f}_p$) and control ($\mathbf{f}_c$) force vectors as

$$\mathbf{f}=\mathbf{f}_p+\mathbf{f}_c\,. \tag{5}$$

Using the above equation in Eq. (4) and by setting $\partial W/\partial\mathbf{f}_c=0$ (see Ref. 10 for more details), a set of actuator forces ($\mathbf{f}_c$) are found that result in minimum radiated sound power. To avoid the trivial solution, primary forces and the unknown control forces are chosen such that they are not coincident. It is important to note that these actuator control forces yield minimum total radiated sound power only for the specific acoustic model of the structure. Different acoustic models may result in slightly different control actuator forces. However, these are expected to be within the same range and for all practical purposes the same in their implementation on the actual structure.

B. Minimization of the truncated power expression

Since the M-by-M acoustic resistance matrix $\mathfrak{R}$ is Hermitian, it can be expanded[11] in terms of its M eigenvalues λ_i and eigenvectors $\mathbf{q}_i$ as

$$\mathfrak{R}=\sum_{i=1}^{M}\lambda_i\mathbf{q}_i\mathbf{q}_i^H, \tag{6}$$

where these eigenvalues and the eigenvectors of the matrix $\mathfrak{R}$ are found from the following equation:

$$\mathfrak{R}Q=Q\mathbf{\Lambda}. \tag{7}$$

In the above equation, $\mathbf{Q}$ is an M-by-M matrix $\mathbf{Q}=[\mathbf{q}_1,\mathbf{q}_2,\mathbf{q}_3,\ldots,\mathbf{q}_M]$ and $\mathbf{\Lambda}$ is an M-by-M diagonal matrix $\mathbf{\Lambda}=\mathrm{diag}(\lambda_1,\lambda_2,\lambda_3,\ldots,\lambda_M)$. Since the matrix $\mathfrak{R}$ is Hermitian, all of its eigenvalues are real and its eigenvectors are orthogonal[11] ($\mathbf{Q}^H\mathbf{Q}=I$, where I is the identity matrix). Thus Eq. (7) can be rewritten as

$$\mathfrak{R}=\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}=\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H. \tag{8}$$

Substituting Eq. (8) in Eq. (4) yields

$$W=\tfrac{1}{2}\,\mathbf{f}^H[(\mathbf{Q}^H\mathbf{Y})^H\ \ \mathbf{\Lambda}(\mathbf{Q}^H\mathbf{Y})]\mathbf{f}, \tag{9}$$

where we have previously shown[11] that since $\mathbf{v}=\mathbf{Y}\mathbf{f}$, the $\mathbf{Q}^H\mathbf{Y}\mathbf{f}$ terms represent the coupling between the structural surface velocity vector with the acoustic basis functions $\mathbf{Q}$. Because of their unique filtering effect, these basis functions have been called the ''velocity filtering functions'' by Borgiotti.[20] The product $\mathbf{Q}^H\mathbf{Y}\mathbf{f}=\mathbf{Q}^H\mathbf{v}$ is weighted by $\mathbf{\Lambda}$, the eigenvalues of the acoustic resistance matrix $\mathfrak{R}$, which we have called the acoustic weighting coefficients. These weighting coefficients range from large to very small. A structure that exhibits a high radiation efficiency is one that exhibits a surface velocity that strongly couples with the acoustic basis functions $\mathbf{Q}$, corresponding to large weighting coefficients in the $\mathbf{\Lambda}$ matrix. Expanding the structural surface velocity vector $\mathbf{v}$ in terms of the acoustic basis functions $\mathbf{Q}$, we can write

$$W=\tfrac{1}{2}\,\lambda^T\mathbf{a}^H\mathbf{a}=\tfrac{1}{2}\,\lambda^T|\mathbf{a}|^2, \tag{10}$$

where $\mathbf{a}=\mathbf{Q}^H\mathbf{v}$ represents a set of velocity-to-basis function coupling coefficients and λ is a vector of the acoustic weighting coefficients. It has been shown[11,20] that only few of these acoustic basis functions will be significant contributors to the total sound power radiated from the structure. The contribution of the rest of these acoustic basis functions to the total radiated sound power will be negligible. Therefore, Eq. (10) can be written[20] in terms of the following truncated summation:

$$W=\frac{1}{2}\,\lambda^T|\mathbf{a}|^2\cong\frac{1}{2}\sum_{i=1}^{N}\lambda_i|a_i|^2, \tag{11}$$

where $N<M$ (recall $\mathfrak{R}$ was an M-by-M matrix). Thus to obtain an accurate estimate of the total radiated acoustic power, one needs to compute the first N coefficients a_i. Another representation of this approximation is to rewrite Eq. (6) as

$$\mathfrak{R} \cong \sum_{i=1}^{N} \lambda_i \mathbf{q}_i \mathbf{q}_i^H, \tag{12}$$

before substitution in Eq. (4). Using a limited number of terms in the above summation, the power minimization carried out in the previous section can be performed in a more computationally efficient manner. This also presents numerous advantages in the experimental implementation of power minimization.[11]

II. IMPLEMENTATION ON A CYLINDRICAL SHELL

To implement the above strategy on a cylindrical shell with endcaps, initially it was necessary to compute the radiation resistance matrix $\mathfrak{R}$. This was done using the wave superposition method. The structural mobility matrix Y was determined using an experimental modal analysis of the subject cylindrical shell. The control strategies mentioned above were then utilized to determine the actuator forces required to minimize the total radiated acoustic power from this cylinder. Finally, these results were experimentally verified by comparison to the predicted ''controlled'' cylinder structural velocity. Each of the above steps will be described in following sections.

A. Cylinder acoustic model

The wave superposition method as presented by Koopmann *et al.*[21] and Song *et al.*[22] is very similar in its numerical formulation to the boundary element method. However, while the boundary element method replaces the radiator surface with an infinite distribution of monopoles and dipoles, the wave superposition method places a relatively small number of monopole sources within the volume enclosed by the radiating surface (interior points). The unknown strengths of these monopole sources are then determined such that the prescribed structural surface normal velocity is reproduced at discrete surface points. Once the source strengths have been determined, the pressure radiated from the structure is computed by superimposing the pressures due to each of these interior monopole sources. Using this formulation, the total power radiated from any structure can be written in form of Eq. (1).

For a harmonically vibrating structure, the velocity amplitude at any point on the surface is written in terms of the interior sources as

$$v(\mathbf{r}_s) = \int_V s(\mathbf{r}_0) \frac{\partial}{\partial \mathbf{n}} G(\mathbf{r}_s, \mathbf{r}_0) dV, \tag{13}$$

where $\mathbf{r}_s$ is a position vector pointing to the surface from the origin, $\mathbf{r}_0$ is a position vector pointing to a source, $\mathbf{n}$ is the outward normal to the surface at $\mathbf{r}_0$, s is the complex source strength of the monopole source at point $\mathbf{r}_0$ inside the volume V, and $G(\mathbf{r}_s, \mathbf{r}_0)$ is the Green's function defined as $G(\mathbf{r}_s, \mathbf{r}_0) = -e^{ik|r_s - r_0|}/4\pi|\mathbf{r}_s - \mathbf{r}_0|$, where k is the acoustic wave number. Equation (13) can be approximated in discrete form in terms of the structural surface velocity vector $\mathbf{v}$ and the monopole source strengths vector $\mathbf{s}$ as

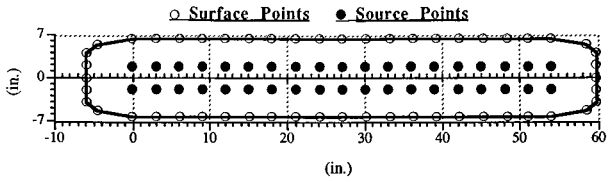


FIG. 1. Cross-sectional view of the aluminum cylindrical shell.

$$\mathbf{v} = \mathbf{E}\mathbf{s}, \tag{14}$$

where

$$E_{ij} = -\frac{1}{4} \frac{ik|\mathbf{r}_j - \mathbf{r}_i| - 1}{|\mathbf{r}_j - \mathbf{r}_i|} e^{ik|r_j - r_i|} \cos\ \theta_{ij}. \tag{15}$$

In the above equation, the subscript j refers to the surface point, subscript i refers to the monopole interior source, and θ_{ij} is the angle between the resultant vector $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ and the surface outward normal at point j. Using Eq. (14) the source strengths can be computed once a velocity vector is given ($\mathbf{s} = \mathbf{E}^{-1}\mathbf{v}$). These source strengths can then be used to compute the acoustic pressure at any point on the surface of the structure using the following equation

$$p(\mathbf{r}) = i\omega\rho \sum_{i=1}^{l} s(\mathbf{r}_i) G(\mathbf{r}_j, \mathbf{r}_i) = \mathbf{M}\mathbf{s}, \tag{16}$$

where

$$M_{ij} = -i\omega\rho\ e^{ik|r_j - r_i|}/4\pi|\mathbf{r}_j - \mathbf{r}_i|. \tag{17}$$

Using the above equations, the total power radiated from any structure can be written in form of the quadratic expression of Eq. (1) by setting the acoustic resistance matrix

$$\mathfrak{R} = 2(\mathbf{E}^{-1})^T \mathbf{R} (\mathbf{E}^{-1})^*, \tag{18}$$

where $\mathbf{E}$ is defined above and

$$R_{ij} = \frac{k\omega\rho}{8\pi} \frac{\sin(k|\mathbf{r}_i - \mathbf{r}_j|)}{k|\mathbf{r}_i - \mathbf{r}_j|}, \tag{19}$$

where the i and j subscripts refer to the ith and the jth monopole source point locations.

The wave superposition method was applied to the case of a cylindrical shell made of an aluminum alloy with heavy endcaps. Measurements along the cylindrical section of this shell showed the outer radius of the shell to be 162 mm (6.375 in.) and its thickness 5 mm (0.1955 in.). The length of the cylindrical section was 1448 mm (57 in.) whereas the total shell length (including the endcaps) was 1670 mm (65.75 in.). Since the endcaps were not hemispherical, careful measurements of the shell profile had to be conducted at the two endcaps before this work could proceed. A representation of the shell profile is shown in Fig. 1. A frequency of 666 Hz was chosen for all analysis presented in this paper since this frequency was coincident with the 2,3 mode of the cylinder (2 half-waves along the axial direction, 3 full waves along the circumferential direction). The shell surface was

TABLE I. Experimental modal analysis results. Mode numbers M,N correspond to N half-axial waves and M full circumferential waves.

Mode number, M,N	Natural frequency, $f_j=\omega_j/2\pi$	Damping loss factor, η_j
1,2	284.2	0.0086
1,3	461.8	0.0058
2,3	666	0.0113
1,4	828.7	0.0017
2,4	915.5	0.0045
3,3	1010	0.0221
3,4	1110	0.0101
1,5	1320	0.0036
2,5	1380	0.0026
3,5	1490	0.0053

divided into 19 axial and 24 circumferential slices. Thus a total of 456 surface points were used along the cylindrical section of this shell. In addition, each endcaps surface was modeled using 145 points (24 circumferential slices of 6 points each plus a single point at the tip of the endcap). Only 456 interior points were required to model the shell velocity along its cylindrical section and the endcaps accurately. As noted by Koopmann *et al.*[21] these interior source points were placed such that the prescribed structural surface velocity could be reproduced identically at specified points on the surface. The interior source points were placed at 48.6 mm (1.9125 in.) away from the cylinder center line. Figure 1 shows an illustration of a single circumferential slice of this shell along with the source and the surface point locations. Since the shell endcaps were very heavy and exhibited very small velocities, no interior source points were required for the endcaps.

B. Cylinder mobility

An experimental modal analysis of the above shell was performed by Sommerlatte.[23] The shell natural frequencies, their corresponding damping loss factors, and mode shapes were found. Detailed analysis of the shell mode shapes showed that the shell mode shapes were essentially those of a cylindrical shell with simply supported ends. This is due to the high impedance discontinuity between the thin cylindrical section and the heavy (massive) endcaps. Table I gives a summary of these results. Using the experimentally obtained structural information in Eq. (3), the shell mobility matrix $\mathbf{Y}$ was computed at 666 Hz.

Given the shell mobility matrix $\mathbf{Y}$, the shell response to a point harmonic force (primary actuator) located at a single point on top of the shell at the sixth axial slice along the shell cylindrical section was computed. This response (shown in terms of displacement vector$=\mathbf{Yf}/i\omega$) is illustrated in Fig. 2 for a 4.44 N (1 lb) harmonic excitation at 666 Hz. Since the shell endcaps were for all practical purposes stationary, their displacement/velocity is assumed to be zero and therefore not included in this plot.

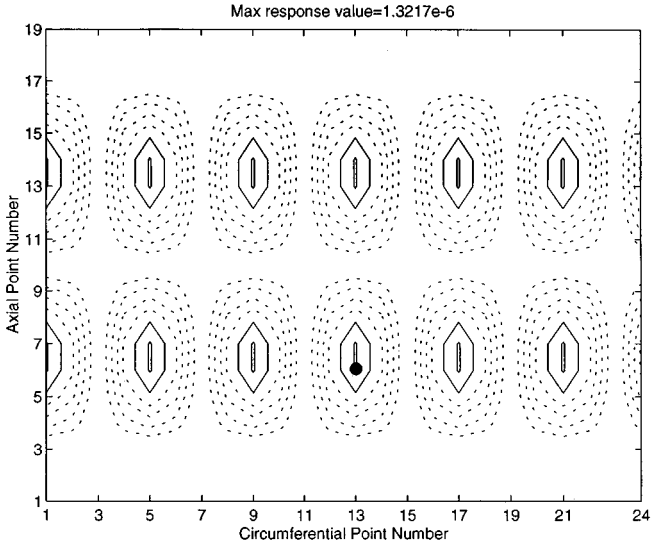


FIG. 2. Vibration response of the shell due to a primary force of 4.44 N (located at ●) at 666 Hz. The solid lines represent the high points of the response.

C. Control via minimization of complete power expression

Using the strategy described in Sec. I A the control actuator force that minimized the radiated acoustic power from this shell was found. For the purpose of this study, the control actuator was assumed to be a point force located along the same axial line as the primary actuator described in the previous section, and located at the fourteenth axial slice of the cylindrical section. Thus the primary and the control force actuators are symmetric with respect to a plane dividing the shell in two halves in the axial direction. Because of the symmetrical nature of the dominant modes at this frequency, the control actuator force was expected to be nearly equal to the primary actuator force. For a primary force of 4.44 N (1 lb), the control force was computed at 4.35 N

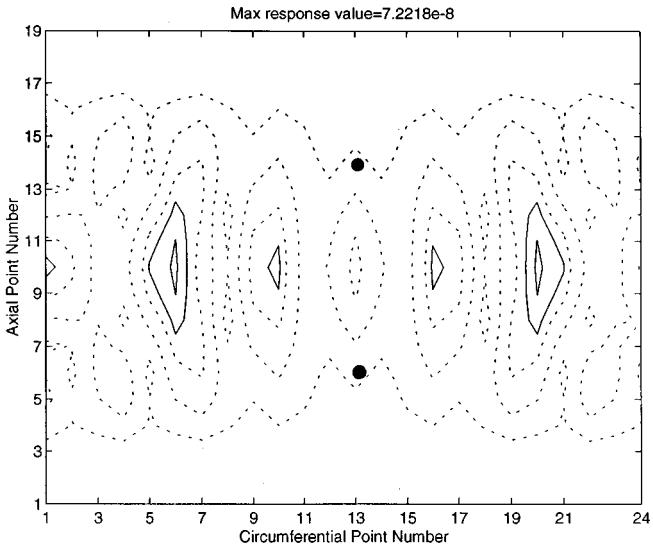


FIG. 3. Vibration response of the shell due to a primary and control forces (located at ●) at 666 Hz. The solid lines represent the high points of the response.

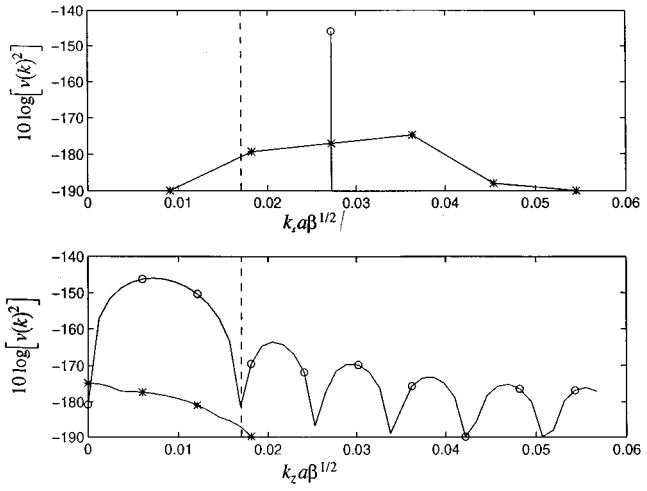


FIG. 4. Maximum dimensionless axial ($k_z a\beta^{1/2}$) and circumferential ($k_s a\beta^{1/2}$) wave numbers ($\beta = h^2/12a^2$, h = shell thickness) at 666 Hz for the uncontrolled (⊖), controlled (⊛) shell responses. The dashed line represents the dimensionless acoustic wave number $k_{acs}a\beta^{1/2}$.

(0.998 lbs). The displacement response of the shell due to both the primary and the control force actuators is shown in Fig. 3. As shown, the controlled response of the shell is no longer symmetric along the circumferential direction. In addition, the controlled response exhibits drastic reduction in amplitude (18 times smaller max amplitude) compared to the uncontrolled response (see Fig. 2). The power radiated from the shell with the primary actuator only (uncontrolled power) was found to be 24.5 dB higher than the power radiated when both the primary and the control actuators were active (controlled power).

To investigate the physical mechanisms behind such drastic reductions in sound power, a wave number analysis of the shell response with and without control was conducted. A two-dimensional discrete fourier transform of the shell response was conducted and used in conjunction with the dimensionless wave number terminology used by Fahy.[24] Zero padding was employed in the axial direction only. This analysis resulted in a three-dimensional surface diagram of dimensionless circumferential and axial wave numbers versus wave number amplitude. Instead of comparing the wave number surface diagrams for the controlled versus the uncontrolled shell response, we decided to compare the maximum circumferential wave numbers of the controlled versus the uncontrolled responses. Similarly, the maximum axial wave numbers were compared. This simplification resulted in two line plots of Fig. 4. As shown, the circumferential wave number exhibits a large (over 20 dB) drop in amplitude. Also, the controlled response is smeared over a wide range of wave numbers as opposed to the uncontrolled wave number. Similarly, the axial wave number has dropped by over 20 dB.

D. Control via minimization of the truncated power expression

To use the strategy described in Sec. I B, the eigenvalues and the eigenvectors of the shell radiation resistance matrix $\mathfrak{R}$ were found at 666 Hz. Because the radiation resistance matrix is Hermitian, its eigenvalues are real and its eigenvectors are complex. As mentioned before, these eigenvalues are referred to as the acoustic weighting coefficients and the eigenvectors are referred to as the acoustic basis functions. Figure 5 shows the first 100 acoustic weighting coefficients in descending order. The rest of these weighting coefficients are considered too small to have any significant effect on the shell radiation at this frequency.

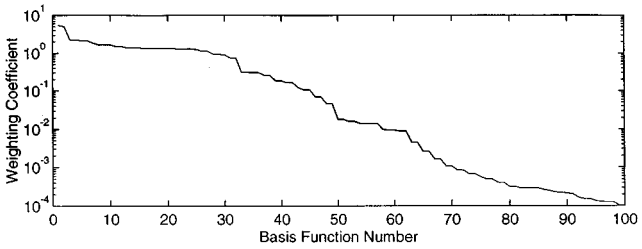


FIG. 5. Acoustic weighting coefficients, λ_i for the shell radiation resistance matrix $\mathfrak{R}$ at 666 Hz.

Using $\mathbf{a} = \mathbf{Q}^H\mathbf{v}$, the coupling coefficients between the uncontrolled shell response and the acoustic basis functions are computed. These coefficients are plotted in Fig. 6. As shown, significant coupling occurs between the 37th, 38th, 45th, 57th, and 81th acoustic basis functions and the velocity exhibited by the shell at 666 Hz. The real and the imaginary components of these acoustic basis functions (not shown here) were found to have a strong resemblance to the shell (2,3), (3,3), and (4,3) mode shapes. Only the first 85 acoustic basis functions were used in the minimization of total radiated power. This number was chosen based on comparison of the relative values of these weighting coefficients as shown in Fig. 5. Only the acoustic basis functions exhibiting a weighting coefficient within 40 dB from the highest value of this plot were deemed significant. Thus in Eq. (11) N was set to 85. Power minimization was carried out using the same force configuration employed in the previous section. The control force, the shell response due to the control plus the primary forces, and the total sound power reductions were found to be identical to those found in the previous section (see Fig. 3). Figure 7 illustrates a plot of the components of the series sum of Eq. (11) with and without control. As shown, as the result of the above minimization, the power radiated by the 38th and 45th acoustic basis functions has been reduced by over 20 dB. This was accomplished by changing the structural vibration response such that the cou-

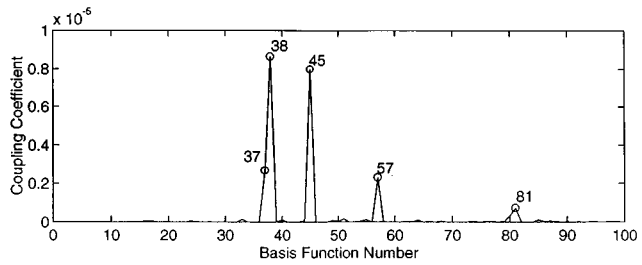


FIG. 6. The velocity-to-basis function coupling coefficients, a_i for the first 100 acoustic basis functions of the *uncontrolled* shell response at 666 Hz.

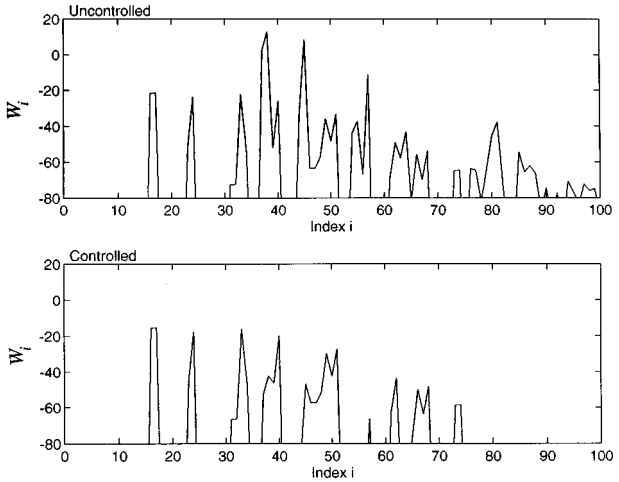


FIG. 7. The power ($W_i = 1/2\ \lambda_i |a_i|^2$ in dB *re*: 10^{-12} W) radiated from the first 100 acoustic basis functions for the uncontrolled and the controlled shell response at 666 Hz.

pling coefficients a_i were altered (see Fig. 8 for plot of first 100 basis functions with control).

E. Experimental verification

To experimentally verify the above predictions, two PCB Model 710M02 actuators (one for the primary excitation and one for the control) were placed at two points on top of the shell. The primary and the control actuators were located at the sixth and the fourteenth axial slice along the shell cylindrical section, respectively. The locations of these actuators are shown in Figs. 2 and 3 at the ●. The force transmitted to the shell was measured by placing a single PCB Model 208B force gage between each actuator and the shell surface. An array of PCB Flexcel modal accelerometers was placed along a single axial line and a single circumferential ring at locations coincident with the surface points used in the model described in Sec. II A. The actuators were driven such that the forces computed previously were applied to the shell. The response of the shell along the axial direction was measured without control (only the primary actuator was active) and with control (both the primary and the control actuator were active). Figures 9 and 10 show a comparison of the measured response versus that of the predicted response along the axial direction (in the circumferential slice that contains the primary and the control actuators). As shown, the predicted uncontrolled and controlled shell

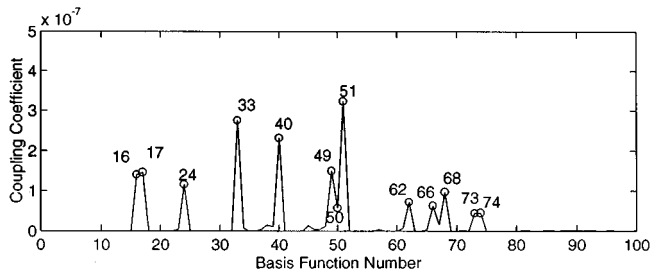


FIG. 8. The velocity-to-basis function coupling coefficients, a_i for the first 100 acoustic basis functions of the *controlled* shell response at 666 Hz.

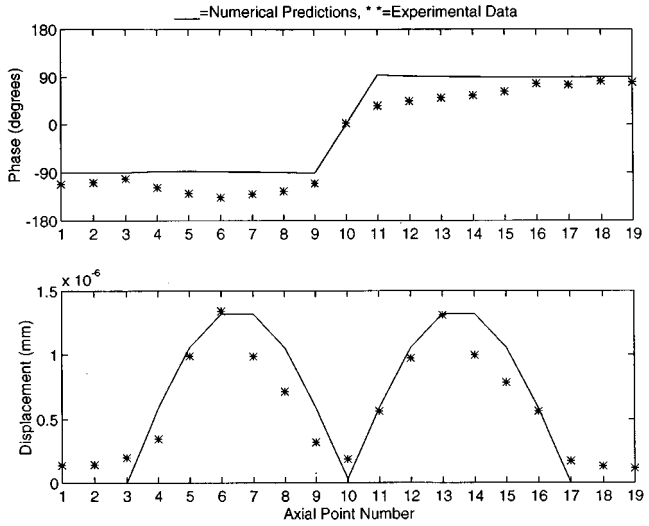


FIG. 9. Comparison of the predicted versus the experimental *uncontrolled* shell response at 666 Hz along the axial line of the two actuators (primary actuator is located at axial point number 6).

response match the measured values. The sound power radiated from the shell was not measured. Since the measured and the predicted structural velocities match so well, the sound power radiated from the shell is assumed to also match the predicted values.

III. CONCLUSIONS

Two strategies for reducing the sound power radiated from structures subjected to a harmonic excitation force were discussed. Both of these strategies are based on the minimization of the quadratic expression for the total radiated sound power. In the first strategy, quadratic optimization of the complete power expression is carried out and a single control force vector is found which minimizes the radiated sound power. In the second strategy, the power expression is writ-

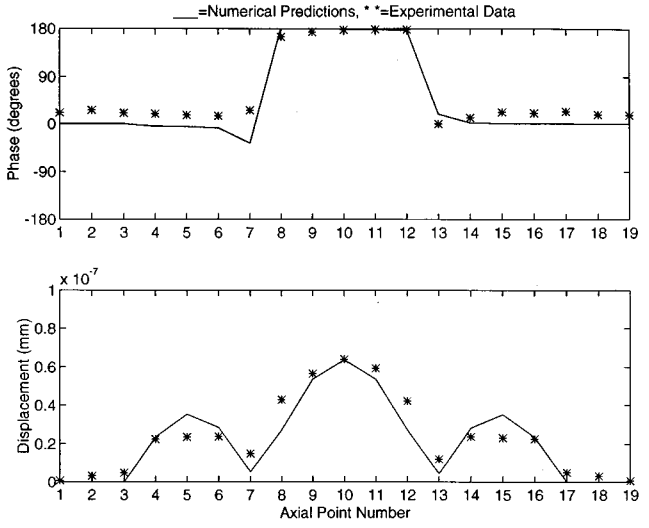


FIG. 10. Comparison of the predicted versus the experimental *controlled* shell response at 666 HZ along the axial line of the two actuators (primary and control actuators are located at axial point numbers 6 and 14, respectively).

ten in terms of a truncated series. Using a limited number of terms in this series, it is possible to minimize the power and arrive at the same control force vector as that found via the first method. These strategies were applied to the case of a cylindrical shell with endcaps. The wave superposition method was used to obtain the shell radiation resistance matrix required for the power formulation. Predictions showed that using a single control actuator it is possible to reduce the total power radiated from this shell by 24.5 dB. This was mainly due to the drastic drop in shell vibration amplitudes resulting from the symmetrical nature of the shell modes and the symmetrical placement of actuators. These predictions were verified experimentally. Future work will concentrate on the use of such strategies for active control of broadband noise.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Ken Cunefare (Georgia Institute of Technology) and Dr. John Fahnline (Penn State Applied Research Laboratory) for their valuable input to this paper. Their willingness to openly share their expertise in the various nuances of numerical modeling techniques as related to the computation of the radiation resistance matrices using different modeling techniques, as well as the eigenvalues and eigenvectors of such matrices, is very much appreciated.

[1] G. E. Warnaka, ''Active attenuation of noise- the state of the art,'' J. Noise Control Eng. **18**, 100–110 (1982).
[2] J. E. Ffowcs Williams, ''Anti-sound,'' Proc. R. Soc. London, Ser. A **395**, 63–88 (1984).
[3] R. R. Leitch and M. O. Tokhi, ''Active noise control systems,'' IEEE Proc. **134(6A)**, 525–546 (1987).
[4] D. C. Swanson, ''Active attenuation of acoustic noise: past, present, and future,'' ASHRAE Trans. **95**, 63–76 (1989).
[5] P. A. Nelson and S. J. Elliott, ''Active noise control: A tutorial review,'' Proceedings of the International Symposium on Active Control of Sound and Vibration held on April 9–11 in Tokyo, Japan pp. 45–74 (1991).
[6] J. C. Stevens and K. K. Ahuja, ''Recent advances in active noise control,'' AIAA J. **29**, 1058–1067 (1991).
[7] D. Guicking, ''Recent advances in active noise control,'' Proceedings of the Second International Congress on Recent Developments in Air- and Structure-Born Sound and Vibration held on March 4–6 at Auburn University, pp. 313–320 (1992).
[8] S. J. Elliott and P. A. Nelson, ''Active noise control,'' IEEE Signal Process Mag. 12–35 (October 1993).
[9] A. I. Vyalyshev and A. I. Dubinin, ''Active acoustic reduction of a plate,'' Sov. Phys. Acoust. **32**, 96–98 (1986).
[10] K. Naghshineh and G. H. Koopmann, ''A design method for achieving weak radiator structures using active vibration control,'' J. Acoust. Soc. Am. **92**, 856–870 (1992).
[11] K. Naghshineh and G. H. Koopmann, ''Active control of sound power using acoustic basis functions as surface velocity filters,'' J. Acoust. Soc. Am. **93**, 2740–2752 (1993).
[12] K. Naghshineh and G. H. Koopmann, ''An active control strategy for achieving weak radiator structures,'' ASME J. Vib. Acoust. **116**, 31–37 (1994).
[13] R. L. Clark and C. R. Fuller, ''Experiments on active control of structurally radiated sound using multiple peizoceramic actuators,'' J. Acoust. Soc. Am. **91**, 3313–3320 (1992).
[14] G. V. Borgiotti and K. E. Jones, ''Wideband spatial filters for the active control of the radiation of elastic shells in an acoustic field,'' Proceedings of the 15th Biennial Conference on Mechanical Vibration and Noise held in Boston, MA, pp. 291–299 (1995).
[15] S. J. Elliott and M. E. Johnson, ''Radiation modes and the active control of sound power,'' J. Acoust. Soc. Am. **94**, 2194–2204 (1993).
[16] K. A. Cunefare and G. H. Koopmann, ''A boundary element approach to optimization of active noise control sources on three-dimensional structures,'' ASME J. Vib. Acoust. **113**, 387 (1991).
[17] K. A. Cunefare and G. H. Koopmann, ''Global optimum active noise control: surface and far field effects,'' J. Acoust. Soc. Am. **90**, 365–373 (1991).
[18] C. G. Mollo and R. J. Bernhard, ''Generalized method of predicting optimal performance of active noise controllers,'' AIAA J. **27**, 1473–1478 (1989).
[19] C. G. Mollo and R. J. Bernhard, ''Numerical evaluation of the performance of active noise controllers,'' ASME J. Vib. Acoust. **112**, 230–236 (1990).
[20] G. V. Borgiotti, ''The power radiated by a vibrating body in an acoustic fluid and its determination from boundary measurements,'' J. Acoust. Soc. Am. **88**, 1884–1893 (1990).
[21] G. H. Koopmann, L. Song, and J. B. Fahnline, ''A method for computing acoustic fields based on the principle of wave superposition,'' J. Acoust. Soc. Am. **86**, 2433–2438 (1989).
[22] L. Song, G. H. Koopmann, and J. B. Fahnline, ''Active control of the acoustic radiation of a vibrating structure using a superposition formulation,'' J. Acoust. Soc. Am. **89**, 2786–2792 (1991).
[23] K. W. Sommerlatte, ''The Effects of Fluid Loading on the Vibration of a Circular Cylindrical Shell,'' Master's thesis, The Pennsylvania State University, 1992.
[24] F. Fahy, *Sound and Structural Vibration Radiation, Transmission and Response* (Academic, New York, 1985), pp. 101–108.

The PVDF-based wave number domain sensing techniques for active sound radiation control from a simply supported beam

Bor-Tsuen Wang
Department of Mechanical Engineering, National Pingtung University of Science and Technology, Pingtung, Taiwan 91207, Republic of China

(Received 22 August 1996; accepted for publication 10 December 1997)

This work presents a novel wave number domain sensing approach which applies polyvinylidene fluoride (PVDF) films for active structural acoustic control. An array of PVDF strip films is equally distributed over a beam to measure the structural response. The PVDF voltage wave number transform functions can then be obtained by performing the discrete wave number transform on the sensing signals. The cost function can be defined as the sum of mean square values of the supersonic components of the wave number transform function. The linear quadratic optimization process can be applied to obtain the optimal control inputs, thereby minimizing the cost function. This work also considers the simply supported beam with infinite rigid baffle subjected to a harmonic point force disturbance. The PVDF-based wave number domain sensing approach is applied, while piezoelectric actuators are used as control inputs. In addition, the sensing method's performance is evaluated by comparing with an accelerometer based wave number domain sensing approach. Both continuous and discrete wave number transform approaches are presented as well. Analysis results demonstrate that the wave number domain sensing approach cannot only provide efficient signals that are related to the structure-borne sound, but is also quite effective in controlling sound radiation. This work advances current efforts to apply the sensing ability of intelligent material structure systems to actively control structural sound radiation. © *1998 Acoustical Society of America.* [S0001-4966(98)05103-0]

PACS numbers: 43.40.Vn [PJR]

LIST OF SYMBOLS

$\widetilde{A}$	acceleration wave number transform function derived from the CWT approach
$\widetilde{A}_{\mathrm{DWT}}$	acceleration wave number transform function derived from the DWT approach
A	PVDF film area
b	beam width
b_p	PVDF width
c	phase speed of sound in air
E_b	beam Young's modulus
e_{31}	piezoelectric strain constant of PVDF sensor
F	point force magnitude
$g_n(r,\theta,\phi)$	the nth pressure modal amplitude
$H_n(\omega)$	the nth mode transfer function
i	spatial sequence for the DWT approach
I	moment of inertia of the beam
K_p	some constant related to PVDF material properties
L	beam length
l_c	length of PZT actuator
l_p	length of PVDF sensor
M_{eq}	equivalent moment induced by PZT actuator
m	wave number sequence for the DWT approach
N	number of spatial samples or number of sensors
N_x	minimum number of wave number sequence such that $\min(N_k)>(\kappa/\Delta\kappa_x)+1$
p_c, p_f, p_t	sound pressure due to PZT actuator, point force disturbance and total
$q_n(\omega)$	the nth modal amplitude
$q_n^c(\omega)$	the nth modal amplitude due to PZT actuators
$q_n^f(\omega)$	the nth modal amplitude due to point force (disturbance)
r,θ,ϕ	radiation field coordinates
t_b	beam thickness
t_p	PVDF thickness
$V(x,t)$	generated voltage from PVDF sensor
$\widetilde{V}$	PVDF voltage wave number transform function derived from the CWT approach
$\widetilde{V}_{\mathrm{DWT}}$	PVDF voltage wave number transform function derived from the DWT approach
$w(x,t)$	beam lateral displacement
x_a	accelerometer location (Fig. 1)
x_c, x_1, x_2	PZT actuator location (Fig. 1)
x_f	force location (Fig. 1)
x_p, x_{p1}, x_{p2}	PVDF sensor location (Fig. 1)
α_n	normal mode wave number
Δx	spacing between sensors
$\Delta\kappa_x$	wave number resolution
$\delta(x)$	Kronecker delta function
ϵ	permittivity of PVDF film
$\Gamma(x)$	shape function of the PVDF sensor
κ	acoustic wave number
κ_x, κ_y	structural wave number
$\kappa_{x,\mathrm{Nyq}}$	Nyquist wave number
$\omega=2\pi f$	excitation frequency
ω_n	the nth natural frequency
Φ_p	radiated sound power

 0001-4966/98/103(4)/1904/12/$10.00

$\Phi_{\kappa,\tilde{V}}$ cost function for PVDF-based wave number sensing by the CWT approach

$\Phi_{\kappa,\tilde{A}}$ cost function for accelerometer based wave number sensing by the CWT approach

$\Phi_{\kappa,\tilde{V}_{\mathrm{DWT}}}$ cost function for PVDF-based wave number sensing by the DWT approach

$\Phi_{\kappa,\tilde{A}_{\mathrm{DWT}}}$ cost function for accelerometer based wave number sensing by the DWT approach

$\phi_n(x)$ the nth displacement mode shape of the simply supported beam

$\phi_n^c(x)$ the nth PZT actuator mode shape of the simply supported beam

$\phi_n^f(x)$ the nth point force mode shape of the simply supported beam

$\phi_n^p(x)$ the nth PVDF sensor mode shape of the simply supported beam

$\tilde{\phi}_n(\kappa_x)$ wave number transform of $\phi_n(x)$

$\tilde{\phi}_n^p(\kappa_x)$ wave number transform of $\phi_n^p(x)$

ρ air density

ρ_b beam density

INTRODUCTION

Structural-borne sound radiation control has received extensive interest, particularly in terms of developing actuators,[1–4] sensors,[5–7] and control algorithms.[8–10] This work presents a novel PVDF based wave number domain sensing technique.

Microphones are conventional error sensors used for active noise control.[11,12] However, such a far-field sensor cannot be implemented for certain applications. In attempting to resolve this problem, some researchers have used the structural sensors such as accelerometers[6] instead of acoustic sensors. Although sufficient control can be achieved, using many accelerometers not only increases the cost, but also influences the original structure's physical characteristics. The feasibility of using polyvinylidene fluoride (PVDF) sensors for structural sound or vibration control has received increasing interest.[13–15] Either a strip of PVDF film[14,15] or a preshaped PVDF modal sensor[13] can be effectively applied to active structural sound radiation control.[13–15] The performance of accelerometers, microphones, and PVDF strip sensors in active structural acoustic control has been studied.[16–18] Microphones generally perform better sound radiation control than the other two; however, microphones have some drawbacks as mentioned above. Among the merits of PVDF sensors include its lower cost, thinner film, and easier implementation than accelerometers. Most importantly, the PVDF films do not induce a mass effect on structures such as accelerometers. Various types of design of PVDF film sensors have been proposed[19–23] and quite promising for controlling structural sound radiation.

As generally known, structural sound radiation can be determined by wave number domain analysis.[24] Only structural wave number components less than the acoustic wave number radiate sound into the far field and, therefore, are termed as supersonic waves. In contrast, the subsonic waves do not contribute to sound radiation. According to previous investigations, wave number domain analysis is a highly effective means of examining the mechanism of structural sound radiation control.[18,25] Such analysis also induces the notion of a wave number domain approach to actively control structure-borne sound. Fuller and Burdisso[26] formulated a cost function on the basis of the wave number domain components associated with the angle of sound radiation to be controlled. The sound radiation can subsequently be effectively controlled. Clark and Fuller[27] analytically demonstrated the model reference control approach while considering the wave number transform of accelerations, corresponding to the desired far field acoustic directivity pattern. According to their results, the wave number distributions can be largely reduced in the supersonic region, thereby allowing for efficiently controlling sound radiation. Maillard and Fuller[28,29] employed accelerometers as error sensors to construct the cost function, i.e., the wave number component associated with a prescribed radiation angle based on the discrete wave number transform. Scott and Sommerfeldt[30] recently investigated the use of an array of shaped PVDF sensors to act as low-pass filters to estimate the radiated power of clamped beams. They showed that the distributed shaped sensors can provide better predictions than the point sensors.

In light of the promising nature of PVDF sensors and wave number domain sensing approaches, this work presents a novel wave number domain sensing technique, in which an array of PVDF strip films attached to the beam is applied. The PVDF sensors are evenly spaced over the beam. The structural response can then be measured and taken by discrete wave number transform to derive the wave number transform function of the PVDF voltage. The cost function, defined as the sum of mean square values of the supersonic wave number components of the wave number transform functions, is related to the radiated power. The linear quadratic optimal control theory can then be applied to determine the control input, thereby minimizing the cost function. Analysis results demonstrates that the sensing technique, in conjunction with the LMS feedforward control algorithm, efficiently reduces the sound radiation from the beam into the far field.

I. THEORETICAL ANALYSIS

A. Lateral vibration of uniform beam

Consider a uniform simply supported beam with length L, as depicted in Fig. 1. The beam lateral displacement, due to harmonic excitation, can be expressed as follows:

$$w(x,t)=e^{j\omega t}\sum_{n=1}^{\infty} q_n(\omega)\phi_n(x), \tag{1}$$

where

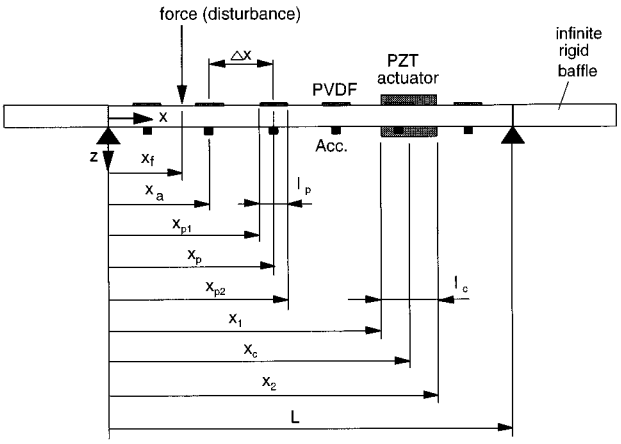


FIG. 1. The arrangement and coordinates of simply supported beam.

$$\phi_n(x)=\sin\ \alpha_n x, \tag{2}$$

$$\alpha_n=\frac{n\pi}{L}. \tag{3}$$

The modal amplitudes $q_n(\omega)$ rely on the form of the excitation force. For a harmonic point force F applying at x_f, the modal amplitude can be shown as

$$q_n^f(\omega)=H_n(\omega)\phi_n^f(x_f)F, \tag{4}$$

where

$$H_n(\omega)=\frac{1}{\rho_b b t_b(\omega_n^2-\omega^2)}, \tag{5}$$

$$\omega_n=\alpha_n^2\sqrt{\frac{E_b I}{\rho_b b t_b}}, \tag{6}$$

$$\phi_n^f(x_f)=\sin\ \alpha_n x_f. \tag{7}$$

Notably, $\phi_n^f(x_f)$ denotes the displacement mode shape of the beam at coordinate x_f. For an actuator consisting of two identical piezoceramic patches, bonded symmetrically on the two opposite beam surfaces, and activated 180° out-of-phase, the corresponding modal amplitude can be derived as follows:

$$q_n^c(\omega)=H_n(\omega)\phi_n^c(x_f)M_{\rm eq}, \tag{8}$$

where

$$\phi_n^c(x_c)=\phi_n'(x_1)-\phi_n'(x_2)=2\alpha_n\ \sin(\alpha_n l_c/2)\phi_n(x_c). \tag{9}$$

Notably, $\phi_n^c(x_c)$ is the PZT actuator mode shape and represents the slope difference between the two edges of the PZT actuator at coordinates x_1, x_2. Also, $\phi_n^c(x_c)$ can be expressed as the displacement mode shape of the beam at the central location of the PZT patch, x_c, for the simply supported beam. For a PVDF film arranged as Fig. 1 illustrates, the generated voltage can be written as[14]

$$V(x_p,t)=e^{j\omega t}K_p\sum_{n=1}^{\infty}\ q_n(\omega)\phi_n^p(x_p), \tag{10}$$

where

$$\phi_n^p(x_p)=\phi_n'(x_{p1})-\phi_n'(x_{p2})=2\alpha_n\ \sin(\alpha_n l_p/2)\phi_n(x_p), \tag{11}$$

$$K_p=\frac{t_p}{\epsilon A}\ \frac{t_b+t_p}{2}\ e_{31}b_p. \tag{12}$$

Notably, the generated voltage is proportional to the slope difference between the two edges of the PVDF film. In addition, Eq. (11) indicates that the PVDF sensor mode shape $\phi_n^p(x_p)$ is related to the displacement mode shape at the central location of the PVDF sensor and weighted by the factor $2\alpha_n\ \sin(\alpha_n l_p/2)$.

B. Sound radiation from the beam

The far-field radiated sound pressure from the beam with infinite rigid baffle due to the point force and PZT actuator can be derived from the Rayleigh integral, respectively,

$$p_f(r,\theta,\phi,t)=e^{j\omega t}\sum_{n=1}^{\infty}\ H_n(\omega)\phi_n^f(x_f)Fg_n(r,\theta,\phi), \tag{13}$$

$$p_c(r,\theta,\phi,t)=e^{j\omega t}\sum_{n=1}^{\infty}\ H_n(\omega)\phi_n^c(x_c)M_{eq}g_n(r,\theta,\phi), \tag{14}$$

where

$$g_n(r,\theta,\phi)=-j\omega\ \frac{\rho c b}{\pi}\ \frac{\kappa}{\alpha_n}\ \frac{e^{-j\kappa r}}{2r}\left[\frac{1-(-1)^n e^{-j\alpha}}{1+(\alpha/n\pi)^2}\right]\times\left[\frac{1-e^{-j\beta}}{\beta}\right], \tag{15}$$

$$\alpha=\kappa L\ \sin\ \theta\ \cos\ \phi, \tag{16}$$

$$\beta=\kappa b\ \sin\ \theta\ \sin\ \phi. \tag{17}$$

Under the assumption of superposition, the total radiated sound pressure denotes the sum of the sound pressures attributed to the disturbance and control inputs

$$p_t=p_f+p_c=e^{j\omega t}\sum_{n=1}^{\infty}\ [\phi_n^f(x_f)F+\phi_n^c(x_c)M_{eq}]H_n(\omega)g_n(r,\theta,\phi). \tag{18}$$

The total radiated power, defined as the integral of the square of the radiated sound pressure over the hemisphere of the radiating field, can then be obtained as

$$\Phi_p=\frac{1}{2\rho c}\int_s|p_t|^2dS=\frac{r^2}{2\rho c}\int_0^{2\pi}\int_0^{\pi/2}|p_t|^2\ \sin\ \theta\ d\theta\ d\phi. \tag{19}$$

The total radiated sound power can be an index to evaluate the effectiveness of controlling the sound radiation.

C. Wave number analysis

By performing continuous wave number transform (CWT) in terms of spatial coordinates for a two-dimensional

rectangular radiator, the acceleration wave number transform function can be obtained by neglecting the harmonic time function, i.e.,

$$\tilde{A}(\kappa_x,\kappa_y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ddot{w}(x,y)e^{-j(\kappa_x x+\kappa_y y)}\,dx\,dy, \tag{20}$$

where

$$\kappa_x=\kappa\,\sin\,\theta\,\cos\,\phi, \tag{21}$$

$$\kappa_y=\kappa\,\sin\,\theta\,\sin\,\phi. \tag{22}$$

For a one-dimensional beam application, the y-direction response is negligible. Therefore, the beam acceleration wave number transform function (WTF), by performing the continuous wave number transform and neglecting the time component, can be expressed as

$$\tilde{A}(\kappa_x,\kappa_y)=-\omega^2\left[\frac{e^{-j\kappa_y b}-1}{-j\kappa_y}\right]\sum_{n=1}^{\infty}q_n(\omega)\tilde{\phi}_n(\kappa_x), \tag{23}$$

where

$$\tilde{\phi}_n(\kappa_x)=\int_{-\infty}^{\infty}\phi_n(x)e^{-j\kappa_x x}\,dx$$

$$=\int_0^L\phi_n(x)e^{-j\kappa_x x}\,dx=\alpha_n\left[\frac{1-(-1)^n e^{-j\kappa_x L}}{\alpha_n^2-\kappa_x^2}\right]. \tag{24}$$

$\tilde{\phi}_n(\kappa_x)$ is the continuous wave number transform of $\phi_n(x)$, and can be approximated by the discrete wave number transform (DWT) as follows:

$$\tilde{\phi}(m\Delta\kappa_x)=\sum_{i-1}^{N}\phi_n(i\Delta x)e^{-j(mi/N)}\Delta x,$$

$$m=-(N-1),-(N-2),\ldots,-1,0,1,...,N-1, \tag{25}$$

where

$$\Delta\kappa_x=\frac{1}{L}, \tag{26}$$

$$N=\frac{L}{\Delta x}. \tag{27}$$

N denotes the number of discretization, i.e., the number of sensors applied in practical measurement; and Δx represent the equal distance between accelerometers. The acceleration WTF by performing the DWT for $\kappa_y=0$, i.e., only the κ_x components are considered in this work, can then be derived as

$$\tilde{A}_{\mathrm{DWT}}(m\Delta\kappa_x)=-b\omega^2\sum_{n=1}^{\infty}q_n(\omega)\tilde{\phi}_n(m\Delta\kappa_x)$$

$$m=-(N-1),-(N-2),\ldots,-1,0,1,...,N-2,N-1. \tag{28}$$

Notably, the negative wave number components (m=negative integer) are also included and simply correspond to radiation in opposite directions from the positive wave number ones.

Similarly, the generated voltages from the PVDF film shown in Eq. (10) can also be taken by the CWT in κ-plane and derived as follows:

$$\tilde{V}(\kappa_x,\kappa_y)=K_p\left[\frac{e^{-j\kappa_y b}-1}{-j\kappa_y}\right]\sum_{n=1}^{\infty}q_n(\omega)\tilde{\phi}_n^p(\kappa_x), \tag{29}$$

where

$$\tilde{\phi}_n^p(\kappa_x)=\int_0^L\phi_n^p(x)e^{-j\kappa_x x}\,dx$$

$$=2\alpha_n^2\,\sin(\alpha_n l_p/2)\left[\frac{1-(-1)^n e^{-j\kappa_x L}}{\alpha_n^2-\kappa_x^2}\right]. \tag{30}$$

Again, the continuous wave number transform of the PVDF voltage shown in Eq. (29) can be approximated by the discrete wave number transform; therefore, the PVDF voltage WTF by performing the DWT for $\kappa_y=0$ can be obtained as

$$\tilde{V}_{\mathrm{DWT}}(m\Delta\kappa_x)=-b\sum_{n=1}^{\infty}q_n(\omega)\tilde{\phi}_n^p(m\Delta\kappa_x),$$

$$m=-(N-1),-(N-2),\ldots,-1,0,1,...,N-2,N-1, \tag{31}$$

where

$$\tilde{\phi}_n^p(m\Delta\kappa_x)=\sum_{i=1}^{N}\phi_n^p(i\Delta x)e^{-j(mi/N)}\Delta x,$$

$$m=-(N-1),-(N-2),\ldots,-1,0,1,...,N-2,N-1. \tag{32}$$

$\tilde{\phi}_n^p(m\Delta\kappa_x)$ denotes the mth wave number components of the PVDF mode shape $\phi_n^p(x)$ by performing the discrete wave number transform; and $i\Delta x$ represents the spatial coordinate of the ith sensor location.

D. Cost functions

Notably, the mean-square value of the acceleration wave number transform function, i.e., $|\tilde{A}|^2$, integrated over the supersonic region is related to the radiated power.[24] Only the wavenumber components satisfying $\kappa_x^2+\kappa_y^2<\kappa^2$ contribute to sound radiation into the far field, and are termed herein as the supersonic waves. The other wave number components associated with subsonic waves do not radiate into the far field. For a one-dimensional beam with infinite rigid baffle, the radiated power can be expressed in terms of the wave number transform of beam acceleration as follows:[31]

$$\Phi_p=\frac{\rho c\kappa}{4\pi\omega^2}\int_{-\kappa}^{\kappa}\frac{|\tilde{A}(\kappa_x)|^2}{\sqrt{\kappa^2-\kappa_x^2}}\,d\kappa_x. \tag{33}$$

Precisely evaluate the above integral is extremely difficult. Cremer and Heck[31] proposed an approximate method to estimate the radiated power. Herein, the cost function can be constructed on the basis of the wave number transform functions introduced in Eqs. (23) and (29). Two types of cost

functions associated with the wave number components in supersonic region are defined as follows;

$$\Phi_{\kappa,\tilde{A}}=\int_{-\kappa}^{\kappa}|\tilde{A}(\kappa_x,\kappa_y=0)|^2\,d\kappa_x, \tag{34}$$

$$\Phi_{\kappa,\tilde{V}}=\int_{-\kappa}^{\kappa}|\tilde{V}(\kappa_x,\kappa_y=0)|^2\,d\kappa_x, \tag{35}$$

where $\Phi_{\kappa,\tilde{A}}$ denotes the integration of supersonic wave number components that contribute to the sound radiation into the far field, and is strongly related to the radiated power. Also, $\Phi_{\kappa,\tilde{V}}$ represents the integration of the mean square value of PVDF voltage wave number transform functions over the supersonic region. Closely examining Eqs. (23) and (29) reveals two main differences. First, constants $-\omega^2$ and K_p [defined in Eq. (12)] are premultiplied for $\tilde{A}$ and $\tilde{V}$, respectively. Second, $\tilde{\phi}_n(\kappa_x)$ and $\tilde{\phi}_n^p(\kappa_x)$ represent $\tilde{A}$ and $\tilde{V}$ in the summation respectively. $\tilde{\phi}_n(\kappa_x)$ and $\tilde{\phi}_n^p(\kappa_x)$ are obtained by performing the wave number transform on $\phi_n(x)$ and $\phi_n^p(x)$. Notably, $\phi_n^p(x)$ as defined in Eq. (11) is the slope difference between the two edges of the PVDF film.[32,33] In particular, $\phi_n^p(x)$ is related to the displacement mode shape $\phi_n(x)$ for the case of a simply supported beam. Comparing Eqs. (24) and (30) indicates that $\tilde{\phi}_n(\kappa_x)$ and $\tilde{\phi}_n^p(\kappa_x)$ have the same form of functions in terms of κ_x. Therefore, cost functions $\Phi_{\tilde{A}}$ and $\Phi_{\tilde{V}}$ can be strongly correlated and also related to the radiated power for the simply supported beam, as shown in Eq. (33). Both $\Phi_{\kappa,\tilde{A}}$ and $\Phi_{\kappa,\tilde{V}}$ are continuous functions obtained from the acceleration and PVDF voltage wave number transform functions, based on the CWT approach. Importantly, these two cost functions can not be directly measured in practice, despite the fact that these cost functions are highly correlated with the structural sound radiation.

Employing an array of evenly spaced accelerometers or PVDF strip film sensors in conjunction with the DWT approach allows us to derive the discrete wave number transform functions based on accelerometers and PVDF sensors through Eqs. (23) and (29), as shown in Eqs. (28) and (31). Both negative and positive wave number components can subsequently be obtained. Moreover, the symmetry of $|\tilde{V}(\kappa_x)|^2$ and $|\tilde{A}(\kappa_x)|^2$ and the folding effect of discrete wave number transform (analogous to the time domain Fourier transform), enable us to also obtain negative wave number components without the discrete wave number transform on the negative integer m. Notably, only the wave number components below and above the Nyquist wave number $\pm\kappa_{x,\mathrm{Nyq}}=(N/2)\Delta\kappa_x$ are valid. Analogous to Eqs. (34) and (35) for the CWT approach, the cost functions based on the DWT approach for accelerometers and PVDF sensors can be defined as

$$\Phi_{\kappa,\tilde{A}_{\mathrm{DWT}}}=\sum_{m=-N_k}^{N_k}|\tilde{A}_{\mathrm{DWT}}(m\Delta\kappa_x)|^2,$$

$$\min(N_k)>\frac{\kappa}{\Delta\kappa_x}+1, \tag{36}$$

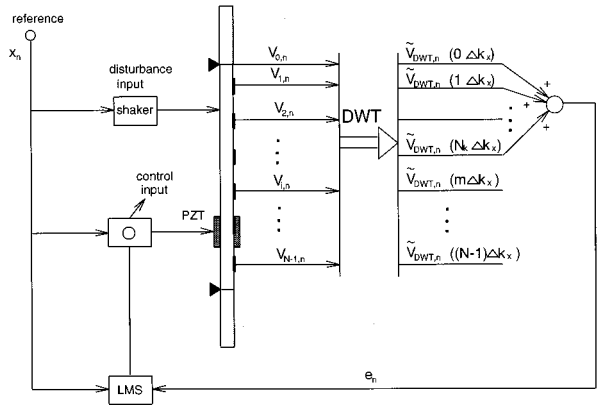


FIG. 2. LMS model for PVDF-based wave number sensing approach.

$$\Phi_{\kappa,\tilde{V}_{\mathrm{DWT}}}=\sum_{m=-N_k}^{N_k}|\tilde{V}_{\mathrm{DWT}}(m\Delta\kappa_x)|^2,$$

$$\min(N_k)>\frac{\kappa}{\Delta\kappa_x}+1. \tag{37}$$

$\Phi_{\kappa,\tilde{A}_{\mathrm{DWT}}}$ and $\Phi_{\kappa,\tilde{V}_{\mathrm{DWT}}}$ represent the sum of mean square values of the supersonic wave number components of the acceleration and PVDF voltage WTFs, respectively. In addition, N_k denotes the minimum number of wave number components such that $\min(N_k)>(\kappa/\Delta\kappa_x)+1$. Interestingly, $\Phi_{\kappa,\tilde{A}_{\mathrm{DWT}}}$ and $\Phi_{\kappa,\tilde{V}_{\mathrm{DWT}}}$ are measurable, and can be used to approximate the cost functions $\Phi_{\kappa,\tilde{A}}$ and $\Phi_{\kappa,\tilde{V}}$. In summary, $\Phi_{\kappa,\tilde{A}}$ and $\Phi_{\kappa,\tilde{V}}$ derived from the CWT approach are ideal cost functions, which are not measurable. $\Phi_{\kappa,\tilde{A}_{\mathrm{DWT}}}$ and $\Phi_{\kappa,\tilde{V}_{\mathrm{DWT}}}$ are defined from the acceleration and PVDF voltage WTFs by the DWT approach, and can be obtained by using an array of equally spaced accelerometers and PVDF sensors, respectively. Figure 2 depicts the control block diagram for the PVDF based wave number sensing approach, in conjunction with the use of LMS feedforward control algorithm. Maillard and Fuller provide further detail regarding the implementation.[28,29]

Any of the abovementioned cost functions is obviously quadratic and positively definite; such a function possesses a unique minimum as well. The linear quadratic optimal control theory (LQOCT) can then be applied to minimize the cost function, thereby allowing us derive the optimal control voltages input to the piezoelectric actuators. Minimization procedures[12] can easily calculate the optimal control parameters, subsequently minimizing the cost function. A complete analysis can be referred to[34] and, therefore, is omitted here.

II. NUMERICAL RESULTS AND DISCUSSIONS

A steel beam having dimensions of a length of 0.38 m, width of 0.04 m, and thickness of 2 mm is used in the simulations. The first few natural frequencies are 33.2, 128.8, 289.9, 515.4, 805.3, and 1159.6 Hz. Notably, no damping occurs in the following analysis. Although the optimal process can adequately control multiple primary sources, only one harmonic point force with input parameters, $F=0.3N$ and $x_f=0.067$ m, is considered in the following analysis.

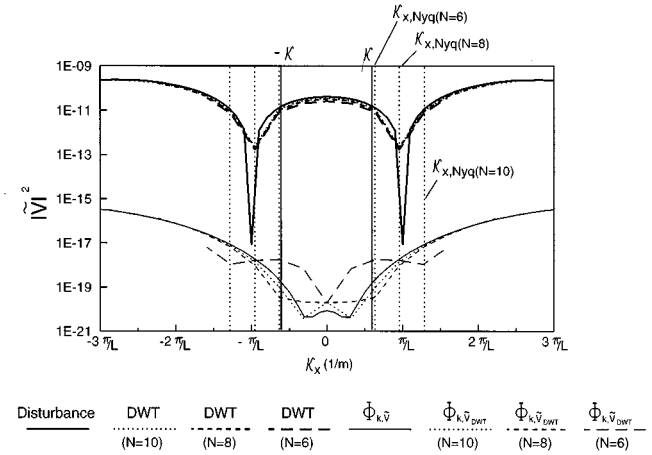


FIG. 3. PVDF voltage wave number spectra before and after being controlled for PVDF sensors, $f=290$ Hz.

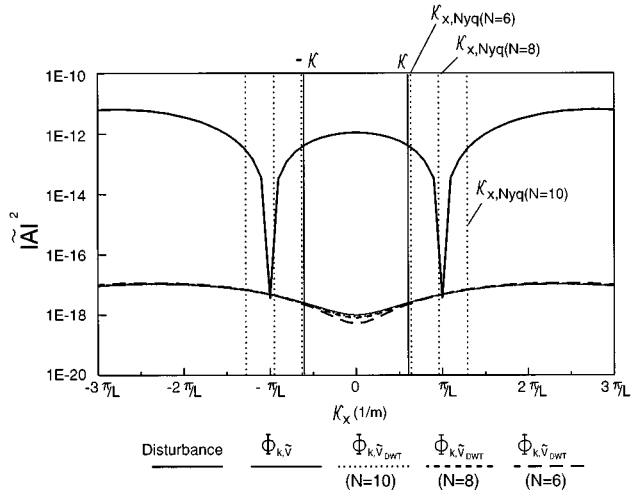


FIG. 4. Acceleration wave number spectra before and after being controlled for PVDF sensors, $f=290$ Hz.

The piezoceramic patch (G-1195)[35] and PVDF films (LDT−28 μm)[36] are used, respectively. The piezoceramic patch is located at $x_1=0.285$ m, $x_2=0.3485$ m. To implement the wave number sensing approach, an array of PVDF strip films are evenly distributed over the beam, length as shown in Fig. 1. Notably, the rectangular shapes of PVDF films, rather than other shapes of PVDF films,[15,21,30] are adopted herein to implement the wave number sensing approach. The PVDF strip film is assumed to be 0.02 m in length for wave number sensing approaches.

For comparison, an array of accelerometers are also used and assumed to be located at the PVDF films' central location, thereby allowing us to implement the wave number sensing approach. Next, the controllability of the wave number sensing approaches is examined, by numerically simulating the cost function derived from the CWT approach in Eqs. (34) and (35). According to those equations, on- and off-resonance excitation cases demonstrate the effectiveness of controlling sound radiation. Either accelerometers or PVDF sensors are respectively used; the CWT and DWT approaches are compared as well. Simulation results of the radiation directivity and WTF distributions demonstrate the control mechanisms of the wave number sensing approaches. The radiated sound pressure is calculated at a radial distance of 3 m from the beam, and plotted in dB *re:* 20×10^{-6} Pa over $\theta=-90°$ to 90° for $\phi=0°$. To calculate the beam response and radiated sound pressure, the modal sums in Eq. (1) must be truncated. While considering computational time and accuracy, the first ten modes are considered and can provide sufficient convergence of the series. Finally, the cases involving different excitation frequencies up to 1000 Hz are also simulated. The cost functions and the radiated powers are plotted against the excitation frequency for both before and after being controlled, respectively.

A. On-resonance excitation case, $f=290$ Hz

Figure 3 depicts the PVDF voltage wave number spectrum before and after being controlled for $f=290$ Hz, i.e., near the third structural resonance. The thick lines represent the response of disturbance, while the thin lines denote the response after being controlled. The solid line is obtained on the basis of Eq. (29) which is a continuous function. This figure also contains PVDF voltage wave number spectra derived from the DWT approach for $N=10$, $N=8$, and $N=6$. The acoustic wave number and the Nyquist wave number for different N values are indicated as well. According to this figure, the PVDF voltage wave number spectra obtained from the DWT approach correlate well with those from the CWT approach. If more sensors are applied, i.e., the larger the N implies a closer PVDF voltage wave number spectra derived from the DWT approach to that derived from the CWT approach. Therefore, using an array of PVDF strip films as error sensors, the continuous wave number spectrum can be well approximated. To select either $\Phi_{\kappa,\tilde{V}}$ or $\Phi_{\kappa,\tilde{V}_{\mathrm{DWT}}}$ as the cost function, the optimal control voltage input to the PZT actuator can be obtained by LQOCT, respectively. After being controlled, the minimized cost functions are also shown in Fig. 3. This figure indicates that the wavenumber components are globally reduced over the structural wave number. The more sensors used for $\Phi_{\kappa,\tilde{V}_{\mathrm{DWT}}}$ implies (a) more reduction of wave number components in the supersonic region, and (b) the closer the remaining wave number spectra to those for $\Phi_{\kappa,\tilde{V}}$.

To further examine the control performance, Fig. 4 depicts the acceleration WTFs before and after being controlled for the PVDF wave number sensing approach. Notably, the sum of the mean square value of the acceleration WTF in the supersonic region is related to the radiated power. The acceleration wave number spectra for either the case $\Phi_{\kappa,\tilde{V}_{\mathrm{DWT}}}$ of $N=10$, $N=8$, or $N=6$ after being controlled are close to each other, as well as the approach to that for the case of $\Phi_{\kappa,\tilde{V}}$. The remaining wave number spectra of $|\tilde{A}|^2$ in the supersonic region, which contributes to the far field sound radiations, reveal a combination of the first and second mode response. This finding implies that the most significant radiation mode, i.e., the third mode in this case, is well attenuated and leaves less efficient radiation modes after being controlled. The control mechanism is termed as ''modal

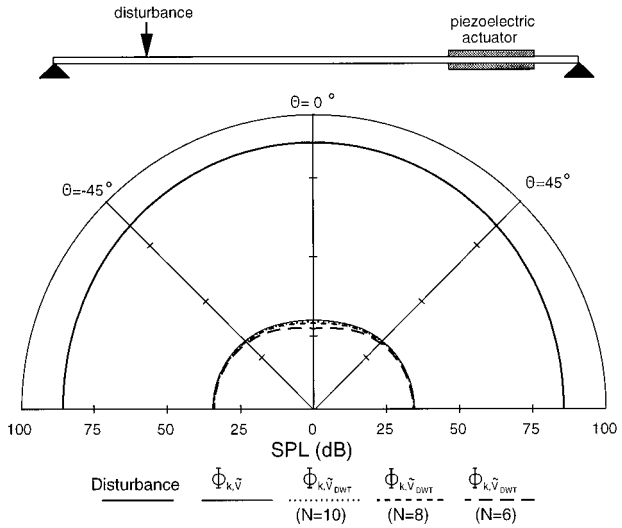


FIG. 5. Sound pressure level distribution for PVDF sensors, $f=290$ Hz.

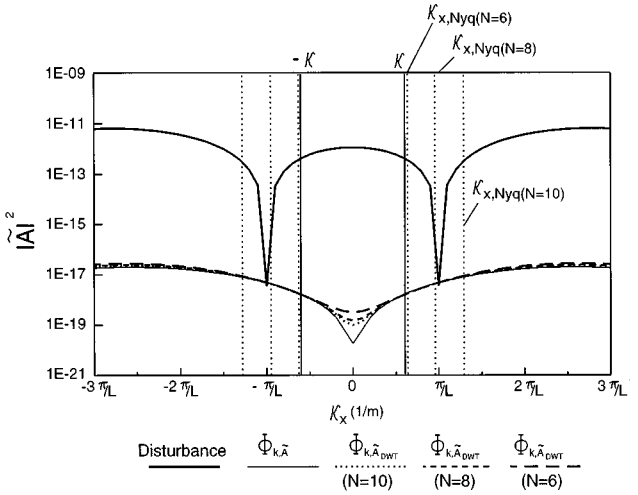


FIG. 7. Acceleration wave number spectra before and after being controlled for accelerometers by CWT approach, $f=290$ Hz.

suppression.''[37] Figure 5 depicts the sound pressure level (SPL) distributions corresponding to the cases in Fig. 4. The SPL distribution due to the disturbance reveals a monopole type radiation pattern. However, the SPL amplitude near $\theta=0°$ is slightly higher than the other amplitudes, thereby confirming the existence of the third mode because the excitation frequency $f=290$ Hz is near the third structural resonance. The remaining SPL distributions after being controlled reveal an omnidirectional radiation pattern and a small dip near $\theta=0°$. This occurrence can be attributed to that the third mode response is well controlled, leaving less significant radiation modes that are the first and second modes as discussed in Fig. 4.

Above numerical results are based on the use of an array of PVDF strip sensors to implement the wave number sensing approach for beam sound radiation control. Of particular interest is to use an array of accelerometers as error sensors to approximate the wave number transform functions in order to construct the cost functions $\Phi_{\kappa,\tilde{A}_{\mathrm{DWT}}}$. Figure 6 presents the acceleration wave number spectra, as derived from the DWT approach before (thin lines) and after (thick lines) being controlled for $N=10$, $N=8$, and $N=6$. The solid line denotes the disturbance. The acoustic wave number and the Nyquist wave number for different N values are also indicated. The acceleration wave number spectra due to the disturbance derived from the DWT approach correlate well with that from the CWT approach. As expected, the more sensors that are applied implies a closer approximation. For selecting $\Phi_{\kappa,\tilde{A}_{\mathrm{DWT}}}$ as the cost function to perform wave number sensing approach for controlling sound radiation, according to our results, the acceleration wave number spectra after being controlled derived from the DWT approach are globally reduced. In particular, the value at a point near $\kappa_x=0$ is relatively small. This small value of κ_x can be referred to as the radiation angle $\phi=0°$, $\theta=0°$. Figure 7 provides the corresponding acceleration wave number spectra derived from the CWT approach after being controlled. The more sensors that are applied imply a smaller wave number component at the dip. If $\Phi_{\kappa,\tilde{A}}$ is selected as the cost function, the acceleration wave number spectrum after being controlled can be seen as a dip at $\kappa_x=0$, as a dipole response. Figure 8 displays the sound pressure level distributions corresponding to the cases in Fig. 7. The remaining sound pressure level distributions reveal a dipole for the case of $\Phi_{\kappa,\tilde{A}}$, while revealing a small dip near $\theta=0°$ for the cases of $\Phi_{\kappa,\tilde{A}_{\mathrm{DWT}}}$. Notably, using more sensors to implement the DWT approach would cause the results to be closer to the CWT approach.

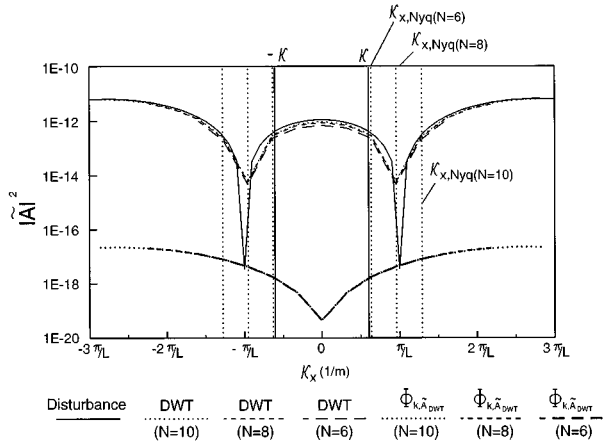


FIG. 6. Acceleration wave number spectra before and after being controlled for accelerometers by DWT approach, $f=290$ Hz.

Table I summarizes the reduction of radiated power for both accelerometers and PVDF wave number sensing approach for $f=290$ Hz. For the case of $\Phi_{\kappa,\tilde{A}}$ which attempts to minimize the acceleration WTF derived from the CWT approach, the reduction of radiated power is the maximum up to 62.5 dB. For the case of $\Phi_{\kappa,\tilde{V}}$ which aims to minimize the PVDF voltage WTF derived from the CWT approach, 57.8 dB is obtained slightly less than that for the case of $\Phi_{\kappa,\tilde{A}}$. For the DWT approach, the accelerometers generally achieve more reduction of radiated power than the PVDF sensors because the accelerometers provide more direct and efficient information regarding radiated sound than the

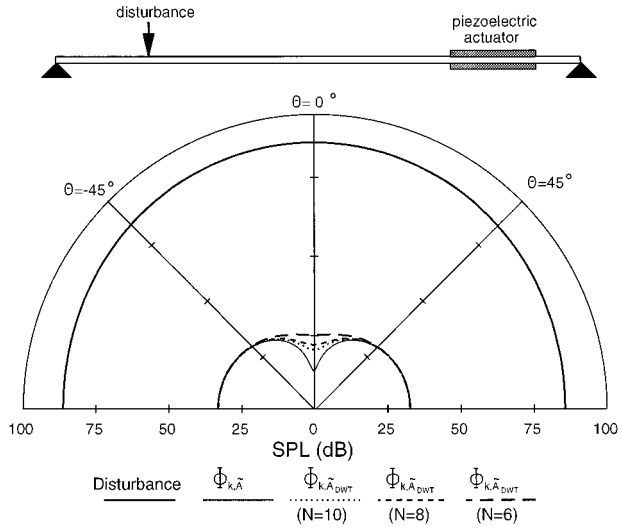


FIG. 8. Sound pressure level distribution for accelerometers, $f=290$ Hz.

PVDF sensors. For different numbers of sensors applied, the control effectiveness is nearly the same, and more sensors applied for the DWT approach more accurately approximates the results from the CWT approach.

B. Off-resonance excitation case, $f=350$ Hz

Figure 9 presents the PVDF voltage wave number spectrum before and after being controlled for $f=350$ Hz, i.e., between the third and fourth structural resonant modes, which is similar to Fig. 3. Again, the PVDF voltage wave number spectra for disturbance from the DWT approach correlate well with those spectra from the CWT approach. More sensors applied can more accurately estimate the wave number spectrum. For using the cost function either $\Phi_{\kappa,\tilde{V}}$ or $\Phi_{\kappa,\tilde{V}_{\mathrm{DWT}}}$, the wave number components in the supersonic region can be markedly reduced; however, spillover may occur in the subsonic region. Corresponding to the cases in Fig. 9, Fig. 10 displays the acceleration wave number spectra before and after being controlled for the PVDF wave number sensing approaches. According to this figure, the wave number components in the supersonic region, which contribute the sound radiation to the far field, are attenuated. In the subsonic region, a spillover can be observed. The control mechanism can be realized in which the supersonic wave number components are well reduced; meanwhile, spillovers

TABLE I. Reduction of the radiated power (dB).

Cost function	$f=290$ Hz	$f=350$ Hz
$\Phi_{\kappa,\tilde{V}}$	57.98	9.18
$\Phi_{\kappa,\tilde{V}_{\mathrm{DWT}}}$ $(N=10)$	58.30	8.92
$\Phi_{\kappa,\tilde{V}_{\mathrm{DWT}}}$ $(N=8)$	58.53	8.90
$\Phi_{\kappa,\tilde{V}_{\mathrm{DWT}}}$ $(N=6)$	59.47	7.49
$\Phi_{\kappa,\tilde{A}}$	62.57	16.28
$\Phi_{\kappa,\tilde{A}_{\mathrm{DWT}}}$ $(N=10)$	62.23	15.94
$\Phi_{\kappa,\tilde{A}_{\mathrm{DWT}}}$ $(N=8)$	62.01	15.72
$\Phi_{\kappa,\tilde{A}_{\mathrm{DWT}}}$ $(N=6)$	61.22	14.85

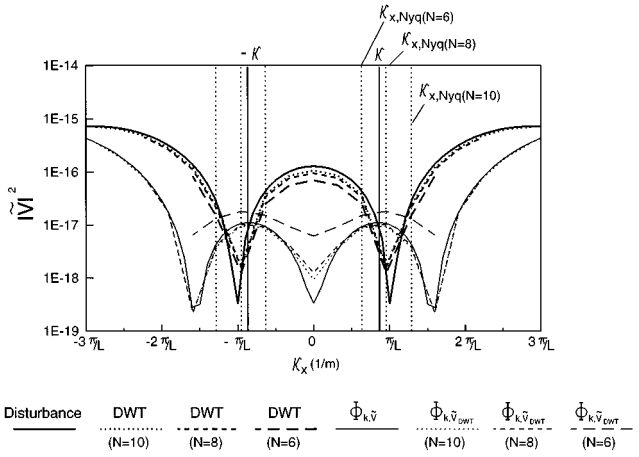


FIG. 9. PVDF voltage wave number spectra before and after being controlled for PVDF sensors, $f=350$ Hz.

occur for the subsonic wave number components that do not radiate. This control mechanism is called ''modal restructuring.[37]'' Notably, for the DWT approach $N=6$, the Nyquist wave number is below the acoustic wave number as shown in Fig. 10. A certain amount of supersonic wave number components can still be reduced because the cost function can be approximately constructed for $N=6$ as shown in Fig. 9. Figure 11 depicts the sound pressure level distributions corresponding to the cases in Fig. 10.

Examining the performance of using an array of accelerometers for the off-resonance excitation case is of worthwhile interest. Figures 12–14 show the acceleration wave number spectra and sound pressure level distributions, which are similar to Figs. 6–8 except $f=350$ Hz. Similar characteristics can be observed for off-resonance excitation as those discussed for the PVDF sensors. Table I also indicates that the radiated power reduces for the off-resonance excitation case. The CWT approach is more efficient than the DWT approach. In general, more sensors that are applied for the DWT approach gain more reduction of the radiated power.

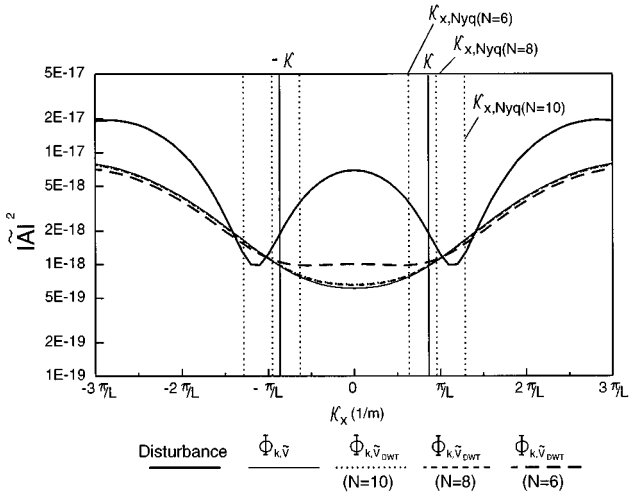


FIG. 10. Acceleration wave number spectra before and after being controlled for PVDF sensors, $f=350$ Hz.

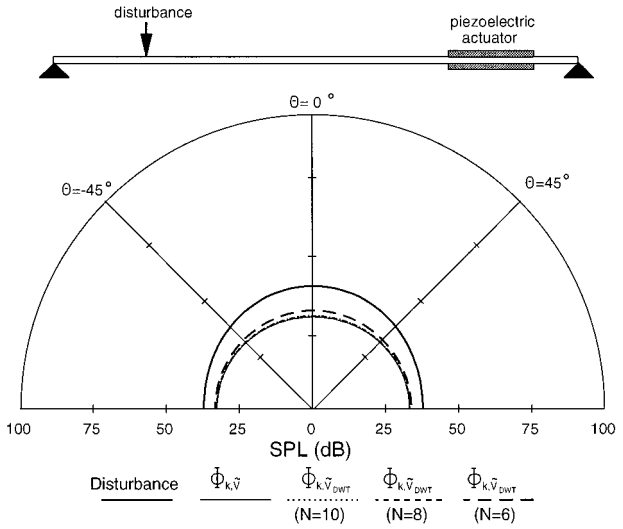


FIG. 11. Sound pressure level distribution for PVDF sensors, $f=350$ Hz.

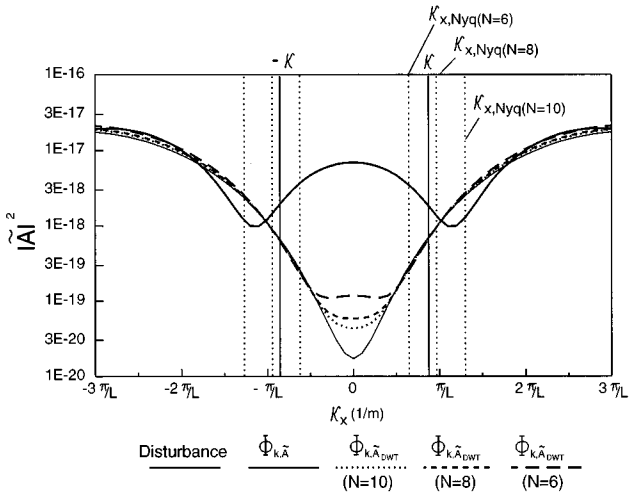


FIG. 13. Acceleration wave number spectra before and after being controlled for accelerometers by CWT approach, $f=350$Hz.

C. Different excitation frequency performance

As mentioned earlier, only the wave number components below or above the positive or negative Nyquist wave number are valid for discrete wave number transform. If wave number components in supersonic region lie above the Nyquist wave number, increasing the number of sensors is inherently required. Another alternative involves properly designing the PVDF sensor to suppress the higher wave numbers so that the effect of subsonic wave numbers will not be eliminated.[15,30] Further investigating in the shape design of PVDF sensor for the wave number sensing approach would be a worthwhile pursuit. Table II lists the minimum number of sensors for different excitation frequencies. For $f=300$ Hz, the acoustic wave number is 5.495, and the minimum N_κ is then 4 in order to include all of the supersonic wave number components; therefore, the minimum number of sensors N is 6, and its corresponding Nyquist wave number is 7.894. If $N=10$, then the Nyquist wave number is 13.15. Sound radiation control can be effective up to $f=700$ Hz, i.e., the corresponding acoustic wave number is 12.82.

Moreover, the performance of the PVDF and accelerometers wave number sensing approaches is further evaluated for different excitation frequencies. Figure 15(a)–(d), respectively, plots the cost functions against excitation frequency before and after being controlled for the cases of $\Phi_{\kappa,\tilde{A}}$, $\Phi_{\kappa,\tilde{V}}$, $\Phi_{\kappa,\tilde{A}_{\mathrm{DWT}}}$, and $\Phi_{\kappa,\tilde{V}_{\mathrm{DWT}}}$. As those figures reveal, the cost functions can be attenuated significantly after being controlled. This finding suggests that the minimization process effectively reduces the cost functions while applying controls. The effectiveness of sound radiation control can be studied by examining the radiated power versus excitation frequency, as shown in Fig. 16. Several observations are summarized as follows:

(1) For accelerometers by the CWT approach, i.e., the cost

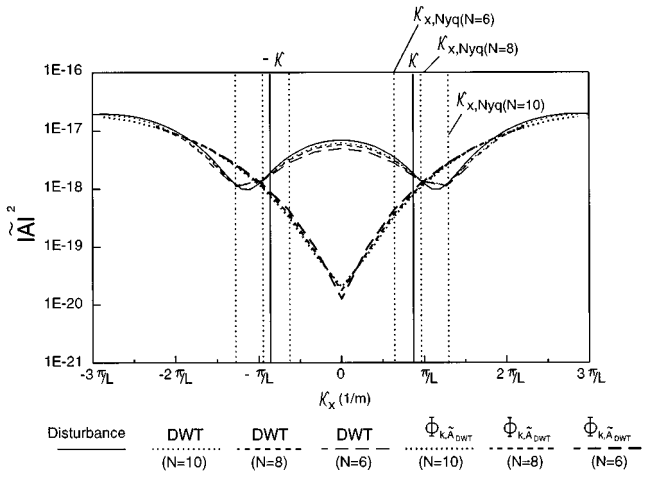


FIG. 12. Acceleration wave number spectra before and after being controlled for accelerometers by DWT approach, $f=350$ Hz.

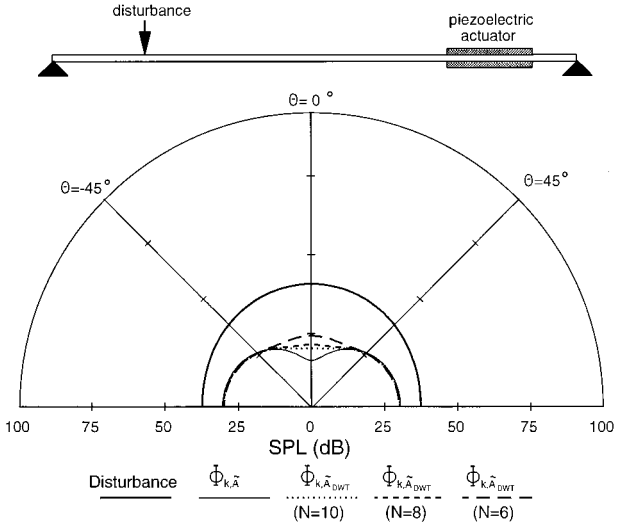


FIG. 14. Sound pressure level distribution for accelerometers, $f=350$ Hz.

TABLE II. Minimum number of sensors for different excitation frequencies. Note $\Delta\kappa_x = 1/L = 2.6315$ 1/m. $\kappa_{x,\mathrm{Nyq}} = (N/2)\Delta\kappa_x$.

f (Hz)	100	200	300	400	500	600	700	800
$\kappa = 2\pi f/c$	1.831	3.663	5.495	7.327	9.159	10.99	12.82	14.65
N_κ	2	3	4	4	5	6	6	7
$(N_\kappa - 1)\Delta\kappa_x$	2.631	5.263	7.894	7.894	10.52	13.15	13.15	15.78
min(N)	2	4	6	6	8	10	10	12
$\kappa_{x,\mathrm{Nyq}}$	2.631	5.263	7.894	7.894	10.52	13.15	13.15	15.78

function $\Phi_{\kappa,\tilde{A}}$ is chosen, the reduction of the radiated power is most effective throughout the frequency range. However, it cannot be implemented in practice.

(2) For either the CWT or DWT approach, the accelerometers based wave number sensing approach generally controls sound radiation more effectively than the PVDF sensors because the defined cost function for the accelerometers is more directly related to the radiated power than that for the PVDF sensors.

(3) For both accelerometer and PVDF based wave number sensing approaches, the DWT approach can achieve approximately the same control effectiveness as the CWT approach. The DWT approach can be practically implemented by using an array of evenly spaced accelerometers[28,29] or PVDF strip sensors.

(4) With the use of the PVDF sensors for wave number sensing approach, the control is ineffective for some off-resonance excitation cases. The optimal location of the control PZT actuator can be a critical factor. The optimal location of the PZT actuator, in conjunction with the use of the PVDF wave number sensing technique, is under investigation.

III. CONCLUSIONS

This study presents a novel wave number sensing approach which uses an array of PVDF strip sensors. By per-

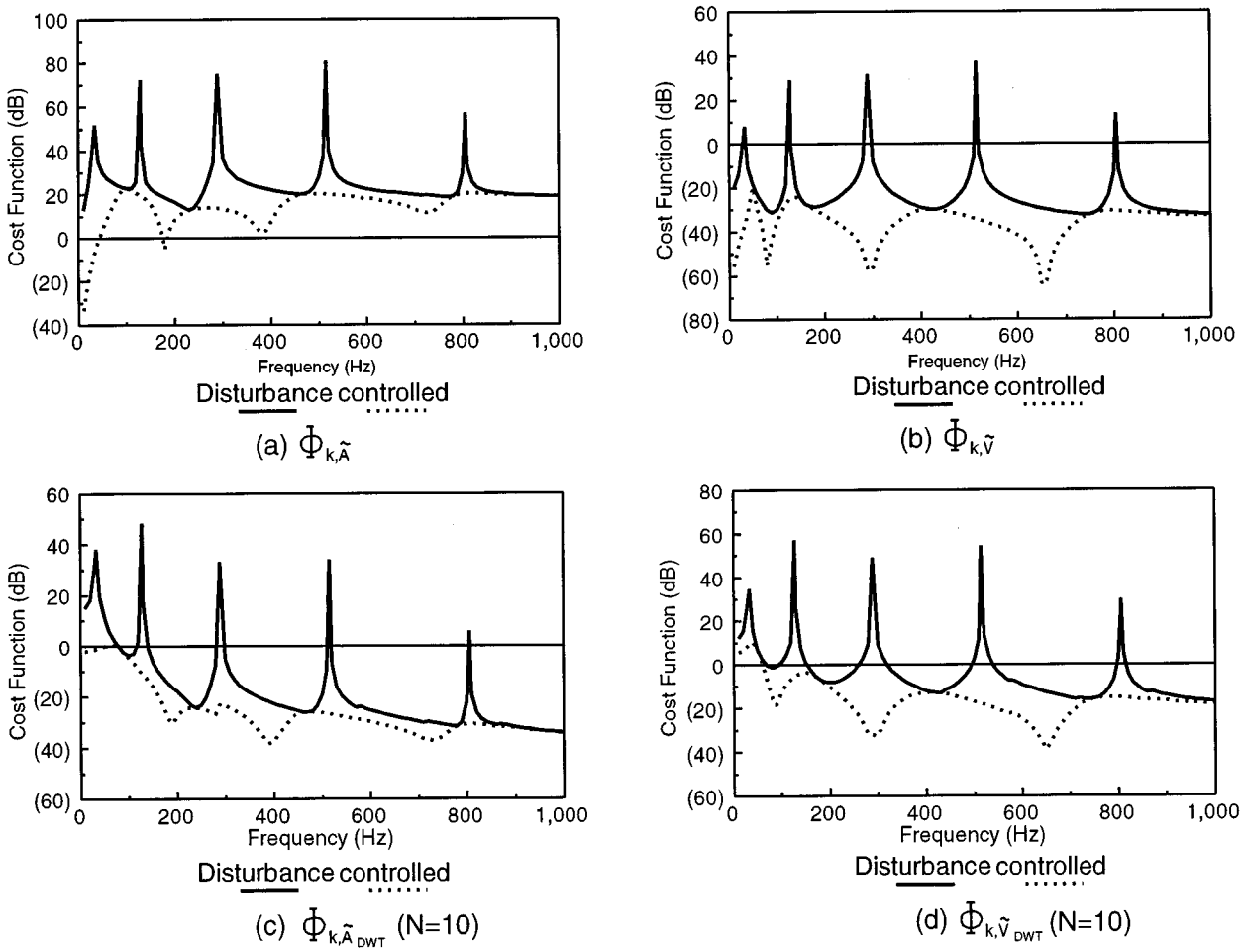


FIG. 15. Cost functions versus excitation frequency.

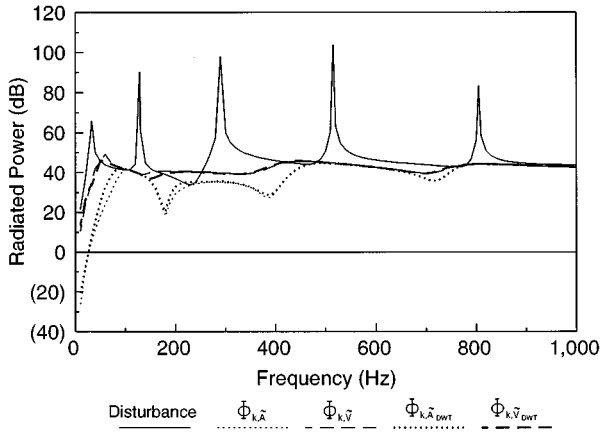


FIG. 16. Radiated power versus excitation frequency.

forming the discrete wave number transform on the sensor signals allows us to obtain the wave number transform functions. The cost function can then be defined as the sum of mean square values of PVDF voltage wave number transform functions in the supersonic region. The PVDF based wave number sensing approach, in conjunction with the use of the LMS feedforward control algorithm, is also applied to control the sound radiation of a baffled simply supported beam. The beam is assumed to be subjected to a harmonic point force disturbance; meanwhile, the PZT patches are applied as control actuators. Analysis results demonstrate that the PVDF based wave number sensing approach can aptly control sound radiation. Sound pressure level distribution and wave number analysis is also performed to examine the control mechanism. The accelerometers as error sensors for the wave number sensing approach, are also considered. Both continuous and discrete wave number transform approaches are compared as well. According to that comparison, the wave number transform function derived from the CWT approach can be well approximated by the DWT approach, thereby allowing the wave number transform functions to be practically measured by either an array of accelerometers or PVDF sensors. The accelerometers generally perform better sound radiation control than the PVDF sensors. In particular, for some off-resonance frequency excitation, the PVDF wave number sensing approach may not be efficient. To improve the control effectiveness for the off-resonance excitation, the optimal placement of the control actuator can be a feasible solution. This work not only enhances the sensing abilities of intelligent materials structure systems, but also provides a valuable reference for further efforts to control structural sound radiation control.

ACKNOWLEDGMENTS

The author would like to thank the National Science Council, Republic of China, for financially supporting this work under Contract No. NSC84-2212-E-020-005. The reviewers are also appreciated for their valuable comments and suggestions, particularly with the formulations of the negative wave numbers.

[1] J. Jones and C. R. Fuller, ''Active control of sound fields in elastic cylinders by multiple forces,'' AIAA J. **27**, 845–852 (1989).
[2] E. K. Dimitriadis, C. R. Fuller, and C. A. Rogers, ''Piezoelectric actuators for distributed vibration excitation of thin plates,'' J. Vib. Acoust. **113**, 100–107 (1991).
[3] B. T. Wang, E. K. Dimitriadis, and C. R. Fuller, ''Active control of structurally radiated noise using multiple piezoelectric actuators,'' AIAA J. **29**, 1802–1809 (1991).
[4] C. Liang, C. A. Rogers, and C. R. Fuller, ''Acoustic transmission and radiation analysis of adaptive shape memory alloy reinforced laminated plates,'' J. Sound Vib. **144**, 475–482 (1991).
[5] R. L. Clark and C. R. Fuller, ''Control of sound radiation with adaptive structures,'' J. Intell. Mater. Syst. Struct. **2**, 431–452 (1992).
[6] L. Meirovitch and S. Thangjitham, ''Active control of sound radiation pressure,'' J. Vib. Acoust. **112**, 237–244 (1990).
[7] B. T. Wang, ''Active control of far-field sound radiated by a beam with piezoelectric control transducers: Physical system analysis,'' Smart Mater. Struct. **3**, 476–484 (1994).
[8] S. J. Elliott, I. M. Stothers, and P. A. Nelson, ''A multiple error LMS algorithm and its application to the active control of sound and vibration,'' IEEE Trans. Acoust., Speech, Signal Process. **35**, 1423–1434 (1987).
[9] R. A. Burdisso and C. R. Fuller, ''Theory of feedforward controlled system eigenproperties,'' J. Sound Vib. **153**, 437–451 (1992).
[10] W. T. Baumann, W. R. Saunders, and H. H. Robertshaw, ''Active suppression of acoustic radiation from impulsively excited structures,'' J. Acoust. Soc. Am. **90**, 3202–3208 (1991).
[11] C. Deffayet and P. A. Nelson, ''Active control of low frequency harmonic sound radiated by a finite panel,'' J. Acoust. Soc. Am. **84**, 2192–2199 (1988).
[12] H. C. Lester and C. R. Fuller, ''Active control of propeller induced noise fields inside a flexible cylinder,'' AIAA J. **28**, 1374–1380 (1990).
[13] C. K. Lee and F. C. Moon, ''Modal sensors/actuator,'' J. Appl. Mech. **57**, 434–441 (1990).
[14] J. E. Hubbard, Jr., ''Distributed sensors and actuator for vibration control in elastic components,'' Noise-Con 87, 407–412 (1987).
[15] R. L. Clark, C. R. Fuller, and R. A. Burdisso, ''Design approaches for shaping polyvinylidene fluoride sensors in active structural acoustic control,'' J. Intell. Mater. Syst. Struct. **4**, 354–365 (1993).
[16] C. R. Fuller and J. D. Jones, ''Influence of sensor and actuator location on the performance of active control systems,'' 87-WA/NCA-9, ASME Annual Meeting, Boston, MA (1987).
[17] B. T. Wang and C. R. Fuller, ''Evaluation of different forms of cost functions in the design of active structural-acoustic control systems,'' Noise-Con 87 (1991).
[18] B. T. Wang, ''The performance of accelerometers, microphones and PVDF sensors in active structural acoustic control: Theoretical analysis,'' Chin. J. Mech. **10**, 205–212 (1993).
[19] R. L. Clark and C. R. Fuller, ''Modal sensing of efficient acoustic radiators with polyvinylidene fluoride distributed sensors in active structural acoustic control approaches,'' J. Acoust. Soc. Am. **91**, 3321–3329 (1992).
[20] R. L. Clark and C. R. Fuller, ''Optimal placement of piezoelectric actuators and polyvinylidene fluoride error sensors in active structural acoustic control approaches,'' J. Acoust. Soc. Am. **92**, 1521–1533 (1992).
[21] Y. Gu, R. L. Clark, C. R. Fuller, and A. C. Zander, ''Experiments on active control of plate vibration using piezoelectric actuators and polyvinylidene fluoride (PVDF) modal sensors,'' J. Vib. Acoust. **116**, 303–308 (1994).
[22] S. D. Snyder, C. H. Hansen, and N. Tanaka, ''Shaped vibration sensors for feedforward control of structural radiation,'' Proceedings of the Second Conference on the Recent Advances in Active Control of Sound and Vibration, 177–188 (1993).
[23] F. Charette, C. Guigou, and A. Berry, ''Development of volume velocity sensors for plates using PVDF film,'' Proceedings of ACTIVE 95, 241–252 (1995).
[24] F. Fahy, *Sound and Structural Vibration: Radiation, Transmission, and Response* (Academic, Orlando, FL, 1985), pp. 72–81.
[25] B. T. Wang and C. R. Fuller, ''Near-field pressure, intensity, and wave number distributions for active structural acoustic control of plate radiation: Theoretical analysis,'' J. Acoust. Soc. Am. **92**, 1489–1498 (1992).
[26] C. R. Fuller and R. A. Burdisso, ''A wave number domain approach to the active control of structure-borne sound,'' J. Sound Vib. **148**, 335–360 (1991).
[27] R. L. Clark and C. R. Fuller, ''A model reference approach for imple-

menting active structural acoustic control,'' J. Acoust. Soc. Am. **92**, 1534–1544 (1992).
[28] J. P. Maillard and C. R. Fuller, ''Advanced time domain wave-number sensing for structural acoustic systems. I. Theory and design,'' J. Acoust. Soc. Am. **95**, 3252–3261 (1994).
[29] J. P. Maillard and C. R. Fuller, ''Advanced time domain wave-number sensing for structural acoustic systems. II. Active radiation control of a simply supported beam,'' J. Acoust. Soc. Am. **95**, 3261–3272 (1994).
[30] B. L. Scott and S. D. Sommerfeldt, ''Estimating acoustic radiation from a Bernoulli–Euler beam using shaped polyvinylidene flouride film,'' J. Acoust. Soc. Am. **101**, 3475–3485 (1997).
[31] L. Cremer and M. Heckl, *Structure-Borne Sound: Structural Vibrations and Sound Radiation at Audio Frequencies* (Springer-Verlag, Berlin, 1988), pp. 528–529.
[32] B. T. Wang, ''Characterization of transfer functions for piezoceramic and conventional transducers,'' J. Intell. Mater. Syst. Struct. **7**, 390–398 (1996).
[33] B. T. Wang, ''Feasibility analysis of using piezoceramic transducers for cantilever beam modal testing,'' Smart Mater. Struct. **6**, 106–116 (1997).
[34] B. T. Wang, *Active Structural Acoustic Control of Intelligent Material Structure System* (NSC Report No. NSC82-0410-E-020-001, Taiwan, R.O.C., 1993), pp. 16–23.
[35] Piezo Systems, Inc., *Product Catalog* (Cambridge, MA, 1990).
[36] Pennwalt Corporation, *Piezo Film Sensor Application Notes* (Valley Forge, PA, 1990).
[37] C. R. Fuller, C. H. Hansen, and S. D. Snyder, ''Active control of sound radiation from a vibrating rectangular panel by sound sources and vibration inputs: An experimental comparison,'' J. Sound Vib. **145**, 195–215 (1991).

Analytical study of different approaches for active control of sound transmission through double walls

J. Pan and C. Bao
Department of Mechanical and Materials Engineering, The University of Western Australia, Nedlands, WA 6907, Australia

(Received 30 April 1997; accepted for publication 12 December 1997)

Sound transmission through double plate structures into an enclosure and its active attenuation are modeled and analyzed in this paper. Using the modal acoustic transfer impedance and mobility matrices, the excitation and interaction in the double plate sound transmission system can be described with clear physical significance. Three active control arrangements are considered for a double wall sound transmission model. They are: (1) to directly attenuate the room modes by using one acoustic control source in the room; (2) to control the sound radiation into the room using a vibrational actuator on the radiating structure; and (3) to block the noise transmission path by inserting one acoustic control source within double wall partition. The analytical solution of the system response is used to explain the mechanisms of attenuation associated with these three control arrangements. The effectiveness and performance of the control system corresponding to each arrangement are discussed. © *1998 Acoustical Society of America.* [S0001-4966(98)00304-X]

PACS numbers: 43.50.Ki, 43.55.Rg [GD]

INTRODUCTION

Active control of sound transmission through a single plate has been the subject of several analytical and experimental investigations.[1–5] Previous work in this area demonstrated that noise transmission can be reduced by applying control forces on partition structures. After the examination of the system performance before and after control, two interesting mechanisms involved in the active structural acoustic control (ASAC) of sound transmission into a room[1–3] were discovered. Numerical and experimental results demonstrated that if the system response is dominated by a panel controlled mode (where the transmitted energy is carried mainly by one dominating panel mode), minimum potential energy is obtained by suppressing the panel-vibration amplitude of the controlling mode. If the system response is dominated by a room-controlled mode (where energy is transmitted from several plate modes into one dominating room mode), the control force is used to adjust (or rearrange) the plate-velocity distribution so that the radiated power from the plate into the room is reduced. In the latter case, there may be an increase in panel-vibration level. Similar mechanisms were also observed in the ASAC of structural sound radiation into a free space.[6] Further work on ASAC of sound transmission has discussed the selection of optimal actuator locations for the effective use of the two control mechanisms in order to achieve the maximum sound transmission loss.[7]

Recently, active noise control has been used to increase the sound transmission loss of double wall structures. As the double wall sound transmission system has many subsystems (two plates, one partition cavity, and a room), several arrangements of control actuators can be selected for the same purpose: minimization of acoustic potential energy in the room. In this paper three control arrangements are considered. They are: (1) to directly attenuate the sound pressure field in the room using one acoustic control source in the room; (2) to control the sound radiation into the room using a vibrational control source on the radiating plate; and (3) to block the noise transmission path by inserting one acoustic control source between the double walls. Details of the previous experimental investigation on active control of sound transmission through double plates has been reviewed in an accompanying experimental paper.[8] The existing analytical work[9,10] related to active control of sound transmission through double walls using acoustic control sources in the double wall cavity are largely for the demonstration of possible noise reduction and the comparison with experimental results. In their paper,[9] Sas *et al.* also used experimental modal analysis to observe the system modal response before and after control and used the mechanisms of fluid-structure interaction of double panel structures to explain the experimental results.

This paper presents a theoretical model for the analysis and design of an actively controlled double wall system with all possible arrangement of control sources. Although the discussion considers one control source at a time, the model allows for multiple control inputs and simultaneous use of different arrangements of control sources. By examining the minimum acoustic potential energy in the room due to an optimal control input, the condition for achieving a significant reduction of the potential energy was obtained. For each control source arrangement, the examination of how to satisfy the condition enables the understanding of the control mechanisms. Because the analysis uses the modal acoustic transfer impedance and mobility matrices to express the system response, it is possible to reveal the control mechanisms implied in the mathematical expressions of the conditions for each arrangement.

I. DESCRIPTION OF THE CONTROL SYSTEM

Figure 1 shows a double wall partition model, where an incident sound wave p_{in}, with elevation angle ϕ and azimuth angle θ,

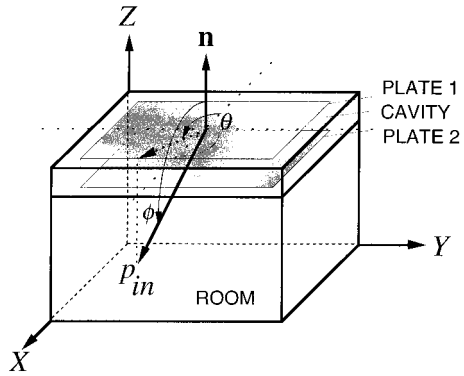


FIG. 1. A double wall model for the active control of sound transmission.

$$p_{\text{in}}(x,y,z;t)=A\ \exp(j\omega t-j\ \cos\ \phi kz - j\ \sin\ \phi\ \cos\ \theta kx-j\ \sin\ \phi\ \sin\ \theta ky), \tag{1}$$

excites the top plate (plate 1). The local coordinates of the cavity and room in the Z direction are denoted respectively as z_c and z_r. The displacement $W^{(1)}$ of plate 1 is described by

$$-m_1\omega^2W^{(1)}+D_1\nabla^4W^{(1)}=-p^{(1)}(z_c=0)+p_{\text{ext}}, \tag{2}$$

where m_1 and D_1 are, respectively, the surface mass and bending stiffness of plate 1. p_{ext} is the external sound pressure on plate 1, which includes the block sound pressure of the incident sound wave and the back reaction of the radiated sound waves. $p^{(1)}$ is the cavity sound pressure. As the effect of sound radiation on the plate vibration is negligible, we have

$$p_{\text{ext}}=2A\ \exp(-j\ \sin\ \phi\ \cos\ \theta kx-j\ \sin\ \phi\ \sin\ \theta ky). \tag{3}$$

The sound pressure in the cavity is composed of contributions from the acoustical control sources in the cavity, and the vibration of plates 1 and 2:

$$k^2p^{(1)}+\nabla^2p^{(1)}=-j\omega\rho_0\sum_{i_c}\ q_{i_c}^{(1)}-\rho_0\omega^2[W^{(2)}\delta(z_c-L_{z_c}) - W^{(1)}\delta(z_c)], \tag{4}$$

where $q_{i_c}^{(1)}=Q_{i_c}^{(1)}\delta(\mathbf{r}_c-\mathbf{r}_{i_c})$ is the source strength of the i_cth control source in the cavity. L_{z_c} is the distance between the two plates.

The displacement of plate 2 is determined by following equation:

$$-m_2\omega^2W^{(2)}+D_2\nabla^4W^{(2)}=p^{(1)}(z_c=L_{z_c})-p^{(2)}(z_r=0) + p_{\text{con}}^{(2)}, \tag{5}$$

where $p_{\text{con}}^{(2)}=\Sigma_{i_{p2}}F_{i_{p2}}^{(2)}\delta(\sigma^{(2)}-\sigma_{i_{p2}}^{(2)})$ is the pressure due to point control forces at locations $\sigma_{i_{p2}}^{(2)}$ of plate 2. $p^{(2)}$ is the sound pressure in the room and is described by

$$k^2p^{(2)}+\nabla^2p^{(2)}=-j\omega\rho_0\sum_{i_r}\ q_{i_r}^{(2)} + \rho_0\omega^2W^{(2)}\delta(z_r)H_W(\sigma^{(2)}), \tag{6}$$

where $q_{i_r}=Q_{i_r}\delta(\mathbf{r}_r-\mathbf{r}_{i_r})$ is the source strength of the i_rth control source in the room and $H_W(\sigma^{(2)})$ is a spatial window function specifying the geometrical area of plate 2.

In this analysis, the plate velocities $\nu^{(1)}=j\omega W^{(1)}$ and $\nu^{(2)}=j\omega W^{(2)}$ are expanded as

$$\nu^{(1)}=[\mathbf{S}_{\mathbf{M}_1}^{(1)}][\mathbf{V}_{\mathbf{M}_1}^{(1)}], \tag{7}$$

$$\nu^{(2)}=[\mathbf{S}_{\mathbf{M}_2}^{(2)}][\mathbf{V}_{\mathbf{M}_2}^{(2)}], \tag{8}$$

where $[\mathbf{V}_{\mathbf{M}_1}^{(1)}]=[V_1^{(1)},V_2^{(1)},\ldots,V_{M_1}^{(1)}]^T$ and $[S_{\mathbf{M}_1}^{(1)}]=[S_1^{(1)},S_2^{(1)},\ldots,S_{M_1}^{(1)}]$ are, respectively, the modal amplitude and modal shape vectors of the velocity in plate 1. The sound pressure in the cavity and room can be expressed in terms of the cavity and room modes as well:

$$p^{(1)}=[\Phi_{N_2}^{(1)}][\mathbf{P}_{\mathbf{N}_1}^{(1)}], \tag{9}$$

$$p^{(2)}=[\Phi_{N_2}^{(2)}][\mathbf{P}_{\mathbf{N}_2}^{(2)}], \tag{10}$$

where $[\mathbf{P}_{\mathbf{N}_1}^{(1)}]=[P_1^{(1)},P_2^{(1)},\ldots,P_{N_1}^{(1)}]^T$ and $[\Phi_{N_1}^{(1)}]=[\Phi_1^{(1)},\Phi_2^{(1)},\ldots,\Phi_{N_1}^{(1)}]$ are, respectively, the modal amplitude and mode shape vectors of the sound field in the cavity.

Substituting Eqs. (7)–(10) into Eqs. (2) and (4)–(6), and using the orthogonal properties of the modal shape functions, we obtained the modal equations for each subsystem in matrix form. For example, the modal amplitudes of the sound pressure in the cavity are expressed as

$$[\mathbf{P}_{\mathbf{N}_1}^{(1)}]=[Z_{A_2}^{(1)}][\mathbf{V}_{\mathbf{M}_2}^{(2)}]-[Z_{A_1}^{(1)}][\mathbf{V}_{\mathbf{M}_1}^{(1)}]+[Z_{M_c}^{(1)}][\mathbf{V}_{\mathbf{M}_c}]. \tag{11}$$

$[Z_{A_1}^{(1)}]$ is the modal acoustic transfer impedance matrix ($N_1\times M_1$) from plate 1 to the cavity:

$$[Z_{A_1}^{(1)}]=\rho_0c_0\begin{bmatrix} B_{1,1}^{(1,1)}/\chi_{A_1}^{(1)} & \cdots & B_{1,M_1}^{(1,1)}/\chi_{A_1}^{(1)} \\ \vdots & \ddots & \\ B_{N_1,1}^{(1,1)}/\chi_{A_{N_1}}^{(1)} & \cdots & B_{N_1,M_1}^{(1,1)}/\chi_{A_{N_1}}^{(1)} \end{bmatrix}, \tag{12}$$

where

$$B_{I,J}^{(1,1)}=B_{J,I}^{(1,1)}=\frac{1}{A_f}\int_{A_f}S_J^{(1)}\Phi_I^{(1)}(z_c=0)\,d\sigma \tag{13}$$

is the modal coupling coefficient and

$$\chi_{A_I}^{(1)}=j\ \frac{M_{A_I}^{(1)}}{\rho_0kA_f}\ (k_{A,I}^2+j\ \eta_{A_1l}k_{A_1I}^2-k^2). \tag{14}$$

In Eq. (14), $M_{A_1}^{(1)}=\rho_0\Lambda_{Al}^{(1)}V_0^{(1)}$ is the cavity modal mass, where $V_a^{(1)}$ is the cavity volume, $\Lambda_{Al}^{(1)}=(1/V_0^{(1)})\times\int_{V_0^{(1)}}[\Phi_I^{(1)}]^2dv$. k_{A_1I} and η_{A_1I} are, respectively, the modal wave number and loss factor of the Ith cavity mode. A_f is the surface area of the plates. The loss factors of the cavity modes are related to the acoustical impedance of cavity walls. $[Z_{A_2}^{(1)}]$ in Eq. (11) is the modal acoustic transfer impedance ($N_1\times M_2$) from plate 2 to the cavity:

$$[Z_{A_2}^{(1)}]=\rho_0 c_0\begin{bmatrix} B_{1,1}^{(2,1)}/\chi_{A_1}^{(1)} & \cdots & B_{1,M_1}^{(2,1)}/\chi_{A_1}^{(1)} \\ \vdots & \ddots & \\ B_{N_1,1}^{(2,1)}/\chi_{A_{N_1}}^{(1)} & \cdots & B_{N_1,M_1}^{(2,1)}/\chi_{A_{N_1}}^{(1)} \end{bmatrix}, \quad (15)$$

where the coupling coefficients are

$$B_{I,J}^{(2,1)}=B_{J,I}^{(2,1)}=\frac{1}{A_f}\int_{A_f} S_j^{(2)}\Phi_I^{(1)}(z_c=L_{z_c})d\sigma. \quad (16)$$

$[\mathbf{V}_{\mathbf{M}_c}]$ in Eq. (11) is the control source vector $(M_c\times 1)$ in the cavity:

$$[\mathbf{V}_{\mathbf{M}_c}]=-[Q_1^{(1)},Q_2^{(1)},\ldots,Q_{M_c}^{(1)}]^T/A_f \quad (17)$$

and the corresponding modal acoustic transfer impedance matrix $(N_1\times M_c)$ is

$$[Z_{M_C}^{(1)}]=\rho_0 c_0\begin{bmatrix} \Phi_1^{(1)}(r_1^{(1)})/\chi_{A_1}^{(1)} & \cdots & \Phi_1^{(1)}(r_{M_c}^{(1)})/\chi_{A_1}^{(1)} \\ \vdots & \ddots & \\ \Phi_{N_1}^{(1)}(r_1^{(1)})/\chi_{A_{N_1}}^{(1)} & \cdots & \Phi_{N_1}^{(1)}(r_{M_c}^{(1)})/\chi_{A_{N_1}}^{(1)} \end{bmatrix}. \quad (18)$$

The pressure in the room is obtained as

$$[\mathbf{P}_{\mathbf{N}_2}^{(2)}]=-[Z_A^{(2)}][\mathbf{V}_{\mathbf{M}_2}^{(2)}]+[Z_{M_r}^{(2)}][\mathbf{V}_{\mathbf{M}_r}], \quad (19)$$

where $[Z_A^{(2)}]$ is the modal acoustic transfer impedance matrix $(N_2\times M_2)$ from plate 2 to the room:

$$[Z_A^{(2)}]=\rho_0 c_0\begin{bmatrix} B_{1,1}^{(2,2)}/\chi_{A_1}^{(2)} & \cdots & B_{1,M_2}^{(2,2)}/\chi_{A_1}^{(2)} \\ \vdots & \ddots & \\ B_{N_2,1}^{(2,2)}/\chi_{A_{N_2}}^{(2)} & \cdots & B_{N_2,M_2}^{(2,2)}/\chi_{A_{N_2}}^{(2)} \end{bmatrix}, \quad (20)$$

here

$$B_{I,J}^{(2,2)}=B_{J,I}^{(2,2)}=\frac{1}{A_f}\int_{A_f} S_J^{(2)}\Phi_I^{(2)}(z_r=0)d\sigma^{(2)}, \quad (21)$$

and

$$\chi_{A_1}^{(2)}=j\,\frac{M_{A_1}^{(2)}}{\rho_0 k A_f}\,(k_{A_2 I}^2+j\,\eta_{A_2 I}k_{A_2 I}^2-k^2). \quad (22)$$

In Eq. (22), $M_{A_1}^{(2)}=\rho_0\Lambda_{AI}^{(2)}V_0^{(2)}$ is the room modal mass where $V_0^{(2)}$ is the room volume, $\Lambda_{AI}^{(2)}=(1/V_0^{(2)})\times\int_{V_0^{(1)}}[\Phi_I^{(2)}]^2 dv$, and $k_{A_2 I}$ and $\eta_{A_2 I}$ are, respectively, the modal wave number and loss factor of the Ith room mode.

$[Z_{M_r}^{(2)}]$ is the modal transfer impedance matrix $(N_2\times M_r)$ from the acoustical control sources in the room to $[\mathbf{P}_{\mathbf{N}_2}^{(2)}]$:

$$[Z_{M_r}^{(2)}]=\rho_0 c_0\begin{bmatrix} \Phi_1^{(2)}(\mathbf{r}_1^{(2)})/\chi_{A_1}^{(2)} & \cdots & \Phi_1^{(2)}(\mathbf{r}_{\mathbf{M}_r}^{(2)})/\chi_{A_1}^{(2)} \\ \vdots & \ddots & \\ \Phi_{N_2}^{(2)}(\mathbf{r}_1^{(2)})/\chi_{A_{N_2}}^{(2)} & \cdots & \Phi_{N_2}^{(2)}(\mathbf{r}_{\mathbf{M}_r}^{(2)})/\chi_{A_{N_2}}^{(2)} \end{bmatrix} \quad (23)$$

and the velocity vector $(M_r\times 1)$ of the acoustical control sources is

$$[\mathbf{V}_{\mathbf{M}_r}]=-[Q_1^{(2)},Q_2^{(2)},\ldots,Q_{M_r}^{(2)}]^T/A_f. \quad (24)$$

The velocity components of plate 1 are related to $[\mathbf{P}_{\mathbf{N}_1}^{(1)}]$ and external sound pressure:

$$[\mathbf{V}_{\mathbf{M}_1}^{(1)}]=-[Y_{P_1}^{(1)}][\mathbf{P}_{\mathbf{N}_1}^{(1)}]+[Y_{M_1}^{(\mathrm{ext})}][\mathbf{P}_{\mathbf{M}_1}^{(\mathrm{ext})}], \quad (25)$$

where $[Y_{P_1}^{(1)}]$ is the modal acoustic transfer mobility matrix $(M_1\times N_1)$ from the cavity to plate 1:

$$[Y_{P_1}^{(1)}]=-\frac{1}{\rho_0 c_0}\begin{bmatrix} B_{1,1}^{(1,1)}/\chi_{P_1}^{(1)} & \cdots & B_{1,N_1}^{(1,1)}/\chi_{P_1}^{(1)} \\ \vdots & \ddots & \\ B_{M_1,1}^{(1,1)}/\chi_{P_{M_1}}^{(1)} & \cdots & B_{M_1,N_1}^{(1,1)}/\chi_{P_{M_1}}^{(1)} \end{bmatrix}, \quad (26)$$

where

$$\chi_{P_j}^{(1)}=j\,\frac{M_{P_J}^{(1)}}{\rho_0 k A_f}\,(k_{P_1 J}^2+j\,\eta_{P_1 J}k_{P_1 J}^2-k^2). \quad (27)$$

In Eq. (27), $M_{P_j}^{(1)}=m_1\Lambda_{P_J}^{(1)}A_f$ is the modal mass of plate 1 where $\Lambda_{P_J}^{(1)}=(1/A_f)\int_{A_f}[S_j^{(1)}]^2\,d\sigma$. $k_{P_i J}$ and $\eta_{P_1 J}$ are, respectively, the modal ''wave number'' and loss factor of the Jth mode of plate 1. The loss factors of the plate modes are related to the imaginary part of the complex Young's modulus.

The modal acoustic transfer mobility matrix $(M_1\times M_1)$ from the external sound pressure to plate 1 is

$$[Y_{M_1}^{(\mathrm{ext})}]=-\frac{1}{\rho_0 c_0}\begin{bmatrix} 1/\chi_{P_1}^{(1)} & \cdots & 0 \\ \vdots & \ddots & \\ 0 & \cdots & 1/\chi_{P_{M_1}}^{(1)} \end{bmatrix}. \quad (28)$$

The modal pressure components due to the external sound pressure is expressed by an $(M_1\times 1)$ vector:

$$[\mathbf{P}_{\mathbf{M}_1}^{(\mathrm{ext})}]=[P_1^{(\mathrm{ext})},P_2^{(\mathrm{ext})},\ldots,P_{M_1}^{(\mathrm{ext})}]^T, \quad (29)$$

where

$$P_J^{(\mathrm{ext})}=\frac{1}{A_f}\int_{A_f} p_{\mathrm{ext}}S_J^{(1)}d\sigma^{(1)}. \quad (30)$$

The velocity in plate 2 is due to $[\mathbf{P}_{\mathbf{N}_1}^{(1)}]$, $[\mathbf{P}_{\mathbf{N}_2}^{(2)}]$, and to the pressure from the vibration control actuators $([\mathbf{P}_{N_{p2}}^{(2)}])$ on the panel surface:

$$[\mathbf{V}_{\mathbf{M}_2}^{(2)}]=[Y_{P_1}^{(2)}][\mathbf{P}_{\mathbf{N}_1}^{(1)}]-[Y_{P_2}^{(2)}][\mathbf{P}_{\mathbf{N}_2}^{(2)}]+[Y_{M_2,M_{p2}}^{(c)}][\mathbf{P}_{\mathbf{N}_{p2}}^{(2)}], \quad (31)$$

where $[Y_{P_1}^{(2)}]$ $(M_2\times N_1)$ and $[Y_{P_2}^{(2)}]$ $(M_2\times N_2)$ are, respectively, the modal acoustic transfer mobility matrices from the cavity to plate 2 and from the room to plate 2:

$$[Y_{P_1}^{(2)}]=-\frac{1}{\rho_0 c_0}\begin{bmatrix} B_{1,1}^{(2,1)}/\chi_{P_1}^{(2)} & \cdots & B_{1,N_1}^{(2,1)}/\chi_{P_1}^{(2)} \\ \vdots & \ddots & \\ B_{M_2,1}^{(2,1)}/\chi_{P_{M_2}}^{(2)} & \cdots & B_{M_2,N_1}^{(2,1)}/\chi_{P_{M_2}}^{(2)} \end{bmatrix}, \quad (32)$$

$$[Y_{P_2}^{(2)}]=-\frac{1}{\rho_0 c_0}\begin{bmatrix} B_{1,1}^{(2,2)}/\chi_{P_1}^{(2)} & \cdots & B_{1,N_2}^{(2,2)}/\chi_{P_1}^{(2)} \\ \vdots & \ddots & \\ B_{M_2,1}^{(2,2)}/\chi_{P_{M_2}}^{(2)} & \cdots & B_{M_2,N_2}^{(2,2)}/\chi_{P_{M_2}}^{(2)} \end{bmatrix}, \tag{33}$$

where

$$\chi_{P_2J}^{(2)}=j\,\frac{M_{P_J}^{(2)}}{\rho_0 k A_f}\,(k_{P_2J}^2+j\,\eta_{P_2J}k_{P_2J}^2-k^2). \tag{34}$$

$[Y_F^{(c)}]$ is the modal acoustic transfer mobility matrix ($M_2 \times M_{p^2}$) from the control force on plate 2 top the velocity of plate 2:

$$[Y_{M_2,M_{p2}}^{(c)}]=-\frac{1}{\rho_0 c_0}\times\begin{bmatrix} S_1^{(2)}(\sigma_1^{(2)})/\chi_{P_1}^{(2)} & \cdots & S_1^{(2)}(\sigma_{M_{p2}}^{(2)})/\chi_{P_1}^{(2)} \\ \vdots & \ddots & \\ S_{M_2}^{(2)}(\sigma_1^{(2)})\chi_{P_{M_2}}^{(2)} & \cdots & S_{M_2}^{(2)}(\sigma_{M_{p2}}^{(2)})/\chi_{P_{M_2}}^{(2)} \end{bmatrix}, \tag{35}$$

and $[\mathbf{P}_{N_{p2}}^{(2)}]$ is the surface pressure vector due to point control forces:

$$[\mathbf{P}_{N_{p2}}^{(2)}]=[F_1^{(2)},F_2^{(2)},\ldots,F_{M_{p2}}^{(2)}]^T/A_f. \tag{36}$$

II. ANALYSIS OF CONTROL MECHANISMS

Equations (11), (19), (25), and (31) present 4 coupled matrix equations from which the complete system response $[\mathbf{P}_{\mathbf{N}_1}^{(1)}]$, $[\mathbf{P}_{\mathbf{N}_2}^{(2)}]$, $[\mathbf{V}_{\mathbf{M}_1}^{(1)}]$, and $[\mathbf{V}_{\mathbf{M}_2}^{(2)}]$ can be determined. In addition, for any given cost function, the optimal feedforward control law for the control inputs $[\mathbf{V}_{\mathbf{M}_c}]$, $[\mathbf{V}_{\mathbf{M}_r}]$, or $[\mathbf{P}_{\mathbf{N}_2}^{(2)}]$ can be obtained when cost function (acoustic potential energy in the room) is minimized. For the simplicity of analysis, only one control source is used at a time. However, the results of the analysis will shed light on the performance and mechanisms of the control system with multiple control sources.

A. Acoustic control source in room

When only one acoustic source is used in the room to control the transmitted sound field, Eq. (19) can be expressed as

$$[\mathbf{P}_{\mathbf{N}_2}^{(2)}]=[\mathbf{P}_{\mathbf{N}_2}^{(\mathbf{p})}]+[\mathbf{Z}_{\mathbf{N}_2}^{(\mathbf{c})}]V_r^{(c)}, \tag{37}$$

where the vector $[\mathbf{P}_{\mathbf{N}_2}^{(p)}]$ is the modal components of the uncontrolled sound pressure in the room due to primary boundary structural vibration:

$$P_I^{(p)}=-\frac{\rho_0 c_0}{\chi_{A_2I}^{(2)}}\sum_{J=1}^{M_2} B_{I,J}^{(2,2)}V_J^{(2)} \tag{38}$$

and $[\mathbf{Z}_{\mathbf{N}_2}^{(c)}]$ is the impedance vector for the point sound source in the room:

$$Z_I^{(c)}=-\frac{\rho_0 c_0}{\chi_{AI}^{(2)}}\,\phi_I^{(2)}(r_c^{(2)}). \tag{39}$$

Because all the elements in $[\mathbf{P}_{\mathbf{N}_2}^{(2)}]$ are orthogonal to each other in terms of their modal shape functions, the mechanism involved in this arrangement is purely modal suppression (i.e., by attenuating the amplitude of each individual element). It is well know that the total time-averaged acoustic potential energy in the room:[11]

$$E_p^{(2)}=\frac{V_0^{(2)}}{4\rho_0 c_0^2}\{(V_r^{(c)})^H[\mathbf{Z}_{\mathbf{N}_2}^{(\mathbf{c})}]^H[\Lambda_{N_2}^{(2)}][\mathbf{Z}_{\mathbf{N}_2}^{(\mathbf{c})}]V_r^{(c)} + (V_r^{(c)})^H[\mathbf{Z}_{\mathbf{N}_2}^{(\mathbf{c})}]^H[\Lambda_{N_2}^{(2)}][\mathbf{P}_{\mathbf{N}_2}^{(\mathbf{p})}]+[\mathbf{P}_{\mathbf{N}_2}^{(\mathbf{p})}]^H[\Lambda_{N_2}^{(2)}] \times[\mathbf{Z}_{\mathbf{N}_2}^{(\mathbf{c})}]V_r^{(c)}+[\mathbf{P}_{\mathbf{N}_2}^{(p)}]^H[\Lambda_{N_2}^{(2)}][\mathbf{P}_{\mathbf{N}_2}^{(\mathbf{p})}]\},$$

has a minimum value corresponding to an optimal control input $(V_r^{(c)})_{\text{opt}}$:

$$\frac{(E_p^{(2)})_{\min}}{V_0^{(2)}/4\rho_0 c_0^2}=[\mathbf{P}_{\mathbf{N}_2}^{(\mathbf{p})}]^H[\Lambda_{N_2}^{(2)}][\mathbf{P}_{\mathbf{N}_2}^{(\mathbf{p})}] - \frac{|[\mathbf{P}_{\mathbf{N}_2}^{(\mathbf{p})}]^H[\Lambda_{N_2}^{(2)}][\mathbf{Z}_{\mathbf{N}_2}^{(\mathbf{c})}]|^2}{[\mathbf{Z}_{\mathbf{N}_2}^{(\mathbf{c})}]^H[\Lambda_{N_2}^{(2)}][\mathbf{Z}_{\mathbf{N}_2}^{(\mathbf{c})}]}, \tag{40}$$

where $[\Lambda_{N_2}^{(2)}]$ is a diagonal normalization matrix of room modes. Equation (40) only presents the possible minimum potential energy as a result of control. To achieve a large reduction in the potential energy [i.e., $(E_p^{(2)})_{\min}/E_p^{(2)}\to 0$], the following relationship needs to be satisfied:

$$|[\mathbf{P}_{\mathbf{N}_2}^{(\mathbf{p})}][\Lambda_{N_2}^{(2)}][\mathbf{Z}_{\mathbf{N}_2}^{(\mathbf{c})}]^H|^2\cong\{[\mathbf{P}_{\mathbf{N}_2}^{(\mathbf{p})}][\Lambda_{N_2}^{(2)}][\mathbf{P}_{\mathbf{N}_2}^{(\mathbf{p})}]^H\}\{[\mathbf{Z}_{\mathbf{N}_2}^{(\mathbf{c})}] \times[\Lambda_{N_2}^{(2)}][\mathbf{Z}_{\mathbf{N}_2}^{(\mathbf{c})}]^H\}. \tag{41}$$

Mathematically, the above relationship is satisfied only if the vectors $[\mathbf{P}_{\mathbf{N}_2}^{(p)}]$ and $[\mathbf{Z}_{\mathbf{N}_2}^{(c)}]$ are in proportion:

$$[\mathbf{P}_{\mathbf{N}_2}^{(\mathbf{p})}]=c[\mathbf{Z}_{\mathbf{N}_2}^{(\mathbf{c})}], \tag{42}$$

where c is a complex constant with velocity unit. Physically, the satisfaction of Eq. (42) requires similar excitation of all modes from the primary and control sources. Successful attenuation of low-frequency potential energy in a room by placing a control point source near the primary point source is an example where Eq. (42) is approximately satisfied. However, in this analysis the elements in $[\mathbf{Z}_{\mathbf{N}_2}^{(\mathbf{c})}]$ depend on the mode shape functions evaluated at the control source location, while that in $[\mathbf{P}_{\mathbf{N}_2}^{(\mathbf{p})}]$ are due to the coupling between the room and plate modes. As a result, Eq. (42) can only be satisfied at those frequencies where the primary sound field is dominated by a single room mode and the modal overlap in the room is low. For this case, if the control source is located so as to excite the dominating mode only, Eq. (42) will be approximately satisfied and large reduction of potential energy at this frequency is expected.

In practical rooms, only at very low frequencies is the primary sound field dominated by a single mode and the contribution of nonresonance modes small. Above this fre-

quency range, increases in modal density and damping often make it difficult for control source to generate the required modal components without exciting others. Numerical results[12] from the active control of sound radiation from a steel panel into in a highly damped room show that a large reduction of potential energy is only achieved at the very low frequencies. In the low-frequency range, the arrangement with the acoustic control source near the radiating panel gives larger reduction than that with a control source located at the far corner of the room.

B. Control force applied to plate 2

When only one point vibrational control actuator is applied to plate 2 to control the transmitted sound field, Eq. (19) can be expressed as

$$[\mathbf{P}_{\mathbf{N}_2}^{(\mathbf{2})}]=-[Z_A^{(2)}]\{[\mathbf{V}_{\mathbf{M}_2}^{(\mathbf{P})}]+[\alpha_{M_2}^{(2)}]^{-1}[Y_{M_2,M_{p2}}^{(c)}]P_{\sigma_2}^{(C)}\}, \tag{43}$$

where the primary velocity vector in plate 2 is

$$[\mathbf{V}_{\mathbf{M}_2}^{(\mathbf{P})}]=-[\alpha_{M_2}^{(2)}]^{-1}[Y_{P_1}^{(2)}][Z_{A_1}^{(1)}][\alpha_{M_1}^{(1)}]^{-1}[Y_{M_1}^{(\mathrm{ext})}][\mathbf{P}_{\mathbf{M}_1}^{(\mathrm{ext})}], \tag{44}$$

and $P_{\sigma_2}^{(C)}$ is pressure generated by the point control force at plate 2.

In Eq. (44), $[\alpha_{M_1}^{(1)}]$ and $[\alpha_{M_2}^{(2)}]$ are respectively $(M_1\times M_1)$ and $(M_2\times M_2)$ matrices:

$$[\alpha_{M_1}^{(1)}]=[I]-[Y_{P_1}^{(1)}][Z_{A_1}^{(1)}], \tag{45}$$

$$[\alpha_{M_2}^{(2)}]=[I]-[Y_{P_1}^{(2)}][Z_{A_2}^{(1)}]-[Y_{P_2}^{(2)}][Z_A^{(2)}]+[Y_{P_1}^{(2)}][Z_{A_1}^{(1)}]\times[\alpha_{M_1}^{(1)}]^{-1}[Y_{P_1}^{(1)}][Z_{A_2}^{(1)}] \tag{46}$$

$[Y_{M_2,M_{p2}}^{(c)}]$ in Eq. (43) becomes a vector with its elements expressed as

$$Y_{I,1}^{(c)}=-\frac{1}{\rho_0 c_0 \chi_{p1}^{(2)}}\,S_1^{(2)}(\sigma_1^{(2)}). \tag{47}$$

Using a similar analysis as for the acoustical control in the room, it can be shown that total reduction of the potential energy in the room can be achieved only if the following relationship is satisfied:

$$[Z_A^{(2)}][\mathbf{V}_{\mathbf{M}_2}^{(\mathbf{P})}]=b[Z_A^{(2)}][\alpha_{M_2}^{(2)}]^{-1}[Y_{M_2,M_{p2}}^{(c)}], \tag{48}$$

where b is a complex constant. Although Eq. (48) shows a result similar to that shown in Eq. (42), it suggests that attenuation of the sound field in the room (to reduce the magnitudes of all elements in $[\mathbf{P}_{\mathbf{N}_2}^{(\mathbf{2})}]$) by controlling the vibration in plate 2 can be achieved by two mechanisms:

(1) Equation (48) can be satisfied if the control system is arranged such that $[\mathbf{V}_{\mathbf{M}_2}^{(p)}]$ and $[\alpha_{M_2}^{(2)}]^{-1}[Y_{M_2,M_{p2}}^{(c)}]$ are in proportion. That is

$$[\mathbf{V}_{\mathbf{M}_2}^{(\mathbf{P})}]=b[\alpha_{M_2}^{(2)}]^{-1}[Y_{M_2,M_{p2}}^{(c)}]. \tag{49}$$

For this case, the optimal control force can be used to effectively suppress the modal amplitudes of in plate 2. As a result, the source term in Eq. (43) (i.e., $[\mathbf{V}_{\mathbf{M}_2}^{(\mathbf{2})}]=[\mathbf{V}_{\mathbf{M}_2}^{(\mathbf{P})}]+[\alpha_{M_2}^{(2)}]^{-1}[Y_{M_2,M_{p2}}^{(c)}](P_{\sigma_2}^{(C)})_{\mathrm{opt}}$) is significantly reduced, and so is the resultant total sound pressure.

(2) Equation (48) can also be satisfied when the controlled plate velocity vector $[\mathbf{V}_{\mathbf{M}_2}^{(\mathbf{2})}]=[\mathbf{V}_{\mathbf{M}_2}^{(\mathbf{P})}]+[\alpha_{M_2}^{(2)}]^{-1}[Y_{M_2,M_{p2}}^{(c)}](P_{\sigma_2}^{(C)})_{\mathrm{opt}}$ is orthogonal with the row vectors in $[Z_A^{(2)}]$ (or is orthogonal with those row vectors corresponding to the dominating pressure components in $[\mathbf{P}_{\mathbf{N}_2}^{(\mathbf{2})}]$). For this case, the controlled plate velocity is not necessarily attenuated. The magnitude and phase of each mode in plate 2 are rearranged such that the superimposed contribution of all the elements in $[\mathbf{V}_{\mathbf{M}_2}^{(\mathbf{2})}]$ to the sound pressure components in the room is significantly reduced.

Previous work suggested that the modal suppressing mechanism is effective when the transmitted energy is carried mainly by one dominating panel mode.[1] On the other hand, the modal rearrangement mechanism is associated with the sound transmission into the room controlled modes where energy is transmitted from several plate modes. As $[\alpha_{M_2}^{(2)}]^{-1}[Y_{M_2,M_{p2}}^{(c)}]P_{\sigma_2}^{(C)}$ is due to the point control force and $[\mathbf{V}_{\mathbf{M}_2}^{(\mathbf{P})}]$ is generated by the distributed sound field, Eq. (49) can be approximately satisfied at the frequencies where plate 2 is dominated by a single mode (e.g., panel controlled mode). For this case, although the room sound pressure field may consist of multiple components coupled with the dominating plate mode, the effective suppression of the dominating plate mode will result in significant reduction of all the pressure components.

When the plate response is dominated by several modes and suppression of the response of these modes using a single control force becomes impossible, the second mechanism may be at work. The optimal control force will adjust the magnitude and phase of each plate mode, such that the sound pressure components due to the superimposed excitation of all the plate modes $[\mathbf{V}_{M_2}^{(\mathbf{2})}]$ through the modal acoustic transfer impedance matrix $[Z_A^{(2)}]$ are suppressed.

C. Acoustic control source in cavity

When only one acoustic actuator is used in the double wall cavity to control the transmitted sound field, the sound pressure vector in the room can be expressed as:

$$[\mathbf{P}_{\mathbf{N}_2}^{(\mathbf{2})}]=-[Z_A^{(2)}][\mathbf{V}_{\mathbf{M}_2}^{(\mathbf{2})}], \tag{50}$$

where

$$[\mathbf{V}_{\mathbf{M}_2}^{(\mathbf{2})}]=[\beta_{N_2}^{(2)}]^{-1}[Y_{P_1}^{(2)}][\mathbf{P}_{\mathbf{N}_1}^{(\mathbf{1})}], \tag{51}$$

$$[\mathbf{P}_{\mathbf{N}_1}^{(\mathbf{1})}]=-[\beta_{N_1}^{(1)}]^{-1}\{[Z_{A_1}^{(1)}][Y_{M_1}^{(\mathrm{ext})}][\mathbf{P}_{\mathbf{M}_1}^{(\mathbf{ext})}]-[Z_{M_c}^{(1)}][\mathbf{V}_{\mathbf{M}_c}]\}. \tag{52}$$

$[\beta_{N^1}^{(1)}]$ and $[\beta_{N_2}^{(2)}]$ are, respectively, $(N_1\times N_1)$ and $(N_2\times N_2)$ matrices

$$[\beta_{N_1}^{(1)}]=[I]-[Z_{A_1}^{(1)}][Y_{P_1}^{(1)}]-[Z_{A_2}^{(1)}][\beta_{N_2}^{(2)}]^{-1}[Y_{P_1}^{(2)}], \tag{53}$$

$$[\beta_{N_2}^{(2)}]=[I]-[Y_{P_2}^{(2)}][Z_A^{(2)}]. \tag{54}$$

Equation (49) shows that the ''source'' of the room sound pressure is the velocity of plate 2 ($[\mathbf{V}_{\mathbf{M}_2}^{(2)}]$), while the ''source'' of the velocity components of plate 2 is sound pressure in the cavity [see Eq. (51)].

It has been shown that a significant reduction of the acoustic potential energy in the room requires primary sound pressure components in the room to be proportional to that due to the control sources:

$$[Z_A^{(2)}][\beta_{N_2}^{(2)}]^{-1}[Y_{P_1}^{(2)}][\beta_{N_1}^{(1)}]^{-1}[Z_{A_1}^{(1)}]\langle[Y_{M_1}^{(\mathrm{ext})}][\mathbf{P}_{\mathbf{M}_1}^{(\mathbf{ext})}]\rangle$$

$$= a[Z_A^{(2)}][\beta_{N_2}^{(2)}]^{-1}[Y_{P_1}^{(2)}][\beta_{N_1}^{(1)}]^{-1}[Z_{M_c}^{(1)}][\mathbf{V}_{\mathbf{M}_c}], \qquad (55)$$

where a is a complex constant. Equation (55) suggests that the proportional condition can be realized by following three possible mechanisms:

(1) The control system may be arranged such that the primary sound pressure components in the cavity, $[\beta_{N_1}^{(1)}]^{-1}[Z_{A_1}^{(1)}]\langle[Y_{M_1}^{(\mathrm{ext})}][\mathbf{P}_{\mathbf{M}_1}^{(\mathbf{ext})}]\rangle$ is proportional to $[\beta_{N_1}^{(1)}]^{-1}[Z_{M_c}^{(1)}]$. That is,

$$[\beta_{N_1}^{(1)}]^{-1}[Z_{A_1}^{(1)}]\langle[Y_{M_1}^{(\mathrm{ext})}][\mathbf{P}_{\mathbf{M}_1}^{(\mathbf{ext})}]\rangle = a[\beta_{N_1}^{(1)}]^{-1}[Z_{M_r}^{(1)}]. \qquad (56)$$

If the condition in Eq. (56) can be satisfied, the sound pressure in the cavity will be effectively suppressed. For this case, the suppression of the sound pressure modal response in the cavity results in a significant noise reduction in the room.

(2) The optimal volume velocity of the control source in the cavity might be selected such that the controlled pressure component vector in the cavity $[\mathbf{P}_{\mathbf{N}_1}^{(\mathbf{1})}]$ is orthogonal to the corresponding row vectors in the modified modal acoustic transfer mobility matrix $[\beta_{N^2}^{(2)}]^{-1}[Y_{P_1}^{(2)}]$. As a result of the modal rearrangement of the cavity pressure components (for this case the pressure components may not be reduced), the magnitudes of the dominating modes in plate 2 are effectively reduced, and so are the sound pressure components in the room.

(3) The optimal volume velocity of the control source in the cavity might be selected such that the controlled modal component vector in plate 2 (due to the excitation of the controlled pressure components in the cavity) $[\mathbf{V}_{\mathbf{M}_2}^{(\mathbf{2})}]$ is orthogonal to the row vectors in $[Z_A^{(2)}]$. For this case, the controlled plate velocity may not necessarily be attenuated. The magnitude and phase of each mode in plate 2 are rearranged by the control source in the cavity such that the superimposed contribution by all the elements of in $[\mathbf{V}_{\mathbf{M}_2}^{(\mathbf{2})}]$ to the dominating sound pressure components in the room is significantly reduced.

In both discussions of (2) and (3) above, the modal rearrangement mechanism was at work. However, for the former case, the cavity sound pressure is rearranged directly to give an attenuated velocity in plate 2. For the latter case, the modal components in plate 2 are rearranged indirectly through the rearrangement of the cavity pressure components to obtain an attenuated sound pressure field in the room.

III. CONCLUSIONS

This paper presents a model for sound transmission through double plate structures in such a way that the mechanisms of the several different active control arrangements can be analysed. For either acoustic control in the room, or structural acoustic control on plate 2 or acoustic control in the partition cavity, the condition for effective reduction of acoustic potential energy in the room is that the primary and secondary pressure vectors in the room are proportional as described by Eqs. (42), (48), and (55). Using the optimal control input, the system may choose different mechanisms to achieve the reduction of acoustic potential energy in the room.

The analysis of the physical mechanisms involved in each control arrangement shows:

(1) When the transmitted sound field in the room is directly controlled by using one acoustic control source in the room, the suppression of the room modal response is the only possible mechanism. It is difficult for the secondary sound field generated by a point acoustic control source to satisfy the required proportional relationship with the primary field generated by distributed structural vibration. Control can be achieved only at those frequencies where the primary sound field is dominated by a single room mode and the control source can excite this mode without the strong excitation of other modes.

(2) The sound pressure in the room can be attenuated through the control of sound radiation into the room using a vibrational control source on the radiating structure. The control source may directly suppress the amplitude of the dominating mode corresponding to the sound pressure field in the room. However, when several plate modes participate in the coupling with a dominating room modes, the control source will rearrange the magnitudes and phase of the plate modes such that the superimposed sound radiation from the plate to the room is significantly reduced. Because there are two possible control mechanisms, it is expected that better attenuation may be achieved than for the case of room control for which only one mechanism applies..

(3) The common understanding of the control of sound transmission through double plates by inserting one acoustic control source between the double wall cavity is based on the fact that the low-frequency sound transmission is due to the (0,0,0) cavity mode. If the cavity is relatively small, the (0,0,0) mode may dominate a reasonably wide frequency range. The suppression of this dominating acoustic mode in the cavity will block the sound transmission path and therefore transmitted sound pressure will be reduced. The analysis of this paper shows three possible mechanisms involved in this control arrangement, one of which is suppression of the cavity modal response. The other two mechanisms are (1) the direct rearrangement of the cavity sound pressure components to minimise the amplitude of the dominating radiating modes in plate 2; (2) the indirect rearrangement of the modal components in plate 2 by adjusting the cavity pressure components, such that the superimposed sound radiation into the room is reduced. These two modal rearrangement mechanisms may be accompanied by an increase of sound pressure in the cavity. In addition, the three possible mechanisms may

explain the improved performance of the control using this arrangement[9] relative to room or panel control.

Although previous numerical calculation and experiments have shown the existence of the modal suppression and modal rearrangement mechanisms, further study are needed to investigate how to effectively use this knowledge in the design of active control systems for sound transmission through double walls.

[1] J. Pan, ''A study of the active control of sound transmission through a panel into a cavity,'' J. Acoust. Soc. Am. **84**, 182 (1988).

[2] J. Pan, C. H. Hansen, and D. A. Bies, ''Active control of noise transmission through a panel into a cavity, I. Analytical study,'' J. Acoust. Soc. Am. **87**, 2098–2108 (1990).

[3] J. Pan and C. H. Hansen, ''Active control of noise transmission through a panel into a cavity, II. Experimental study,'' J. Acoust. Soc. Am. **90**, 1488–1492 (1991).

[4] C. R. Fuller, ''Active control of sound transmission/radiation from elastic plates by vibration inputs: I. Analysis,'' J. Sound Vib. **136**, 1–15 (1990).

[5] D. R. Thomas, P. A. Nelson, R. J. Pinnington, and S. J. Elliott, ''Active control of sound transmission through stiff lightweight composite fuselage constructions,'' Proceeding of the 14th AIAA Aeroacoustics Conference II, 552–560 (1992).

[6] J. Pan, S. D. Snyder, C. H. Hansen, and C. R. Fuller, ''Active control of farfield sound radiated by a rectangular panel—A general analysis,'' J. Acoust. Soc. Am. **91**, 2056–2066 (1992).

[7] J. Pan and C. H. Hansen, ''Active control of noise transmission through a panel into a cavity, III. Effect of actuator location,'' J. Acoust. Soc. Am. **90**, 1493–1501 (1991).

[8] C. Bao and J. Pan, ''Experimental study of different approaches for active control of sound transmission through double walls,'' J. Acoust. Soc. Am. **102**, 1664–1670 (1997).

[9] P. Sas, C. Bao, F. Augusztinovicz, and W. Desmet, ''Active control of sound transmission through a double panel partition,'' J. Sound Vib. **180**, 609–625 (1995).

[10] L. Gagliardini and P. Bouvet, ''DAP: the active controlled double wall,'' Proc. Inter-Noise 93, 107–110 (1993).

[11] P. A. Nelson and S. J. Elliott, *Active Control of Sound* (Academic, New York, 1992).

[12] J. Pan and C. Bao, ''Active attenuation of noise transmission through elastic partitions with high modal densities,'' Proc. Inter-Noise 96, Liverpool, UK, III, pp. 1055–1060 (1996).

A discussion of modal uncoupling and an approximate closed-form solution for weakly coupled systems with application to acoustics

Linda P. Franzoni
Center for Sound and Vibration, Mechanical and Aerospace Engineering, North Carolina State University, Raleigh, North Carolina 27695

Donald B. Bliss
Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27706

(Received 9 September 1996; revised manuscript received 20 September 1997; accepted 31 December 1997)

Modal analysis is often used to solve problems in acoustics, leading to a system of coupled equations for the modal amplitudes. A common practice in analytical work utilizing modal analysis has been to assume that weak modal coupling is negligible, thereby enabling the modal coefficients to be solved independently in closed form. The validity of this assumption, as well as the order of the error from neglecting modal coupling, is discussed. It is possible to incorporate the principal effects of weak modal coupling in a very simple way without solving the fully coupled system. An approximate closed-form solution for weakly coupled systems of equations is developed. The procedure gives insight into the errors incurred when coupling is neglected, and shows that these errors may be unacceptably large in systems of practical interest. A model problem involving a pipe with an impedance boundary condition is solved when the one-dimensional sound field is harmonically driven, and when it undergoes reverberant decay from initial conditions. The approximate solution derived in this paper is compared with results for the fully coupled and fully uncoupled equivalent problems. The approximation works well even for systems where the coupling is fairly strong. The results show that modal coupling must be included, at least approximately, if certain salient features of the sound field, such as intensity flow and detailed reverberant structure, are to be predicted correctly. © *1998 Acoustical Society of America.* [S0001-4966(98)02104-3]

PACS numbers: 43.55.Br, 43.20.Ks [JDQ]

INTRODUCTION

Many important problems in acoustics, vibration, and dynamics are solved by modal analysis. Solutions are expanded in the normal modes of free vibration, leading to a coupled system of equations for the modal amplitudes. These equations may be either differential or algebraic depending on whether the problem is solved in the time or frequency domain. Coupling between equations typically arises through terms associated with certain boundary conditions, or through the generalized force terms. Thus the equations for modal amplitudes in an acoustic enclosure are coupled through the impedance boundary condition on absorptive walls.[1] Similarly, the modal amplitude equations for a fluid loaded structure are coupled through generalized forces of acoustic and hydrodynamic origin.

The system of modal amplitude equations, if not too large, can be solved readily by matrix methods. However, sometimes there are advantages to having closed-form solutions, even if these solutions are approximate. For instance, simple expressions for the modal amplitudes are desirable in certain types of theoretical work in which these expressions are required for further analytical development. Examples of this situation include analytical prediction of high-frequency sound fields by asymptotic modal analysis,[2–4] and statistical characterization of complex systems using fuzzy structure theory.[5–7] A common approach, when coupling is weak, is to neglect off-diagonal terms and solve the resulting uncoupled system.[8,9] Unfortunately, recent work has shown that this approach is less accurate than formerly believed for cases with realistic amounts of coupling.[9,10] This shortcoming may limit the range of applicability of high-frequency asymptotic modal methods where uncoupling the system facilitates further analytical development.

The errors due to uncoupling the system of modal equations are examined in this paper. Then, an approximate method to include weak coupling effects in a closed-form solution is developed formally. This solution provides physical insight into the primary effect of weak coupling. An appealing aspect of the new approximate method is that it increases the range of applicability of simple expressions for modal amplitude for use in subsequent theoretical analyses. Also, because of the formal nature of the procedure, it is possible to extend the approximate solution to higher order, although the advantage of doing so may be of limited utility, since a more complicated expression results. The lowest order approximate solution is shown to be reasonably accurate even when the coupling is fairly strong. The accuracy of these approximate solutions is compared to fully coupled and fully uncoupled solutions for a model problem involving the driven response and reverberant decay of an air column with an impedance boundary condition. In the model problem, the error is shown to scale inversely with impedance.

I. TECHNICAL APPROACH

Numerous physical problems are governed by partial differential equations of the form:

$$\mathscr{L}[w]+\rho\ddot{w}=f_m+f_e, \tag{1}$$

where w is a system property (displacement, pressure, etc.), $\mathscr{L}$ is a linear operator, and $\rho\ddot{w}$ is the inertia term. The prescribed external forcing is f_e, and the force due to system motion is f_m. The unknown w can be expanded in the normal modes of free vibration, Φ_n, multiplied by generalized coordinates, q_n, namely;

$$w=\sum_{n=0}^{\infty} q_n(t)\Phi_n(x,y,z). \tag{2}$$

The normal modes $\Phi_n(x)$ satisfy the homogeneous form of Eq. (1) with harmonic time dependence,

$$\mathscr{L}[\Phi_n]-\rho\omega_n^2\Phi_n=0, \tag{3}$$

where the eigenfrequencies are ω_n. The boundary conditions associated with Eq. (3) need not be those of the physical problem associated with Eq. (1). The important point is that the normal modes $\Phi_n(x)$ must form a complete orthogonal set.

Substituting Eq. (2) into Eq. (1) leads to a system of coupled ordinary differential equations for the generalized coordinates $q_n(t)$,

$$\ddot{q}_n+\omega_n^2 q_n=Q_{m_n}/M_n+Q_{e_n}/M_n,$$

where the generalized forces are

$$Q_{m_n}=\int_V f_m\Phi_n\,dV \quad \text{and} \quad Q_{e_n}=\int_V f_e\Phi_n\,dV,$$

and the generalized mass is

$$M_n=\int_V \rho\Phi_n^2\,dV.$$

For harmonic external forcing the equations become

$$[\omega_n^2-\omega^2]\overline{q_n}=\overline{Q}_{m_n}/M_n+\overline{Q}_{e_n}/M_n, \tag{4}$$

where $q_n=\overline{q_n}e^{i\omega t}$, etc. The generalized force, $\overline{Q}_{m_n}$, associated with system motion may depend on a linear combination of all the generalized coordinates. It is convenient to write this generalized forcing term as follows:

$$\overline{Q}_{m_n}/M_n=\sum_r i\omega\omega_r C_{nr}(\omega)\overline{q_r}, \tag{5}$$

where the quantities C_{nr} are dimensionless coupling coefficients. Note that this generalized force couples the equations for modal amplitude, namely,

$$[\omega_n^2-\omega^2]\overline{q_n}+\sum_{r=0}^{\infty} i\omega\omega_r C_{nr}(\omega)\overline{q_r}=\overline{F_n}(\omega), \tag{6}$$

where $\overline{F_n}$ is the external generalized force normalized with the generalized mass.

When the modal equations are derived for an acoustic enclosure, as in the subsequent example, the modal coordinates q_n are the pressure amplitudes P_n, and the coupling term arises from the impedance of the enclosure boundaries.[1] In this case the generalized force associated with Q_m is applied at the system boundaries, rather than being distributed throughout, and a special mathematical step involving Green's theorem is used to recast the equation in the above form. Modal equations for a fluid-loaded shell also have a similar form, although the details of the coupling term are different. In the fluid-loaded case, the coupling term contains information about the fluid loading, and the structural interdependencies of the shell.

The form of Eq. (6) resembles that of a spring-mass-damper system. For instance, if Q_{m_n} is associated simply with system damping then $\omega_r C_{nr}$ equals 2 $\omega_n\zeta_{nr}$, where ζ_{nr} is an intermodal damping ratio. This notation is not adopted here because the generalized force term can be interpreted more broadly to include additional stiffness and inertial effects. For example, in the modal equations derived for an acoustic enclosure, the coupling term due to impedance has both resistive and reactive components. Thus in general, the coupling coefficients C_{nr} will be complex.

When modal coupling is weak because the coupling coefficients C_{nr} are small, the equations have been solved approximately by neglecting the off-diagonal coupling. A two-mode model problem solved by Dowell[8] seems to support this approach. However, recent results for an acoustic enclosure model problem show that modal coupling is surprisingly important, even when it would seem to be weak based on the size of the coupling coefficients.[9,10]

A. Estimation of coupling effects

To better quantify the effects of modal coupling, Eq. (6) is re-expressed in the following dimensionless form:

$$\hat{q}_n(\omega)+\sum_{r=0}^{\infty} \frac{i\epsilon\hat{C}_{nr}b_{nr}}{\omega_r/\omega-\omega/\omega_r+i\epsilon\hat{C}_{rr}}\hat{q}_r(\omega)=1, \tag{7}$$

where the new dimensionless modal amplitude is

$$\hat{q}_n\equiv[\omega_n^2-\omega^2+i\omega\omega_n\epsilon\hat{C}_{nn}]\overline{q_n}(\omega)/\overline{F_n}(\omega), \tag{8}$$

and $\epsilon\hat{C}_{nr}\equiv C_{nr}\overline{F_r}/\overline{F_n}$. The parameter ε is introduced to order the coupling coefficients; for weak coupling $\epsilon\ll 1$. Note that the magnitude of ϵ is chosen so that the normalized coupling coefficients $\hat{C}_{nr}$ are $O[1]$. The $r=n$ term has been removed from the summation in Eq. (7) and absorbed into the new dimensionless modal amplitude. This removal requires the term b_{nr} inside the summation; the quantity b_{nr} is defined to be equal to zero when r equals n, and otherwise equal to unity. It is worth noting that Eq. (8) can be interpreted as $\hat{q}_n\equiv\overline{q_n}/\overline{q_n}_{\text{uncoupled}}$.

Note that the definition of $\hat{q}_n$ in Eq. (8) excludes the possibility that one or more of the quantities $\overline{F_n}$ may be zero. This exclusion facilitates the theoretical development. In fact, when later results are returned to dimensional form, they can be shown to be correct even if some of the quantities $\overline{F_n}$ are zero. A more serious restriction occurs when an on-diagonal coupling coefficient C_{nn} is much smaller than the other coefficients, or zero. Treatment of this special case is given in Appendix B.

The virtue of Eq. (7) is that the order of magnitude of the terms can be clearly seen. The forcing term on the right hand side is $O[1]$. The new coupling term in the summation is either $O[1]$ or $O[\epsilon]$ depending on whether or not the frequency ω is near a modal resonance ω_r:

$$\frac{i\epsilon\hat{C}_{nr}}{\omega_r/\omega-\omega/\omega_r+i\epsilon\hat{C}_{rr}}=\begin{cases} O[\epsilon], & \omega\neq\omega_r \\ O[1], & \omega\approx\omega_r \end{cases}. \tag{9}$$

The conclusion is that the dimensionless modal coordinates $\hat{q}_n$ must be $O[1]$ for all frequencies, even at resonance. Furthermore, because the modal coupling term is always $O[1]$ for certain values of ω, modal coupling *always* has an important effect on the dimensionless modal amplitudes $\hat{q}_n$, *regardless* of the size of ϵ.

Now, suppose $\epsilon\ll 1$, restricting the physical system to weak coupling. For sufficiently small ε the modal peaks will be distinctly separate, i.e., the width of individual peaks, $O[\epsilon\ \omega_r]$, will be small compared to the spacing between peaks, $\omega_{r+1}-\omega_r$. In summing over all the modes, either ω will never be equal to any of the modal resonant frequencies, ω_r, or ω will be very near a specific resonant frequency, say ω_R. If ω is not very near a resonant frequency, then Eq. (7) reduces to

$$\hat{q}_n=1+O[\epsilon] \quad \text{and} \quad \text{when} \ \ \epsilon\to 0,\ \hat{q}_n=1, \tag{10}$$

which is equivalent to neglecting coupling entirely. However, if ω is very close to the resonance frequency ω_R, namely $|\omega-\omega_R|\leqslant O[\epsilon\ \omega_R]$, then Eq. (7) becomes

$$\begin{aligned} &\hat{q}_n+[\hat{C}_{nR}/\hat{C}_{RR}]\hat{q}_R+O[\epsilon]=1, \quad n\neq R, \\ &\hat{q}_R+0+O[\epsilon]=1, \quad n=R. \end{aligned} \tag{11}$$

For this case, it is clear that when $n\neq R$, the dimensionless quantity $\hat{q}_n$ is always affected by modal coupling, even as $\epsilon\to 0$. An expression that embodies all these conditions to $O[1]$ accuracy is given by

$$\hat{q}_n(\omega)=1-\frac{i\epsilon\hat{C}_{nR}b_{nR}}{\omega_R/\omega-\omega/\omega_R+i\epsilon\hat{C}_{RR}}1. \tag{12}$$

Returning Eq. (12) to the original dimensional form gives

$$\begin{aligned} \overline{q_n}(\omega)=&\frac{1}{[\omega_n^2-\omega^2+i\omega_n\omega C_{nn}]} \\ &\times\left[\overline{F_n}(\omega)-\frac{i\omega_R\omega C_{nR}b_{nR}}{\omega_R^2-\omega^2+i\omega_R\omega C_{RR}}\overline{F_R}(\omega)\right]. \end{aligned} \tag{13}$$

This expression is the approximate *closed-form* result for $q_n(\omega)$ incorporating the effects of weak modal coupling. The coupling effect is in the second term on the right of Eq. (13), and the inclusion of this simple term will be shown to significantly improve results. Physically, the Rth mode, which has natural frequency ω_R closest to the driving frequency ω influences the surrounding modes as was indicated by the discussion preceeding Eq. (12). Coupling between all other modes, and backcoupling on the Rth mode, is unimportant by comparison.

In earlier work, the effect of modal coupling has been included through a straightforward iterative approach sometimes referred to as the method of successive substitutions.[11,12] A ''zeroth order'' solution is obtained by completely neglecting the coupling. Then the coupling terms can be approximated using this solution, leading to a ''first order'' solution. The process can be repeated to refine the solution. However, this process does not necessarily retain the order of magnitude of terms in a consistent manner. A significant contribution of this paper is the development of a closed-form solution in which all terms retained are of the same order of magnitude. This result was achieved through the use of the dimensionless variables that correctly account for the order of the terms whether on or off resonance. Appendix A compares the results of the present method with those obtained by successive substitutions.

A more detailed investigation reveals that the applicability of Eq. (13) is limited to cases where the on-diagonal coupling coefficient C_{RR} has a nonvanishing real part that is not extremely small compared to that of the other coupling coefficients C_{nr}. The special case of C_{RR} small or zero leads to a balance of terms somewhat different than in the above development. This special case, which also has a simple closed-form solution, is analyzed in Appendix B.

B. Higher order solutions

The expression for modal amplitudes in the normalized variables, as given by Eq. (12), is accurate in these variables to $O[1]$ and contains errors of order ϵ. In fact, in order to satisfy both the conditions of ω near a resonance and ω not near a resonance, it was necessary to introduce an error of order ϵ. Higher order solutions can be obtained by following the procedures outlined above and expressing the modal amplitudes as a power series in ϵ, namely, $\hat{q}_n=\hat{q}_n^0+\epsilon\hat{q}_n^1+O[\epsilon^2]+O[\epsilon^3]$, etc. For the case where the frequency, ω, is near one of the natural resonances of the system, ω_R, the first terms of the series are given by:

$$\begin{aligned} \hat{q}_n(\omega)=&1-b_{nR}\frac{\hat{C}_{nR}}{\hat{C}_{RR}}-\sum_{m=0}^{\infty}b_{mR}b_{nm} \\ &\times\frac{i\epsilon\hat{C}_{nm}}{\omega_m/\omega-\omega/\omega_m}\left(1-b_{mR}\frac{\hat{C}_{mR}}{\hat{C}_{RR}}\right)+O[\epsilon^2]. \end{aligned}$$

However, if ω is *never* near a resonance then the solution would be:

$$\hat{q}_n(\omega)=1-\sum_{m=0}^{\infty}b_{nm}\frac{i\epsilon\hat{C}_{nm}}{\omega_m/\omega-\omega/\omega_m}+O[\epsilon^2].$$

As before, it is possible to construct one solution which is valid over the entire range of ω. A solution for $\hat{q}_n$ which is valid on and off resonance, and which is accurate to Order$[\epsilon]$ is:

$$\hat{q}_n(\omega)=1-b_{nR}\frac{i\epsilon\hat{C}_{nR}}{\omega_R/\omega-\omega/\omega_R+i\epsilon\hat{C}_{RR}}-\sum_{m=0}^{\infty}b_{mR}b_{nm}\frac{i\epsilon\hat{C}_{nm}}{\omega_m/\omega-\omega/\omega_m}\times\left(1-b_{mR}\frac{i\epsilon\hat{C}_{mR}}{\omega_R/\omega-\omega/\omega_R+i\epsilon\hat{C}_{RR}}\right)+O[\epsilon^2].$$

Note that the expression inside the parenthesis (in the summation over m) is simply the previous lower order $\hat{q}_n$ from Eq. (12). Returning to the dimensional variables:

$$\bar{q}_n(\omega)=\frac{1}{\omega_n^2-\omega^2+i\omega\omega_nC_{nn}}\times\left\{\bar{F}_n-b_{nR}\frac{i\omega\omega_RC_{nR}\bar{F}_R}{\omega_R^2-\omega^2+i\omega\omega_RC_{RR}}-\sum_{m=0}^{\infty}b_{mR}b_{nm}\frac{i\omega\omega_mC_{nm}}{\omega_m^2-\omega^2}\times\left(\bar{F}_m-b_{mR}\frac{i\omega\omega_RC_{mR}\bar{F}_R}{\omega_R^2-\omega^2+i\omega\omega_RC_{RR}}\right)+O[\epsilon^2]\right\}. \tag{14}$$

In a similar manner, solutions which are accurate to $O[\epsilon^2]$, $O[\epsilon^3]$, etc. can be obtained. However, since the advantage of having closed-form solutions for the modal amplitudes lies primarily in their simplicity and usefulness for further analytical work, the higher order solutions may not be as valuable as the lower order solutions.

II. MODEL PROBLEM, RESULTS, AND DISCUSSION

The efficacy of the approximate solutions developed above is now illustrated by a simple model problem. Acoustic standing waves are set up in a 1-D pipe of length L excited by a piston at $x=0$ vibrating at velocity $v_0(\omega)e^{i\omega t}$. The specific acoustic impedance at $x=L$ is Z_L. This problem, which contains the appropriate physics, was chosen because it was identical to that in Refs. 9 and 10, and because a closed-form exact solution can be found without modal analysis. The model problem is sketched in Fig. 1.

In the modal analysis approach, the pressure variable is expanded in ''hard box modes,'' that is, the normal modes for which the boundary conditions are rigid walls. Using this expansion and invoking Green's theorem, the homogeneous wave equation and applicable boundary conditions can be re-expressed as a set of ordinary differential equations (as in Refs. 1, 9, and 10) for the pressure mode amplitudes P_n:

$$\ddot{P}_n+\omega_n^2P_n+2\sum_{r=0}^{\infty}\omega_n\zeta_{nr}\dot{P}_r=i\omega\rho_0c_0^2v_0e^{i\omega t}. \tag{15}$$

Note that the process of applying Green's theorem has effectively transformed the velocity boundary condition into a generalized force and the impedance boundary condition into a generalized force coupling term. The coefficients in the coupling term are formally the product of acoustic modal functions divided by the impedance, integrated over the absorbing area, and then normalized by the area of the absorbing surface. The coupling coefficients are given by

$$2i\omega\omega_n\zeta_{nr}=\rho_0c_0^2\frac{i\omega(-1)^{n+r}}{Z_L\int_0^L\cos^2(n\pi x/L)\,dx}, \tag{16}$$

where $\omega_n=n\pi c_0/L$ is the natural frequency of the nth mode.

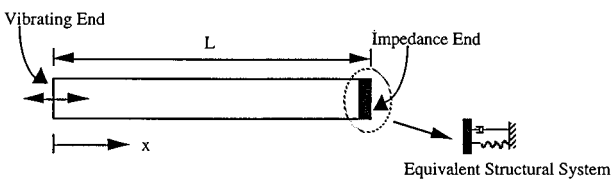


FIG. 1. The model problem.

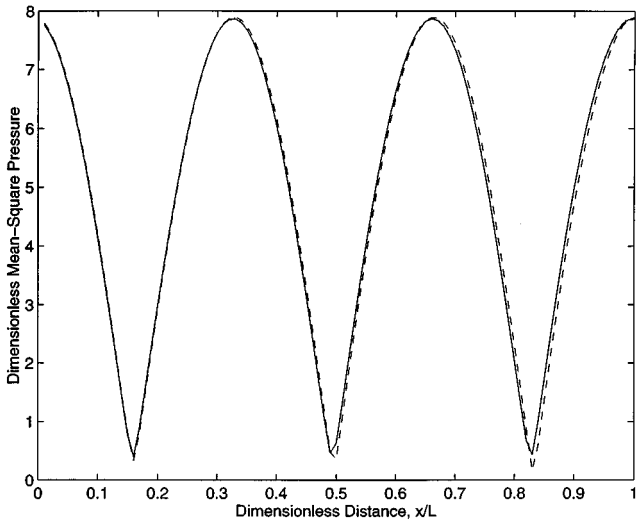


FIG. 2. Dimensionless mean-square pressure $\overline{p^2}/(\rho_0c_0v_0)^2$ vs distance, x/L for an endwall impedance of $z_L=10+10i$, at a fixed dimensionless frequency of $\omega L/c_0=9.4$. $x=0$ corresponds to the forced wall, $x=L$ corresponds to the impedance wall. Three curves plotted: Fully coupled=solid line; uncoupled=dashed line; approximate=dotted line.

A. Forced-response results

Figures 2, 3, and 4 show standing waves in the pipe for impedances $10+10i$, $5+1i$, $2+2i$, respectively. The corresponding absorption coefficients are 0.181, 0.541, and 0.615. Neglecting modal coupling works well only for the largest impedance and fails dramatically for the two smaller impedances. Characteristically the error appears similar to a distributed damping along the pipe. This characterization was first made by Pan,[9] but attributed to modal analysis itself, rather than to the modal *uncoupling*, as pointed out in subsequent work by Franzoni and Dowell.[10] In the uncoupled case, levels fall along the pipe, there is a phase shift, and a pressure maximum occurs at the impedance wall. This behavior is explained in Appendix C. In contrast, the approximate method for modal coupling, Eq. (13), works very well for the two larger impedances, being virtually indistinguish-

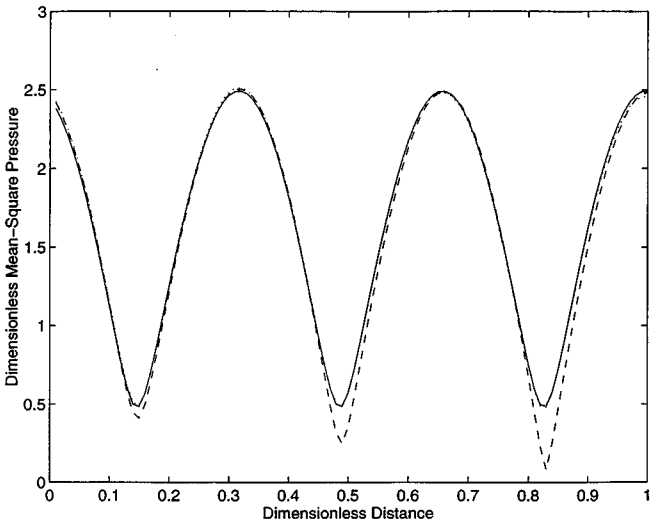


FIG. 3. Dimensionless mean-square pressure $\overline{p^2}/(\rho_0 c_0 v_0)^2$ vs distance, x/L for an endwall impedance of $z_L=5+i$, at a fixed dimensionless frequency of $\omega L/c_0=9.4$. $x=0$ corresponds to the forced wall, $x=L$ corresponds to the impedance wall. Three curves plotted: Fully coupled=solid line; uncoupled=dashed line; approximate=dotted line.

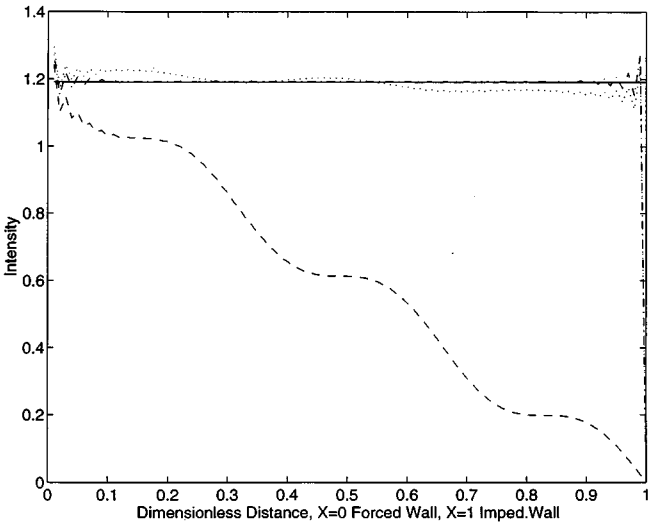


FIG. 5. Dimensionless intensity vs dimensionless distance (x/L). $x/L=0$ corresponds to the forced wall, $x/L=1$ corresponds to the impedance wall. Four curves are plotted: Fully coupled=solid line; uncoupled =dashed line; lowest order approximate=dotted line; higher order approximate=dash-dot line. The endwall impedance, $z_L=5+i$, and the dimensionless frequency $=9.25$.

able from the exact result. It shows discrepancies for the smallest impedance, but does remarkably well considering the high degree of coupling for this impedance.

Errors due to uncoupling are much more pronounced for the intensity, as shown in Fig. 5. The intensity versus distance is plotted for a normalized wall impedance of $5+1i$ at a fixed frequency which is near a resonance. The neglect of modal coupling is clearly seen to produce an effect similar to the introduction of a fictitious spatial damping. This conclusion is shown to be precisely true in Appendix C. When the modes are uncoupled, the intensity always vanishes at the impedance wall. Note, however, that the approximate method developed in this paper captures the correct physics of the problem.

In Fig. 6, the root-mean-square dimensionless pressure versus frequency $(\omega L/c)$ is shown, for an impedance at the end wall of $z_L=2+2i$. This case is fairly well damped, and corresponds to an absorption coefficient of 0.615. Four curves are shown: the fully coupled solution, the uncoupled solution, the $O[1]$ solution [the lowest order approximate solution, Eq. (13)], and the $O[\epsilon]$ solution [referred to as the ''higher order solution'' in the technical discussion, Eq. (14)]. The uncoupled solution misses several of the peaks and valleys of the frequency response. The lowest order ap-

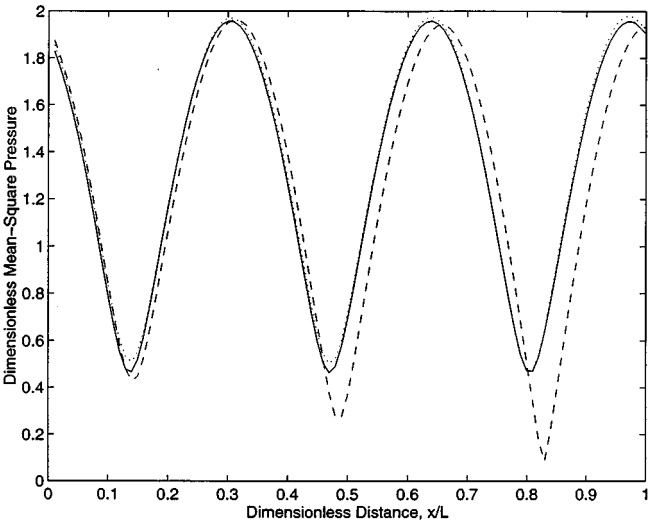


FIG. 4. Dimensionless mean-square pressure $\overline{p^2}/(\rho_0 c_0 v_0)^2$ vs distance, x/L for an endwall impedance of $z_L=2+2i$, at a fixed dimensionless frequency of $\omega L/c_0=9.4$. $x=0$ corresponds to the forced wall, $x=L$ corresponds to the impedance wall. Three curves plotted: Fully coupled=solid line; uncoupled=dashed line; approximate=dotted line.

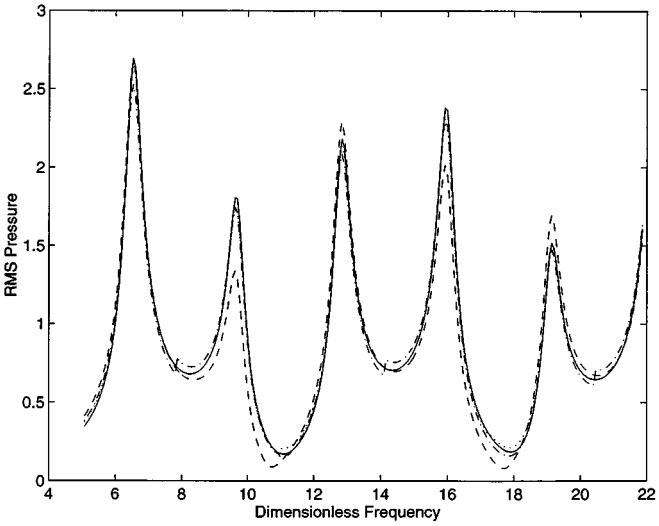


FIG. 6. Dimensionless root-mean-square pressure, $\overline{p}/(\rho_0 c_0 v_0)$, vs dimensionless frequency, $\omega L/c_0$. Endwall impedance, $z_L=2+2i$. Four curves are plotted: Fully coupled=solid line; uncoupled=dashed line; lowest order approximate=dotted line; higher order approximate=dash-dot line.

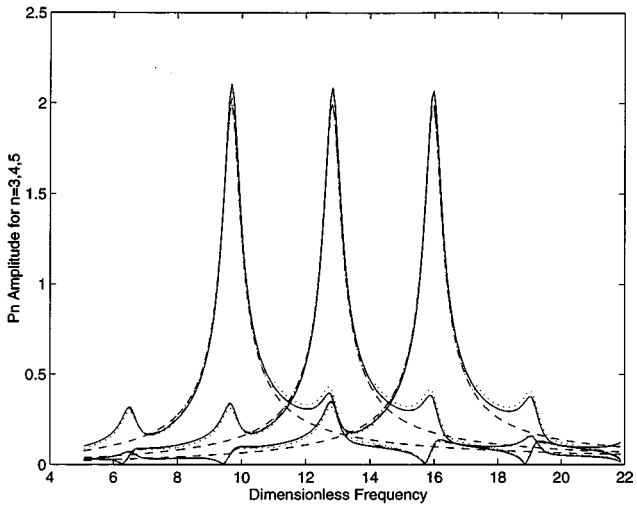


FIG. 7. Individual modal amplitudes, P_n for modes $n=3$, 4, and 5 vs dimensionless frequency, $\omega L/c_0$. Endwall impedance, $z_L=2+2i$. Four curves are plotted: Fully coupled=solid line; uncoupled=dashed line; lowest order approximate=dotted line; higher order approximate=dash-dot line.

proximate solution agrees with the fully coupled solution, except half-way between resonance peaks, where a slight jump is seen in the curve. The higher order approximate solution is most accurate and does not ''jump'' between resonance peaks. The reasons for the agreements/disagreements with the exact (fully coupled) solution are best elaborated upon by examining the modal amplitudes individually, as in the next figure.

The approximate method performs well because it incorporates the primary effect of modal coupling, which is the contribution of the most nearly resonant mode on the surrounding modes. Figure 7 illustrates this effect by showing individual adjacent modal amplitudes as a function of frequency which correspond to the previous figure. Four curves are plotted: fully coupled, uncoupled, approximately coupled to Order[1], and approximately coupled to Order[ϵ]. At frequencies *near* a modal resonance the structure of the *off-resonant* modes is altered. Neglecting modal coupling completely omits this alteration. The individual modal amplitudes for the approximate method agree quite well with the fully coupled solution, even though they retain the closed-form simplicity of the uncoupled case.

In Fig. 7, the approximate Order[1] solution makes small jumps between resonance peaks, as it switches from one ''nearest'' resonant mode to the next. These jumps, which are actually higher order, are caused by the abrupt steplike behavior of the function b_{nR}. This behavior could be improved upon at this order by taking the resonant modes above and below the frequency of interest, or by going to higher order, as demonstrated by the Order[ϵ] solution.

B. Time-domain results

The significance of modal coupling and the errors associated with uncoupling are also quite pronounced in the following reverberation problem. Consider the 1-D pipe with an impedance boundary at one end, as shown in Fig. 1, but instead of a velocity boundary condition at the other end, it is closed by a ''hard wall.'' At $t=0$ (initial condition), a given pressure pulse begins to propagate in the pipe. In the D'Alembert sense, the wave splits into two half-strength right and left traveling waves that reflect alternately off the hard and the impedance walls, undergoing a stepwise decay. This problem is solved three ways: as a fully coupled modal analysis, as an uncoupled modal analysis, and as a partially coupled modal analysis (similar to the previous approximate method). Because this reverberation problem differs significantly from the previous forced problem, the mathematical formulation for this case will be briefly outlined below before the results are discussed.

The time-domain equation representing the physical system is written in terms of the modal coordinates, P_n:

$$\ddot{P}_n+\omega_n^2 P_n+\sum_{r=0}^{\infty} C_{nr}\dot{P}_r=0, \tag{17}$$

where each C_{nr} differs from the coupling coefficients in the forced problem by a factor of $(i\omega)$ as a result of transforming the boundary condition from frequency to time domain, assuming the wall impedance is independent of frequency [otherwise the form of Eq. (17) would be more complicated].

Since Eq. (17) is second order in time, two initial conditions must be specified. These correspond to the set of modal amplitudes that describe the initial pressure pulse and to the set of their time derivatives. However, the choice of modal amplitudes, and the time derivative of the zeroth modal amplitude cannot be arbitrarily chosen. In order to satisfy mass conservation, a relationship must exist between $\dot{P}_0$ and the P_n's. The mass conservation requirement states that: $d(\text{mass})/dt=0$. The system mass can be written as

$$\int_0^L \rho A\, dx+\rho_0 A(L+\xi_{x=L})+\text{higher order terms},$$

where ρ is the acoustic density perturbation. Then, differentiating with respect to time

$$\int_0^L \frac{\partial \rho}{\partial t}\, dx+\rho_0 u\Big|_{x=L}=0.$$

This expression can be written in terms of pressure, since $p=c^2\rho$ and $u|_{x=L}=p|_{x=L}/z_L$. Expanding the pressure in hard box modes:

$$p=\sum_{n=0}^{\infty} P_n\Phi_n=P_0 1+\sum_{n=1}^{\infty} P_n\Phi_n$$

and noting that

$$\int_0^L \Phi_n\, dx=0, \quad \text{for } n\neq 0,$$

the mass conservation expression becomes

$$\frac{L}{c^2}\dot{P}_0+\frac{1}{z_L}\sum_{n=0}^{\infty} P_n\Phi_n(L)=0. \tag{18}$$

Note that if the time derivative of this equation is taken, it is equivalent to the original time-domain equation [Eq. (17)] for modal amplitude P_0 with $n=0$, since $\omega_0=0$. Therefore, mass is always conserved if Eq. (18) is satisfied initially, and

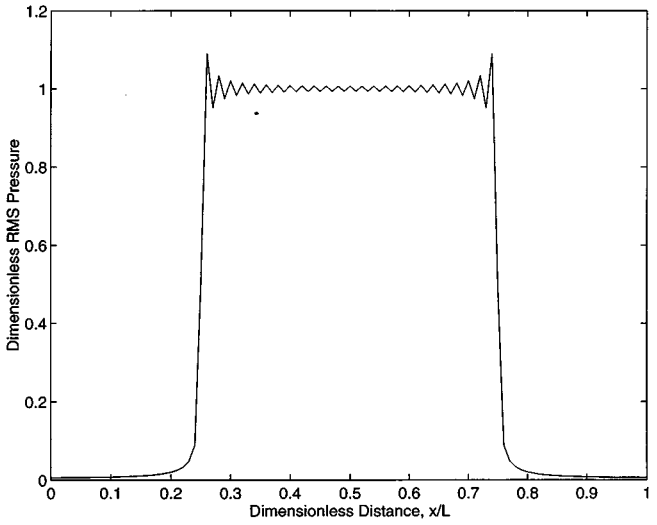


FIG. 8. Initial condition shown spatially. Dimensionless root-mean-square pressure versus dimensionless distance at time=0. $x/L=0$ corresponds to the hard wall, $x/L=1$ corresponds to the impedance wall. The impedance, $z_L=5$. Three curves plotted: Fully coupled=solid line; uncoupled=dashed line; approximate=dotted line.

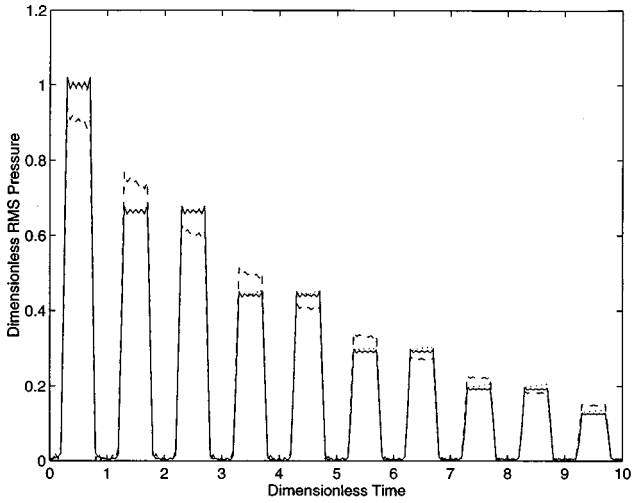


FIG. 9. Pressure level at $x=0$ (the hard wall) versus time. Three curves plotted: Fully coupled=solid line; uncoupled=dashed line; approximate =dotted line. The endwall impedance, $z_L=5$.

the P_n's solve the original fully coupled equation.

The constraint implied by Eq. (18) is that if $\dot{P}_0$ is specified independent of the P_n's, the physical law of mass conservation is violated. This consequence must be considered carefully in the case of modal uncoupling. The appropriate initial conditions must be chosen which correspond to the same physical problem as the fully coupled case. However, because the uncoupled modal equation for $n=0$ does not contain a summation over the $n\neq 0$ modes, mass conservation will be violated. If the initial P_n's and $\dot{P}_n$'s are identical in both cases, coupled and uncoupled, the corresponding solutions will not decay to the same value (namely, zero). However, if $\dot{P}_0$ for the uncoupled case is adjusted to satisfy the uncoupled mass conservation equation, the two solutions will decay to the same value.

In the uncoupled case, the equation for the $n=0$ modal amplitude is: $\ddot{P}_0+C_{00}\dot{P}_0=0$. This is a second order ODE that can be solved exactly: $P_0(t)=[-\dot{P}_0(t=0)/C_{00}]e^{-C_{00}t}+\text{Constant}$. In order for the mode to decay to zero, the constant must equal zero. From this equation, at $t=0$, for a given initial P_0 the appropriate initial condition for $\dot{P}_0$ is

$$\dot{P}_0(t=0)=-C_{00}P_0(t=0). \tag{19}$$

In an initial value problem, the P_n's and $\dot{P}_n$'s will be prescribed by the pressure and velocity distribution at $t=0$, for all n, including $n=0$. If initially there is no velocity at the $x=L$ wall, $\dot{P}_0=0$ on physical grounds [or mathematically from the above discussion of mass conservation, see Eq. (18)]. However, if $\dot{P}_0=0$ in the uncoupled case, P_0 will equal a constant and will not decay. In order to avoid this, the uncoupled initial condition vector should have a nonzero $\dot{P}_0$ term, and it should correspond to Eq. (19).

In the approximate coupling case, the initial pulse is assumed to be represented by a superposition of modes, each of which is treated separately. An uncoupled equation for the modal amplitude of one of these modes ($n=R$) is solved exactly, i.e., P_R becomes known. It is assumed in the approximate coupling approach that $P_R \gg P_r$ for all r's, other than n and R. Therefore, the summation over r is replaced by a single term $r=R$. The equation for the modal amplitudes P_n becomes

$$\ddot{P}_n+C_{nn}\dot{P}_n+\omega_n^2P_n=-C_{nR}\dot{P}_R \quad \text{for} \quad n\neq R. \tag{20}$$

This second order ODE can be solved for P_n, which has both a particular solution and a homogeneous solution. The homogeneous solution is identical to the solution of the P_R equation. The final solution for $P_n(t)$ is a superposition of these solutions for all R's which were present in the original pressure distribution.

Results comparing fully coupled, uncoupled, and approximately coupled approaches for a time-domain problem

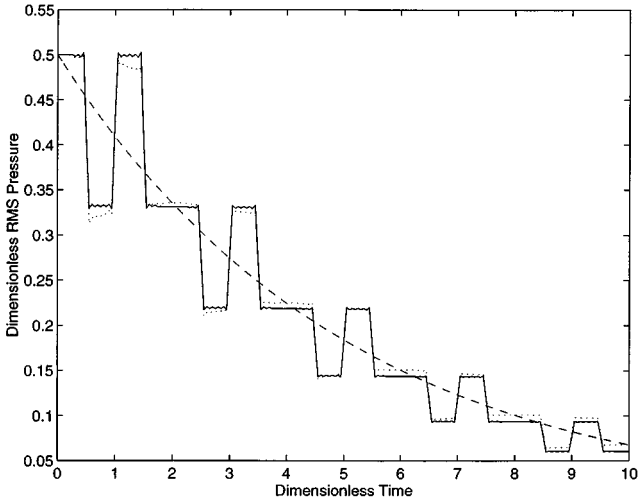


FIG. 10. Pressure level at $x=3L/4$ versus time. Three curves plotted: Fully coupled=solid line; uncoupled=dashed line; approximate=dotted line. The endwall impedance, $z_L=5$.

are shown in the following figures. The model problem is represented by the 1-D enclosure of Fig. 1, with the velocity end condition replaced by a hard wall. The initial pulse corresponds to a uniform pressure distribution at the center of the air column, which extends over half the column's length. The outer quarter of the enclosure on each side is at $p=0$, initially. This initial condition is shown in Fig. 8.

Figures 9 and 10 show the decay of the sound pressure level as a function of time for two different spatial locations: at the hard wall and at a cavity midpoint (one-quarter of the length away from the impedance wall), respectively. As shown in Fig. 9, at the hard wall the uncoupled solution alternates between overpredicting and underpredicting the exact level as each pulse goes by. In Fig. 10, at $x=3L/4$ (one-quarter length from the impedance wall) the uncoupled solution picks up the average levels, but misses the detailed behavior of the stepwise pressure changes. The uncoupled modal solution decays in a more gradual way than the exact solution. The uncoupled solution behaves as if the absorption is distributed throughout the tube rather than being located at the impedance end of the tube, as was mentioned earlier for the forced problem. The simple approximate method developed in this paper works extremely well at all locations in the enclosure, and picks up the detailed behavior of the pressure pulse.

III. SUMMARY AND CONCLUSION

Modal uncoupling leads to reasonable answers only if the frequency range does not include a modal resonance and if the coupling is extremely weak, e.g., high impedance surfaces. Since resonant behavior and absorbing surfaces are of practical importance, methods that completely neglect coupling may be of limited interest. This paper has developed an approximate method to include modal coupling and has shown formally the errors involved in uncoupling. To lowest order, the approximate method is analytically simple and performs extremely well in sample calculations, even when coupling is not small.

The approximate method was demonstrated for a one-dimensional model problem. The results were compared to an uncoupled analysis and to a fully coupled analysis for the model problem. Qualitatively, the uncoupled results appear to have distributed damping over the entire length of the duct, where in fact, the damping was only on an endwall. This behavior is exhibited in the standing wave pattern and intensity distribution for the forced problem, and in the temporal decay for the initial value problem.

Furthermore, it was shown that the individual modal amplitudes are effected by one another. This effect is ignored when the modes are uncoupled. The approximate solution includes the effect of the mode closest to resonance on the modal amplitude under consideration. The inclusion of this resonant mode is shown to greatly improve the results.

A high degree of modal overlap is often present in high-frequency sound fields. The implications for modal uncoupling under these conditions was not considered in this work, but should be investigated in future work. In addition, this work will enable theories like asymptotic modal analysis to be extended to more highly damped systems, which are of great practical importance.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Science Foundation.

APPENDIX A: RELATIONSHIP OF THE PRESENT WORK TO THE ITERATIVE METHOD OF SUCCESSIVE SUBSTITUTIONS

In previous work, the effect of weak modal coupling has been approximated by an iterative approach sometimes called the method of successive substitutions (Refs. 11 and 12). In this approach, the ''zeroth order'' iteration is obtained by neglecting the coupling altogether, and solving for each uncoupled modal response in closed form. This approximation is then used to estimate the previously neglected coupling terms, and the modal response can again be computed in closed form incorporating this estimate of modal coupling. The process can be repeated indefinitely, but the resulting expressions rapidly become unwieldy due to the infinite sum in the modal coupling term. The result of the first three iterations applied to Eq. (6) is as follows:

$$\overline{q}_n^{(0)}(\omega)=\frac{\overline{F}_n(\omega)}{[\omega_n^2-\omega^2+i\omega_n\omega C_{nn}]}, \tag{A1}$$

$$\overline{q}_n^{(1)}(\omega)=\frac{1}{[\omega_n^2-\omega^2+i\omega_n\omega C_{nn}]}\times\left[\overline{F}_n(\omega)-\sum_{r\neq n}\frac{i\omega_r\omega C_{nr}\overline{F}_r(\omega)}{\omega_r^2-\omega^2+i\omega_r\omega C_{rr}}\right], \tag{A2}$$

$$\overline{q}_n^{(2)}(\omega)=\frac{1}{[\omega_n^2-\omega^2+i\omega_n\omega C_{nn}]}\times\left[\overline{F}_n(\omega)-\sum_{r\neq n} i\omega_r\omega C_{nr}\overline{q}_r^{(1)}(\omega)\right]. \tag{A3}$$

For comparison, the corresponding results of the current method, Eqs. (13) and (14), are reproduced below:

$$\overline{q}_n(\omega)=\frac{1}{[\omega_n^2-\omega^2+i\omega_n\omega C_{nn}]}\times\left[\overline{F}_n(\omega)-\frac{i\omega_R\omega C_{nR}b_{nR}}{\omega_R^2-\omega^2+i\omega_R\omega C_{RR}}\overline{F}_R(\omega)\right], \tag{13}$$

$$\overline{q}_n(\omega)=\frac{1}{\omega_n^2-\omega^2+i\omega\omega_n C_{nn}}\times\left\{\overline{F}_n-b_{nR}\frac{i\omega\omega_R C_{nR}\overline{F}_R}{\omega_R^2-\omega^2+i\omega\omega_R C_{RR}}-\sum_{m=0}^{\infty} b_{mR}b_{nm}\frac{i\omega\omega_m C_{nm}}{\omega_m^2-\omega^2}\times\left(\overline{F}_m-b_{mR}\frac{i\omega\omega_R C_{mR}\overline{F}_R}{\omega_R^2-\omega^2+i\omega\omega_R C_{RR}}\right)+O[\epsilon^2]\right\}. \tag{14}$$

In the main text of this paper the purpose of the section associated with Eqs. (6)–(14) is to correctly identify the order of magnitude of terms, and then formally retain the appropriate terms at each order of a small parameter expansion. This correctly ordered development is one of the major contributions of this paper. The approach is a perturbation expansion method applied to the equations cast in a special dimensionless form [Eq. (8)] that correctly accounts for the order of magnitude of terms, *whether on or off resonance*. This approach is different from the iterative approach described above.

A comparison of Eq. (A1) and Eq. (13) shows that Eq. (13) contains an extra term. This term arises because, as shown in the main text, modal coupling is always important when the frequency is near one of the modal resonance frequencies. Physically, the response of the resonant mode is sufficiently large to counteract the small size of the coupling coefficient, making this one coupling term not negligible to lowest approximation.

On the other hand, the iterative procedure is not a formal perturbation method and as applied to this problem does not retain the orders of magnitude of terms in a consistent manner. This is why Eqs. (13) and (A1) do not agree. Indeed, the ''zeroth order'' iteration (which neglects modal coupling entirely and is called ''uncoupled'' in our comparison of results) is never strictly correct because modal coupling is always important at certain frequencies, as shown in the main text in the discussion of Eq. (9).

The important point is that Eq. (13), which has only one additional term, is hardly more complicated than the previously used uncoupled solution (A1), and, unlike that approach, it picks up the essential physical behavior. Figures 4, 5, and 10 are good illustrations of the comparison between uncoupled [''zeroth order'' iteration, Eq. (A1)], approximate [the lowest order result from this paper, Eq. (13)], and fully coupled (exact).

Finally, comparing the higher level approximations, Eq. (A2) performs no better than Eq. (13) because all the terms in the infinite sum except one are actually higher order. That one term corresponds to the additional term in Eq. (13) which accounts for being near a modal resonance. On the other hand, Eq. (14) properly accounts for all effects to the next highest order, and is hardly more complicated than Eq. (A2). Furthermore, Eq. (14) performs as well as (A3), which has an embedded double sum containing mostly superfluous terms, and Eq. (14) is much simpler.

APPENDIX B: EFFECT OF AN ON-DIAGONAL COUPLING COEFFICIENT BEING VERY SMALL OR EQUAL TO ZERO

The previous discussion defined dimensionless modal amplitudes which were equal to: $\hat{q}_n \equiv [\omega_n^2 - \omega^2 + i\omega\omega_n \varepsilon \hat{C}_{nn}]\overline{q_n}(\omega)/F_n(\omega)$ [this is Eq. (8) in the text]. However, in the original variables this would imply that the amplitudes must be infinite at resonance ($\omega = \omega_n$) if the on-diagonal coupling coefficient, $C_{nn} = 0$. This special case requires a separate analysis, but the modal amplitudes can still be found in closed form.

Suppose $C_{RR} = \delta \hat{C}_{RR}$ where $\hat{C}_{RR}$ is $O[1]$, and all other $C_{nr} = \varepsilon \hat{C}_{nr}$, with $\delta \ll \varepsilon$. Then Eq. (9) becomes

$$\frac{i\varepsilon \hat{C}_{nr}}{\omega_r/\omega - \omega/\omega_r + i\varepsilon \hat{C}_{rr}} = \begin{cases} O[\varepsilon] & \omega \neq \omega_r \\ O[1] & \omega \approx \omega_r \neq \omega_R \end{cases}$$

and

$$\frac{i\varepsilon \hat{C}_{nR}}{\omega_R/\omega - \omega/\omega_R + i\delta \hat{C}_{RR}} = \begin{cases} O[\varepsilon] & \omega \neq \omega_R \\ O[\varepsilon/\delta] & \omega = \omega_R \end{cases}.$$

From Eq. (7),

$$\hat{q}_n(\omega) + \sum_{r=0}^{\infty} \frac{i\varepsilon \hat{C}_{nr} b_{nr}}{\omega_r/\omega - \omega/\omega_r + i\varepsilon \hat{C}_{rr}} \hat{q}_r(\omega) = 1.$$

If $\omega \neq \omega_R$ then $\hat{q}_n + O[\varepsilon] = 1$, in other words, Eq. (10) holds. However, if $\omega = \omega_R$ then Eq. (11) becomes

$$\hat{q}_n + \frac{\varepsilon}{\delta}[\hat{C}_{nR}/\hat{C}_{RR}]\hat{q}_R + O[\varepsilon] = 1 \quad n \neq R,$$

$$\hat{q}_R + \sum_{r=0}^{\infty} \frac{i\varepsilon \hat{C}_{Rr} b_{Rr}}{\omega_r/\omega - \omega/\omega_r + i\varepsilon \hat{C}_{rr}} \hat{q}_r = 1, \quad n = R.$$

In this pair of equations, $\hat{q}_r$ in the summation can be replaced by $\hat{q}_n$ from the first equation of the pair, with $n = r$. Thus the second equation becomes

$$\hat{q}_R - \frac{\varepsilon^2}{\delta} \hat{q}_R \sum_{r=0}^{\infty} \frac{i\hat{C}_{Rr}\hat{C}_{rR} b_{Rr}}{\omega_r/\omega - \omega/\omega_r + i\varepsilon \hat{C}_{rr}} \frac{1}{\hat{C}_{RR}} + O[\varepsilon] = 1.$$

Now, $\hat{q}_R \equiv [\omega_R^2 - \omega^2 + i\omega\omega_R \delta \hat{C}_{RR}]\overline{q_R}/\overline{F_R}$ and if $\omega \approx \omega_R$, $\hat{q}_R \approx i\omega_R^2 \delta \hat{C}_{RR}\overline{q_R}/\overline{F_R}$, therefore,

$$i\omega_R^2 \delta \hat{C}_{RR}\overline{q_R}\left[1 - \frac{\varepsilon^2}{\delta}\frac{1}{\hat{C}_{RR}} \sum_{r=0}^{\infty} \frac{i\hat{C}_{Rr}\hat{C}_{rR} b_{Rr}}{\omega_r/\omega - \omega/\omega_r + i\varepsilon \hat{C}_{rr}}\right] = \overline{F_R}.$$

Letting $\delta \to 0$ yields

$$i\omega_R^2 \varepsilon^2 \overline{q_R} \sum_{r=0}^{\infty} \frac{i\hat{C}_{Rr}\hat{C}_{rR} b_{Rr}}{\omega_r/\omega - \omega/\omega_r + [\text{H.O.T.}]} = \overline{F_R},$$

which can now be solved for $\overline{q_R}$, which is $O[1/\varepsilon^2]$ instead of $O[1/\varepsilon]$ due to the remoteness of the backcoupling effect.

APPENDIX C: PHYSICAL BEHAVIOR OF UNCOUPLED SOLUTIONS AND EQUIVALENCE TO SPATIALLY DISTRIBUTED ABSORPTION

In the main text, it was noted that model problem solutions with modal uncoupling appear similar to a physical case in which absorption is distributed throughout the acoustic field, rather than at the impedance boundary. This Appendix shows this to be precisely true, and demonstrates that modal uncoupling changes the physical reality of the problem. This result gives considerable insight into the effect of modal uncoupling.

Consider a model problem that is a variation on the configuration shown in Fig. 1. Now the impedance end is replaced by a rigid termination ($Z_L = \infty$) at $x = L$, and the volume of the tube is filled with a porous media having flow resistance Φ. In porous media theory[13,14] the governing equa-

tions with harmonic forcing take on the familiar form for a simple fluid if the density and sound speed are replaced by equivalent complex quantities $c_e = c_p(1 - i\Phi/\rho_p c_p)^{-1/2}$ and $\rho_e = \rho_p(1 - i\Phi/\rho_p c_p)$. The quantities c_p and ρ_p are dependent on a number of details, with c_p lying between the isentropic and isothermal sound speeds (depending on the frequency range), and with ρ_p fairly near to ρ_0. If the fluid is assumed nonconducting and if secondary inertia effects (apparent mass) are neglected, then $c_p = c$ and $\rho_p = \rho_0$. In this case, only the effect of flow resistance is retained, which is sufficient for present purposes. Harmonic waves propagating in this medium exhibit exponential spatial decay due to the distributed dissipative effect of the flow resistance.

In carrying out the derivation for this case, the pressure field is again expanded in the hard box modes of the nondissipative case, namely the same modal functions used in the main text. The details of this straightforward derivation are omitted, with the result equivalent to Eq. (15) for harmonic forcing now being

$$(\omega_n^2 - \omega^2)P_n + \frac{\Phi}{\rho_0}\, i\omega P_n = i\omega\rho_0 c^2 v_0. \tag{C1}$$

It is important to note that these equations are uncoupled in their full formulation (not as an approximation).

For comparison, the approximate uncoupled equation to the original model problem is found from simplifying Eqs. (15) and (16) to be

$$(\omega_n^2 - \omega^2)P_n + \frac{i\omega\rho_0 c^2}{Z_L \Lambda_n}\, P_n = i\omega\rho_0 c^2 v_0, \tag{C2}$$

where $\Lambda_n = L$ if $n = 0$, and $\Lambda_n = L/2$ for $n > 0$. Comparing the above two equations (excluding $n = 0$) shows them to be identical if

$$\frac{\Phi}{\rho_0} = \frac{2\rho_0 c^2}{Z_L L}. \tag{C3}$$

This result shows that the effect of wall impedance in the original problem becomes equivalent to the effect of a distributed absorption when the equations are uncoupled.

The above analogy between uncoupling and distributed absorption applies except for $n = 0$, which differs by a factor of 2 from the above (due to the difference in Λ_n, which came from integrating $\cos^2 n\pi x/L$). In any event, the fact that this one mode has a different amplitude has no impact on the important conclusions.

The important point is that uncoupling the modes effectively changes the physical problem being solved. Removing the coupling prevents energy exchange between modes and results in significant changes in overall behavior, particularly as regards the acoustic intensity (flow of energy). The uncoupled problem can be understood in terms of the modified model problem with distributed absorption and a hard wall termination at $x = L$. As shown in the simulations presented in the main text, particularly Figs. 2–5, when the modes are *uncoupled* the following behavior is observed: energy is lost along the tube; the intensity vanishes at the impedance wall (rigid in the equivalent distributed problem, see Fig. 5); a pressure maximum always occurs at the impedance wall (compare the curves in Fig. 4, at $x = L$).

A final implication of this result, when extended to three-dimensional room acoustics, is that analyses using modal uncoupling will be unable to predict the correct distribution of acoustic intensity, and will be unable to predict the structure of special correlation zones (intensification and de-intensification zones[15]) near boundaries.

[1] E. H. Dowell, G. F. Gorman III, and D. A. Smith, ''Acousto-elasticity: General theory. Acoustic natural modes and forced response to sinusoidal excitation, including comparisons with experiment,'' J. Sound Vib. **52**, 519–542 (1977).

[2] Y. Kubota and E. H. Dowell, ''Asymptotic modal analysis for sound fields of a reverberant chamber,'' J. Acoust. Soc. Am. **92**, 1106–1112 (1992).

[3] L. Peretti and E. H. Dowell, ''Asymptotic modal analysis of a rectangular acoustic cavity excited by wall vibration,'' AIAA J. **30**, 1191–1198 (1992).

[4] Y. Kubota, H. D. Dionne, and E. H. Dowell, ''Asymptotic modal analysis and statistical energy analysis of an acoustic cavity,'' J. Vib. Acoust. **110**, 371–376 (1988).

[5] A. D. Pierce, V. W. Sparrow, and D. A. Russell, ''Fundamental structural-acoustic idealizations for structures with fuzzy internals,'' J. Vib. Acoust. **117**, 339–348 (1995).

[6] M. Strasberg and D. Feit, ''Vibration damping of large structures induced by attached small resonant structures,'' J. Acoust. Soc. Am. **94**, 1814–1815 (1993).

[7] D. A. Russell and V. W. Sparrow, ''Backscattering from a baffled finite plate strip with fuzzy attachments,'' J. Acoust. Soc. Am. **98**, 1527–1533 (1995).

[8] E. H. Dowell, ''Reverberation time, absorption and impedance,'' J. Acoust. Soc. Am. **64**, 181–191 (1978).

[9] J. Pan, ''A note on the prediction of sound intensity,'' J. Acoust. Soc. Am. **93**, 1641–1644 (1993).

[10] L. P. Franzoni and E. H. Dowell, ''On the accuracy of modal analysis in systems with damping,'' J. Acoust. Soc. Am. **97**, 687–690 (1995).

[11] H. Feshbach, ''On the perturbation of boundary conditions,'' Phys. Rev. **65**, 307–318 (1944).

[12] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).

[13] D. B. Bliss, ''Study of bulk reacting porous sound absorbers and a new boundary condition for thin porous layers,'' J. Acoust. Soc. Am. **71**, 533–545 (1982).

[14] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).

[15] L. F. Peretti and E. H. Dowell, ''A Study of Intensification Zones in a Rectangular Acoustic Cavity,'' AIAA J. **30**, 1199–1206 (1992).

Experimental evaluation of simplified models for predicting noise levels in industrial workrooms

Murray Hodgson
Occupational Hygiene Program and Department of Mechanical Engineering, University of British Columbia, 3rd Floor, 2206 East Mall, Vancouver, British Columbia V6T 1Z3, Canada

(Received 11 April 1997; revised 20 October 1997; accepted 17 November 1997)

Simplified models for predicting noise levels in industrial workrooms have been developed by Friberg, Thompson *et al.*, Wilson, Embleton and Russell, Kuttruff (''diffuse'' and ''specular'' models applicable to fitted rooms only), Zetterling, Sergeyev *et al.* (applicable only to untreated workrooms), and Hodgson. They predict octave-band or A-weighted steady-state sound-pressure level as a function of source/receiver distance. These models have been programmed and evaluated by comparing predicted sound-propagation curves with those measured in 30 empty and fitted industrial workrooms with and without absorptive ceiling treatments. In empty workrooms the Sergeyev *et al.*, Thompson, and Hodgson models worked quite well. The Zetterling model performed moderately well. The other models were inaccurate. Models underestimated levels in most cases. With the addition of absorbent treatments the accuracy of the Friberg, Wilson, Zetterling, and Embleton and Russell models improved; that of the Thompson and Hodgson models worsened. In fitted workrooms the Hodgson and Kuttruff (diffuse) models were accurate. The Friberg and Zetterling models were moderately accurate. The other models were inaccurate. The Thompson and Kuttruff (specular) models generally overestimated levels; the other models tended to underestimate levels. With absorbent treatment the accuracy of the Embleton and Russell model improved. [S0001-4966(98)02403-5]

PACS numbers: 43.55.Ka, 43.50.Jh [JDQ]

INTRODUCTION

Simplified models for predicting industrial noise levels have the advantage of simplicity and negligible calculation time over more generally applicable and comprehensive methods such as ray tracing. Simplified models predict octave-band or A-weighted steady-state sound-pressure level as a function of distance from a single omni-directional sound source of known output sound-power level. The industrial workroom is typically described by model parameters describing the room geometry, surface absorption, and contents (fittings—the numerous machines, stockpile, benches, etc. in the workroom).

In recent work,[1] existing simplified models for predicting noise levels in industrial workrooms were reviewed and critiqued. Most models were found to be conceptually inadequate; for example, some ignored a key parameter—the fittings. Preliminary attempts were subsequently made to develop an improved model—the Hodgson model.[2]

In the present work, nine simplified models were evaluated by comparing predicted sound-propagation curves—$\mathrm{SP}(r)$, the variation with distance r of the sound-pressure level L_p minus the source sound-power level L_w—with those measured in 30 empty and fitted industrial workrooms with and without absorptive surface treatments.

I. PREDICTION MODELS

Following are details of the nine models considered in this study. Generally the models apply to long, parallelepiped workrooms with no large internal barriers, with isotropically horizontally distributed fittings and with omni-directional sound sources.

A. The Friberg model

Friberg[3] developed an empirical formula for predicting the slope, assumed constant, of the A-weighted sound-propagation curve. The workroom fittings are assumed to be located on the floor and to have some average height. Surface absorption is quantified by the average mid-frequency ceiling-absorption coefficient, α'. The slope, in dB/dd (dd=distance doubling), is given by $-(a\alpha'+b)$, in which a and b are constants whose values depend on the fitting density, as determined by qualitative descriptors, and the room height-to-width ratio, as shown in Table I.

B. The ''Thompson'' model

Thompson *et al.*[4] proposed a modification to the expression describing steady-state levels according to diffuse field theory to allow its application to irregularly proportioned workrooms which, for omni-directional sources, is as follows:

$$\mathrm{SP}(r)=10\log[\exp(-mr)/4\pi r^2+4V/rS(\alpha S_w+4mV)]+10\log[(T+460)/527+30/\mathrm{BP}] \quad (1)$$

in which r is the source/receiver distance; α is the average room absorption coefficient; S_w is the wall surface area; m is the air-absorption exponent; V is the room volume; S is the room surface area, T is the temperature, and BP is the barometric pressure.

TABLE I. Constants required for predictions according to the Friberg model (Ref. 3). The room shape and fittings are categorized as follows (h = fitting height, H = room height, W = room width):

Room shape

N—rooms with $W<4H$

B—rooms with $W>6H$

Room contents

L—rooms with zero or low fitting density, or densely furnished with $h<H/8$

M—rooms with medium density of high fittings, or high density of fittings with $H/8<h<H/4$

H—rooms with high density of fittings with $h>H/4$

Room category	a	b
BH	−3.0	−4.0
BM	−2.5	−3.75
BL	−2.0	−3.5
NH	−3.0	−3.0
NM	−2.75	−2.75
NL	−2.5	−2.5

C. The Wilson model

Wilson[5] proposed a very simple method for predicting the slope of the A-weighted sound-propagation curve in workrooms with width and length at least four times the height. The slope is assumed to be constant with the following values:

(i) −3 dB/dd in acoustically hard, empty workrooms;
(ii) −4 dB/dd in acoustically hard, fitted workrooms or in absorbent-lined, empty workrooms;
(iii) −5 dB/dd in absorbent-lined workrooms with fittings.

D. The "*E* & *R*" model

Embleton and Russell[6] developed a model which constitutes a Canadian standard. It predicts levels on the assumption that SP(1 m) = −10.3 dB. It applies to empty rooms with rectangular plan shape and relatively constant height, with length and width which are at least four times the height, and which contain no internal barriers. Levels as a function of source/receiver distance are first determined for the room without ceiling and walls using Fig. 1. These levels are then corrected for ceiling and wall reflections using correction factors, presented in Table II. The corrections depend on the source/receiver distance, the room height, the distance from the source to side walls, and whether the reflecting surfaces are "poorly," "partly," or "highly" absorptive. In the present work these were interpreted as meaning $\alpha\leqslant 0.1$, $0.1<\alpha\leqslant 0.5$, and $\alpha>0.5$, respectively, where α is the surface absorption coefficient.

TABLE II. Corrections to noise levels for ceiling and wall reflections according to the Embleton and Russell model (Ref. 6).

A. Ceiling corrections

Distance / Height	Highly absorbing ceiling (dB)	Partly absorbing ceiling (dB)	Poorly absorbing ceiling (dB)
1.0	0	0	0
1.25	0	0.25	1
1.6	0	1.0	2
2.0	0	1.5	3
2.5	0	2.0	4
3.2	0	2.5	5
4.0	0	3.0	6
5.0	0	3.5	7
6.3	0	4.0	8
8.0	0	4.5	9
10.0	0	5.0	10
12.5	0	5.5	11
16.0	0	6.0	12
20.0	0	6.5	13

B. Wall corrections

Side-wall absorption	dB
Poorly absorptive	3
Partly absorptive	2
Highly absorptive	1

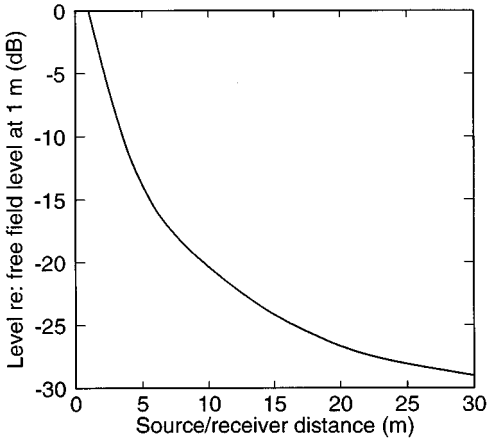


FIG. 1. Noise levels, relative to free-field levels at 1 m from a source, in workrooms without ceiling and walls, according to the E&R model (Ref. 6).

E. The Kuttruff models

Kuttruff[7] proposed two models for predicting octave-band sound propagation which only apply to wide and long workrooms which are fitted. The models are based on the assumption that a workroom with fittings on the floor can be modeled as an empty workroom with a diffusely reflecting floor. One model further assumes that the ceiling is also diffusely reflecting—for example, in the case of a nonflat roof. This "diffuse" model can be expressed as follows:

$$\mathrm{SP}_{\mathrm{tot}}(r)=10\log[A(r,h,\alpha)] \tag{2}$$

with

$$A(r,h,\alpha)=1/(4\pi r^2)+(1-\alpha)\{[(1+r^2/h^2)^{1.5}+b(1-\alpha)(b^2+r^2/h^2)^{1.5}/\alpha]/\pi h^2\} \tag{3}$$

in which $h=H$, the room height; r is the source/receiver distance; α is the average surface absorption coefficient; and β is a tabulated factor, approximated by an equation in the present work. The second "specular" model assumes a specularly reflecting ceiling. In this case,

$$\mathrm{SP}_{\mathrm{tot}}(r)=10\log[A(r,H,\alpha)+(1-\alpha_c)A(r',2H,\alpha_f)] \tag{4}$$

in which $r'=[r^2+(2H-3)^2]^{0.5}$, and α_c and α_f are the ceiling and floor absorption coefficients, respectively.

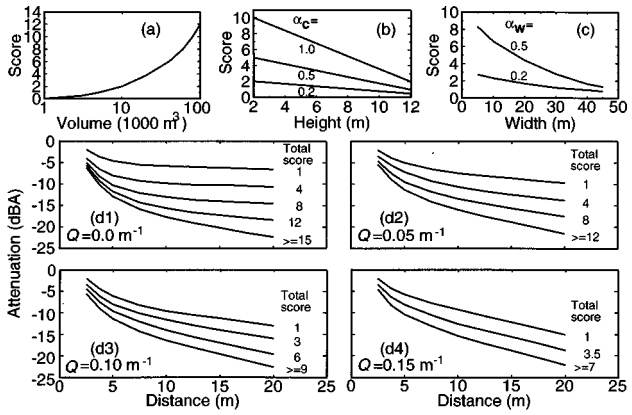


FIG. 2. Curves for predicting noise levels using the Zetterling model (Ref. 8). α_c and α_w are the average mid-frequency ceiling and wall absorption coefficients, respectively. Q is the average fitting density in l/m.

F. The Zetterling model

Zetterling[8] proposed a model to predict the reduction of A-weighted noise level relative to the level at 1 m from a source in parallelepiped workrooms of any shape without internal barriers. First, the ''acoustic quality of the room'' is quantified by a ''total score'' using Fig. 2(a)–(c) which relate this quantity to, respectively, the room volume, the ceiling height, and average mid-frequency absorption coefficient α_c, and the room width and average mid-frequency wall-absorption coefficients α_w. The reduction is then determined from the total score, the source/receiver distance, and the estimated fitting density Q, using the appropriate version of Fig. 2(d), or extrapolating between them. In the present work, the curves of Fig. 2(a)–(d) were approximated by equations. Note that the model only applies for $2<r<20$ m.

G. The ''Sergeyev'' model

Osipov, Sergeyev, and Shubin[9] developed a model which predicts 250–4000-Hz octave-band noise levels. It applies to parallelepiped workrooms, typical of those found in three industrial sectors (textile, metal-working, printing), which are untreated. Noise levels are determined using the following formula:

$$\mathrm{SP}(r)=10\log\{1/2\pi r^2+(1-\alpha)(r+W)J(\alpha,\rho)/[HW(r+H)]\} \quad (5)$$

with

$$J(\alpha,\rho)=0.1/[\alpha+\rho^2\ \exp(0.65\rho)] \quad (6)$$

in which $\rho=-r\ \ln(1-\alpha)/\mathrm{MFP}$; r is the source/receiver distance; α is the ''effective'' absorption coefficient; W is the room width; H is the room height; $\mathrm{MFP}=4V/S$ is the mean free path; V is the room volume; and S is the room surface area. Effective absorption coefficients—shown in Table III—were determined for typical empty workrooms and for fitted workrooms in the three industrial sectors.

TABLE III. Effective octave-band absorption coefficients for typical empty workrooms and furnished workrooms in various industries for use in the Sergeyev model (Ref. 9).

Case	Octave band (Hz)				
	250	500	1000	2000	4000
Empty workrooms	0.09	0.09	0.09	0.08	0.09
Textile industry	0.25	0.29	0.40	0.40	0.43
Printing industry	0.31	0.27	0.26	0.31	0.31
Metalworking industry	0.32	0.30	0.34	0.34	0.38

H. The Hodgson model

Using a database similar to that used in this evaluation, Hodgson[2,10] developed an empirical model for predicting 125–4000-Hz octave-band sound-propagation curves in workrooms which are either empty or ''typically'' fitted, without and with a ''typical'' sound-absorptive acoustical treatment. It takes the following form:

$$\mathrm{SP}(r)=I_E+\Delta I_F-3.3(S_E+\Delta S_F+\Delta S_A)\log r \quad (7)$$

in which r is the source/receiver distance, I_E and S_E are the intercept and slope applicable to empty, untreated workrooms, ΔI_F and ΔS_F are intercept and slope corrections applied to account for the fittings, and ΔS_A is the slope correction applied to account for absorbent treatment. Octave-band values for these parameters are given in Table IV.

II. WORKROOMS, SP MEASUREMENTS, AND PREDICTION PROCEDURES

The study involved 30 industrial workrooms. Their dimensions and aspect ratios are shown by workroom category in Table V. Note that some of the workrooms have aspect ratios which do not respect the limitations of some of the prediction models. All workrooms were of typical modern construction, with a steel-deck roof, concrete floor, masonry/glazing/metal-cladding walls, and horizontally uniform fitting distribution. The workrooms were in four categories: empty (10 workrooms); empty with absorbent surface treatment (5); fitted (10); fitted with absorbent surface treatment (5). The absorbent treatments consisted of various sound-absorbing materials applied to or suspended from the ceiling and/or all or part of the walls.

In each workroom sound-propagation curves were measured in octave bands from 125–4000 Hz. A calibrated omni-directional loudspeaker array was placed near one end wall, at half-width. Steady-state sound-pressure levels were

TABLE IV. Parameter values associated with Hodgson's frequency-dependent empirical prediction model (Ref. 2).

Band (Hz)	I_E	ΔI_F	S_E	ΔS_F	ΔS_A
125	−11.6	1.9	2.2	1.7	0.6
250	−11.3	2.1	2.1	1.7	1.0
500	−11.5	2.6	2.2	1.3	1.5
1000	−11.1	3.3	1.9	1.5	1.4
2000	−11.4	2.4	2.1	1.3	0.6
4000	−11.2	1.7	2.6	1.7	0

TABLE V. Main workroom input data: E=empty; F=fitted; A=absorbent; P=print shop; M=metalwork shop; h=fitting height; Q=fitting density.

Workroom	Length (m)	Width (m)	Height (m)	$\alpha_{c,\mathrm{mid}}$	$\alpha_{w,\mathrm{mid}}$	$\alpha_{\mathrm{av,mid}}$	Industry type	h (m)	Q (l/m)
E1	76	47	7.9	0.09	0.09	0.09	E	⋯	⋯
E2	70	49	6.6	0.11	0.11	0.08	E	⋯	⋯
E3	38	26	6.6	0.11	0.03	0.06	E	⋯	⋯
E4	63	30	6.3	0.13	0.03	0.06	E	⋯	⋯
E5	33	15	6.2	0.21	0.03	0.08	E	⋯	⋯
E6	62	39	6.1	0.17	0.03	0.08	E	⋯	⋯
E7	62	48	6.8	0.08	0.08	0.06	E	⋯	⋯
E8	86	30	7.0	0.13	0.03	0.08	E	⋯	⋯
E9	48	34	5.9	0.10	0.03	0.05	E	⋯	⋯
E10	52	25	7.0	0.14	0.03	0.07	E	⋯	⋯
EA1	64	58	8.2	0.27	0.03	0.12	E	⋯	⋯
EA2	62	49	6.8	0.82	0.03	0.13	E	⋯	⋯
EA3	128	24	8.2	0.37	0.03	0.17	E	⋯	⋯
EA4	58	20	7.7	0.36	0.03	0.15	E	⋯	⋯
EA5	104	45	8.0	0.25	0.03	0.11	E	⋯	⋯
F1	46	15	7.2	0.13	0.03	0.06	M	1.4	0.06
F2	68	21	4.1	0.13	0.03	0.07	P	1.1	0.08
F3	23	10	4.6	0.13	0.03	0.06	M	1.3	0.09
F4	37	27	7.3	0.13	0.03	0.06	P	1.2	0.10
F5	64	45	6.0	0.13	0.03	0.07	P	1.3	0.07
F6	38	31	5.1	0.13	0.03	0.07	P	1.5	0.12
F7	26	21	4.8	0.13	0.03	0.06	M	1.5	0.15
F8	30	22	7.2	0.13	0.03	0.06	P	1.0	0.05
F9	26	14	7.2	0.13	0.03	0.06	M	1.0	0.05
F10	27	21	6.2	0.06	0.03	0.06	M	1.4	0.07
FA1	48	36	7.0	0.38	0.03	0.18	M	1.3	0.08
FA2	46	27	5.6	0.38	0.40	0.16	M	2.0	0.10
FA3	91	38	5.1	0.38	0.03	0.18	P	1.4	0.08
FA4	26	14	7.2	0.38	0.03	0.13	M	1.0	0.05
FA5	30	22	7.2	0.38	0.03	0.14	P	1.0	0.05

measured at convenient distances along the workroom at half-width. Reverberation times were also measured.

The input data required for each of the models were estimated for each workroom. Many of these data—for example, absorption coefficients—cannot be measured directly and must, therefore, be estimated by various means. Thus, they have significant uncertainties associated with them; these result in uncertainties in the model evaluations. In fact, it would be most accurate to say that the present work is an evaluation of simplified prediction models and their input data. In any case, details of the input parameters, and how their values were estimated, when not obvious, are as follows:

(i) Workroom major dimensions;

(ii) Air absorption data—air absorption exponents in Np/m were determined from the temperature and relative humidities recorded at the time of the workroom measurements;

(iii) Surface absorption data—four surface absorption coefficients were required by the nine prediction models: wall, ceiling, floor, and average. In the case of empty workrooms, average octave-band absorption coefficients were obtained from the measured reverberation times using diffuse-field theory. Since the workroom floors were constructed of concrete, it was assumed that all floor coefficients took values increasing with frequency from 0.01 to 0.03, typical of dense concrete. Similarly, the walls of most of the workrooms were made of blockwork or plasterboard. Values for these surfaces varying between 0.02 and 0.04 were assumed. All workroom ceilings were of steel-deck construction. Some walls were of metal-cladding construction. The absorption coefficients of these surfaces—assumed the same—were calculated for each workroom from the average, blockwork, plasterboard, and concrete coefficients and associated surface areas. Resulting values typically decreased with frequency from about 0.25 to 0.1, typical of metal cladding. In empty workrooms with absorptive surface treatments, absorption coefficients were determined from the average absorption coefficients and those of the untreated surfaces. Resulting values varied from about 0.4 at mid-frequency to 0.6 at low frequency. In fitted workrooms the presence of the fittings reduces reverberation times, due to scattering effects, giving the illusion that there is more absorption than is actually present.[11] Thus, the use of diffuse field theory to obtain surface absorption coefficients is no longer accurate. Furthermore, in the case of fitted workrooms with acoustical treatments neither the co-

efficients of the cladding nor of the treatments could be determined reliably using the procedures described above. To obtain values for these coefficients, the same coefficients for the untreated, noncladding surfaces were used as in the empty case. Values for metal cladding and the acoustical treatments were, respectively, the average of the values calculated for the empty workrooms and their acoustical treatments. Mid-frequency (average of the 500-, 1000-, and 2000-Hz octave-band values) absorption coefficients were also calculated for each workroom, for use in the A-weighted models;

(iv) Fitting data—average-fitting heights and densities were estimated from a knowledge of the fittings involved and from experience.

Table V shows the main input data for the 30 workrooms.

The nine models were used to predict SP curves for each of the 30 workrooms, with the exception of the two Kuttruff models which were only compared to the fitted workrooms, and the Sergeyev model which was not compared to the workrooms with absorbent treatment. The 125–4000-Hz octave-band or A-weighted total levels were predicted as applicable. In the case of the Friberg, Wilson, and Zetterling models which do not predict the absolute level of the sound-propagation curve, this was normalized to SP(1 m)=−11 dB; i.e., free-field levels at 1 m from the source, according to common engineering practice, supported by recent results.[10] Predicted and measured results were compared graphically and statistically. Models which predict A-weighted total levels were compared with the energy-averaged SP levels in the 500-, 1000-, and 2000-Hz octave bands. Note that this essentially assumes a flat source output spectrum.

III. RESULTS

Figures 3 and 4 show the 1000-Hz octave-band and A-weighted results for untreated typical empty (workroom E4) and fitted (workroom F5) workrooms, respectively. Table VI shows, for the four categories of workroom, the average and standard deviation of the differences between the predicted and measured SP curves, averaged over all source/receiver distances and all frequencies for each model and workroom. Note that the results for workroom EA1 were highly atypical, for an unknown reason; thus, these results were not included in the statistics. Keep in mind also that the results regarding accuracy depend as much on the accuracy of the input data as on the inherent accuracy of the models, as discussed above. This is particularly true for the variation of the results with frequency; these are energy-based models which ignore wave effects and, thus, only consider frequency artificially by the variation with frequency of the input data. Averaging over all frequencies has the advantage of reducing errors due to parameter estimation. In any case, the results indicate the apparent accuracy of the models when used with typical input data. Note also that results averaged over all source/receiver distances provide no direct information on how the models perform, for example, near to or far from sources.

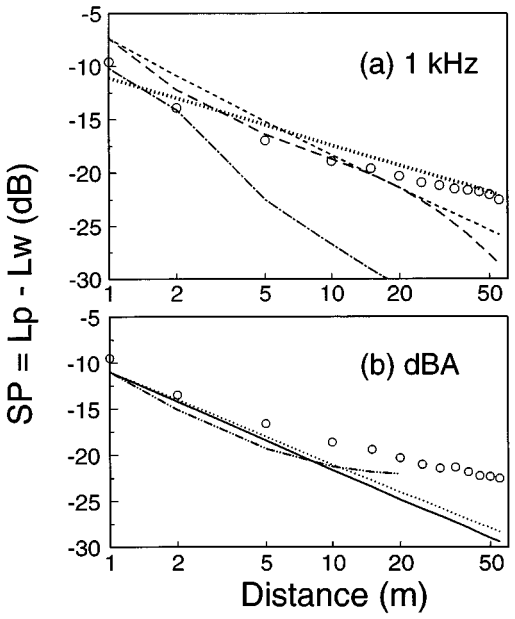


FIG. 3. (a) The 1-kHz octave-band, and (b) A-weighted total SP levels (○) as measured in empty workroom E4 and as predicted by seven models: (———) Friberg; (— — —) Thompson; (···) Wilson; (-–-–) Sergeyev; (·-·-·) Zetterling; (.–.–) E & R; (••••) Hodgson.

In the following, the main conclusions are generalized by model and by workroom type; accuracy is described in terms of whether the average difference between prediction and experiment was <1 dB (accurate), 1–3 dB (moderately accurate), or >3 dB (inaccurate).

A. Accuracy by model

The Friberg model is accurate in empty+absorbent workrooms, moderately accurate in fitted and

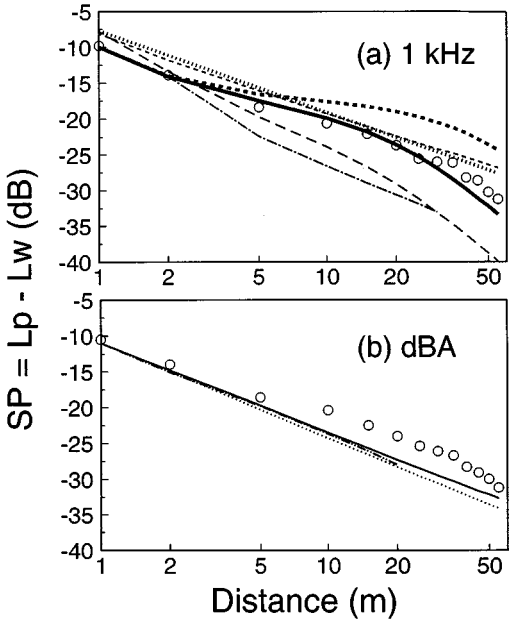


FIG. 4. (a) The 1-kHz octave-band and (b) A-weighted total SP levels (○) as measured in fitted workroom F5 and as predicted by nine models: (———) Friberg; (— — —) Thompson; (···) Wilson; (-–-–) Sergeyev; (·-·-·) Zetterling; (·–·–) E & R; (———) Kuttruff (diffuse); (- - - -) Kuttruff (specular); (••••) Hodgson.

TABLE VI. Difference between measured and predicted SP levels for nine simplified models and 30 workrooms, averaged over all source/receiver distances, and average and standard deviation of the average differences, averaged over all workrooms of each type.

Workroom	Friberg	Thompson	Wilson	Sergeyev	Zetterling	E&R	Kuttruff (diffuse)	Kuttruff (specular)	Hodgson
E1	−5.0	−1.1	−2.3	−0.5	−2.7	−6.4	⋯	⋯	0.9
E2	−3.5	0.3	−4.0	−1.0	−2.0	−7.7	⋯	⋯	−1.1
E3	−4.4	−1.1	−3.7	−1.6	−2.1	−7.1	⋯	⋯	−0.4
E4	−3.9	0.7	−3.8	−0.6	−1.9	−7.6	⋯	⋯	−1.4
E5	−3.8	0.9	−1.0	1.4	−0.9	−4.1	⋯	⋯	1.8
E6	−4.0	0.0	−1.5	−0.1	−1.4	−6.4	⋯	⋯	1.4
E7	−2.4	−1.5	−2.5	0.1	−2.7	−6.1	⋯	⋯	0.4
E8	−5.6	0.0	−3.5	−1.1	−2.0	−7.6	⋯	⋯	−0.5
E9	−4.0	−1.2	−4.6	−1.8	−2.8	−8.0	⋯	⋯	−1.1
E10	−5.1	−1.2	−2.6	−0.3	−3.1	−5.8	⋯	⋯	2.1
Avg (s.d.)	−4.2 (0.9)	−0.4 (0.9)	−3.0 (1.2)	−0.6 (0.9)	−2.2 (0.7)	−6.7 (1.2)	⋯	⋯	0.2 (1.3)
EA1	−8.8	−3.5	−6.0	⋯	−7.3	−7.6	⋯	⋯	−4.2
EA2	−1.2	1.6	−1.0	⋯	−0.4	−2.9	⋯	⋯	2.1
EA3	0.6	−1.0	−1.7	⋯	−0.5	−3.3	⋯	⋯	1.8
EA4	−0.9	0.7	−3.2	⋯	0.2	−4.5	⋯	⋯	0.3
EA5	−0.2	1.5	−0.7	⋯	−0.1	−2.0	⋯	⋯	2.6
Avg (s.d.)	−0.4 (0.8)	0.7 (1.2)	−1.7 (1.1)	⋯	−0.2 (0.3)	−3.2 (1.0)	⋯	⋯	1.7 (1.0)
F1	0.2	7.3	−1.3	−2.6	−1.9	−2.4	2.4	10.0	3.9
F2	−3.6	5.5	−5.8	−4.8	−3.2	−4.8	−0.3	2.7	−1.4
F3	−1.8	3.1	−4.5	−4.2	−3.0	−5.3	−1.1	2.0	−1.6
F4	−2.4	2.7	−3.3	−5.1	−2.3	−3.6	−0.2	5.9	1.1
F5	−1.4	6.3	−1.2	−0.5	−1.3	−1.4	2.8	7.1	2.7
F6	−4.2	4.3	−5.1	−4.3	−3.8	−4.4	−0.2	2.9	−0.9
F7	−0.9	4.2	−3.4	−2.3	−0.8	−3.9	−0.7	0.8	0.0
F8	−2.0	4.0	−4.9	−4.8	−1.4	−5.8	−1.9	0.0	−1.5
F9	−2.2	3.2	−5.1	−4.9	−2.4	−5.7	−1.6	0.7	−1.6
F10	−1.0	2.5	−4.6	−5.4	−1.2	−3.6	−1.9	1.0	−1.6
Avg (s.d.)	−1.9 (1.3)	4.3 (1.6)	−3.9 (1.6)	−3.9 (1.6)	−2.1 (1.0)	−4.1 (1.4)	−0.3 (1.7)	3.3 (3.3)	−0.1 (2.0)
FA1	−1.2	6.9	−5.3	⋯	−3.2	−2.5	−0.1	7.4	0.4
FA2	−0.7	4.4	−1.7	⋯	−1.3	0.1	1.8	10.1	2.7
FA3	−0.7	3.7	−3.0	⋯	−0.9	−1.6	−0.4	6.6	0.7
FA4	−1.0	3.1	−5.3	⋯	−0.5	−3.6	−1.6	2.6	−1.4
FA5	−1.8	3.2	−4.6	⋯	−0.8	−4.2	−2.0	1.1	−2.0
Avg (s.d.)	−1.1 (0.5)	4.3 (1.6)	−4.0 (1.6)	⋯	−1.3 (1.1)	−2.4 (1.7)	−0.5 (1.5)	5.6 (3.7)	0.1 (1.9)

fitted+absorbent workrooms, and inaccurate in empty workrooms. It underestimates levels in most cases. The underestimate increases with source/receiver distance; that is, the model overestimated the SP curve slope, especially in empty workrooms.

The Thompson model is accurate in empty and empty+absorbent workrooms, generally underestimating levels in the former and overestimating them in the latter. It is inaccurate in all fitted workrooms, generally overestimating levels. Levels are high at short distances. In empty workrooms this, and the fact that the slope is overestimated, results in the model incorrectly appearing to be accurate when the results are averaged over all source/receiver distances.

The Wilson model is inaccurate in all workroom types, with levels generally underestimated. The SP curve slope is overestimated.

The Sergeyev model is accurate in empty workrooms, but inaccurate in fitted rooms; levels are underestimated in these cases. In general, short-distance levels are too high—presumably because of the use of the $1/2\pi r^2$ term to describe the direct sound—and long-distance levels are too low, particularly in fitted rooms.

The Zetterling model is accurate in empty+absorbent workrooms, and moderately accurate in other room types. Levels are generally underestimated.

The Embleton and Russell model is moderately accurate in fitted+absorbent workrooms, but inaccurate in other cases. Levels are accurately predicted near the source, but are underestimated at larger distances.

The Kuttruff (diffuse) model is accurate in all fitted rooms, but tends to underestimate levels, especially at larger distances. The Kuttruff (specular) model is inaccurate, generally overestimating levels at all but the shortest distances.

The Hodgson model is accurate in all workroom types, tending to overestimate levels in fitted rooms and underestimate them in other room types.

B. Accuracy by workroom type

In empty workrooms the Sergeyev, Thompson, and Hodgson models worked quite well. However, the apparent

accuracy of the Sergeyev and Thompson models is deceptive; they overestimate at short distances and underestimate at large distances. The Zetterling model performed moderately well. The other models were inaccurate. Models underestimated levels in most cases. With the addition of absorbent treatments, the accuracy of the Friberg, Wilson, Zetterling, and Embleton and Russell models improved; that of the Thompson and Hodgson models worsened. In fitted workrooms the Hodgson and Kuttruff (diffuse) models were accurate. The Friberg and Zetterling models were moderately accurate. The other models were inaccurate. The Thompson and Kuttruff (specular) models generally overestimated levels; the other models tended to underestimate levels. With absorbent treatment the accuracy of the Embleton and Russell model improved.

IV. CONCLUSION

Predictions of workroom sound-propagation curves by nine simplified models have been compared to values measured in 30 empty and fitted workrooms without and with absorbent treatment. The Kuttruff (diffuse) and Hodgson models were found to be the most accurate. The Kuttruff model has the advantage that it is analytic; however, it does not apply to empty workrooms. The Hodgson model applies to both empty and fitted workrooms, but has the disadvantage of being an empirical model. That the Hodgson model performs well is not surprising, since it was developed from essentially the same database as that used to test the models in this work. The E & R model was found to be the least accurate.

ACKNOWLEDGMENT

The author would like to acknowledge with thanks the valuable contribution of Paul Li to the execution of this work.

[1] M. R. Hodgson, ''Review and critique of existing simplified models for predicting workroom noise levels,'' Can. Acoust. **19**(1), 15–23 (1991).
[2] M. R. Hodgson, ''Preliminary simplified models for predicting sound-propagation curves in workrooms,'' Can. Acoust. **20**(3), 37–38 (1992) and Erratum, Can. Acoust. **20**(4), 19 (1993).
[3] R. Friberg, ''Noise reduction in industrial halls obtained by acoustical treatment of ceilings and walls,'' Noise Control Vib. Reduct. **6**, 75–79 (1975).
[4] J. K. Thompson, L. D. Mitchell, and C. J. Hurst, ''A modified room acoustics approach to determine sound-pressure levels in irregularly-proportioned workroom spaces,'' Proc. Inter-Noise '76, 465–468 (1976).
[5] P. M. Wilson, ''A pragmatic look at sound propagation in real workroom spaces,'' Proc. IOA Conference on Noise Control in Workroom Buildings, 24–27 (1982).
[6] Canadian Standard Z107.52-M1983, ''Recommended practice for the prediction of sound-pressure levels in large rooms containing sound sources.''
[7] H. Kuttruff, ''Sound propagation in working environments,'' Proc. 5th FASE Symposium, Thessaloniki, 17–32 (1985).
[8] T. Zetterling, ''Simplified calculation model for noise propagation in large workrooms,'' Proc. Inter-Noise '86, 767–770 (1986).
[9] G. L. Osipov, M. V. Sergeyev, and I. L. Shubin, ''Optimum location of sound absorbing material and estimation of its noise-reduction efficiency in industrial spaces,'' Proc. InterNoise '87, Beijing, 683–686 (1987).
[10] M. R. Hodgson, ''Sound-propagation curves in industrial workrooms: statistical trends and empirical prediction models,'' J. Building Acoust. **3**(1), 25–32 (1996).
[11] M. R. Hodgson, ''Measurements of the influence of fittings and roof pitch on sound propagation in panel-roof factories,'' Appl. Acoust. **16**(5), 369–391 (1983).

Mechanisms of the cochlear nucleus octopus cell's onset response: Synaptic effectiveness and threshold

Kenneth L. Levy and Daryl R. Kipke
Bioengineering Program, Arizona State University, Tempe, Arizona 85287-6006

(Received 17 February 1997; revised 8 December 1997; accepted 19 December 1997)

Octopus cells are one of the principal cell types in the mammalian posteroventral cochlear nucleus. These cells respond to the onset of a toneburst with a precisely timed spike followed by little, if any, sustained activity. While experimental studies have partially characterized the cell, the mechanisms of this onset response are not well understood. The present study involved a model-based investigation that analyzed the responses of a compartmental model of the octopus cell in terms of synaptic effectiveness and dynamic spike threshold. The simulations demonstrate that properties of the onset response (first-spike latency, temporal precision of the first spike, and sustained firing rate) can be predicted from the values of these cell properties for a wide range of model configurations. These relationships were further analyzed through the development of mathematical expressions for synaptic effectiveness and dynamic spike threshold. This computational analysis resulted in a relatively simple explanation of the onset response, as well as predictions of the responses of octopus cells to nontonal, complex stimuli. © *1998 Acoustical Society of America.*
[S0001-4966(98)01304-6]

PACS numbers: 43.64.Bt, 43.64.Qh, 43.64.Ri [RDF]

INTRODUCTION

Octopus cells comprise one of the principal cell types in the posteroventral cochlear nucleus (PVCN). Each octopus cell receives afferent inputs from numerous auditory-nerve fibers which form excitatory synapses on its soma and its few short, thick dendrites that extend along the tonotopic axis (Osen, 1969; Kane, 1973; Rhode *et al.*, 1983; Liberman, 1993; Golding *et al.*, 1995). Several studies have associated octopus cells with PVCN units having an onset-type response pattern to short-duration tonebursts presented at the characteristic frequency of the unit (Godfrey *et al.*, 1975; Ritz and Brownell, 1982; Rouiller and Ryugo, 1984; Rhode *et al.*, 1983; Smith *et al.*, 1993; Feng *et al.*, 1994). The defining features of this onset response are the single, precisely timed spike that occurs with great precision shortly after the onset of nearly each stimulus presentation and the low or absent sustained firing during the steady-state portion of the stimulus. These responses have led to the identification of O_I and O_L unit types. While the experimental findings indicate that not all O_I and O_L units in PVCN can be attributed to octopus cells, some of the variability in onset responses probably reflects variations within the octopus cell population. Namely, O_I and O_L units have first-spike latencies between two to four milliseconds, with standard deviations of 0.1–0.3 ms, and sustained firing rates from zero to about 200 spikes/s (Rhode and Smith, 1986; Winter and Palmer, 1995). The tendency of these units to respond at the onset of tonebursts over a broad frequency range suggests, in part, that octopus cells trade spectral resolution for temporal precision (Golding *et al.*, 1995). This could have implications in the encoding of temporal information, such as fundamental frequency (Kim *et al.*, 1986).

The mechanisms of the onset response of octopus cells are unknown, but several properties have been discussed as possible factors, including: (1) a brief synaptic integration time that requires near coincidence of inputs for spiking (Golding *et al.*, 1995); (2) many weak excitatory inputs that must temporally and spatially summate to be effective (Rhode and Smith, 1986; Golding *et al.*, 1995); and (3) increases in spike threshold related to increased depolarization (Ritz and Brownell, 1982; Feng *et al.*, 1994). In addition, based on observations of GABAergic terminals on octopus cells (Saint Marie *et al.*, 1989) and enhanced responses in PVCN onset units to GABA antagonists (Palombi and Caspary, 1992), inhibitory inputs have been suggested as a mechanism of reducing the sustained activity of octopus cells (Palombi and Caspary, 1992). While inhibitory responses from octopus cells have not been reported in an *in vivo* preparation using tonal stimuli (Feng *et al.*, 1994), nor in a slice (Golding *et al.*, 1995), inhibitory effects cannot be ruled out.

In a previous study (Levy and Kipke, 1997), a biologically plausible computational model of an octopus cell was developed to investigate relationships between morphological and electrophysiological properties of octopus cells and the corresponding onset responses recorded from octopus cells and PVCN onset units. The model was found to accurately simulate octopus cell responses to a wide range of stimuli, including current injections, CF tonebursts, amplitude modulated tonebursts, frequency modulated tonebursts, and wide-band noise. Inhibitory inputs were not required for the model to produce reasonable onset responses. A parameter sensitivity analysis was also undertaken to investigate the underlying mechanisms that gave rise to the onset response of the model (Kipke and Levy, 1997). It was found that the onset response was an emergent response of the model that resulted from the interactions of many parameters. There was no obvious relationship between the time constant of the model or the level of sustained depolarization and the onset response. There was a strong relationship,

however, between synaptic effectiveness (considered the ratio of the synaptic conductance to the leakage conductance) and the onset response. This ratio affected the amplitude of the synaptic potential as well as the efficiency in which it was transmitted to the soma. These computational studies also demonstrated that an increased sustained somatic depolarization usually corresponded to an increased sustained firing rate, and that increasing the active-channel sodium conductance density increased the sustained firing rate. This implies that sustained somatic depolarization affects the spike threshold and the onset response by a limited amount.

The objective of this study was to quantitatively investigate synaptic effectiveness and the dynamic spike threshold in the octopus cell model. Definitions of synaptic effectiveness and threshold were established and simulations were run to examine the effects of various model parameters on these two measures. This computational analysis resulted in a relatively simple explanation of the onset response, as well as predictions of the responses of octopus cells to nontonal, complex stimuli.

I. METHODS

A. Computational model of the octopus cell

The octopus cell model has been fully described in a previous paper (Levy and Kipke, 1997) and will only be summarized in this section. The compartmental model consisted of a $25\times25\ \mu$m cylindrical lumped soma, a $5\times15\ \mu$m axon hillock, a $3\times300\ \mu$m cylindrical axon, and a $17.6\times238\ \mu$m dendrite, which was electronically equivalent to four $7\times150\ \mu$m dendrites. The model had one independent excitatory synaptic conductance in each of the 14 dendrite compartments and 4 independent conductances on the single soma compartment. The conductances were driven by independent spike inputs whose firing sequences were precisely controlled. Following an input spike, the conductance waveforms were alpha functions with a time-to-peak of 0.5 ms and a peak density of $1.2\ \mathrm{mS/cm^2}$ (dendrite) or $0.3\ \mathrm{mS/cm^2}$ (soma). The synaptic transmission delay was represented by a constant 0.5 delay between the input spike and the onset of the post-synaptic conductance change. The axon hillock was represented as one compartment while the axon was divided into 21 compartments. These compartments each contained active sodium and potassium channels for spike generation. In contrast to the previous studies in which the axon hillock and axon were removed as part of the simplification process, these structures were retained in this study to allow spikes to be discriminated in the distal axon. The model parameters are described further in the Appendix. The model was implemented using the GENESIS neurosimulator (Bower and Beeman, 1995) running on a Sun Sparc 20 workstation (Sun Microsystems, Inc.).

B. Model parameter variations

From the reference configuration of the model, selected parameters were varied (halved and doubled) to study the effects on synaptic effectiveness and threshold. In order to isolate the primary from secondary effects of parameter variations, the dendrite length was adjusted to maintain a constant electronic length (0.6λ) and constant relative synaptic locations. The synaptic density was also changed to maintain a constant total synaptic conductance. In relation to a previous study (Kipke and Levy, 1997), changes to the soma and axon diameters in this study corresponded to changing the soma diameter while holding the active channel densities constant. Changes to the soma diameter alone corresponded to changing the active channel densities inversely to the soma diameter.

C. Measuring synaptic effectiveness

The first method of measuring synaptic effectiveness (method 1) emphasized the effects of sustained somatic depolarization. Current was injected into the soma to produce a desired sustained depolarization.[1] After a delay of nine milliseconds, an input was applied to the middle of the dendrite. This delay equaled about three time constants of the sodium inactivation parameter, the slowest parameter in the model, which effectively put the cell in steady state when the input was applied. The synaptic effectiveness for the given sustained depolarization was defined as the voltage difference between the sustained depolarization and the peak somatic depolarization caused by the input spike.

In method 2 of measuring synaptic effectiveness, somatic depolarization was produced by synchronous synaptic activity rather than injected current. At the onset of the simulation, a single input was applied in the middle of the dendrite. After a delay of 8 ms, two inputs were applied, each offset $20\ \mu$m from the middle of the dendrite. After another 8-ms delay, the paired inputs and the input to the middle of the dendrite were synchronously applied. This pattern continued with additional paired inputs added at additional 20-μm offsets. The 8-ms delay between inputs allowed the model to reach steady state before the next input pattern was applied. For a given input pattern, synaptic effectiveness was defined as the difference between the peak somatic potential when the middle input was included and the peak potential when it was not included. The somatic depolarization resulting from each set of paired inputs was also measured.

D. Measuring the dynamic threshold

Characterizing the dynamic threshold of the octopus cell model consisted of measuring the steady-state value and the time constant, each as a function of sustained somatic depolarization. Both measurements were facilitated by the property that the octopus cell model fired just once at the onset of injected current pulses. The steady-state threshold was measured by providing a two-level current pulse (Fig. 1). A current step was first injected into the soma to produce a desired subthreshold sustained depolarization. After the model reached steady state (about 9 ms), the current was increased to the level required to elicit a spike (Fig. 1A). The steady-state threshold voltage for the given sustained depolarization, $\theta_{ss}(\nu)$, was defined as the somatic potential measured at least nine milliseconds after the increase in injected current. The minimum injected current required to bring the model to threshold was determined with a resolution of 0.5 nA and is referred to as threshold current (I_1) below. The spike was

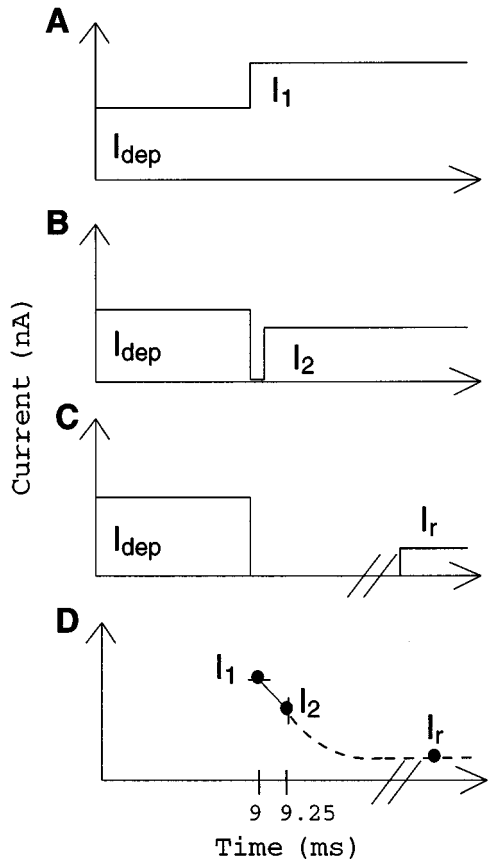


FIG. 1. Schematic diagram of the protocol used to measure the threshold time constant. Refer to the text for a description.

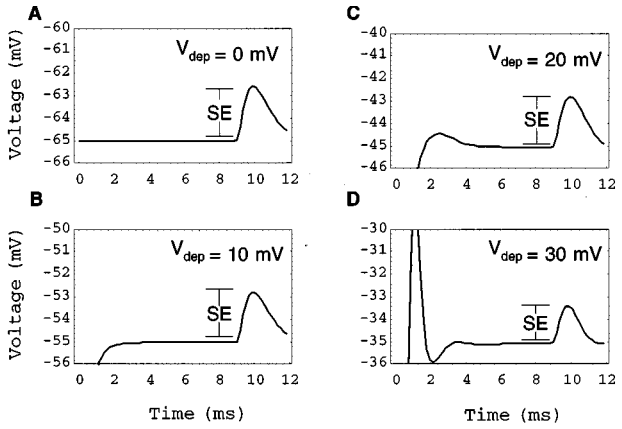


FIG. 2. Measurement of synaptic effectiveness using method 1 for four levels of depolarization: (A) 0 mV; (B) 10 mV; (C) 20 mV; and (D) 30 mV. Synaptic effectiveness is defined as the voltage change caused by applying one input spike to the center of the dendrite when the model is in steady state (9 ms in these simulations). These simulations used the reference model. Note that the ordinate range is different for each plot.

measured in the distal axon because during large somatic depolarizations, it was difficult to discriminate a somatic spike.

Measurement of the threshold time constant assumed that the threshold voltage adapted according to a first-order differential equation and that the injected current was linearly related to the threshold voltage. The measurement method involved first increasing the threshold by depolarizing the model with injected current. After steady state was reached, the current was removed and the threshold was measured at two points in time (Fig. 1D). At the cessation of the initial current pulse (I_{dep}), the dynamic threshold was assumed to decay toward its resting value according to

$$\tau_\theta \theta'(t) + (\theta(t) - \theta_r) = 0, \tag{1}$$

where $\theta(t)$ is the threshold voltage and τ_θ is the threshold time constant. Here, θ_r is the threshold voltage when the membrane is initially at rest, i.e., $\theta_r = \theta_{\text{ss}}(0)$. Equation (1) was solved to yield an expression for the threshold time constant as a function of the dynamic threshold measured at two points in time

$$\tau_\theta = (t_2 - t_1)\left[\ln\left(\frac{\theta(t_1) - \theta_r}{\theta(t_2) - \theta_r}\right)\right]^{-1}. \tag{2}$$

Under the linearity assumption, $\theta(t_1)$, $\theta(t_2)$, and θ_r could be replaced with the corresponding currents, I_1, I_2, and I_r, respectively. I_2 was the minimum amplitude required to elicit a spike for a current step that started $\Delta t = t_2 - t_1 = 0.25$ ms after the initial injected current pulse was turned off (Fig. 1B). This 0.25-ms delay was long enough to minimize digitization errors in the threshold current measurements, but short enough (less than one time constant) such that the threshold was still adapting. I_1 was previously defined. I_r was the threshold current when the membrane potential was at rest (Fig. 1C).

II. RESULTS

A. Synaptic effectiveness

Using method 1, synaptic effectiveness was found to decrease by a small amount as a function of sustained depolarization. Starting from the resting membrane potential to a sustained depolarization of 30 mV, synaptic effectiveness decreased from 2.4 mV to 1.7 mV (Fig. 2 and Table I). This effect was simply a result of the smaller differences between the local dendritic membrane potential and the synaptic reversal potential, i.e., the driving force for the synaptic current.

Most of the responses for each of the parameter variations showed a similar trend of synaptic effectiveness decreasing slightly with increases in sustained depolarization (Table I). Exceptions to this occurred at a depolarization of 20 mV, in which several parameter variations resulted in increased synaptic effectiveness that was probably due to the amplification of the synaptic potential by the active sodium channels as the somatic potential neared threshold. This idea was supported by the finding that in all such cases the steady-state threshold was relatively small (described below). Synaptic effectiveness could not be measured for a number of parameter variations because the model spiked continuously at the given sustained depolarization. Relatively large changes in synaptic effectiveness were observed for different parameter variations. Most notably, synaptic effectiveness increased with increased R_m and/or increased synaptic conductance density.

TABLE I. Synaptic effectiveness with sustained depolarization (method 1).

Sustained depolarization	Synaptic effectiveness (mV)				Estimate by equation (mV)				Mean error (%)
	0	10	20	30	0	10	20	30	
Reference									
⋯	2.4	2.3	2.3	1.7	2.4	2.2	2.0	1.8	5.8
Dendrite diameter									
1/2	4.8	4.5	5.5	3	5.3	4.8	4.3	3.9	17.2
2	1.0	0.9	0.9	0.8	1.0	0.9	0.8	0.7	5.9
Soma and axon diameter									
1/2	2.7	2.5	2.5	2.3	2.6	2.4	2.1	1.9	10.2
2	1.9	1.8	2.1	[a]	2.1	1.9	1.7	1.6	11.7
Soma diameter (not axon diameter)									
1/2	2.6	2.4	2.4	1.9	2.6	2.4	2.1	1.9	3.1
2	2.2	2.0	2.0	1.6	2.1	1.9	1.7	1.6	6.1
R_m									
1/2	1.9	1.8	1.7	1.4	2.1	1.9	1.7	1.5	5.8
2	2.7	2.5	2.8	[a]	2.2	2.0	1.8	1.6	24.7
C_m									
1/2	2.9	2.7	2.8	2.2	3.1	2.8	2.6	2.3	5.6
2	1.8	1.6	1.6	1.3	1.5	1.4	1.2	1.1	17.4
R_a									
1/2	1.8	1.7	1.7	1.4	1.8	1.6	1.5	1.3	6.2
2	3.2	3.0	3.0	2.4	3.2	2.9	2.6	2.3	5.2
$G_{\text{exc\_max}}$									
1/2	1.2	1.1	1.0	0.9	1.2	1.1	1.0	0.9	0.0
2	4.7	4.4	4.5	3.3	4.7	4.3	3.9	3.4	4.7
$\bar{G}_{Na}$									
1/2	2.4	2.2	2.1	1.6	2.4	2.2	2.0	1.8	4.3
2	2.4	2.3	[a]	[a]	2.4	2.2	2.0	1.8	2.2
$\bar{G}_K$									
1/2	2.4	2.2	2.4	[a]	2.4	2.2	2.0	1.8	5.6
2	2.4	2.2	2.2	1.4	2.4	2.2	2.0	1.8	9.4
Both $\bar{G}_{Na}$ and $\bar{G}_K$									
1/2	2.4	2.2	2.1	1.8	2.4	2.2	2.0	1.8	1.2
2	2.4	2.2	[a]	[a]	2.4	2.2	2.0	1.8	0.0

[a]Synaptic effectiveness could not be measured because the cell spiked continuously.

To gain a better understanding of the effects of each parameter on synaptic effectiveness, the results of the simulations were used to develop a mathematical expression for synaptic effectiveness as a function of somatic depolarization and selected model parameters. This set of equations was based on the lumped equivalent circuit of the compartmental model (Fig. 3), which consisted of four parallel branches representing: (1) the peak synaptic conductance for one excitatory input; (2) the input resistance for the dendrite; (3) the input resistance for the soma; and (4) the reactance for the lumped dendrite and soma. The peak synaptic conductance, g_{exc}, was given by the total synaptic conductance for the dendrite surface area divided by the number of synaptic inputs, $g_{\text{exc}}=\bar{G}_{\text{exc}}\pi d_d l_d/15$. Standard cable theory provided the expression for the input resistance for the dendrite (Rall, 1989), $re_{\text{in\_}d}=2\sqrt{R_a R_m}\coth[L]/\pi(d_d)^{3/2}$. The input resistance for the soma was given by $re_{\text{in\_}s}=(4R_a l_s/\pi d^2 s)+(R_m/\pi d_s l_s)$. Finally, the reactance of the dendrite and soma was given by $im_{\text{in}}=\tau_{\text{exc}}/C_m\pi(d_d l_d+d_s l_s)$, in which it was assumed that the major frequency component of the reactance produced by the input spike occurred at the time constant of the excitatory synaptic channel, τ_{exc}. The latter three branches were further combined into an equivalent conductance, $y_{\text{in}}=\sqrt{re_{\text{in}}^2+im_{\text{in}}^2}/re_{\text{in}}im_{\text{in}}$, where re_{in} is the equivalent resistance of the two input impedances in parallel.

From these terms, an expression for synaptic effectiveness (a voltage) was provided by the scaled voltage-divider circuit,

$$\hat{V}_{SE1}=ed\,\frac{(E_{\text{exc}}-V_m)g_{\text{exc}}}{g_{\text{exc}}+y_{\text{in}}}, \tag{3}$$

where ed is the electronic decay for a finite-length sealed-end dendrite (Rall, 1989) and V_m is the reduced voltage,

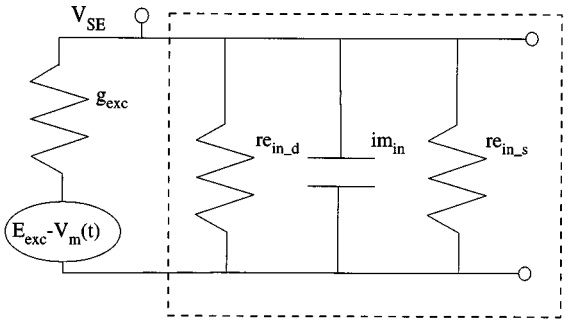


FIG. 3. The equivalent circuit used in deriving the equations that estimate synaptic effectiveness. V_{se} represents the synaptic effectiveness. Refer to text for description.

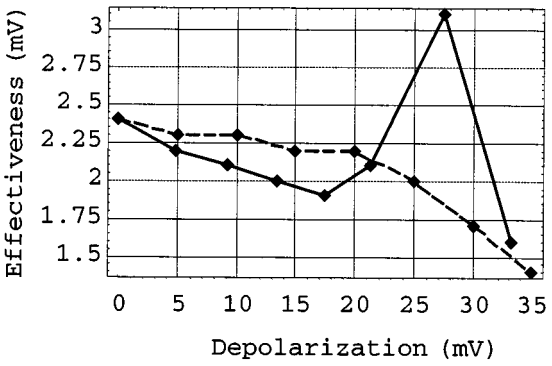


FIG. 5. Synaptic effectiveness versus sustained (dashed-line) and initial (solid-line) depolarization using method 2. Refer to the text for details.

$V_m = V_r + V_{\text{dep}}$. $V_r = -65$ mV is the resting membrane potential and V_{dep} is the sustained depolarization. The remaining variables in these equations are described in the Appendix. This set of equations closely predicts synaptic effectiveness for both the reference model and most of the parameter variations that were evaluated (Table I). The average error between the calculated synaptic effectiveness and the measured synaptic effectiveness from the compartmental model was 6% for the reference model configuration and 7% for all parameter variations.

Using method 2 to measure synaptic effectiveness demonstrated that as more inputs became active and increased the somatic depolarization, synaptic effectiveness decreased (Fig. 4). While this was similar to the previous result, the slope of synaptic effectiveness as a function of depolarization was larger over the first 17 mV of depolarization (-0.025 mV/mV vs -0.01 mV/mV; Fig. 5). This discrepancy between the two measurement methods was caused by the additional synaptic inputs in method 2 increasing the membrane conductance, which increased the passive decay

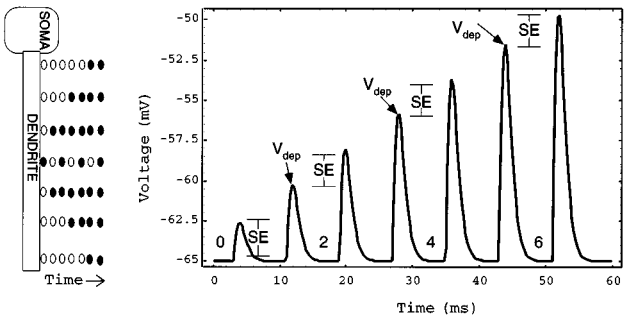


FIG. 4. Measurement of synaptic effectiveness using method 2. (Left) Illustration of the simulation protocol. The small circles indicate the placement of synaptic inputs along the modeled dendrite. Each column of circles represents a particular input configuration, with filled circles indicating active inputs and open circles indicating inactive inputs. The protocol involves seven input configurations presented sequentially. (Right) Somatic potential resulting from the input configurations. The first set of inputs was applied at 3 ms, with subsequent sets of inputs applied at 8-ms intervals. Synaptic effectiveness (SE) is measured by subtracting the somatic potential produced by the activation of the input at the center of the dendrite with the additional pair(s) of active inputs from the activation of the additional pair(s) alone. The initial depolarization produced by the initially active inputs is labeled Vdep. These simulations used the reference model.

of the synaptic potential as it was transmitted to the soma. The nonmonotonic behavior of synaptic effectiveness as the initial depolarization exceeded 20 mV reflected the effects of active sodium and potassium channels and spiking (Fig. 5, solid line). The positive slope resulted from amplification by the active sodium channels. The peak resulted from a spike. The falling phase indicated that the spike amplitude was not effected by additional active inputs.

Consistent with the trend seen in method 1, for nearly all parameter variations, synaptic effectiveness decreased with the number of input spikes (Table II). Exceptions to this trend were observed when the dendrite diameter was halved or the synaptic density was doubled (marked with asterisks in Table II). The exceptions were probably caused by the dynamics of the active channels amplifying the somatic potentials.

These data were accounted for in the mathematical expression for synaptic effectiveness by modifying Eq. (3) to include a scale factor,

$$\hat{V}_{SE2} = ed \frac{(E_{\text{exc}} - V_m) g_{\text{exc}}}{y_{\text{in}} + g_{\text{exc}}} \frac{y_{\text{in}}}{y_{\text{in}} + n g_{\text{exc}}}, \tag{4}$$

where n is the number of active input spikes. In this case, V_m is the initial depolarization caused by the active inputs without the active center input. This scale factor accounted for the extra decay in the synaptic potential produced by the increase in membrane conductance as more inputs became active. The average error between the calculated and measured synaptic effectiveness for method 2 was 1% for the reference model and 5% for all variations.

B. Threshold

The steady-state threshold and threshold time constant were measured for sustained depolarizations from 0 to 50 mV in 5-mV steps for the reference model and from 0 to 40 mV in 10-mV steps for each parameter variation. Figure 6 shows the responses of the reference model just below (left column) and just above threshold (right column) for a sustained depolarization of 20 mV. In both cases, a constant current pulse of 6 nA was applied to provide the 20-mV depolarization. A step increase to 7.5 nA caused increased depolarization to 21 mV in the soma (Fig. 6A) but no spike in the axon (Fig. 6B). A step increase to 8 nA was sufficient

TABLE II. Synaptic effectiveness with synaptic inputs (method 2).

Number of initially active synaptic inputs	Initial depolarization (mV)			Synaptic effectiveness (mV)				Estimate with equation (mV)				Mean error (%)
	2	4	6	0	2	4	6	0	2	4	6	
Reference												
⋯	4.8	9.2	13.5	2.4	2.2	2.1	2.0	2.4	2.3	2.1	2.0	1.1
Dendrite diameter												
1/2	9.1	17.3	35.0	4.8	4.1	4.3	3.9[a]	5.3	4.6	4.0	3.1	9.9
2	2.0	4.0	5.9	1.0	1.0	0.9	0.9	1.0	0.9	0.9	0.9	2.5
Soma and axon diameter												
1/2	5.4	10.5	15.2	2.7	2.5	2.3	2.2	2.6	2.4	2.2	2.1	4.1
2	3.7	7.3	10.7	1.9	1.8	1.7	1.6	2.1	2.0	1.9	1.8	11.5
Soma diameter (not axon diameter)												
1/2	5.1	9.8	14.4	2.6	2.4	2.2	2.1	2.6	2.4	2.2	2.1	0.0
2	4.2	8.2	12.0	2.1	2.0	1.9	1.8	2.1	2.0	1.9	1.8	0.0
R_m												
1/2	3.8	7.4	10.9	1.9	1.8	1.7	1.6	2.1	2.0	1.9	1.8	11.5
2	5.3	10.3	15.2	2.7	2.5	2.3	2.3	2.2	2.0	1.9	1.8	19.4
C_m												
1/2	5.6	10.8	15.7	2.9	2.6	2.4	2.3	3.1	2.9	2.6	2.4	7.8
2	3.7	7.2	10.6	1.8	1.7	1.7	1.6	1.5	1.4	1.4	1.3	17.7
R_a												
1/2	3.6	7.0	10.3	1.8	1.7	1.6	1.5	1.8	1.7	1.6	1.5	0.0
2	6.3	12.1	17.5	3.2	2.9	2.7	2.6	3.2	2.9	2.7	2.4	1.9
$G_{\text{exc\_max}}$												
1/2	2.4	4.8	7.1	1.2	1.2	1.1	1.1	1.2	1.2	1.1	1.1	0.0
2	9.2	17.4	27.2	4.7	4.1	3.9	5.4[a]	4.7	4.1	3.6	3.1	2.6
$\overline{G}_{Na}$												
1/2	4.8	9.2	13.5	2.4	2.2	2.1	2.0	2.4	2.3	2.1	2.0	1.1
2	4.8	9.2	13.6	2.4	2.2	2.1	2.0	2.4	2.3	2.1	2.0	1.1
$\overline{G}_K$												
1/2	4.8	9.2	13.5	2.4	2.2	2.1	2.0	2.4	2.3	2.1	2.0	1.1
2	4.8	9.2	13.5	2.4	2.2	2.1	2.0	2.4	2.3	2.1	2.0	1.1
Both $\overline{G}_{Na}$ and $\overline{G}_K$												
1/2	4.8	9.2	13.5	2.4	2.2	2.1	2.0	2.4	2.3	2.1	2.0	1.1
2	4.8	9.2	13.6	2.4	2.2	2.1	2.0	2.4	2.3	2.1	2.0	1.1

[a]A spike was produced by the simultaneous presentation of the inputs. In this case, synaptic effectiveness is artificially large due to the amplification effect of the active channels.

to elicit a spike followed by a sustained somatic depolarization of 25 mV (Fig. 6C and D). From these data, the steady-state threshold was found to be $\theta_{ss}(20) = -65 + 25 = -40$ mV. While these current levels were relatively high for the purpose of estimating the threshold parameters, the model was previously verified to be accurate for smaller currents; an injected current of 1 nA was found to maintain the soma at less than 5 mV depolarized (Levy and Kipke, 1997).

Figure 7 summarizes the measured steady-state thresholds for the reference model over the entire depolarization range tested. The threshold remained relatively constant at -42 mV (23 mV above the resting potential) from 0 to 18 mV depolarized and then increased approximately linearly with a slope of one and offset of 5 mV for larger depolarizations. This suggests that the increase in sodium activation balanced the increases in sodium inactivation and potassium activation. When the membrane was initially at rest, doubling R_m or $\overline{G}_{Na}$, or halving R_a or $\overline{G}_K$, caused the steady-state threshold to decrease. The effects of R_m, $\overline{G}_{Na}$, and $\overline{G}_K$ can be explained by the fact that for an isolated patch of membrane to reach threshold, the inward sodium current must be greater than the outward potassium and leakage current. R_a comes into play because of axial currents into and out of the active segments of the membrane.

In comparison to the reference model, the general form of the steady-state threshold versus sustained depolarization curves were similar for most of the tested parameter varia-

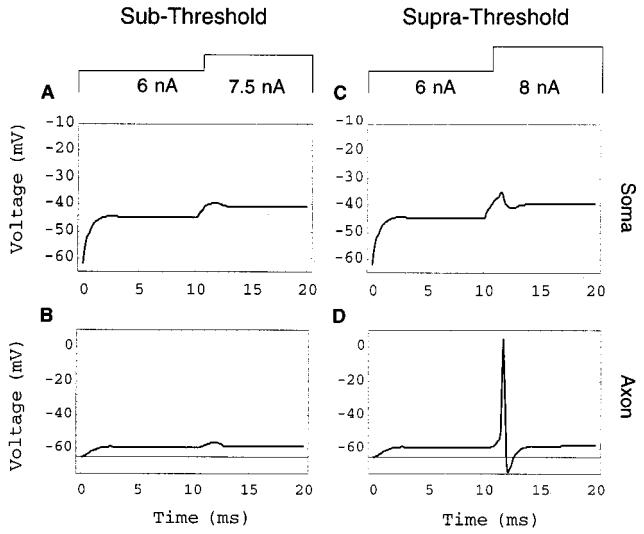


FIG. 6. Determination of the steady-state threshold for 20-mV sustained depolarization. The left column shows subthreshold responses and the right column shows supra-threshold responses. The insets at the top indicate the injected current for each case. The top row shows the somatic potential and the bottom row shows the axonal potential. These simulations used the reference model.

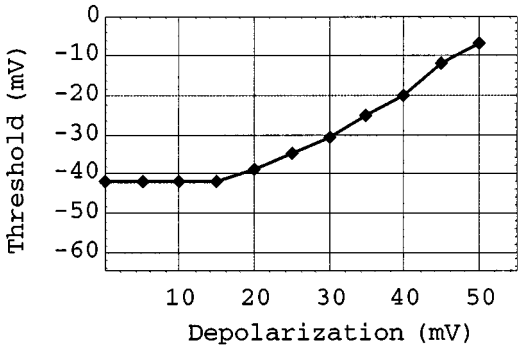


FIG. 7. Steady-state threshold as a function of sustained depolarization for the reference model.

tions (Table III). The slope of the threshold curve was generally more invariant to parameter variations than was the offset value of the threshold. The exception was a slope increase when the maximum active sodium conductance density was halved. The threshold offset changed for all parameters variations, except for changes in dendrite diameter. It was not possible to make a threshold measurement for a few parameter variations that resulted in continuous spiking.

These results for spike threshold were summarized along four main points: (1) the steady-state threshold was relatively constant below 18 mV of depolarization; (2) the threshold increased with a slope of about one for depolarizations greater than 18 mV; (3) changes in $\bar{G}_{Na}$ were more effective than changes in $\bar{G}_K$; and (4) changes in R_m produced relatively large changes in threshold. From these data, the steady-state threshold voltage was expressed as

$$\hat{\theta}_{ss}(\nu)=\begin{cases}\theta_0+18, & \nu<18,\\ \theta_0+\nu, & \nu\geqslant 18,\end{cases} \tag{5}$$

where $\hat{\theta}_{ss}(\nu)$ is the steady-state threshold voltage in millivolts, ν is the sustained depolarization in mV, and θ_0 is a constant given by

$$\theta_0=3\left(\frac{4400}{a_a}+\frac{C_m}{4}+\frac{1}{R_m}+\frac{\bar{G}_K}{200}-\frac{\bar{G}_{Na}}{18}-\frac{1}{14R_a}\right), \tag{6}$$

where a_a is the total surface area of the active membrane. Rather than being a mathematical reduction from the underlying compartmental model, the scale factor for each term in Eq. (6) was determined empirically in the process of fitting the data. While there is a logical rationale for the inclusion of each term, the terms do not necessarily represent a physical property of the model. Using Eq. (5), the estimated steady-state threshold was calculated for all parameter variations and sustained depolarization levels (Table III). The correspondence between the measured and estimated values was generally quite good, with an average error of 19% for all variations (excluding 1/2 G_{Na}).

Using Eq. (2), the threshold time constant, τ_θ, was calculated for sustained depolarizations from 25 to 50 mV in 5-mV steps for the reference model and at 30 mV and 40 mV depolarized for each parameter variation. The threshold could not be estimated for sustained depolarizations less than 25 mV because these initial depolarizations do not appreciably move the threshold from its resting level. The resting threshold level for current (I_r) and voltage (θ_r) were previously determined to be 7 nA and -42 mV, respectively (reference model at sustained depolarization$=0$ mV). Furthermore, for each level of sustained depolarization, the measured steady-state threshold provided the threshold current levels (I_1) at the time immediately following the cessation of sustained current. Determining the I_2 current levels required additional simulations. Using one level of sustained depolarization as an example, a steady injected current of 15 nA caused an initial spike followed by sustained depolarization of 40 mV (Fig. 8, left). After 9 ms the model reached steady state and this current was turned off for 0.25 ms and then turned back on at a lower level of 11.5 nA. The second pulse was not sufficient to cause the model to fire. Increasing the level of the second pulse to 12 nA was sufficient to elicit a spike indicating that $I_2=12$ nA (Fig. 8, right). Combined with $I_1=18$ nA [the current corresponding to $\theta_{ss}(40)$] and $I_r=7$ nA, the threshold time constant was calculated as 0.32 ms. The threshold time constant decreased from 0.49 ms to 0.31 ms as the somatic depolarization increased from 30 mV to 50 mV (Fig. 9). For each parameter variation, the threshold time constants were calculated for depolarizations of 30 and 40 mV (Table IV).

III. DISCUSSION

The onset response of octopus cells is notable for the precisely timed initial spike shortly after the stimulus is turned on, followed by the cell effectively ''turning off'' during the steady-state portion of the stimulus. While the compartmental model used in this study is a relatively simple

TABLE III. Steady-state threshold.

Sustained depolarization (somatic potential)	Threshold voltage (mV)					Estimate by equation (mV)					Mean error (%)
	0 (−65)	10 (−55)	20 (−45)	30 (−35)	40 (−25)	0 (−65)	10 (−55)	20 (−45)	30 (−35)	40 (−25)	
Reference											
⋯	−42	−42	−39	−31	−20	−42	−42	−39	−31	−20	0.0
Dendrite diameter											
1/2	−42	−42	−39	−31	−20	−42	−42	−39	−31	−20	0.0
2	−42	−42	−39	−30	−18	−42	−42	−39	−31	−20	2.9
Soma and axon diameter											
1/2	−41	−41	−37	−29	−15	−38	−38	−36	−26	−16	6.9
2	−45	−44	−42	3spks	4spks	−44	−44	−42	−32	−22	0.7
Soma diameter (not axon diameter)											
1/2	−42	−42	−38	−28	−16	−42	−42	−40	−30	−20	7.5
2	−42	−42	−39	−31	−19	−42	−42	−40	−30	−20	2.2
R_m											
1/2	−39	−39	−36	−24	−12	−36	−36	−34	−24	−14	7.5
2	−45	−45	−43	7spks	8spks	−45	−45	−43	−33	−23	0.0
C_m											
1/2	−43	−43	−41	−33	−22	−43	−43	−41	−31	−21	2.1
2	−40	−40	−37	−26	−12	−41	−41	−39	−29	−19	16.1
R_a											
1/2	−44	−44	−42	−33	4spks	−44	−44	−42	−32	−22	0.8
2	−41	−41	−39	−29	−18	−41	−41	−39	−29	−19	1.1
$\overline{G}_{Na}$											
1/2	−35	−34	−29	−8	14	−39	−39	−37	−27	−17	102.5
2	−46	−46	4 spks	9spks	10spks	−48	−48	−46	−36[a]	−26[a]	4.3
$\overline{G}_K$											
1/2	−43	−43	−41	2spks	3spks	−43	−43	−41	−31	−21	0.0
2	−42	−42	−38	−25	−8	−41	−41	−39	−29	−19	32.2
Both $\overline{G}_{Na}$ and G_K											
1/2	−38	−37	−34	−23	−7	−40	−40	−38	−28	−18	40.8
2	−47	−45	−43	3spks	3spks	−47	−47	−45	−35[a]	−25[a]	3.0

[a]These estimated thresholds would spike continuously for sustained depolarization since the threshold is at or below the depolarization value. Spks=The number of spikes that occurred when trying to estimate the threshold. The underlying assumption is that the more spikes that occurred, the lower the threshold.

representation of the cell, several factors support its relevance to understanding the cell: the membrane and conductance equations have well-known biophysical relevance; many of the model parameters were set from experimental data obtained from octopus cells (Rhode *et al.*, 1983; Feng *et al.*, 1994; Golding *et al.*, 1995; Rouiller and Ryugo, 1984; Romand, 1978; Winter and Palmer, 1995; Godfrey *et al.*, 1975; Ritz and Brownell, 1982; Osen, 1969; Kane, 1973); the response properties of the model were extensively verified against experimental data (Levy and Kipke, 1997); and the model was largely insensitive to small parameter variations (Kipke and Levy, 1997). From this foundation, the model-based quantitative measurements of synaptic effectiveness and threshold obtained in this study can be used to propose an explanation of the underlying mechanism for the onset response of an octopus cell.

The onset peak in the PSTH results from the threshold being low at the onset of the toneburst because the membrane is not depolarized. The precision in the onset peak is due to the fact that many active inputs are needed within a short time window (i.e., near-coincident inputs) to produce a somatic depolarization that reaches threshold. The degree of coincidence needed in the input spikes is a function of the short membrane time constant, threshold time constant and synaptic potential rise time. This summation of inputs provides a mechanism by which temporal variations in individual inputs are removed (i.e., the inputs are averaged) and the timing of the first output spike becomes more precise.

The low sustained firing rate in the octopus cell model is due to both a decrease in synaptic effectiveness and an increase in threshold that occurs during the sustained portion of the toneburst. These changes result from the sustained portion of the toneburst producing a sustained somatic depolarization. During the sustained depolarization, additional synaptic inputs become less effective, with a slope of -0.025 mV/mV depolarized. The decrease in synaptic effec-

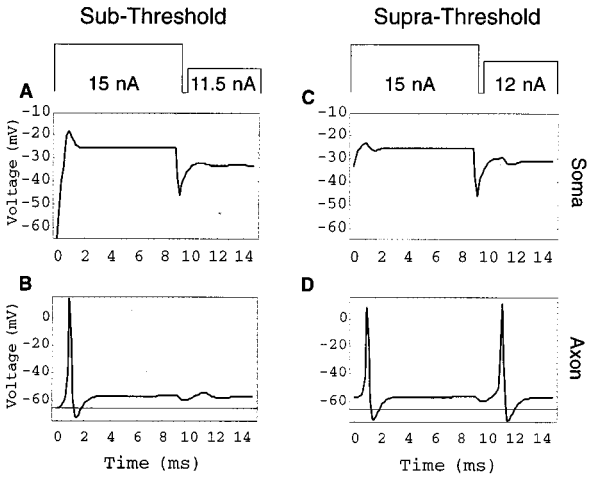


FIG. 8. Determination of the threshold time constant for 40-mV sustained depolarization. Same format as Fig. 6. These simulations used the reference model.

tiveness is due to both a decrease in the driving function (the difference between the synaptic potential and the local membrane potential) and a leakier membrane. The threshold increases with sustained depolarization in a linear fashion with unity slope. When comparing how synaptic effectiveness and threshold vary with sustained depolarization, the threshold is found to increase about 40 times the amount that the synaptic effectiveness decreases. Thus threshold effects dominate changes in sustained rate.

Within the parameter space that was investigated, the precision of the first-spike latency increases or remains the same for parameter variations that decrease synaptic effectiveness and/or increase the threshold, except when the maximum synaptic density is decreased by one-half (see Tables II and III). This indicates that more of the less effective input spikes are required to produce an output spike, leading to increased averaging of the synaptic inputs and, ultimately, to improved precision in the timing of the onset spike. When the maximum synaptic density is decreased, the precision of the first spike decreases because the synaptic effectiveness is too small; by the time the octopus cell model initially spikes,

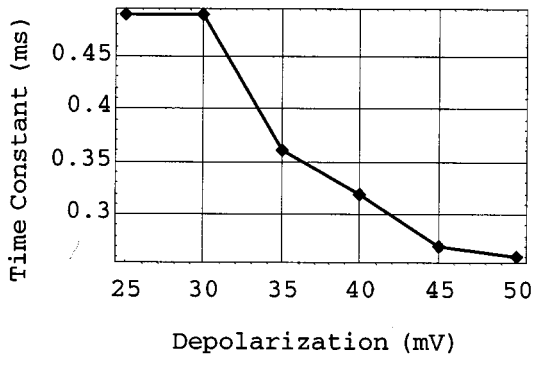


FIG. 9. Threshold time constant as a function of sustained depolarization for the reference model.

TABLE IV. Threshold time constant.

	Time constant (ms)	
Sustained depolarization	30 mV	40 mV
Reference		
…	0.49	0.32
Dendrite diameter		
1/2	0.36	0.23
2	0.54	0.36
Soma and axon diameter		
1/2	0.45	0.28
2	…	…
Soma diameter (not axon diameter)		
1/2	0.43	0.27
2	0.36	0.27
R_m		
1/2	0.54	0.43
2	…	…
C_m		
1/2	0.31	0.21
2	0.68	0.59
R_a		
1/2	…	…
2	0.36	0.33
$\bar{G}_{Na}$		
1/2	0.54	0.64
2	…	…
$\bar{G}_K$		
1/2	…	…
2	0.46	0.32
Both $\bar{G}_{Na}$ and $\bar{G}_K$		
1/2	0.68	0.71
2	…	…

[a]… Indicates that the time constant could not be measured because the model would spike continuously.

the effects on the somatic potential of many of the initial input spikes have decayed, resulting in a less precise first spike. For parameter variations that jointly increase or decrease synaptic effectiveness and threshold, the changes in precision of first-spike latency represents a balance between the two mechanisms.

The changes in the sustained firing rate of the octopus cell model due to parameter variations could also be explained in terms of changes in synaptic effectiveness and threshold. In comparison to the reference model, whenever both the threshold decreased and the synaptic effectiveness increased, or one remained the same, the response became more O_L (higher sustained rate). The following parameter variations produced this effect: halving dendrite diameter, specific membrane capacitance, or potassium density; doubling specific membrane resistance, sodium density, or both sodium and potassium densities. Whenever both the threshold increased and synaptic effectiveness decreased, or one remained the same, the response became more O_I (lower sustained rate). Whenever the threshold and synaptic effec-

tiveness both either increased or decreased, not even the relative magnitudes of each change could be used to predict the response, as evident by variations in soma diameter and axial resistivity.

Equations (3), (4), and (5) effectively summarize the results detailed above. These equations accurately predict synaptic effectiveness and threshold for all the parameter variations. Since an O_I response occurs when synaptic inputs are inefficient, Eqs. (3) and (4) predict that O_I responses occur when $1/R_m > G_{\text{exc}}$. Since an O_I response also occurs when the threshold is large, Eq. (5) predicts that O_I responses occur when $G_K > G_{Na}$ and $R_m < 1$. The responses agree with these predictions. Combining these facts, if G_{exc}, G_{Na}, and R_m are small, then the response will have a strong onset peak. From this perspective, the small time constant does not produce the onset response, but simply is a property of cells with onset responses.

There are several limitations of this analysis that warrant discussion. First, the current injections involve step changes while input spikes in octopus cells are produced by synaptic potentials with rise times of approximately 0.5 ms. When depolarization is produced by input spikes, the threshold has time to adapt, which means that more depolarization must be produced by synaptic inputs than that predicted by the threshold values determined in this study. To better understand this limitation, the reference model was depolarized 20 mV (to -45 mV) and the number of coincident input spikes that were needed to produce an output spike was found and compared to the theoretical calculations. It was predicted that three simultaneous spikes would be needed by subtracting -45 mV from the threshold at 20 mV depolarized (using Table III), and dividing by the synaptic effectiveness at 20 mV depolarized (using Table II). The simulation results agreed with the calculated result.

Second, the measurement of the steady-state threshold involved determining the sustained voltage after the second current step, defined as the steady-state threshold voltage. This parameter is probably different than the true threshold voltage, which is equal to the voltage produced at the onset of the second pulse. This discrepancy occurs since the active channels are in different states at these two times. However, the minimum current to produce a spike is used, the change in active channel state should be small. Thus the error in the threshold should be small. This error is unavoidable since the potential after the onset of the second step is contaminated by the spike and, thus, is not measurable.

Third, the measurement of the threshold time constant is limited because the somatic potential during the interval with no injected current (see Fig. 8) does not instantaneously drop to rest as implicitly assumed in the equations, but rather moves toward rest at a rate dependent on the time constant of the model. Less importantly, the assumption that the threshold current is linearly related to the threshold voltage is only partially valid due to the presence of active sodium and potassium channels in the membrane. Overall, these limitations mean that this calculation is more an estimate of the order of magnitude for the time constant than a precise measurement, and should be considered accordingly. Specifically, it was hypothesized that the threshold time constant should be most dependent on sodium inactivation, since this process is the slowest. The detailed plot of the threshold time constant versus sustained somatic depolarization for the reference model contains evidence that the threshold time constant is similar to sodium inactivation (Fig. 9).

Direct experimental verification of the predictions of this modeling study is desirable. Unfortunately, current physiological preparations cannot activate one synaptic input at a time as desired to measure synaptic effectiveness. In addition, it is not possible to place an electrode in the soma and axon of the octopus cell as needed to measure threshold. Therefore, verification will be based on the accuracy of the predictions of responses to novel stimuli produced by this modeling study.

This computational analysis of the octopus cell predicts that the cell should respond strongly to changes in spike activity of its auditory-nerve inputs rather than to relatively constant steady-state rates. In particular, based on the general properties that auditory-nerve fiber spike rates are related to sound energy near their CFs for steady-state stimuli, and that octopus cells are broadly tuned with relatively high CFs (Kane, 1973; Liberman, 1993), the spike activity of the population of octopus cells should most strongly reflect fluctuations in the high-frequency energy of complex sounds. The cells should be largely insensitive to broadband noise bursts except at the onset of each burst. Furthermore, the small value of the threshold time constant of the model predicts that as the rate of change in the signal energy decreases, the synchronized octopus cell response will decrease. These effects should be seen in responses to tonebursts with varying rise times, modulation transfer functions for amplitude-modulated stimuli, and synchronization to the fundamental frequency of steady-state vowels. Specifically, tonebursts with slow rise times should result in a decrease in the precision of the first spike, modulation transfer functions should be largely high pass, and octopus cells should synchronize to the fundamental frequency of speech in noise. These predictions motivate the use of complex stimuli to fully characterize octopus cell responses.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. BES-9409939 and the Deafness Research Foundation. The authors also thank three anonymous reviewers for providing insightful comments.

APPENDIX

The relevant model parameters are provided in Table AI. The model is fully described in a previous paper (Levy and Kipke, 1997).

TABLE AI. Summary of mode parameters. (NOTE: For impedances and admittances, variables with an initial capital letter have units relative to an area or length, whereas variables with an initial small letter have absolute units.)

Abbreviation	Value	Description
C_m	1 μF/cm^2	Specific membrane capacitance
d_d	17.6 μm	Diameter of the equivalent dendrite represent 4 dendrites with diameter of 7 μm each
d_s	25 μm	Diameter of the cylindrical soma
E_{exc}	45 mV	Reversal potential for excitatory synaptic input
E_L	-65 mV	Reversal potential for leakage conductance
$G_{\text{exc\_max}}$	1.2 mS/cm^2	Peak excitatory synaptic density for glutamate
$G_{\text{K\_bar}}$	70 mS/cm^2	Maximum active potassium density
G_L	2 mS/cm^2	Leakage density ($1/R_m$)
$G_{\text{Na\_bar}}$	35 mS/cm^2	Maximum active sodium density
l_d	237 μm	Length of the equivalent dendrite representing 4 dendrites, each 150 μm long
l_s	25 μm	Length of the cylindrical soma
R_a	150 Ω cm	Axial resistivity
R_m	0.5 kΩ cm^2	Specific membrane resistance
τ_{exc}	0.5 ms	Time constant of excitatory input
τ_m	0.5 ms	Time constant of membrane ($\tau_m = R_m C_m$)

[1]Depolarization values are expressed relative to the resting potential of -65 mV.

Bower, J. M., and Beeman, D. (**1995**). *The Book of GENESIS* (Kluwer Academic, New York).

Feng, J. J., Kuwada, S., Ostapoff, E. M., Batra, R., and Morest, D. K. (**1994**). ''A physiological and structural study of neuron types in the cochlear nucleus. I. Intracellular responses to acoustic stimulation and current injection,'' J. Comp. Neurol. **346**, 1–18.

Godfrey, D. A., Kiang, N. Y. S., and Norris, B. E. (**1975**). ''Single unit activity in the posteroventral cochlear nucleus of the cat,'' J. Comp. Neurol. **162**, 247–268.

Golding, N. L., Robertson, D., and Oertel, D. (**1995**). ''Recordings from slices indicate that octopus cells of the cochlear nucleus detect coincident firing of auditory nerve fibers with temporal precision,'' J. Neurosci. **15**, 3138–3153.

Kane, E. C. (**1973**). ''Octopus cells in the cochlear nucleus of the cat: Heterotypic synapses upon homeotypic neurons,'' Int. J. Neurosc. **5**, 251–279.

Kim, D. O., Rhode, W. S., and Greenberg, S. R. (**1986**). ''Responses of cochlear nucleus neurones to speech signals: neural encoding of pitch, intensity and other parameters,'' in *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. D. Patterson (Plenum, New York), pp. 281–288.

Kipke, D. R., and Levy, K. L. (**1997**). ''Sensitivity of the cochlear nucleus octopus cell to synaptic and membrane properties: A modeling study,'' J. Acoust. Soc. Am. **102**, 403–412.

Levy, K. L., and Kipke, D. R. (**1997**). ''A computational model of the cochlear nucleus octopus cell,'' J. Acoust. Soc. Am. **102**, 391–412.

Liberman, M. C. (**1993**). ''Central projections of auditory nerve fibers of differing spontaneous rate, II: Posteroventral and dorsal cochlear nuclei,'' J. Comp. Neurol. **327**, 17–36.

Osen, K. K. (**1969**). ''Cytoarchitecture of the cochlear nuclei in the cat,'' J. Comp. Neurol. **136**, 453–484.

Palombi, P. S., and Caspary, D. M. (**1992**). ''$GABA_A$ receptor antagonist bicuculline alters response properties of posteroventral cochlear nucleus neurons,'' J. Neurophysiol. **67**, 738–746.

Rall, W. (**1989**). ''Cable theory for dendritic neurons,'' in *Methods in Neuronal Modeling*, edited by C. Koch and I. Segev (MIT, Cambridge, MA), pp. 9–62.

Rhode, W. S., Oertel, D., and Smith, P. H. (**1983**). ''Physiological response properties of cells labeled intracellularly with horseradish peroxidase in cat ventral cochlear nucleus,'' J. Comp. Neurol. **213**, 448–463.

Rhode, W. S., and Smith, P. H. (**1986**). ''Encoding time and intensity in the ventral cochlear nucleus of the cat,'' J. Neurophysiol. **56**, 262–286.

Ritz, L. A., and Brownell, W. E. (**1982**). ''Single unit analysis of the posteroventral cochlear nucleus of the decerebrate cat,'' Neuroscience **7**, 1995–2010.

Romand, R. (**1978**). ''Survey of intracellular recording in the cochlear nucleus of the cat,'' Brain Res. **148**, 43–65.

Rouiller, E. M., and Ryugo, D. K. (**1984**). ''Intracellular marking of physiologically characterized cells in the ventral cochlear nucleus of the cat,'' J. Comp. Neurol. **225**, 167–186.

Saint Marie, R. L., Morest, D. K., and Brandon, C. J. (**1989**). ''The form and distribution of GABAergic synapses on the principal cell types of the ventral cochlear nucleus of the cat,'' Hearing Res. **42**, 97–112.

Smith, P. H., Joris, P. X., Banks, M. I., and Yin, T. C. T. (**1993**). ''Responses of cochlear nucleus cells and projections of their axons,'' in *The Mammalian Cochlear Nuclei: Organization and Function*, edited by M. A. Merchan, J. M. Juiz, D. A. Godfrey, and E. Mugnaini (Plenum, New York), pp. 349–360.

Winter, I. M., and Palmer, A. R. (**1995**). ''Level dependence of cochlear nucleus onset unit responses and facilitation by second tones or broadband noise,'' J. Neurophysiol. **73**, 141–159.

The role of the chinchilla pinna and ear canal in electrophysiological measures of hearing thresholds

W. J. Murphy and R. R. Davis
National Institute for Occupational Safety and Health, Division of Biomedical and Behavioral Science, Bioacoustics and Occupational Vibration Section, MS C-27, 4676 Columbia Parkway, Cincinnati, Ohio 45226

(Received 2 April 1997; revised 16 October 1997; accepted 15 December 1997)

Measurements of the acoustic transfer function (ATF) of the pinnae of 8 chinchillas were compared with the auditory-evoked potential (AEP) thresholds of 16 chinchillas measured in free field and with insert earphones. The ATF was measured in anesthetized chinchillas in a far-field condition in a semi-anechoic room using a logarithmic frequency sweep from 100 Hz to 20 kHz. Probe microphone measurements were collected with the probe opening at the tympanic membrane and in the same approximate position with the chinchilla removed from the sound field. For each animal's acoustic transfer function, the average of five in-the-ear and three free-field measurements were determined. The ATF exhibited a 5-dB passive gain at about 1 kHz and a broad resonance between 2.5 and 6 kHz of about a 10-dB gain. AEP thresholds were obtained from monaural chronically implanted chinchillas at 0.5, 1, 2, 4, and 8 kHz using first free-field and then insert earphone stimuli. The free-field sound pressure was measured with a microphone in the approximate position of the chinchilla's head. The earphone sound pressures were measured with a probe microphone positioned near the tympanic membrane. The free-field AEP hearing thresholds exhibited + 10-dB gain at 4 kHz compared to the insert earphone AEP thresholds. The agreement between the ATF and AEP derived transfer function suggested that the threshold differences at 4 kHz between the two testing configurations can be accounted for by the pinna and ear canal gain.
[S0001-4966(98)00404-4]

PACS numbers: 43.64.Bt, 43.64.Ha, 43.64.Jb [BLM]

INTRODUCTION

Rosowski's analysis (Rosowski, 1991) of previous measurements of middle and external ear function demonstrates the importance of the acoustic properties of the pinna and ear canal on the sound collecting performance of the auditory periphery. He showed that the frequency dependence of the hearing threshold of the human, cat, and chinchilla could be predicted based on the known acoustic properties of the external and middle ear. He also showed the important role of the acoustic properties of the external and middle ear in producing the functional changes in the cochlea due to an acoustic trauma.

For the chinchilla, Rosowski utilized data from von Bismark's thesis (von Bismark, 1967; von Bismark and Pfeiffer, 1967). In his thesis, von Bismark measured the acoustic properties of the chinchilla external ear by cementing a probe-tube microphone into the bony portion of the ear canal, close to the tympanic membrane. Tone sweeps were played in the free field and measurements were taken from the microphone. Although the data have been referenced many times, they have never been published.

Behavioral audiograms indicate that the frequencies between 1 kHz and 6 kHz are the region of maximum sensitivity for the chinchilla (Fay, 1988; Heffner and Heffner, 1991). In humans, Hellström (1996) showed that the frequency of maximum temporary threshold shift (TTS) was correlated with the peak frequency of the acoustic transfer function (ATF) from the free-field to the tympanic membrane. The ATF was calculated as the ratio of pressures measured in the ear canal and measured in free field. The ATF characterized the acoustic response of the pinna and ear canal. Individuals with wide and long ear canals had more TTS due to a low-frequency noise exposure than people with a short and narrow ear canal.

The present study was conducted to replicate von Bismark's earlier data and to compare auditory-evoked potential (AEP) thresholds measured in free field and with insert earphones. An earlier version of these data was reported by Davis (1993) and Murphy and Davis (1996). By using a flexible probe-tube microphone similar to that used in humans (Hellström, 1996), surgery to the ear canal can be avoided and the animal can be used more than once. The probe-tube microphone permits *in situ* calibration of stimuli for evoked potential and otoacoustic emissions testing (Siegel and Hirohata, 1994; Whitehead *et al.*, 1995). The procedure reported here for measuring the ATF of chinchillas was similar to Hellström's technique used with humans.

In this paper, data from complementary experiments are presented. In Sec. I, direct measurements of the ATF are presented. In Sec. II, the estimates of the ATF are inferred through measuring the AEP thresholds using sound field and insert earphone configurations. Sections III and IV present a discussion of the comparison of the two experiments and conclusions, respectively.

I. ACOUSTIC TRANSFER FUNCTION

A. Methods

Four female and four male adult chinchillas with otoscopically verified normal external ears were subjects.[1] A logarithmic frequency sweep from 100 Hz to 20 kHz was generated by a Stanford Research SR780 dual channel network system analyzer. The sweep signal was amplified by a Brüel & Kjær 2706 power amplifier which drove a single element Bose 25 speaker. The signals from an Etymotic Research ER-7C probe tube microphone and a Brüel & Kjær 4165 1/2 in. microphone were measured simultaneously. The ATF was computed by the SR780 and stored on diskette for off-line analysis.

Five in-the-ear measurements and three free-field measurements were collected in an Eckel Industries AN-ECK-OIC semi-anechoic chamber. The chinchilla was placed on a square of stiff wire mesh attached to a floor stand approximately 115 cm from the front of the speaker. The animal's body was oriented along the axis of the speaker and was located at the same elevation as the speaker cone. A horizontal reference bar suspended above the speaker and the animal provided a spatial reference for positioning the probe microphone body and animal. During free-field measurements, the microphone body was suspended from the reference bar and the probe tip was positioned approximately where the chinchilla's tympanic membrane would be located.

To prevent head movement, subjects were administered ketamine (22 mg/kg im), xylazine (1.1 mg/kg im), and atropine sulfate (0.054 mg/kg im). The position of the pinna did not exhibit obvious changes in position due to the anesthesia. The flexible silicone probe tube of the ER-7C microphone was positioned in the ear canal with the probe tube opening approximately 1 mm from the tympanic membrane. The insertion was otoscopically inspected. The body of the ER-7C microphone was laid back over the animal's head. The microphone body did not alter the position of the pinna. Between each measurement, the probe microphone was removed from the ear canal, replaced and reinspected with an otoscope. The Brüel & Kjær 4165 microphone was positioned 80 cm from the front of the speaker and slightly below the chinchilla. The Brüel & Kjær microphone served as a reference microphone which was not moved or adjusted during the course of the measurements.

B. Results

The data from the in-the-ear (ITE) and free-field (FF) measurements were averaged separately for each animal. The transfer function between in-the-ear to free-field measurements was calculated for each animal,

$$T_{\text{Animal,ITE-FF}}(f)=\frac{\overline{T_{\text{Animal,ITE}}(f)}}{\overline{T_{\text{Animal,FF}}(f)}}, \tag{1}$$

where $\overline{T_{\text{Animal,ITE}}(f)}$ was the average of the complex pressure ITE measurements and $\overline{T_{\text{Animal,FF}}(f)}$ was the average of the associated FF measurements. The complex transfer functions from the eight animals were then averaged to produce the overall acoustic transfer function between the ITE and FF measurements,

$$\overline{T_{\text{ITE-FF}}(f)}=\overline{T_{\text{Animal,ITE-FF}}(f)}. \tag{2}$$

Only one of the in-the-ear measurements for one of the animals proved to be considerably different from the rest of the data. The observer had noted in the data record book that the animal's head had rotated to the side giving cause to exclude that particular measurement. After the averages were computed, the dB magnitude was determined.

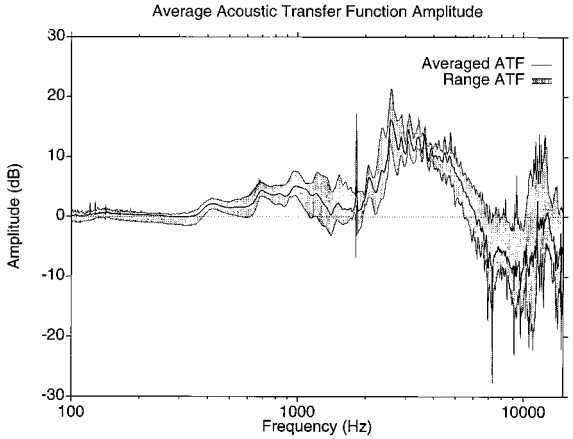


FIG. 1. The average acoustic transfer function and range (light gray) of the measurements collected from eight chinchillas. The transfer function exhibits a broad resonance between 2.5 and 6 kHz with about 10 dB of gain. From 700 Hz to about 1.3 kHz, a gain of about 4–5 dB is evident. From 7 to 10 kHz, an attenuation of about 5 dB is observed.

Figure 1 shows the average acoustic transfer function for the chinchillas (solid black line) and the range of the dB measurements (light gray shaded area). The transfer function exhibits approximately +3-dB gain from 700 Hz to 1.3 kHz, +10-dB gain from 2.5 to 6 kHz and −5-dB attenuation from 7 to 11 kHz. Table I lists the values of the acoustic transfer function and standard deviations at the standard audiometric frequencies. Above 15 kHz, the data exhibit considerable variability induced by the geometry of the pinna and ear canal. Tests in our laboratory of occluded and unoccluded ER-7 probe tubes show differences less than 10 dB for swept-sine stimuli above 14 kHz.

Von Bismark's results for type I and II transfer functions are compared with the averaged data from eight animals in

TABLE I. The average free field to eardrum acoustic transfer function and associated standard deviations at audiometric test frequencies measured from eight chinchillas.

Frequency (Hz)	Transfer function ±standard deviation (dB)
125	+0.4±0.5
250	+0.2±0.5
500	+1.5±0.7
1000	+5.1±1.3
2000	+3.5±1.3
3000	+10.6±1.4
4000	+9.9±1.1
6000	+1.3±2.4
8000	−5.9±3.4

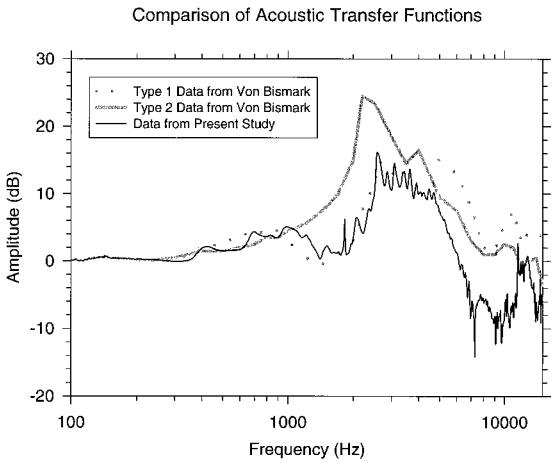


FIG. 2. Von Bismark's transfer functions contrasted with our average transfer function. Our data differs with the type I data primarily at the higher frequencies which could reflect the different placement of the microphone in the ear canal. The type II function is considerably wider and has a lower resonance peak. The pentobarbital anesthetic used by von Bismark was believed to compromise the middle ear ventilation in the type II data.

Fig. 2. Von Bismark attributed the type I data to animals which maintained normal middle ear ventilation. Differences between the present data and von Bismark's are evident at high frequencies and likely result from microphone placement. Whereas von Bismark cemented the microphone in the ear canal wall, placement here was in the inferior portion of the ear canal and the probe tip opening was directed towards the tympanic membrane.

The type II data in Fig. 2 exhibited more gain and a wider resonance. Von Bismark attributed the wider resonance in the type II data to the lack of ventilation of the middle ear due to the use of pentobarbital as the anesthetic agent which resulted in a static pressure difference across the tympanic membrane. A departure from normal quiescent middle ear pressure changes the compliance and middle ear transfer function. Since the ear canal is acoustically terminated with the middle ear, the pinna/ear canal transfer function will be affected by a poorly ventilated middle ear. Distortion product otoacoustic emission (DPOAE) data have shown that different anesthetic agents affect the ventilation of the middle ear of the mongolian gerbil (Zheng *et al.*, 1995). Ketamine did not depress the DPOAE response. However, injection of pentobarbital depressed DPOAE levels as did applying positive middle ear pressure. The ketamine/xylazine mixture used with other chinchillas in our lab has yielded similar ranges for AEP thresholds for awake and anesthetized animals as well as little difference in the DPOAE response. Normal middle ear ventilation was believed to have been maintained in the measurements taken here since they exhibit greater similarity with the type I data.

II. AEP-DERIVED TRANSFER FUNCTION

A. Methods

Sixteen adult chinchillas were monauralized via left cochlear destruction and implanted with chronic electrodes in the inferior colliculus and central sulcus region at SUNY-Buffalo Hearing Research Laboratories.[2] The animals were tested for auditory-evoked potentials in free-field conditions at SUNY-Buffalo and were then transported by automobile to NIOSH Taft Laboratories in Cincinnati. Following the quarantine period, the AEP hearing thresholds were measured using a Bio-Logic Navigator system equipped with insert earphones.[3]

The SUNY-Buffalo procedure has previously been described in Bancroft *et al.* (1991). The tone-burst stimuli (0.5, 1, 2, 4, 8, and 16 kHz) had a 5-ms rise/fall, 10-ms duration, and 100-ms interstimulus interval. Stimuli produced by a Loughborough TMS32020 16-bit digital to analog signal processing board were anti-alias filtered (20 kHz), attenuated and presented through Realistic 1218 speakers. The system was calibrated in SPL with a Larson-Davis 800B sound level meter and a Larson-Davis 2559 1/2-in. condenser microphone placed at the location of the chinchilla's head. Evoked responses from the chronic electrodes were amplified with a Grass P511 amplifier, digitized with a Loughborough TMS32020 16-bit A/D converter, averaged and analyzed. Animals were awake, alert, and restrained during testing. The restraint used by Buffalo had the animal facing the source with its nose pointing along the axis of the speaker, similar to the orientation for the ATF measurements.

The test system at NIOSH Taft Laboratories consisted of a Bio-Logic Navigator system, Etymotic Research ER-3 insert ear phones, ER-7C probe-tube microphone, and a Brüel & Kjær 2133 real-time frequency analyzer. The animals were awake, alert, and restrained during the test period (for a description of the restraint, see Snyder and Salvi, 1994). The restraint oriented the animal's head vertically rather than horizontally as Buffalo's restraint did. Once restrained, the probe-tube tip was positioned about 1 mm from the tympanic membrane. A neonatal foam plug was inserted while the probe tube was held against the pinna.[4] The ER-7C microphone was connected to the probe tube and the animal was placed in a small Industrial Acoustics Corporation sound-isolation booth. The active and reference electrodes were connected to the implant. A gold-cup electrode with electrode gel was connected to the left pinna which served as a ground.

The Bio-Logic system was programmed to collect a series of AEP responses at 5-dB intervals at 0.5, 1, 2, 4, and 8 kHz. The Bio-Logic system was not capable of producing a 16-kHz stimulus. Each response was the average of 512 presentations where the stimulus had a 2-ms rise/fall, 5-ms duration, and 43-ms interstimulus interval. The ear canal signal was calibrated in SPL before and after testing at the tone-burst frequencies using a constant 80 dB HL pure-tone signal which was detected by an ER-7C microphone and measured with the Brüel & Kjær 2133 real-time frequency analyzer. The average of the before and after calibration signal levels were used to correct the Bio-Logic HL measurements to SPL thresholds. One set of AEP thresholds was measured from each animal. AEP thresholds for both labs were determined at the level where an identifiable AEP response was no longer observed. If the AEP response was evident at 10 dB and not at 5 dB, the threshold was determined to be 7.5 dB.

TABLE II. The average free-field to insert earphone AEP derived transfer function, associated standard deviations and t values measured from sixteen chinchillas. Data were tested with a two-tailed Student's t test to identify which frequencies were significantly different from zero.

Frequency (Hz)	Transfer function±standard deviation (dB) Insert earphone–free field (dB)	Significance t values, H_0: $\mu=0$
500	−1.5±4.9	0.1616
1000	−4.5±4.7	0.0001*
2000	−1.1±5.3	0.3110
4000	9.6±6.8	0.0001*
8000	−2.0±9.3	0.3298

B. Results

These measurements were performed primarily to provide a comparison between AEP data collected in the different configurations and also to determine whether any threshold shift occurred during transport between the two labs. The thresholds from the SUNY-Buffalo animals were as good or better than animals which have been implanted and tested at NIOSH. After completion of the exposures conducted at NIOSH (Davis *et al.*, 1996), the animals were returned to SUNY-Buffalo for further testing. Five of the 16 animals reported in this paper were in the Control group for the experimental study. The Control animals exhibited no positive threshold shifts, therefore differences between AEP thresholds were not attributable to transportation. Any differences are likely the result of different testing methods.

After determining the AEP thresholds, the differences between the free-field and insert earphone thresholds were calculated across frequency for all animals. A two-tailed Student's t test was performed to identify those frequencies where the threshold difference was significantly different from 0-dB gain. The averaged differences for the various frequencies are listed in Table II along with their t values. The prominent feature is a gain of 9.6 dB at 4 kHz. The differences at 1 kHz and at 4 kHz were statistically significant.

The statistical significance of the differences between AEP derived transfer function and the acoustic transfer function was tested with a Student's t test (see Fig. 3). The AEP derived transfer function was significantly different from the ATF at 1 and 2 kHz ($|t|<0.05$). The two transfer functions were not significantly different at 0.5, 4, and 8 kHz. Thus the resonance at 4 kHz appears to result from the acoustical gain provided by the pinna and ear canal.

III. DISCUSSION

From Fig. 2, data from this study agreed with von Bismark's type I data below 5 kHz. The resonance in the ATF of the chinchilla pinna significantly affected the level of sound reaching the middle ear and cochlea. In the 2.5–6 kHz range, the pinna and ear canal passively amplified the stimuli by as much as 20 dB for individual animals (see Fig. 1). From 8 to 10 kHz, the stimuli were attenuated by as much as 21 dB for individual animals.

Patterson and Hamernik (1992) have demonstrated increased hearing loss in chinchillas due to impulse noise as the carrier frequency of the impulse approached 1.5 kHz. The amount of hearing loss remained flat from 1.5 to 3.5 kHz, the highest carrier frequency used in their experiment. The small amount of gain seen in the acoustic transfer function between 700 and 1.3 kHz seems to agree with the increased hearing loss they found. The broad gain between 2.5 and 6 kHz would suggest that the region of greatest hearing loss would occur with carrier frequencies in that range. Their stimuli were presented at levels between 127 and 147 dB SPL and most certainly drove the cochlea beyond its mechanical limits (Spoendlin, 1976; Henderson *et al.*, 1991, 1994).

A similar study by Ahroon and Hamernik (1996) investigated the effect of changing the center frequency (0.5, 1, 2, 4, and 8 kHz) of a narrow-band impact presented at 115 dB SPL for either 6 h/day for 20 days or 24 h/day for 5 days. The loss of outer hair cells for these equal energy exposures was greatest for the 4-kHz exposure and next greatest for the 8-kHz exposure. When the energy of the impact at 1 kHz was varied from 109 to 127 dB (Hamernik and Ahroon, 1996), the loss of OHCs due to the 121-dB impact was comparable to the loss for 115-dB impacts at 4 kHz. These results suggest that the 10-dB effective gain for the pinna at 4 kHz enhanced the 115-dB impact. However, the OHC loss due to an 8-kHz 115-dB impact contradict what might be

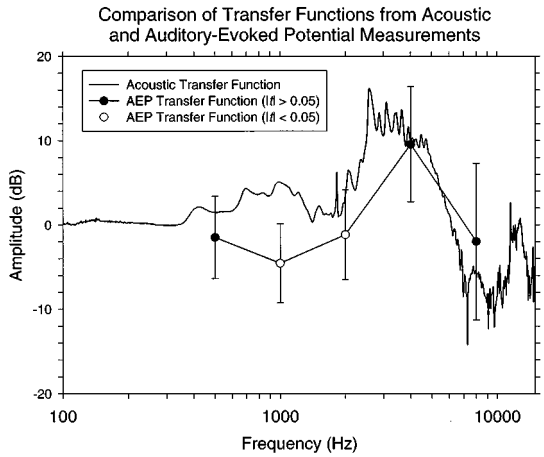


FIG. 3. A comparison of the acoustic transfer function and the auditory-evoked potential transfer function measured between free-field and in-the-ear canal conditions. The AEP and ATF functions agree at 500 Hz, 4 and 8 kHz, while differences at 1 and 2 kHz could be the result of terminating the ear canal with the ER-2 earplug. Error bars for the AEP derived transfer function represent the standard deviation of the measurements from 16 animals. Open symbols indicate that the AEP data are significantly different from the ATF data to a p value $p<0.05$.

expected based upon the acoustic transfer function. Certainly, factors such as the pinna orientation and the location of the sound source relative to the head affect the resonant character of the auditory periphery. Shaw and Teranishi (1968) showed that as the elevation of a source increased, the resonant peak frequency for a blocked meatus condition moved to higher frequencies when examining the human pinna data. Additional data need to be collected to fully interpret hearing loss susceptibility in chinchillas based on the position of the source.

Microstructure regularly spaced at 260 Hz can be identified in the acoustic transfer function (see Fig. 1). These microstructure were produced by the room resonances and were observed in both the free-field and in-the-ear measurements. The use of two microphones, Brüel & Kjær and ER-7C, reduced the microstructure. Without constructing a better anechoic chamber, the microstructure are unavoidable. The resonance at 1 kHz is likely the result of the pinna given that the pinna is the larger structure and would have a lower resonant frequency. However, the AEP transfer function exhibits a dip at 1 kHz.

Several possible explanations might account for the difference between ATF and AEP transfer functions. One explanation might be the orientation of the animal in the testing environment during free-field stimulation. The orientation of the pinna during awake AEP testing could be slightly different than the anesthetized position for ATF measurements. The effect of pinna orientation has been observed in AEP tests of chinchillas in free-field conditions (Davis *et al.*, 1996). In the raw data of that experiment (Davis *et al.*, 1996), an improved threshold (5–10 dB at 8 kHz) in *nonexposed control* animals resulted from experimental procedures. The difference was the effect of propping or not propping the pinna toward the speaker during testing. Because the pinna effects were observed at 8 kHz and not 1 kHz, the orientation of the animal probably is *not* responsible for the differences observed between ATF and AEP derived transfer functions.

A second possibility might be the different stimulus envelope parameters used to determine AEP thresholds in free field and insert earphone configurations. Additional testing of chinchillas using the envelope parameters yielded no significant effect on the AEP thresholds.

A third possible explanation would be a difference between the calibration of the test delivery systems between SUNY-Buffalo and NIOSH. The sound level calibrators at NIOSH have NIST traceable calibrations and are used to calibrate the Brüel & Kjær microphone and ER-7C microphone in the free-field condition. Daily calibrations of the ER-7C microphone during AEP testing are performed using the 94-dB signal provided by the preamplifier on the ER-7C. The microphones at Buffalo are calibrated with a NIST traceable calibrator and the free-field delivery system is checked regularly to ensure it is properly calibrated.

IV. CONCLUSIONS

The 10-dB gain at 4 kHz in the AEP data can be attributed to the effective gain of the pinna and ear canal. The attenuation at 8 kHz exhibits considerably more variation in the AEP data than in the ATF data. The placement of the probe tip for the AEP measurements was somewhat more difficult since the animals were not anesthetized and an insert earphone prohibited inspection of the final placement. Thus the data permit comparison of chinchilla hearing thresholds between free field and insert earphone configurations for frequencies below 6 kHz. The effects of the pinna orientation and the probe tube placement will complicate comparisons above 6 kHz.

The present study confirms von Bismark's earlier measurements and demonstrates the utility of using a flexible probe-tube microphone to measure the acoustic transfer function in small animals. As well, the passive amplification of the pinna produces significant effects that must be considered when when determining the transformation of free-field measurements of stimuli to eardrum sound pressure levels.

ACKNOWLEDGMENTS

The authors wish to thank Yun-Hua Shen (SUNY-Buffalo) for the surgical preparation and the free-field AEP threshold measurements of the chinchillas. The chinchillas were provided as part of a joint NIOSH/SUNY-Buffalo investigation into the combined effects of noise and organic solvents. We also thank Dr. Donald Henderson for his cooperation in allowing these chinchillas to undergo AEP threshold tests at NIOSH. We thank Dr. Derek Dunn, Dr. Greg Lotz, Dr. John Rosowski, and Dr. Arnold Tubis for their careful review of an earlier draft of this manuscript. Curt Sizemore and Brenda Swartzberg provided assistance with data collection and analysis.

[1]Acoustic transfer function measurements were approved by the NIOSH Division of Biomedical and Behavioral Science (DBBS), Animal Care and Use Committee (ACUC), protocol C69DAV.

[2]Surgical procedures were approved by the SUNY-Buffalo ACUC, protocol COM05080N.

[3]SUNY-Buffalo animals were transported to NIOSH Taft Laboratories and measurements performed on them were approved by the NIOSH DBBS ACUC, protocol C73DAV.

[4]Although the chinchillas had been monauralized, a foam insert earphone was inserted in the contralateral ear canal. The standard chronic electrode implant preparation at NIOSH Taft Laboratories does not include monauralization and therefore AEP tests are performed with insert earphones to isolate the response. The neonatal (tan) foam plugs provide an excellent fit for chinchillas and yield >60 dB of isolation from acoustic cross talk.

Ahroon, W. A., and Hamernik, R. P. (**1996**). ''The effects of continuous and interrupted narrow band impact noise exposures on hearing: Frequency considerations,'' J. Acoust. Soc. Am. **100**, 2627.

Bancroft, B. R., Boettcher, F. A., Salvi, R. J., and Wu, J. (**1991**). ''Effects of noise and salicylate on auditory evoked-response thresholds in the chinchilla,'' Hearing Res. **54**, 20–28.

Davis, R. R. (**1993**). ''Acoustic properties of the pinna and ear canal: Chinchilla and Rabbit,'' Abstracts of the 16th Midwinter meeting of the Association for Research in Otolaryngology, St. Petersburg, FL, 7–11 February, p. 12.

Davis, R. R., Murphy, W. J., Zhang, X. Y., Henderson, D., Morata, T. C., and Khan, A. (**1996**). ''The effects of toluene and noise on the chinchilla auditory system,'' Abstracts of the 19th Midwinter Meeting of the Association for Research in Otolaryngology, St. Petersburg, FL, 4–8 February, p. 31.

Fay, R. R. (**1988**). *Hearing in Vertebrates: A Psychophysics Databook* (Hill-Fay, Winnetka).

Hamernik, R. P., and Ahroon, W. A. (**1996**). ''The effects of continuous and interrupted 1 kHz narrow band impact noise exposures on hearing: Energy considerations,'' J. Acoust. Soc. Am. **100**, 2627.
Heffner, R. S., and Heffner, H. E. (**1991**). ''Behavioral hearing range of the chinchilla,'' Hearing Res. **52**, 13–16.
Hellström, P-A. (**1996**). ''Individual differences in peripheral sound transfer function: Relation to NIHL,'' in *Scientific Basis of Noise-Induced Hearing Loss*, edited by A. Axelson, H. M. Borchgrevink, R. P. Hamernik, P-A. Hellström, D. Henderson, and R. J. Salvi (Thieme Medical, New York), pp. 110–116.
Henderson, D., Spongr, V., Subramaniam, M., and Campo, P. (**1994**). ''Anatomical effects of impact noise,'' Hearing Res. **76**, 101–117.
Henderson, D., Subramaniam, M., Gratton, M. A., and Saunders, S. S. (**1991**). ''Impact noise: The importance of level, duration and repetition rate,'' J. Acoust. Soc. Am. **89**, 1350–1357.
Murphy, W. J., and Davis, R. R. (**1996**). ''Acoustic properties of the chinchilla pinna and ear canal,'' J. Acoust. Soc. Am. **99**, 2563.
Patterson, J. H., and Hamernik, R. P. (**1992**). ''An experimental basis for the estimation of auditory system hazard following exposure to impulse noise,'' in *Noise-Induced Hearing Loss*, edited by A. L. Dancer, D. Henderson, R. J. Salvi, and R. P. Hamernik (Mosby YearBook, Boston), pp. 336–348.
Rosowski, J. J. (**1991**). ''The effects of external- and middle-ear filtering on auditory threshold and noise-induced hearing loss,'' J. Acoust. Soc. Am. **90**, 124–135.
Siegel, J. H., and Hirohata, E. T. (**1994**). ''Sound calibration and distortion product otoacoustic emissions at high frequencies,'' Hearing Res. **80**, 146–152.
Shaw, E. A. G., and Teranishi, R. (**1968**). ''Sound pressure generated in an external ear replica and real human ears by a nearby point source,'' J. Acoust. Soc. Am. **44**, 240–249.
Snyder, D. L., and Salvi, R. J. (**1994**). ''A novel chinchilla restraint device,'' Lab Animal **23**, 42–44.
Spoendlin, H. (**1976**). ''Anatomical changes following various noise exposures.'' in *Effects of Noise on Hearing*, edited by D. Henderson, R. P. Hamernik, D. S. Dosanjh, and J. H. Mills (Raven, New York), pp. 69–89.
von Bismark, G. (**1967**). ''The sound pressure transformation function from the free-field to the eardrum of chinchilla,'' Masters thesis, Department of Electrical Engineering, Massachusetts Institute of Technology.
von Bismark, G., and Pfeiffer, R. R. (**1967**). ''On the sound pressure transform from free field to eardrum of chinchilla,'' J. Acoust. Soc. Am. Suppl. 1 **42**, S146.
Whitehead, M. L., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (**1995**). ''Effects of ear-canal standing waves on measurements of distortion-product otoacoustic emissions,'' J. Acoust. Soc. Am. **98**, 3200–3214.
Zheng, Y., Ohyama, K., Hozawa, K., Wada, H., and Takasaka, T. (**1995**). ''Effect of anesthetic agents on distortion product otoacoustic emissions in the mongolian gerbil,'' Abstracts of the 18th Midwinter Meeting of the Association for Research in Otolaryngology, St. Petersburg, FL, 5–9 February, p. 185.

Locus of generation for the $2f_1-f_2$ vs $2f_2-f_1$ distortion-product otoacoustic emissions in normal-hearing humans revealed by suppression tuning, onset latencies, and amplitude correlations

Glen K. Martin, David Jassir, Barden B. Stagner, Martin L. Whitehead, and Brenda L. Lonsbury-Martin
Department of Otolaryngology (M805), University of Miami School of Medicine, P.O. Box 016960, Miami, Florida

(Received 21 May 1997; revised 24 November 1997; accepted 17 December 1997)

The present study used distortion-product otoacoustic emission (DPOAE) suppression tuning curves (STCs), DPOAE onset latencies (OLs), and DPOAE amplitude correlations to investigate the locus of generation of the $2f_1-f_2$ DPOAE versus the $2f_2-f_1$ DPOAE in humans. The results of the tuning study revealed that, for the $2f_1-f_2$ DPOAE, the tips of the STCs tuned consistently below the geometric-mean (GM) frequency of the primary tones. In contrast, for the $2f_2-f_1$ DPOAE, STCs tuned above the GM of the primaries, with 50% of the tip frequencies at, or above, the $2f_2-f_1$ frequency place. When the average ratio of the $2f_2-f_1$ to the $2f_1-f_2$ tip frequencies was computed, a factor of 1.44 provided an estimate of the frequency shift needed to align the two DPOAE generation sites. Other results showed that OLs for the $2f_2-f_1$ DPOAE were uniformly shorter than those for the $2f_1-f_2$, with differences at the low frequencies amounting to as much as 6–7 ms. Further, for both DPOAEs, curves describing latency decreases as a function of increasing GM frequencies were best fit by power functions. Shifting the GM frequency producing the $2f_2-f_1$ DPOAE by a factor of 1.6 caused the latency distributions for both DPOAEs to overlap thus resulting in a single function that described cochlear delay as a function of GM frequency. Finally, for each GM frequency in the DP-gram, sliding correlations from 108 normal ears were performed on both DPOAEs by holding the primaries producing the $2f_1-f_2$ DPOAE constant, while all $2f_2-f_1$ DPOAE amplitudes were successively correlated with the $2f_1-f_2$ amplitudes. This procedure demonstrated that, for a given GM frequency producing the $2f_1-f_2$, the correlations between the two DPOAEs peaked when the primaries of the $2f_2-f_1$ were at a GM frequency that positioned the $2f_2-f_1$ frequency place near the GM of the primaries that produced the $2f_1-f_2$ DPOAE. As a whole, the above findings strongly suggest that the $2f_2-f_1$ DPOAE in humans is generated basal to the primary-tone place on the basilar membrane. © *1998 Acoustical Society of America.* [S0001-4966(98)00504-9]

PACS numbers: 43.64.Jb, 43.64.Bt, 43.64.Kc, 43.64.Ri [RDF]

INTRODUCTION

A basic understanding of distortion-product otoacoustic emissions (DPOAEs) requires the determination of how various odd- and even-order emission components (e.g., f_2-f_1, $3f_1-2f_2$, $2f_1-f_2$, $2f_2-f_1$, etc.) are generated in the cochlea and arise in the ear canal signal. The early nerve-fiber recording studies of Kim and associates (Kim *et al.*, 1980; Siegel *et al.*, 1982) suggested that the $2f_1-f_2$ DPOAE was generated in the region of primary-tone interaction, and, subsequently, was propagated to the DPOAE-frequency place. The later suppression-tuning studies of Brown and Kemp (1984) showing that a suppressor tone produced maximum suppression around the primaries, or near f_2, confirmed the earlier observations in auditory nerve fibers. The results of Kemp and Brown (Kemp and Brown, 1983; Brown and Kemp, 1984) were subsequently corroborated more systematically in humans (Harris *et al.*, 1992; Lonsbury-Martin *et al.*, 1994; Abdala and Sininger, 1996; Abdala *et al.*, 1996; Gaskill and Brown, 1996; Kummer *et al.*, 1995), and extended to a number of laboratory species including the rabbit (Martin *et al.*, 1987), alligator lizard (Koppl and Manley, 1993), bat (Frank and Kossl, 1995), and opossum (Faulstich *et al.*, 1996). These outcomes are also supported by several findings from clinical investigations including the excellent correlation between the level of the $2f_1-f_2$ DPOAE when plotted at the geometric mean (GM) of the primary tones and hearing level at the corresponding audiometric frequency (e.g., Martin *et al.*, 1990), and the superior clinical performance of DPOAEs when the f_2 frequency is related to the audiometric frequency (e.g., Gorga *et al.*, 1993). Together, these results strongly implicate the primary-tone place as the generation site of the $2f_1-f_2$ DPOAE in humans.

Evidence is now accumulating that other odd-order DPOAEs, namely the $2f_2-f_1$ DPOAE in contrast to the $2f_1-f_2$, may also arise from distinct cochlear locations, under circumstances more complicated than originally expected. The early cochlear modeling studies of Hall (1974) showing that the $2f_2-f_1$ DPOAE was confined to the primary-tone place set the stage for accepting the view that

 0001-4966/98/103(4)/1957/15/$10.00

the origin of this particular DPOAE was at the site of the primary tones. Thus in the absence of any evidence to the contrary, it was most parsimonious to assume that, in biological systems, the two DPOAEs arise from the same nonlinearity, which was located at one site on the basilar membrane.

Since the discovery of otoacoustic emissions, it has become well-established that both lower and upper odd-order DPOAE side bands can be recorded in the ear canals of laboratory animals (Brown and Kemp, 1985; Kemp and Brown, 1986; Brown, 1987; Lonsbury-Martin *et al.*, 1987; Brown *et al.*, 1989) and humans (Moulin and Kemp, 1995). Although the appearance of the lower side-band DPOAEs in the ear canal might be anticipated, the prominence of the $2f_2-f_1$ DPOAE in the ear canal signal of both animals (Martin *et al.*, 1987) and humans (Moulin *et al.*, 1993; Moulin and Kemp, 1996; Wable *et al.*, 1996) can be viewed as problematic. That is, before it can be emitted in the ear canal, the $2f_2-f_1$ DPOAE must traverse its best-frequency place on the basilar membrane (Hall, 1980). Thus it would be expected that the $2f_2-f_1$ DPOAE would be largely absorbed as it traversed its characteristic place, and, consequently, that it would be either immeasurable, or extremely small in ear canal recordings. However, in the ears of laboratory mammals and in humans at primary-tone levels around 75 dB SPL, the $2f_1-f_2$ and the $2f_2-f_1$ DPOAEs can be nearly comparable in amplitude. This apparent difficulty can, however, be surmounted, if it can be shown that the $2f_2-f_1$ DPOAE is generated at, or basal to, its characteristic place on the basilar membrane.

Other observations of Brown and Kemp (1985) in gerbils showed that the $2f_2-f_1$ DPOAE latency was shorter than that for the $2f_1-f_2$ DPOAE. They suggested that this latency difference could be accounted for, if the $2f_2-f_1$ DPOAE was generated in a region basal to the primary-tone interaction site. A later study in rabbits by Martin *et al.* (1987) used suppression tuning curves (STCs) and temporary threshold shifts (TTSs) to provide further evidence supporting the above notion. Specifically, by relating the suppression tip frequency to the DPOAE frequency, they discovered that the $2f_1-f_2$ DPOAE appeared to originate from the region of the primary-tone interaction, whereas the $2f_2-f_1$ DPOAE seemed to originate from a region basal to the primary-tone place. In a more recent study, Moulin *et al.* (1993) found some support for this idea in humans at low frequencies in that the $2f_2-f_1$ DPOAE best correlated with the transient-evoked otoacoustic emission (TEOAE) bands at, or nearest to, the $2f_2-f_1$ frequency. However, at high frequencies, the $2f_2-f_1$ emission correlated best with TEOAE-frequency bands that were considerably lower than the frequency place of the $2f_2-f_1$ DPOAE.

Most recently, Whitehead *et al.* (1996) demonstrated shorter onset latencies (OLs) for the $2f_2-f_1$ than for the $2f_1-f_2$ DPOAE in both humans and rabbits. In humans, Moulin and Kemp (1996) also reported consistently shorter group-delay latencies measured with the phase-gradient technique for the $2f_2-f_1$ than for the $2f_1-f_2$ DPOAE. Both DPOAE-latency studies found consistent differences in the latencies of the lower side-band DPOAEs, depending on whether an f_1 or an f_2 sweep was used, but no differences for the upper band $2f_2-f_1$ DPOAE, which also implies different origins for the two types of emissions (Moulin and Kemp, 1995). Wable *et al.* (1996), too, reported detailed phase-gradient latencies for both the $2f_1-f_2$ and $2f_2-f_1$ DPOAEs. They found that when the eliciting primaries were identical, the latencies of the $2f_2-f_1$ DPOAEs were consistently shorter than those for the $2f_1-f_2$. However, when the primary-tone pairs were selected in order to make the $2f_2-f_1$ frequency equal to the f_2 frequency producing the $2f_1-f_2$, the latencies of the two DPOAEs were not the same as would be expected, if this procedure aligned the two generation sites. In fact, the $2f_2-f_1$ latencies were still consistently shorter than those for the $2f_1-f_2$ DPOAE under these conditions. The $2f_2-f_1$ latencies only approached the $2f_1-f_2$ ones when the frequency for the $2f_2-f_1$ DPOAE was set equal to the frequency for the $2f_1-f_2$ DPOAE. Overall, the above DPOAE-latency findings in humans are consistent with the notion that the $2f_2-f_1$ originates basal to the $2f_1-f_2$ site of generation, thus resulting in consistently shorter latencies for this emission.

The research presented here utilized suppression, onset latency, and correlation techniques to more thoroughly investigate the sites of origin of the $2f_1-f_2$ versus the $2f_2-f_1$ DPOAEs in humans with normal hearing. All these methods produced results which were consistent with the most recent latency studies in humans (Moulin and Kemp, 1996; Wable *et al.*, 1996; Whitehead *et al.*, 1996) in that the $2f_2-f_1$ DPOAE appears to originate at, or basal to, the $2f_2-f_1$ frequency place. Portions of these results have been reported elsewhere (Lonsbury-Martin *et al.*, 1994; Lecusay *et al.*, 1996).

I. METHODS

A. Subjects

One ear was tested from each subject, and all subjects had normal hearing sensitivity defined as pure-tone thresholds $\leqslant$20 dB hearing level (HL) for audiometric test frequencies at octave intervals from 0.25 to 8 kHz, and at 3 and 6 kHz. Furthermore, all subjects demonstrated normal tympanograms (i.e., type A) and acoustic-reflex thresholds (i.e., $<$100 dB HL) obtained by routine immittance testing. All subjects also had a negative history of hearing disorders, and exposure to toxic agents as assessed by direct questioning. Seventeen subjects (12 males, 5 females), with a mean age 25.5 ± 8.6 (1 standard deviation) years were used in the STC procedures. From another group of 13 subjects, OLs were obtained from 10 males and 3 females, with a mean age of 23.4 ± 5 years. Correlations between $2f_1-f_2$ and $2f_2-f_1$ DPOAE amplitudes were based on DPOAE versus frequency functions (i.e., DP-grams) from a database (Whitehead *et al.*, 1993) of 149 ears (70 males, 79 females) contributed by 94 normal-hearing humans, with a mean age of 29.2 ± 8.6 years. For the present study, from the total of 149 ears, 108 ears were selected which met the criterion that both the $2f_1-f_2$ and the $2f_2-f_1$ DPOAE amplitudes were 6 dB above the noise floor. This criterion was selected to be cer-

tain that all correlations were based on robust data pairs that were relatively uncontaminated by spurious noise.

B. DPOAE measurement procedures

The DP-grams were collected at GM frequencies [GM $=(f_1\times f_2)^{0.5}$], in 0.1-octave steps, from 0.25 to 8.0 kHz using procedures previously described in detail elsewhere (Whitehead *et al.*, 1995). Briefly, the f_1 and f_2 primary tones were generated by a 16-bit digital signal processing (DSP) board (Digidesign, Audiomedia) in a Macintosh IIci computer, and presented over ER-2 (Etymotic Research) speakers. Ear canal sound pressure was measured by either an ER-10 or ER-10B (Etymotic Research) microphone, and sampled and synchronously averaged ($n=32$) by the DSP board. From a 4096-point fast Fourier transform (FFT) of the time sample, the amplitudes of the $2f_1-f_2$ and $2f_2-f_1$ DPOAEs and their related noise floors were extracted. Noise floors were computed as the average of four FFT bins on either side of the DPOAE frequency. Because the $2f_2-f_1$ DPOAE is at a higher frequency than the $2f_1-f_2$ and low-frequency noise dominates the noise-floor measurements, the $2f_2-f_1$ typically had a slightly lower noise floor (see Fig. 7) than the $2f_1-f_2$ DPOAE. In order to reliably elicit $2f_1-f_2$ and $2f_2-f_1$ DPOAEs in the same ears, all primary-tone stimuli were presented at equilevels (i.e., $L_1=L_2$) of 75 dB SPL, with an f_2/f_1 ratio of 1.21.

C. Suppression-response areas

Suppression-response areas (SRAs) were obtained at the three GM frequencies of 3 ($f_1=2.57$, $f_2=3.11$ kHz), 4 ($f_1=3.636$, $f_2=4.4$ kHz), and 6 kHz ($f_1=5.51$, $f_2=6.67$ kHz) for each ear. The primary tones were held constant at 75 dB SPL while a suppressor tone was swept in the frequency domain in 0.1-octave steps, from 0.25 to 10.0 kHz, and in the level extent in increments of 1 dB, from 55 to 85 dB SPL. The suppressor tone was digitally added to the f_1 channel. Data were collected in a matrix in which the amount of suppression was calculated relative to DPOAE amplitude in the absence of the suppressor tone. This reference DPOAE level was based on the mean of all control DPOAE amplitudes measured at the end of each frequency column in the suppression matrix. Data were plotted as contour plots (Microsoft Excel, v5.0) in which successive contours represented increasing iso-suppression, in 3-dB steps. A spreadsheet algorithm was developed to extract iso-suppression tuning curves from these contour plots, and to automatically calculate STC tip frequency, threshold, and $Q_{10\,\mathrm{dB}}$. Because the amplitude of the $2f_2-f_1$ DPOAE was often much lower than that of the $2f_1-f_2$, thus making STC extraction difficult, the data in the suppression matrix were smoothed once prior to the STC extraction. This smoothing procedure employed a common image-analysis technique in which each point in the matrix was replaced by the average of itself and its eight neighbors. On average, one smoothing pass increased the numbers of usable $2f_1-f_2$ STCs by approximately 16%, and the number of $2f_2-f_1$ STCs by 90%. To be certain that the smoothing procedure did not unduly distort the data, paired Student's t-tests were computed for a 6-dB suppression criterion, across all three STC frequencies, in order to compare curves with no smoothing to those with one degree of smoothing. This comparison showed that the smoothing procedure had no significant effect on tip frequency, but did significantly increase the tip threshold by 2.2 dB ($t=4.8$, $df=39$, $p<0.0001$) and decrease the $Q_{10\,\mathrm{dB}}$ by 0.36 ($t=5.8$, $df=24$, $p<0.0001$). Although comparatively significant, in actuality, smoothing had only minor effects on the STC parameters while greatly increasing the number of usable $2f_2-f_1$ STCs for comparison to the corresponding $2f_1-f_2$ STCs. Consequently, all STC data reported here are based upon one smoothing of the suppression data matrix. For the suppression criteria of 3 and 6 dB, the STC parameters were computed for both $2f_1-f_2$ and $2f_2-f_1$ DPOAEs.

D. Onset latencies

Initially, DP-grams were collected for both DPOAEs using GM frequencies varying from 0.25 to 8 kHz in 0.1-octave intervals. For each subject, a range of frequencies was identified which elicited both the $2f_1-f_2$ and $2f_2-f_1$ DPOAEs that were >3 dB above the related noise floor. For a particular ear, the number of frequencies tested, all of which satisfied this criterion, varied from 4 to 16, but, on average, each subject contributed approximately seven pairs of values to the database. Because different amounts of phase rotation are required to obtain the $2f_1-f_2$ versus the $2f_2-f_1$ OL, each DPOAE latency was measured successively, with the order being randomized across trials.

The procedure for obtaining OLs has recently been described in detail by Whitehead *et al.* (1996). Briefly, it is based on a phase-rotation method in which the phases of f_1 and f_2 are systematically advanced between stimulus presentations. The phase advances are arranged to give equal numbers of in-phase and out-of-phase presentations so that the primary tones and their harmonics, and all DPOAEs, except that of interest, are canceled in the final average. This process uncovers the raw DPOAE time waveform. An FFT is then computed on the waveform, and spectral editing is performed to remove any residual primaries and high-frequency noise by setting the appropriate bins in the FFT to zero. At the final processing step, the time waveform is recovered by an inverse FFT, and is then high-pass filtered to remove low-frequency components. The OL was scored objectively by using a sliding-amplitude correlation procedure in which the DPOAE waveform was correlated over 4-cycle periods, centered around each point in the sample, to a reference sinusoid at the DPOAE frequency. This sinusoid was synthesized so that it was in-phase with the DPOAE at steady state (see Fig. 4). The DPOAE onset was arbitrarily defined as the time for which the correlation coefficient equaled 0.5 or 0.75 for, at least, 2 ms. Across both DPOAEs, a total of 250 time waveforms were collected.

In an attempt to maintain objectivity, final latency values were arrived at in the following manner. First, all DPOAE time waveforms were visually inspected, and those (23%) that were obviously too noisy to be objectively scored by the computer-based correlation procedure were eliminated from the database. Next, scatterplots were computed for both

DPOAE latencies using the 0.5- and 0.75-correlation criteria. Examination of these distributions revealed, not unexpectedly, that the 0.5-criterion latencies for a given frequency tended to be slightly shorter compared to the 0.75-criterion ones, and that the 0.75 distribution was characterized by more obvious outliers. Nevertheless, both distributions completely overlapped, and clearly showed the same differences in latency between the $2f_1-f_2$ and the $2f_2-f_1$ DPOAEs. Further inspection of the time waveforms revealed that at the lowest frequencies the correlation procedure sometimes introduced significant errors, even though alignment with the synthesized waveform was very clear upon visual inspection (see Fig. 4 for an example at 0.871 kHz). Consequently, to obtain the most accurate estimates of OL, all time waveforms for both the $2f_1-f_2$ and $2f_2-f_1$ DPOAEs were visually inspected and judged as either being most accurately scored by the 0.5 (46%) or 0.75 criterion (20%), or they were assigned a latency by visual inspection (34%), if either of the two objective criteria were judged to be erroneous. This latter subjective scoring procedure resulted in overall mean latencies that fell between the two completely objective techniques. This latter outcome reinforced the impression that such a subjective approach resulted in the least number of errors for low-frequency waveforms. Finally, 0.8 ms was subtracted from all latencies to account for the observed acoustic delay in the sound delivery tubes. The final results summarized below are based on 101 $2f_1-f_2$ and 91 $2f_2-f_1$ latency measures consisting of 77 latency pairs in which both latencies were measured from the same ear.

E. DPOAE correlation procedures

The relationship between the $2f_1-f_2$ and $2f_2-f_1$ DPOAEs was further determined by performing Pearson product-moment correlations (r) between the $2f_1-f_2$ amplitude for each GM frequency in the DP-gram ($n=37$) and each $2f_2-f_1$ amplitude for each successive DP-gram frequency. This comparison yielded 37 average r values. For example, taking the DP-gram GM frequency of 2.828 kHz, for each ear ($n=108$), the level of the $2f_1-f_2$ DPOAE at the GM of 2.828 kHz was first paired with each $2f_2-f_1$ DPOAE amplitude for the GM of 0.498 (i.e., the initial frequency of the DP-gram), and the resulting r computed. Then, keeping the $2f_1-f_2$ GM frequency constant at 2.828 kHz, r values were successively computed between each $2f_1-f_2$ and the $2f_2-f_1$ amplitudes from related ears for each subsequent GM frequency in the DP-gram (i.e., from 0.568 to 8 kHz). The resulting r's for each $2f_1-f_2$ GM frequency were then 3-point smoothed, and the GM frequency of the r value giving the maximum correlation between the $2f_1-f_2$ and the $2f_2-f_1$ amplitudes was noted. These peak values were plotted as a function of the $2f_1-f_2$ vs $2f_2-f_1$ GM frequency to determine the relationship of the GM frequencies that yielded the maximum correlation between the two DPOAEs. The two lowest frequencies (0.498 and 0.568 kHz) were not significant, because of the small number of subjects displaying reliable DPOAEs at these frequencies. All of the remaining peak correlations were highly significant ($p<0.001$), with the exception of one frequency (1.313 kHz) with a lower significance level ($p<0.01$).

F. Statistical analysis

For comparisons between groups, a two-tailed Student's t-test (paired observations) was performed using a commercially available software package (Abacus, StatView v4). The adopted level of statistical significance was $p<0.05$.

II. RESULTS

A. Suppression tuning curves

Figure 1 displays individual examples of SRAs at 4 kHz in the form of contour plots (left) and STCs (right) extracted from the suppression matrices. From inspection of this plot, it is clear that the $2f_1-f_2$ DPOAE (top plot) tuned within the region of the primary tones (arrows), whereas the $2f_2-f_1$ DPOAE (lower plot) tuned nearer to the $2f_2-f_1$ frequency place (asterisk), or basal to this location. This general outcome was consistent, regardless of the various suppression criteria (3–12 dB) illustrated in the figure.

In Fig. 2, all of the STC tip frequencies are plotted for the three GM frequencies of 3, 4, and 6 kHz, for both the 3-(left) and 6-dB (right) suppression criteria. These data show that the tip frequencies for the $2f_1-f_2$ DPOAE consistently were aligned at, or slightly below, the GM of the f_1 and f_2 primary tones [i.e., around the 3- (top), 4- (middle), or 6-kHz (bottom) frequency region]. In contrast, the $2f_2-f_1$ STC tip frequencies were scattered, in general, above the GM frequencies. However, there was much less consistency in their exact frequency locations compared to those observed for the $2f_1-f_2$ DPOAE. More specifically, across both suppression criteria, 64% of the 3-kHz, 83% of the 4-kHz, and 50% of the 6-kHz tip frequencies were at, or above, the $2f_2-f_1$ DPOAE frequency indicated by the asterisk on the abscissa. These data also demonstrate that the overall findings were influenced little by whether or not a 3- or 6-dB suppression criterion was utilized.

For both DPOAEs, the average characteristics over all the STCs for 3- and 6-dB suppression criteria are summarized in Table I. Inspection of these data reveals that the $2f_2-f_1$ amplitude was consistently smaller than that for the $2f_1-f_2$ DPOAE, for all GM frequencies tested ($t=12.3$, $df=19$, $p<0.0001$). Further statistical analyses across all three frequencies also demonstrated that the $2f_2-f_1$ DPOAE had a significantly ($t=4.7$, $df=19$, $p<0.0002$) lower tip threshold by 4.7 dB, on average, for the 3-dB criterion, and by 3 dB, on average, for the 6-dB criterion ($t=3.5$, $df=19$, $p<0.003$). The $Q_{10\,\mathrm{dB}}$ also tended to be slightly larger for the $2f_2-f_1$, for both suppression criteria, but this difference did not reach statistical significance. Comparison of the effects of suppression criteria of 3 vs 6 dB indicated that the 6-dB $2f_2-f_1$ tip frequency was, on average, 0.340 kHz lower in frequency ($t=2.9$, $df=18$, $p<0.01$) than the 3-dB tip frequency. However, no differences on tip frequency values were found for the $2f_1-f_2$ DPOAE. For the tip thresholds of both the $2f_1-f_2$ and $2f_2-f_1$ DPOAEs, the 6-dB criterion, as would be expected, resulted, on average, in slightly higher thresholds by 3 ($t=15.5$, $df=45$, $p<0.0001$) and 4.4 dB ($t=6.8$, $df=16$, $p<0.0001$), respectively.

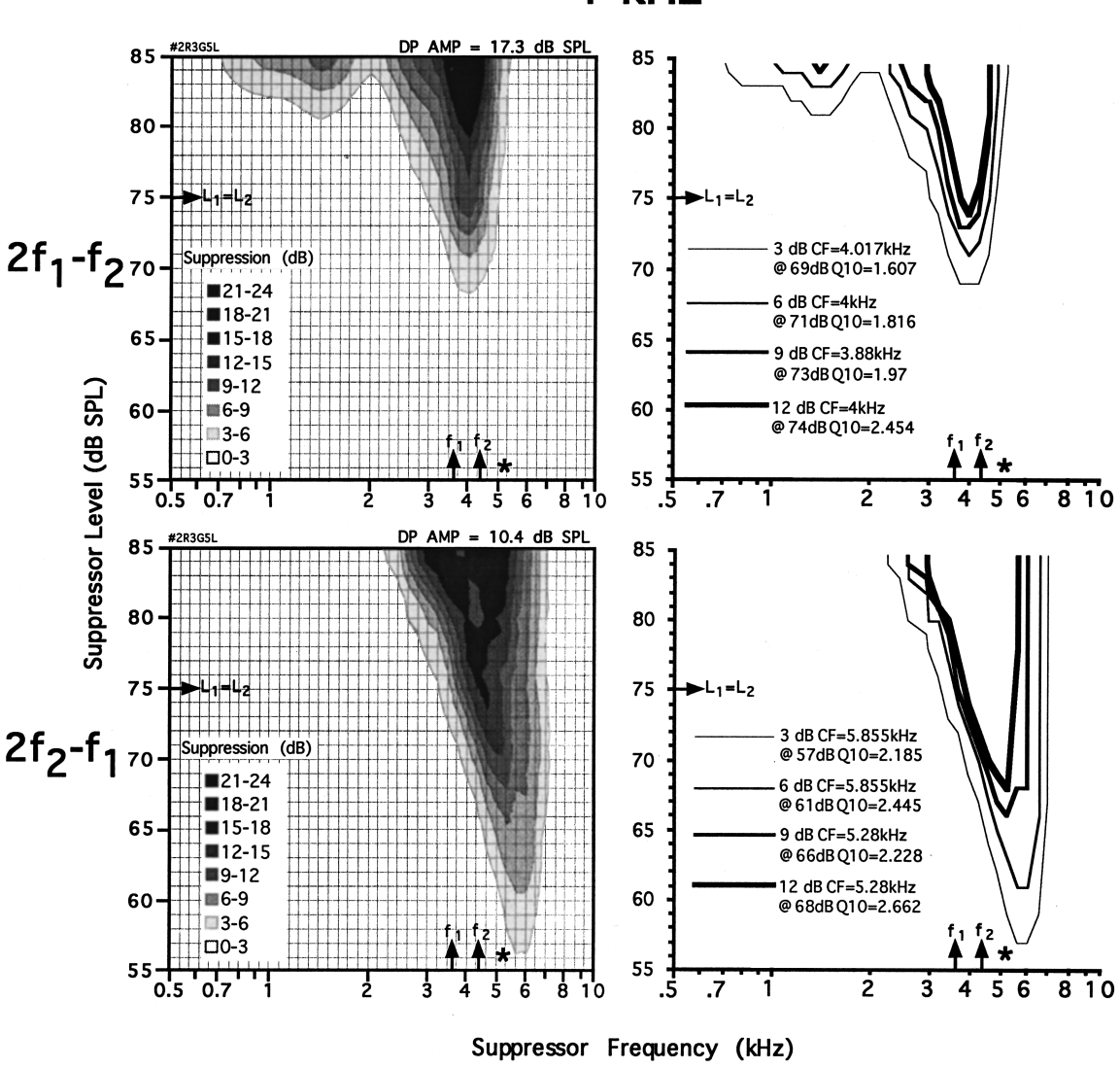


FIG. 1. Suppression-response areas (SRAs) on the left were obtained by sweeping a suppressor tone in 1-dB steps along the level and 0.1-octave intervals along the frequency axes for 75 dB SPL primary tones with a GM of 4 kHz for both the $2f_1-f_2$ (top) and $2f_2-f_1$ (bottom) DPOAEs. Bands of increasing grays to blacks represent 3-dB increments in the amount of suppression elicited by the suppressor tone. Suppression-tuning curves (STCs) on the right were extracted for suppression criteria ranging from 3 to 12 dB. For each suppression criterion, legends indicate tip frequency (CF), threshold in dB SPL at the CF, and $Q_{10\,\mathrm{dB}}$. Note that for the $2f_1-f_2$, STCs were centered at the primary tones (arrows along abscissa) with CFs around 4 kHz whereas, for the $2f_2-f_1$, STCs tuned nearer the $2f_2-f_1$ frequency (asterisk) with CFs of approximately 5.3 kHz.

If it can be assumed that the tip frequencies of the STCs indicate the region of maximal DPOAE generation, then the above results are consistent with the notion that the $2f_2-f_1$ is, in all cases, generated basal to the $2f_1-f_2$ DPOAE. Moreover, it can also be assumed that the ratio of the two DPOAE tip frequencies indicates the factor needed to align the two loci of DPOAE generation. From the CF tip frequency data of Table I, the average tip frequencies for the $2f_2-f_1$ STCs were divided by the average tip frequencies of the $2f_1-f_2$ STCs to calculate the factor needed to place the $2f_1-f_2$ STC tip at the $2f_2-f_1$ STC tip frequency. The resulting values are presented in Table II for all STC frequencies and both suppression criteria. The overall average ratio was 1.44, but these average values ranged from a low of 1.28, for the 6-dB criterion at 3 kHz, to a high of 1.59, for the 3-dB criterion at 4 kHz.

B. Onset latencies

The OLs for the $2f_1-f_2$ and $2f_2-f_1$ were obtained in the regions of the DP-grams where both DPOAEs were clearly above their respective noise floors. Figure 3 shows, for four representative subjects, examples of the $2f_1-f_2$ and $2f_2-f_1$ DP-grams, which were used to determine the frequency region in which good OLs could be obtained for both emissions. The interesting feature about these plots is that although the $2f_1-f_2$ amplitude remained relatively constant across GM frequency, the $2f_2-f_1$ amplitude for the same primary-tone pairs varied dramatically, and, more or less, randomly. Although, on average, the $2f_2-f_1$ DPOAE was much smaller than the $2f_1-f_2$, it is clear that, within restricted frequency regions, the levels of the $2f_2-f_1$ DPOAEs were comparable to the typically more robust

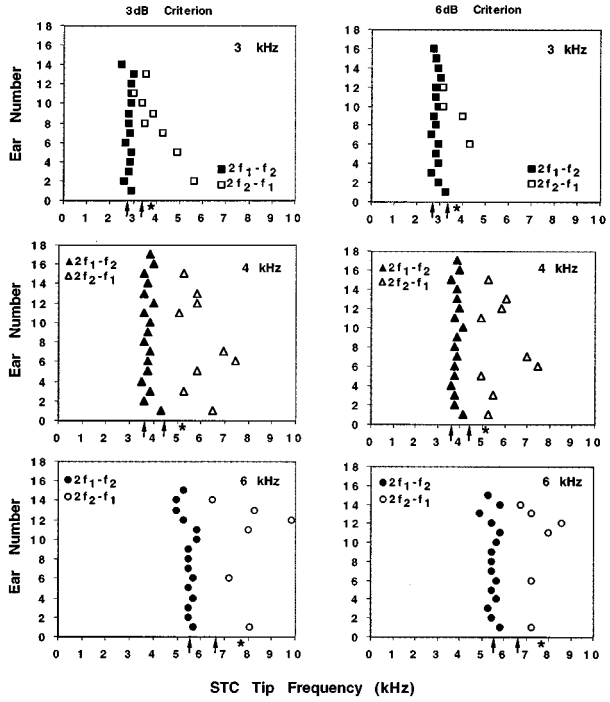


FIG. 2. STC tip frequencies obtained for all subjects at the GM frequencies of 3, 4, and 6 kHz, for the 3- and 6-dB suppression criteria, for the $2f_1-f_2$ (solid symbols) vs $2f_2-f_1$ (open symbols) DPOAEs. All $2f_1-f_2$ DPOAEs clearly tuned at, or slightly below, the GM of the primary tones, whereas the majority of the $2f_2-f_1$ DPOAEs tuned above the $2f_2-f_1$ frequency (asterisks by abscissa). Except at 3 kHz, none of the $2f_2-f_1$ DPOAEs tuned at or below the primary-tone place. The vertical axis is the ear number and the same subject has the same number in each panel. There are missing data points in some cases due to the difficulties in determining a unique tip frequency. Primary-tone frequencies are indicated by arrows on abscissa.

$2f_1-f_2$. Also apparent from these plots is that the locations of large-amplitude $2f_2-f_1$ DPOAEs occurred essentially at random, and, thus, there was no apparent frequency-shifting maneuver that aligned the two DP-grams in a straightforward manner.

Figure 4 shows examples of the raw time waveforms from which OLs were extracted for the $2f_1-f_2$ (left) and $2f_2-f_1$ (right) DPOAEs covering the range of frequencies tested. The top traces of each frequency panel depict the actual DPOAE waveform superimposed upon a reference sinusoid that was aligned with the DPOAE frequency at steady state. The corresponding bottom trace shows the time course of the correlation of the emission trace with the steady-state sinusoid. On this trace, the point marked by the vertically placed cursor signifies the time at which the r value remained >0.5 (solid line) or 0.75 (dashed line) for 2 ms. At the lower frequencies, especially at 0.871 kHz (top), it is obvious that the 0.5 criterion clearly scored the latency too soon, whereas the 0.75-criterion latency appeared to be slightly too long. The arrow on the abscissa of the time waveform above indicates the latency (9 ms) that was determined by visual inspection. At the lowest frequencies (i.e., 0.871 kHz), there is a clear difference between the two emission latencies, with the $2f_2-f_1$ exhibiting a latency of 6.5 ms compared to 9 ms for the $2f_1-f_2$ DPOAE. In these examples, the differences between the two emissions was on the order of 2.5 ms at lower frequencies, and became shorter and more difficult to visually appreciate at the higher frequencies, particularly for the higher-frequency $2f_2-f_1$ DPOAEs.

Figure 5(a) depicts a scatterplot of the OLs for the two DPOAEs determined either by visual verification of the correlation-based latencies, or visual assignment of a final latency value. It can be seen that the latencies of the $2f_2-f_1$ DPOAE (open circles) for the same GM frequency were significantly shorter than the $2f_1-f_2$ latencies (solid circles), particularly at the low frequencies ($t=9.1$, $df=79$, $p<0.0001$). These latency data were best fit by power func-

TABLE II. Mean $2f_2-f_1$ STC tip frequency/mean $2f_1-f_2$ tip frequency.

Geometric-mean frequency (kHz)	3-dB STC criterion	6-dB STC criterion	Mean
3.0	1.41	1.28	1.35
4.0	1.59	1.52	1.56
6.0	1.46	1.37	1.41
Mean	1.49	1.39	1.44

TABLE I. Average STC properties of amplitude (dB SPL), characteristic or best frequency (kHz), tip threshold (dB SPL), and $Q_{10\,dB}$ tuning factor.

GM frequency (kHz)	DPOAE	n	Amplitude (dB SPL)	CF (kHz)	Tip threshold (dB SPL)	$Q_{10\,dB}$ tuning
A. 3-dB criterion						
3.0	$2f_1-f_2$	14	8.08	2.83	69.71	1.56
	$2f_2-f_1$	8	0.66	4.00	64.80	2.14
4.0	$2f_1-f_2$	17	9.84	3.79	69.12	1.91
	$2f_2-f_1$	9	−1.42	6.02	63.33	2.24
6.0	$2f_1-f_2$	15	12.64	5.47	70.20	1.55
	$2f_2-f_1$	6	0.83	7.98	64.67	1.67
B. 6-dB criterion						
3.0	$2f_1-f_2$	15	7.49	2.84	73.07	1.89
	$2f_2-f_1$	4	1.84	3.65	70.50	2.50
4.0	$2f_1-f_2$	17	9.84	3.85	72.35	2.29
	$2f_2-f_1$	9	−1.25	5.87	68.50	2.22
6.0	$2f_1-f_2$	15	12.55	5.52	72.87	1.81
	$2f_2-f_1$	6	1.16	7.55	68.40	1.96

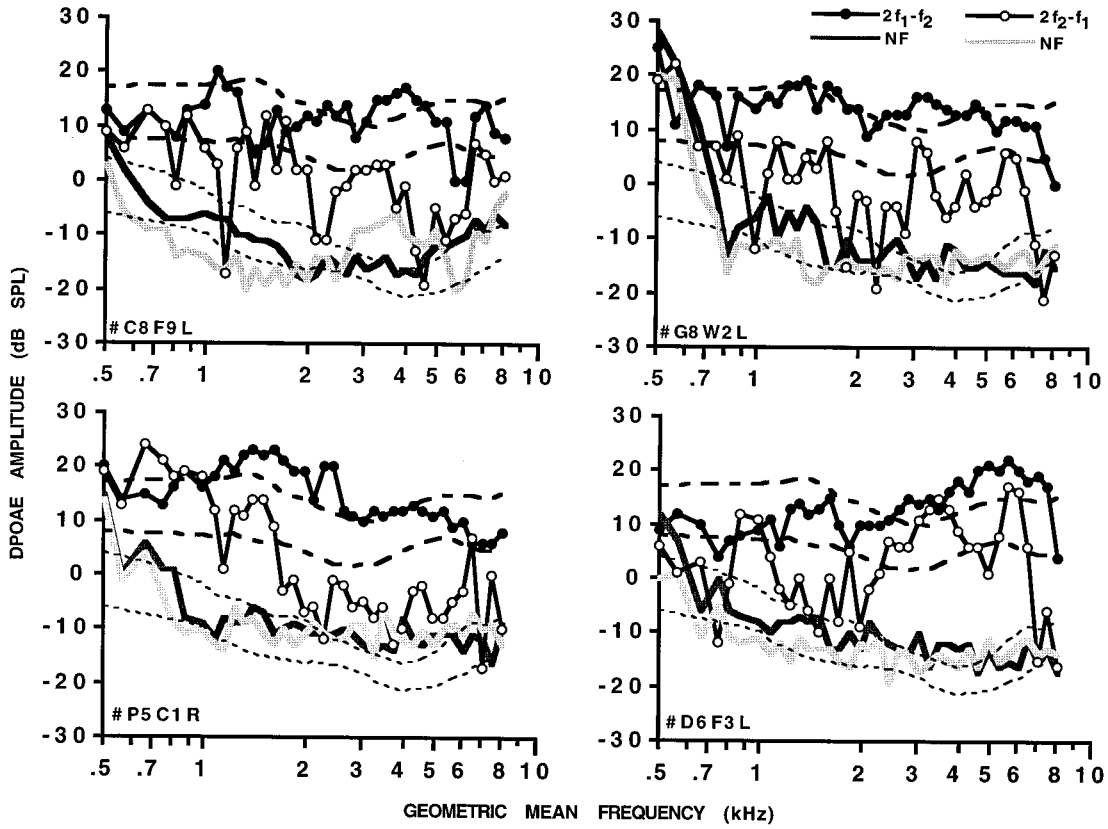


FIG. 3. Individual examples of $2f_1-f_2$ (solid circles) vs $2f_2-f_1$ (open circles) DP-grams used to determine the frequencies at which OLs could be obtained for both DPOAEs. The striking feature of these plots is the relatively stable amplitude of the $2f_1-f_2$ DPOAE across frequency that can be contrasted to the highly variable amplitudes of the $2f_2-f_1$ evoked by the same primaries. Also, note that there is no visually apparent shifting that would align the amplitudes of the two DP-grams. Normal levels of the $2f_1-f_2$ suggest these differences cannot be ascribed to confounding middle-ear factors. Bold dashed lines at $\sim$10–20 dB SPL=laboratory normative amplitudes for $2f_1-f_2$ DPOAEs; thin dotted lines at $\sim$0 to $-$15 dB SPL=laboratory normative noise floors (NFs) for $2f_1-f_2$; bold solid line=individual $2f_1-f_2$ NF; stippled bold line=individual $2f_2-f_1$ NF.

tions, which yielded the following equations, in which the emission latency, in milliseconds, was expressed as a function of the GM frequency of the eliciting primaries, in kHz:

$$\text{Latency}_{2f1-f2}=7.745*\text{GM}^{-0.757},$$

$$\text{Latency}_{2f2-f1}=4.580*\text{GM}^{-0.637}.$$

These statistical fits resulted in correlations of 0.96 for the $2f_1-f_2$ (solid line), and 0.84 for the $2f_2-f_1$ DPOAE (dashed line).

Portion (b) of Fig. 5 depicts latency differences (i.e., $2f_1-f_2$ minus $2f_2-f_1$) between the two DPOAEs in which both measures came from the same ear. This procedure eliminates individual variability that might arise due to, for example, differences in cochlear length across subjects. At the low frequencies, the round-trip cochlear delays between the $2f_1-f_2$ and the $2f_2-f_1$ DPOAEs were very large, i.e., they were on the order of 7–8 ms, with the $2f_2-f_1$ appearing much sooner in the ear canal. In fact, few cases fell on, or below, the dashed line indicating that the latencies were rarely equal, or that the $2f_1-f_2$ DPOAE was very seldom shorter than the latency for the corresponding $2f_2-f_1$ DPOAE.

The $2f_1-f_2$ latency data indicate that the majority of cochlear-travel time occurred for the low-frequency primaries, and diminished rapidly as higher frequencies were approached. Because the two DPOAE frequencies are apportioned at a constant distance when plotted on a logarithmic axis, they should, therefore, be approximately equidistant on the basilar membrane (Liberman, 1982; Tsuji and Liberman, 1997). Thus if the latencies for both DPOAEs lie on the same function, but the $2f_2-f_1$ latencies are simply plotted at the incorrect frequency place, then for a linear latency versus frequency function, the differences should be constant. For nonlinear functions, such as the power functions plotted in Fig. 5(a), latency differences should decrease as a function of frequency. It is obvious that the latency functions did not maintain a constant latency difference between the two DPOAEs as can be seen in Fig. 5(b) in which the latency differences between the two emissions are much greater at lower frequencies. The fact that the latency differences decreased as a function of GM frequency suggests that the two DPOAEs share a common function describing cochlear delay as a function of frequency, but, as plotted in Fig. 5(a), one function is shifted with respect to the other.

If the above assumptions are correct, then it should be possible to plot the GM frequency for the $2f_2-f_1$, which is shifted by a constant factor that aligns the two latency functions. The results of such a procedure are shown in Fig. 6 in

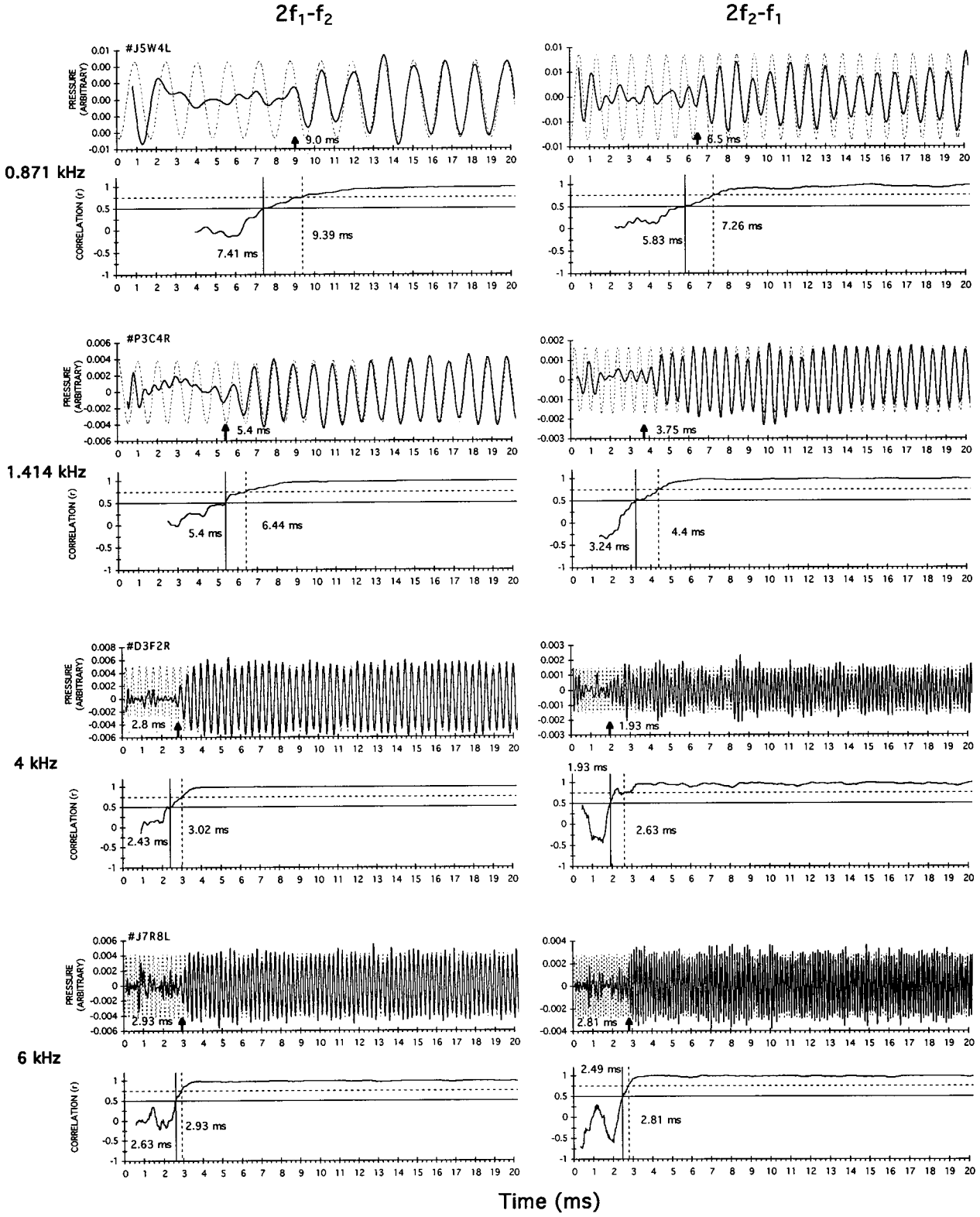


FIG. 4. Examples of OLs for the $2f_1-f_2$ and $2f_2-f_1$ DPOAEs collected at representative frequencies along the DP-gram. Each pair of plots displays the raw DPOAE waveform (bold trace) with the synthesized DPOAE frequency in the background (thin dotted sinusoidal trace) along with the sliding correlations computed over four cycles of the waveform below. On the bottom trace, correlations of 0.5 and 0.75 are marked by the solid and dotted vertical cursors, respectively. Arrows at the bottom of the time waveforms mark either the 0.5, 0.75, or a visually determined latency (see text). At low frequencies (e.g., 0.871 or 1.414 kHz), it is clear that the latency of the $2f_2-f_1$ DPOAE was considerably shorter than that for the $2f_1-f_2$ DPOAEs. These frequencies accounted for the majority of instances in which latencies determined by visual inspection were deemed most accurate. At higher frequencies, the objective correlation methods appeared to be equivalent to, or better than, visual inspection.

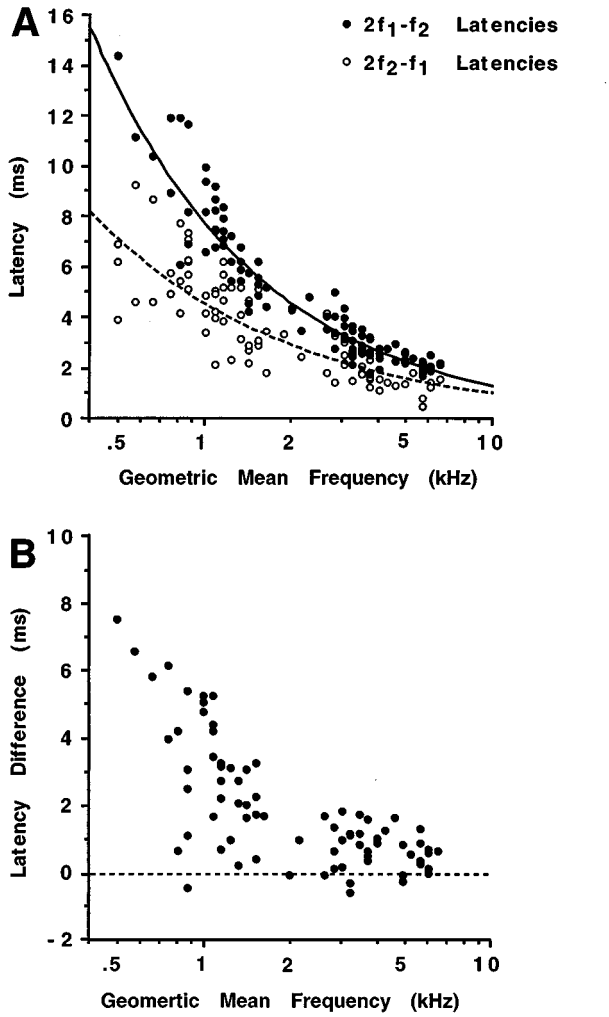


FIG. 5. Scatterplots (A) of OLs for the $2f_1-f_2$ DPOAE (solid circles) compared to OLs for the $2f_2-f_1$ DPOAE (open circles) plotted at the GM of the primary tones. Essentially all of the $2f_2-f_1$ latencies were shorter than those for the $2f_1-f_2$. Because the decrease in latency as a function of frequency was nonlinear, the differences were greatest at low rather than at high frequencies. Both sets of data were best fit by power functions (solid line$=2f_1-f_2$; dashed line$=2f_2-f_1$). Portion (B) plots differences in latency between the two DPOAEs (i.e., $2f_1-f_2$ minus $2f_2-f_1$) based upon measures from the same ears. This plot demonstrates that the differences in latency between the two DPOAEs apparent in panel (A) cannot be accounted for by sampling biases caused by plotting latencies from different subjects.

which the GM frequency producing the corresponding $2f_2-f_1$ was multiplied by 1.6 prior to plotting. Although not perfect, it can be seen that, for the most part, such a transformation does an excellent job of overlapping the two latency distributions. It should also be noted that this fit was performed by visual inspection of the effects of various constants and, consequently, although it is a close fit, it is not necessarily the most accurate 'best fit' to these data.

C. Correlations between $2f_1-f_2$ and $2f_2-f_1$

The average amplitudes of the $2f_1-f_2$ and $2f_2-f_1$ DPOAEs in the 108 normal ears upon which the correlation computations were performed are plotted in Fig. 7. It is clear from this comparison that across all frequencies, the $2f_2-f_1$ DPOAE was smaller, on average, than the $2f_1-f_2$

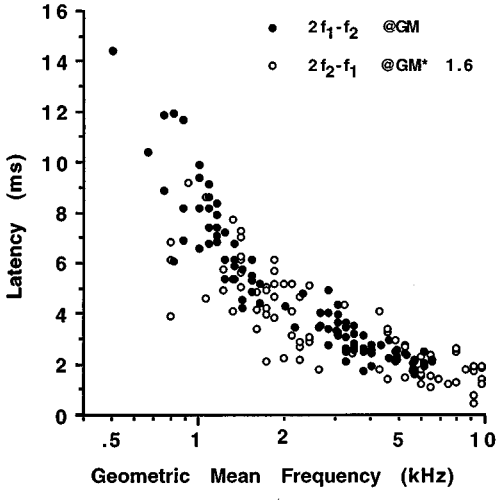


FIG. 6. Latency data of Fig. 5(A) in which the GM frequency producing the corresponding $2f_2-f_1$ was multiplied by 1.6 prior to plotting. This transformation aligned the two distributions and resulted in a single function describing cochlear delay as a function of GM frequency. Such a manipulation assumes that the two DPOAEs do not come from the same frequency place on the basilar membrane and avoids the problem of accounting for the large latency differences depicted in Fig. 5, if both DPOAEs originated at the same location.

DPOAE. Specifically, although the $2f_1-f_2$ tended to maintain a relatively constant amplitude across frequency, the $2f_2-f_1$ was clearly largest from about 0.5 to 1 kHz. However, from 1 to 3 kHz, it declined gradually, and then remained at a fairly constant level of -5 to -10 dB SPL for the frequencies above 3 kHz.

Based on the data of Fig. 7, correlations were performed between the $2f_1-f_2$ amplitudes at a given GM frequency, and all the corresponding $2f_2-f_1$ amplitudes for each GM frequency in the DP-gram. For each $2f_1-f_2$ frequency, as noted above, all r values were 3-point smoothed, and the

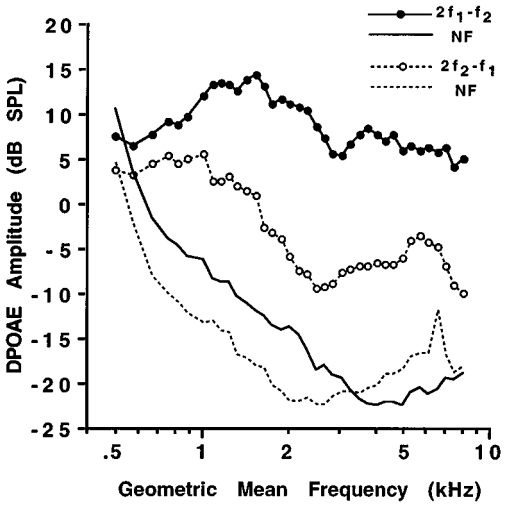


FIG. 7. Average amplitudes of the $2f_1-f_2$ DPOAE (solid circles) and $2f_2-f_1$ DPOAE (open circles) from 108 ears of 94 normally hearing subjects. On average, $2f_2-f_1$ levels were smaller than those of the $2f_1-f_2$, especially at frequencies above 1 kHz. Thin solid line$=2f_1-f_2$ NF; thin dotted line$=2f_2-f_1$ NF.

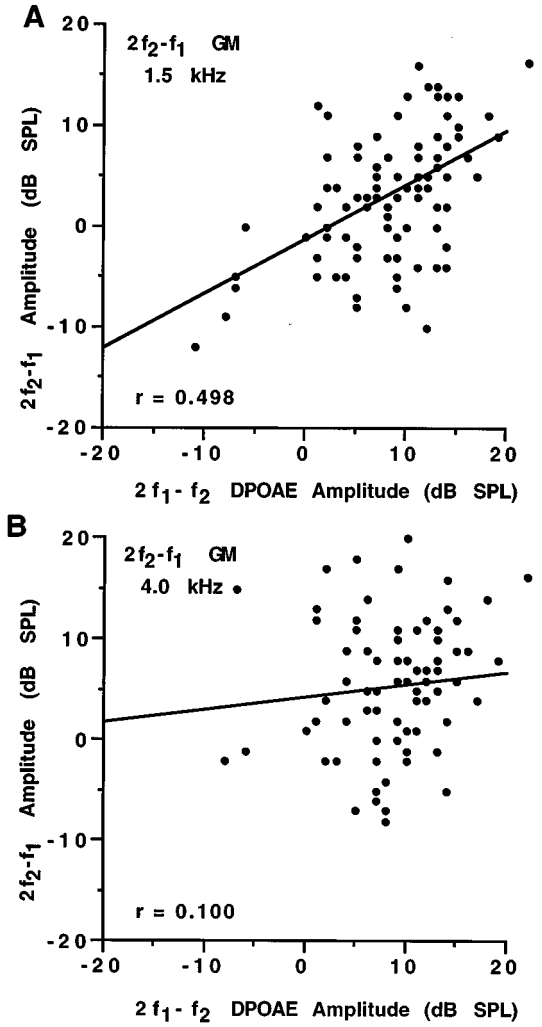


FIG. 8. Example scatterplots showing the relationship between $2f_1-f_2$ and $2f_2-f_1$ amplitudes across 108 ears, when the GM of the primaries generating the $2f_1-f_2$ was held constant at 2.828 kHz, while the $2f_2-f_1$ amplitudes were elicited either by primaries at the GM of 1.5 kHz (A) or 4.0 kHz (B). The obvious correlation in (A) occurs when the primaries were chosen so that the $2f_2-f_1$ frequency place fell near the GM of the primaries producing the $2f_1-f_2$ DPOAE.

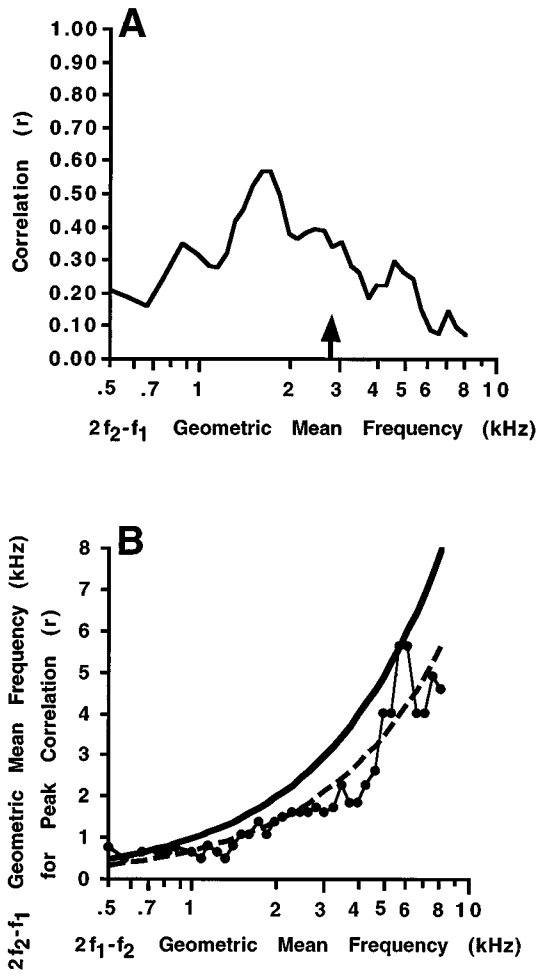


FIG. 9. Example (A) of all correlations ($n=37$) obtained when the GM of the $2f_1-f_2$ was held constant at 2.828 kHz (arrow) and the $2f_1-f_2$ amplitudes were successively correlated with the $2f_2-f_1$ amplitudes for all GM frequencies in the DP-gram ($n=37$). All r values were three-point smoothed and the peak correlation determined. Peak correlations (B) determined as illustrated in (A) for all $2f_2-f_1$ GM frequencies plotted against the $2f_1-f_2$ GM frequency. Solid bold line indicates expected peak correlations, if both DPOAEs originated at the GM of the primaries producing the $2f_1-f_2$. Bold dashed line indicates expected peak correlations, if the $2f_2-f_1$ was generated basal to the $2f_1-f_2$. Peak correlations (solid circles) tended to fall on the dashed line indicating that the amplitudes of the two DPOAEs correlated best when the $2f_2-f_1$ primaries were shifted so that the $2f_2-f_1$ frequency place was near the primaries producing the $2f_1-f_2$ DPOAE.

peak r value noted. The (a) and (b) panels of Fig. 8 illustrate representative scatterplots for the $2f_1-f_2$ DPOAE that resulted from this procedure using unsmoothed data for the GM frequency of 2.828 kHz. Portion (a) of the figure shows a region of high correlation when the 2.828-kHz $2f_1-f_2$ amplitudes were correlated with all the $2f_2-f_1$ amplitudes produced when the primary pairs were at a GM of 1.5 kHz. This frequency resulted in a relatively good correlation of 0.48, whereas a similar procedure performed for $2f_2-f_1$ DPOAEs arising from primaries with a GM of 4 kHz resulted in a correlation of only 0.10. The results of this procedure for the 2.828-kHz GM for all GM pairs for the $2f_2-f_1$ are shown in Fig. 9(a). It is clear from this plot that the peak in the correlation function occurred when the GM of the primaries producing the $2f_2-f_1$ was lower in frequency than the 2.828-kHz GM marked by the arrow. This combination of primaries with a GM of approximately 1.5 kHz would tend to position the $2f_2-f_1$ frequency place near the primary-tone place responsible for producing the $2f_1-f_2$ DPOAE associated with 2.828-kHz GM primaries.

In Fig. 9(b), correlations are plotted for all $2f_1-f_2$ GM pairs on the abscissa against the GM that produced the highest correlation with the $2f_2-f_1$ DPOAE on the ordinate. Curves on the graph indicate the peak-correlation frequencies (solid circles), and the expectations for various generation sites for the two DPOAEs. If both DPOAEs originated at the GM of the $2f_1-f_2$ primary pairs, then it would be expected that the peak correlations would occur when the two GM frequencies were equal, as shown by the solid-heavy line. However, the peak r values did not follow this ''equi'' function, but, rather, they essentially tracked the

dashed line, which plots the GM frequency needed to place the $2f_2-f_1$ frequency at the GM of the corresponding $2f_1-f_2$. It is clear from these data that the peak correlations were obtained when the primary pairs producing the $2f_2-f_1$ were at a lower frequency (i.e., they followed the trajectory of the dashed line), thus placing the $2f_2-f_1$ frequency at, or near to the primary place that produced the $2f_1-f_2$ DPOAE.

III. DISCUSSION

The present study provided evidence in humans for the locus of generation for the $2f_1-f_2$ versus the $2f_2-f_1$ DPOAE. This evidence was based on the outcomes of a series of experiments using suppression tuning, latency, and amplitude-correlation techniques. Using STC procedures, it was discovered that the $2f_1-f_2$ consistently tuned slightly below the GM of the primary tones, whereas the $2f_2-f_1$ tuned basal to the primary-tone place. These distinct findings were supported by the results of further studies of OLs in which the $2f_2-f_1$ OLs were considerably shorter than those for the $2f_1-f_2$. The OLs for the two DPOAEs were made comparable by shifting the GM frequencies so that the $2f_2-f_1$ latencies from the same primary-tone pairs that produced the $2f_1-f_2$ were plotted somewhat above the $2f_2-f_1$ characteristic place. The success of this transformation again reinforced the basal generation hypothesis for the $2f_2-f_1$ DPOAE. Finally, in a population of normal ears, $2f_1-f_2$ amplitudes correlated best with $2f_2-f_1$ amplitudes when the GM of the primaries producing the $2f_2-f_1$ was chosen so that the $2f_2-f_1$ frequency was located at the GM frequency producing the $2f_1-f_2$. Thus the outcomes of all three procedures support the conclusion that the $2f_2-f_1$ DPOAE is generated at, or basal to, its characteristic place on the basilar membrane.

A. General observations on $2f_1-f_2$ vs $2f_2-f_1$

Several general observations on the behavior of the $2f_1-f_2$ vs $2f_2-f_1$ DPOAE provide reasons to believe that there might be significant differences between these two emissions. Most striking is the pattern of DP-grams measured in the same ears (Fig. 3) in which the $2f_1-f_2$ amplitudes remained relatively constant in the presence of dramatic variations in the levels of the $2f_2-f_1$. Because the $2f_1-f_2$ DPOAE remained robust across all frequencies, it does not appear that differences in middle-ear transmission accounted for these differences in amplitude between the two DPOAEs. Considering that both emissions are evoked simultaneously by the same set of primary tones, the lack of a correlation between the amplitudes of the two DPOAEs can be interpreted as an indication of a significant difference in some aspect of their generation.

It is also noteworthy that in individual DP-grams, or over a large population of subjects (Fig. 7), the $2f_2-f_1$ is typically much smaller in amplitude than the $2f_1-f_2$ DPOAE. These differences have been noted previously in humans (Moulin *et al.*, 1993; Moulin and Kemp, 1996; Wable *et al.*, 1996) and animals (Lonsbury-Martin *et al.*, 1987), and also provide evidence for a difference between the two emissions. Unfortunately, the exact basis for this difference remains obscure. These data could be construed as supporting the traditional belief that the $2f_2-f_1$ is generated at the primary-tone place, and is then strongly attenuated as it traverses its characteristic-frequency place on the basilar membrane. However, if this latter conjecture were true, it is difficult to appreciate how, in a significant number of instances, the $2f_2-f_1$ could be as large as the $2f_1-f_2$ (see Fig. 3). An alternative hypothesis, supported by the findings presented here, is that the $2f_2-f_1$ is not generated in the region of maximal primary-tone interactions. In this situation, the input to the cochlear nonlinearity responsible for DPOAE generation would presumably be much smaller, due to the decreased amplitude of the traveling waves basal to the primary-tone place. This decreased basilar-membrane vibration (Johnstone *et al.*, 1986) would be expected, on average, to yield consistently smaller $2f_2-f_1$ emission amplitudes, which would be similar to the effects of lowering primary-tone levels on DPOAEs amplitudes (e.g., Stover *et al.*, 1996a), in general.

B. Suppression tuning curves (STCs)

It is generally assumed (Kummer *et al.*, 1995) that presentation of a third tone in the form of a suppressor interferes with the nonlinearity responsible for DPOAE production and, consequently, STCs provide evidence for the locus of DPOAE generation on the basilar membrane. If this assumption is correct, then a number of STC studies (Kemp and Brown, 1983; Brown and Kemp, 1984; Harris *et al.*, 1992; Lonsbury-Martin *et al.*, 1994; Abdala and Sininger, 1996; Abdala *et al.*, 1996; Gaskill and Brown, 1996; Kummer *et al.*, 1995) have consistently shown that the $2f_1-f_2$ DPOAE in humans is generated in the region of the primary-tone place, or near f_2. The exact location of maximum suppression appears to depend on primary-tone level (Kummer *et al.*, 1995), or on the relative levels of the f_1 and f_2, with STC tips consistently tuned nearer to, or above the f_2 place, when the level of f_2 is considerably lower than the level of f_1 (Lecusay *et al.*, 1996). Additionally, a small contribution to the amplitude of the $2f_1-f_2$ DPOAE appears to arise from the DPOAE place (Kummer *et al.*, 1995; Brown *et al.*, 1996; Gaskill and Brown, 1996).

The present study compared STCs obtained from the same ears at three GM frequencies for the $2f_1-f_2$ and $2f_2-f_1$ DPOAEs. In general agreement with other studies, the $2f_1-f_2$ STCs consistently tuned slightly below the GM frequency of the primaries. In contrast, the $2f_2-f_1$ STCs all tuned basal to the primary-tone place and, for all GM frequencies tested, at least, 50% of the STC tips were at, or above, the $2f_2-f_1$ frequency. These findings support the supposition that the $2f_2-f_1$ DPOAE arises from a region remote from the vicinity of maximal primary-tone interaction.

It is interesting to note that, whereas all $2f_1-f_2$ STC tip frequencies were very closely aligned (see Fig. 2), the $2f_2-f_1$ STC tips showed much greater variability. Although this difference may in part be due to the greater difficulty in measuring the smaller amplitude $2f_2-f_1$, as a whole, this variability appeared to reflect real differences in the location of generator of the $2f_2-f_1$ DPOAE across ears.

The inconsistency of the $2f_2-f_1$ STC tip frequencies across subjects was particularly intriguing in that during inspection of the $2f_1-f_2$ vs $2f_2-f_1$ OLs, it was noted that some of the latency-study individuals showed much greater differences between the latencies for the two DPOAEs than other subjects did for the same frequency. It would be interesting to establish in the same subjects if the location of the $2f_2-f_1$ STC tip frequencies correlated with this variability in OLs. Unfortunately, such a correlation could not be performed in the present study, because the STC and latency measures originated from different subjects.

Assuming that STC tip frequencies indicate the locus of DPOAE generation, then the ratio of the $2f_2-f_1$ to the $2f_1-f_2$ STC tip frequencies should have provided the factor needed to shift the region of $2f_1-f_2$ generation to the locus of the $2f_2-f_1$ production. On average, the mean value of this ratio was 1.44. Because the actual factor required to align the $2f_1-f_2$ GM to the $2f_2-f_1$ frequency place was 1.29, the STC data imply that the $2f_2-f_1$ is generated on the basilar membrane at a frequency region that is above the $2f_1-f_2$ frequency place. However, if the $2f_1-f_2$ is generated at or nearer to f_1 as indicated by many of the STC tips of Fig. 2, then the appropriate ratio becomes 1.42 suggesting that the $2f_2-f_1$ is generated not above, but at its frequency place.

C. Latency

The results of several recent studies using either a phase-gradient method (Moulin and Kemp, 1996; Wable *et al.*, 1996) or an OL procedure (Whitehead *et al.*, 1996) to measure DPOAE latencies showed that the latency of the $2f_2-f_1$ DPOAE in humans is shorter than that for the $2f_1-f_2$. The present study agreed with these previous findings in that OLs for the $2f_2-f_1$ were consistently shorter than those for the $2f_1-f_2$ DPOAE. In addition, both DPOAE latencies were found to systematically decrease as a function of increasing frequency.

Estimates of cochlear travel time based on DPOAEs depend on the particular methods used to measure them (Kimberley *et al.*, 1993; O Mahoney and Kemp, 1995; Moulin and Kemp, 1996; Whitehead *et al.*, 1996; Bowman *et al.*, 1997). Phase-gradient latencies can be obtained with either a swept-f_1 or a swept-f_2 procedure, whereas OLs are determined from the raw DPOAE waveform (Whitehead *et al.*, 1996). However, regardless of the procedure used to measure DPOAE latency, all methods have consistently found shorter latencies for the $2f_2-f_1$ compared to the $2f_1-f_2$ DPOAE. Interestingly, the two phase-gradient procedures do not exhibit differences for the upper side-band $2f_2-f_1$, which has been interpreted as evidence for a different region of generation for this emission (Moulin and Kemp, 1995).

The agreement across studies that $2f_2-f_1$ has a shorter latency than the $2f_1-f_2$ DPOAE is, again, consistent with the assumption of a more basal site of generation for the $2f_2-f_1$ emission. Surprisingly, however, latency differences [Fig. 5(b)] between the two DPOAEs at low frequencies were as much as 6–7 ms in individual ears. To explain such large differences in individual latencies would presumably require significantly different modes of propagation on the basilar membrane, or dramatically different filtering processes, if the two DPOAEs actually arise from the same basilar-membrane location. However, the existing differences can be explained easily by assuming different loci of generation along the cochlear partition, with the latencies of both emissions lying on a single function that describes cochlear delay as a function of frequency. Based on the above assumptions, the GM yielding the $2f_2-f_1$ should be plotted at a higher frequency in order to place this DPOAE at its appropriate generation site. When the $2f_2-f_1$ GM frequencies were shifted by a factor of 1.6 (Fig. 6), the two latency functions were very closely aligned. This transformation resulted in consistent latencies for a given basilar-membrane location, and, also, relates favorably to the data of Wable *et al.* (1996) in which the $2f_1-f_2$ compared to the $2f_2-f_1$ latencies overlapped when each DPOAE was plotted at their respective frequencies.

D. Correlations between $2f_1-f_2$ and $2f_2-f_1$ amplitudes

Although individual DP-grams showed significant differences between the levels of $2f_1-f_2$ and $2f_2-f_1$ DPOAEs across frequency (Fig. 3), there was no visually apparent frequency shift that aligned the two DP-grams to make the amplitude dips correlate with one another. On the other hand, when individual $2f_1-f_2$ vs $2f_2-f_1$ DPOAE levels were correlated across some 108 ears, substantial peak correlations were found. It would be expected that, if both the emissions originated around the primary-tone region, correlations between them would be greatest when the GM of the primaries producing each DPOAE were the same [solid curve in Fig. 9(B)]. In fact, correlations peaked when the GM producing $2f_2-f_1$ DPOAEs were at lower frequencies than the GM producing the $2f_1-f_2$ [dashed line in Fig. 9(B)]. One interpretation of this outcome is that shifting the GM for the $2f_2-f_1$ resulted in aligning the sites of generation for both DPOAEs. If this explanation is correct, then the $2f_2-f_1$ must again be generated basal to the $2f_1-f_2$. Although the source of this correlation cannot be determined from the present study, as noted above, it was not apparent on the individual DP-grams. Presumably the source of the correlation is more subtle, but perhaps, the fine structure of the DP-gram (He and Schmiedt, 1993) seen with much smaller frequency steps of the primaries is the basis of the overall correlations found here.

E. Findings in laboratory animals versus humans

An important issue in the field of otoacoustic emissions is to understand the similarities and differences between findings in laboratory animals compared to humans. The outcome in humans that the $2f_2-f_1$ DPOAE appears to arise basal to the primary-tone place is consistent with the outcomes of several animal studies. Brown and Kemp (1985) first demonstrated in gerbils that latencies for the $2f_2-f_1$ DPOAE were shorter than those for the $2f_1-f_2$. Recently, Whitehead *et al.* (1996) also obtained shorter OLs in rabbits for the $2f_2-f_1$ than for the $2f_1-f_2$ DPOAE, which is consistent with the findings for humans. Earlier, Martin *et al.*

(1987) demonstrated in rabbits, similar to that found here for humans, that the $2f_2-f_1$ STC tips tuned basal to the primary-tone place. Consistent with these latter STC results was the observation that noise overstimulation maximally affected the $2f_2-f_1$ DPOAE when the exposure tone was at the $2f_2-f_1$ frequency. Because noise overexposure typically affects frequencies at, or approximately, 1/2 octave above the exposure frequency, this latter outcome strongly implicates a region in rabbits above the $2f_2-f_1$ frequency place as the site of generation for this emission. With respect to the $2f_1-f_2$ DPOAE, to our knowledge, all laboratory studies in animals agree with the outcomes in humans in that the region of generation of this DPOAE appears to be located at the primary-tone place. In short, it appears that evidence for two generation sites, one for the $2f_1-f_2$ and another for the $2f_2-f_1$, exists both in humans and in laboratory animals. These findings are important for understanding DPOAEs measurable in the ear canal of humans, so that their presence or absence can eventually be interpreted in a clinically meaningful manner. Furthermore, demonstrations of similarities between various types of species strengthens the notion that DPOAEs in laboratory animals mimic those measured in humans.

Laboratory animals have been clearly shown to have two mechanisms of $2f_1-f_2$ DPOAE generation, one at low stimulus levels, which is relatively vulnerable to cochlear insults, and another high-level source that is relatively invulnerable to otoacoustic insults (Whitehead *et al.*, 1992a,b; Mills *et al.*, 1993; Mills and Rubel, 1994). The present results were obtained with 75 dB SPL primary tones and, consequently, may involve some aspects of the high-level source, if, in fact, one exists in humans. Unpublished observations in our laboratory following procedures in rabbits that eliminate the low-level DPOAE generator, such as noise exposure or injections of furosemide, suggest that the low-level source is also involved in generation of the $2f_2-f_1$. Mills *et al.* (1993) also described the reduction in the $2f_2-f_1$ DPOAE following furosemide injections in gerbils. Further, at relatively low primary-tone levels in humans, Moulin and Kemp (1996), at $L_1=65/L_2=60$ dB SPL, and Wable *et al.* (1996), at $L_1=L_2=55$ dB SPL, also found that the $2f_2-f_1$ latency remained shorter than that for the $2f_1-f_2$ DPOAE. Thus overall, it appears that the $2f_2-f_1$ is not dependent on a high-level DPOAE generator and, consequently, the presence of this emission, or the fact that it is generated in a location remote to the $2f_1-f_2$, does not provide evidence for a high-level emission generator in humans.

F. Mechanisms of $2f_2-f_1$ generation

Considering that the amplitude of the traveling-wave envelope is still relatively small in regions basal to the primary-tone place (Johnstone *et al.*, 1986), it is important to understand, if our interpretations of the data are correct, how a substantial $2f_2-f_1$ can be generated at this remote location, in the presence of greatly reduced basilar-membrane vibrations. One possible, although presumably an unlikely hypothesis is that the $2f_2-f_1$ DPOAE is actually generated at the primary-tone place, propagated to its characteristic-frequency place, and then enhanced by the cochlear amplifier after which it appears in the ear canal signal. Under these circumstances, one could propose that suppression of the cochlear amplifier at this site explains the STC findings, and that re-emission from the $2f_2-f_1$ region is responsible for the decreased latencies of the $2f_2-f_1$. This notion could be tested in humans, if primary tones could be placed in a region of cochlear damage that dramatically reduced or eliminated the $2f_1-f_2$ DPOAE, while leaving the $2f_2-f_1$ relatively unchanged. Although we have not searched for such an example in humans, which seemingly would be difficult to find due to the small amplitudes of their $2f_2-f_1$ DPOAEs, such cases have been frequently observed in our laboratory in rabbits undergoing experimental acoustic overstimulation with octave bands of noise. Following such noise damage in rabbits, substantial $2f_2-f_1$ DPOAEs are often observed, even when the primary tones are located in the region of damage, as indicated by the essentially complete elimination of the $2f_1-f_2$ DPOAE. These findings suggest that under such circumstances the $2f_2-f_1$ cannot originate from the primary-tone place.

From the standpoint of the physics of a saturating nonlinearity, the generation of the $2f_1-f_2$ DPOAE slightly apical and the $2f_2-f_1$ slightly basal to f_2 has a firm physical basis. However, both sites should be quite close to the f_2 place on the basilar membrane (Paul F. Fahey, personal communication). Thus the present interpretations appear to be at odds with what one would expect from these physical predictions. This conflict can be resolved for the latency data, if it is assumed that the dominant component of the $2f_1-f_2$ DPOAE, under some circumstances, is not coming from the f_2 place. A likely alternative location, for example, would be the $2f_1-f_2$ DPOAE place (Brown and Beveridge, 1996; Stover *et al.*, 1996b; Fahey and Allen, 1997). Under these circumstances, the $2f_2-f_1$ latency would be shorter than the $2f_1-f_2$, and this emission would ''appear'' to come from a more basal region of generation. Exactly how the suppression and correlation data, or the observations following cochlear damage mentioned above, can be accounted for by this hypothesis still remains problematic. Overall, the present findings appear to contradict the notion that both emissions must arise very near to the f_2 place. Clearly, future experiments are needed to resolve the differences between the theoretical sites of generation, and the findings of the present study.

IV. SUMMARY AND CONCLUSIONS

The present report presents suppression tuning, onset latency, and amplitude-correlation data that convincingly suggest that the $2f_2-f_1$ DPOAE is generated basal to the primary-tone place in humans. These findings are consistent with those obtained for laboratory animals, and support the commonly held notion that most aspects of DPOAE generation and basilar-membrane mechanics are similar across species. Overall, the findings present a complex picture of DPOAE generation that involves contributions from several locations on the basilar membrane. Presumably DPOAEs produced at these distinct sites interact with one another in a complicated manner that further confounds the simple inter-

pretation of distortion products measured from a location in the outer ear canal. If the notion that the generation of some DPOAEs at basilar-membrane locations remote from the primary-tone place can be firmly established by future experimentation, then this knowledge must eventually be incorporated into models of cochlear mechanics.

ACKNOWLEDGMENTS

This research was supported in part by funds from the Public Health Service (DC00313, DC00613, DC/ES03114). The computer programming of Michael E Knapp M.D., the technical assistance of Geoffrey M. Waxman, and the secretarial assistance of Mayte Ruiz are gratefully acknowledged. The authors would also like to acknowledge the helpful suggestions of two anonymous reviewers, and Dr. Paul F. Fahey for comments on the manuscript.

Abdala, C., and Sininger, Y. (**1996**). ''The development of cochlear frequency resolution in the human auditory system,'' Ear Hear. **17**, 374–385.

Abdala, C., Sininger, Y. S., Ekelid, M., and Zeng, F-G. (**1996**). ''Distortion product otoacoustic emission suppression tuning curves in human adults and neonates,'' Hearing Res. **98**, 38–53.

Bowman, D. M., Brown, D. K., Eggermont, J. J., and Kimberley, B. P. (**1997**). ''The effect of sound intensity on f_1-sweep and f_2-sweep distortion product otoacoustic emissions in human adults,'' J. Acoust. Soc. Am. **101**, 1550–1559.

Brown, A. M. (**1987**). ''Acoustic distortion from rodent ears: A comparison of responses from rats, guinea pigs and gerbils,'' Hearing Res. **31**, 25–38.

Brown, A. M., and Kemp, D. T. (**1984**). ''Suppressibility of the $2f_1-f_2$ stimulated acoustic emissions in gerbil and man,'' Hearing Res. **13**, 29–37.

Brown, A. M., and Kemp, D. T. (**1985**). ''Intermodulation distortion in the cochlea: Could basal vibration be the major cause of round window CM distortion?,'' Hearing Res. **19**, 191–198.

Brown, A. M., Harris, F. P., and Beveridge, H. A. (**1996**). ''Two sources of acoustic distortion products from the human cochlea,'' J. Acoust. Soc. Am. **100**, 3260–3267.

Brown, A. M., McDowell, B., and Forge, A. (**1989**). ''Acoustic distortion products can be used to monitor the effects of chronic gentamicin treatment,'' Hearing Res. **42**, 143–156.

Fahey, P. F., and Allen, J. B. (**1997**). ''Measurement of distortion product phase in the ear canal of the cat,'' J. Acoust. Soc. Am. **102**, 2880–2891.

Faulstich, B. M., Kossl, M., and Reimer, K. (**1996**). ''Analysis of non-linear cochlear mechanics in the marsupial *Monodelphis domestica*: Ancestral and modern mammalian features,'' Hearing Res. **94**, 47–53.

Frank, G., and Kossl, M. (**1995**). ''The shape of the $2f_1-f_2$ suppression tuning curves reflects basilar membrane specializations in the mustached bat, *Pteronotus parnellii*,'' Hearing Res. **83**, 151–160.

Gaskill, S. A., and Brown, A. M. (**1996**). ''Suppression of human acoustic distortion product: Dual origin of $2f_1-f_2$,'' J. Acoust. Soc. Am. **100**, 3260–3274.

Gorga, M. P., Neely, S. T., Bergman, B., Beauchaine, K. L., Kaminski, J. R., Peters, J., and Jesteadt, W. (**1993**). ''Otoacoustic emissions from normal-hearing and hearing-impaired subjects: Distortion product responses,'' J. Acoust. Soc. Am. **93**, 2050–2060.

Hall, J. L. (**1974**). ''Two-tone distortion products in a nonlinear model of the basilar membrane,'' J. Acoust. Soc. Am. **56**, 1818–1828.

Hall, J. L. (**1980**). ''Cochlear models: Evidence in support of mechanical nonlinearity and a second filter (A review),'' Hearing Res. **2**, 455–464.

Harris, F. P., Probst, R., and Xu, L. (**1992**). ''Suppression of the $2f_1-f_2$ otoacoustic emission in humans,'' Hearing Res. **64**, 133–141.

He, N-J., and Schmiedt, R. A. (**1993**). ''Fine structure of the $2f_1-f_2$ acoustic distortion product: Changes with primary level,'' J. Acoust. Soc. Am. **94**, 2659–2669.

Johnstone, B. M., Patuzzi, R., and Yates, G. K. (**1986**). ''Basilar membrane measurements and the travelling wave,'' Hearing Res. **22**, 147–153.

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 Bases and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer-Verlag, Berlin), pp. 82–88.

Kemp, D. T., and Brown, A. M. (**1986**). ''Wideband analysis of otoacoustic intermodulation,'' in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer-Verlag, New York), pp. 306–313.

Kim, D. O., Molnar, C. E., and Matthews, J. W. (**1980**). ''Cochlear mechanics: Nonlinear behavior in two-tone responses as reflected in cochlear-nerve-fiber responses and in ear canal sound pressure,'' J. Acoust. Soc. Am. **67**, 1704–1721.

Kimberley, B. P., Brown, D. K., and Eggermont, J. J. (**1993**). ''Measuring human cochlear travelling wave delay using distortion product emission phase responses,'' J. Acoust. Soc. Am. **94**, 1343–1350.

Koppl, C., and Manley, G. A. (**1993**). ''Distortion-product otoacoustic emission in the bobtail lizard. II: Suppression tuning characteristics,'' J. Acoust. Soc. Am. **93**, 2834–44.

Kummer, P., Janssen, T., and Arnold, W. (**1995**). ''Suppression tuning characteristics of the $2f_1-f_2$ distortion-product emissions in humans,'' J. Acoust. Soc. Am. **98**, 197–210.

Lecusay, R. A., Fletcher, C. A., Lonsbury-Martin, B. L., Stagner, B. B., Waxman, G. M., and Martin, G. K. (**1996**). ''Otoacoustic emissions in normal-hearing humans: Musicians vs nonmusicians,'' Assoc. Res. Otolaryngol. Abstr. **19**, 25.

Liberman, M. C. (**1982**). ''The cochlear frequency map for the cat: Labeling auditory-nerve fibers of known characteristic frequency,'' J. Acoust. Soc. Am. **72**, 1441–1449.

Lonsbury-Martin, B. L., Martin, G. K., Whitehead, M. L., and Stagner, B. B. (**1994**). ''$2f_1-f_2$ vs $2f_2-f_1$ distortion-product otoacoustic emission generators,'' Soc. Neurosci. Abstr. **20**, 971.

Lonsbury-Martin, B. L., Martin, G. K., Probst, R., and Coats, A. C. (**1987**). ''Acoustic distortion products in rabbit ear canal. I. Basic features and physiological vulnerability,'' Hearing Res. **28**, 173–189.

Martin, G. K., Ohlms, L. A., Franklin, D. J., Harris, F. P., and Lonsbury-Martin, B. L. (**1990**). ''Distortion product emissions in humans: III. Influence of sensorineural hearing loss,'' Ann. Otol. Rhinol. Laryngol. **99**, Suppl. 147,30–42.

Martin, G. K., Probst, R., Scheinin, S. A., Coats, A. C., and Lonsbury-Martin, B. L. (**1987**). ''Acoustic distortion products in rabbits. II. Sites of origin revealed by suppression and pure-tone exposures,'' Hearing Res. **28**, 191–208.

Martin, G. K., Telischi, F. F., Jassir, D., Stagner, B. B., and Lonsbury-Martin, B. L. (**1997**). ''Effects of cochlear-nerve section on suppression and enhancement of DPOAEs by interference tones in rabbits,'' Assoc. Res. Otolaryngol. Abstr. **20**, 493.

Mills, D. M., and Rubel, E. W. (**1994**). ''Variation of distortion product otoacoustic emissions with furosemide injection,'' Hearing Res. **77**, 183–199.

Mills, D. M., Norton, S. J., and Rubel, E. W. (**1993**). ''Vulnerability and adaptation of distortion product otoacoustic emissions to endocochlear potential variation,'' J. Acoust. Soc. Am. **94**, 2108–2122.

Moulin, A., and Kemp, D. T. (**1995**). ''Interpreting the phase of distortion product otoacoustic emissions in humans,'' Br. J. Audiol. Abstr. **29**, 65.

Moulin, A., and Kemp, D. T. (**1996**). ''Multicomponent acoustic distortion product otoacoustic emission phase in humans. I. General characteristics,'' J. Acoust. Soc. Am. **100**, 1617–1639.

Moulin, A., Collet, L., Veuillet, E., and Morgon, A. (**1993**). ''Interrelations between transiently evoked otoacoustic emissions and acoustic distortion products in normally hearing subjects,'' Hearing Res. **65**, 216–233.

O Mahoney, C. F., and Kemp, D. T. (**1995**). ''Distortion product otoacoustic emission delay measurement in human ears,'' J. Acoust. Soc. Am. **97**, 3721–3735.

Siegel, J. H., Kim, D. O., and Molnar, C. E. (**1982**). ''Effects of altering organ of Corti on cochlear distortion products f_2-f_1 and $2f_1-f_2$,'' J. Neurophysiol. **47**, 303–328.

Stover, L. J., Neely, S. T., and Gorga, M. P. (**1996b**). ''Latency and multiple sources of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **99**, 1016–1024.

Stover, L. J., Gorga, M. P., Neely, S. T., and Montoya, D. (**1996a**). ''Toward optimizing the clinical utility of distortion product otoacoustic emission measurements,'' J. Acoust. Soc. Am. **100**, 956–967.

Tsuji, J., and Liberman, M. C. (**1997**). ''Intracellular labeling of auditory nerve fibers in guinea pig: Central and peripheral projections,'' J. Comp. Neurol. **381**, 188–202.

Wable, J., Collet, L., and Chery-Croze, S. (**1996**). ''Phase delay measure-

ments of distortion product otoacoustic emissions at $2f_1-f_2$ and $2f_2-f_1$ in human ears,'' J. Acoust. Soc. Am. **100**, 2228–2235.

Whitehead, M. L., Lonsbury-Martin, B. L., and Martin, G. K. (**1992a**). ''Evidence for two discrete sources of $2f_1-f_2$ distortion-product otoacoustic emission in rabbit: I. Differential dependence on stimulus parameters,'' J. Acoust. Soc. Am. **91**, 1587–1607.

Whitehead, M. L., Lonsbury-Martin, B. L., and Martin, G. K. (**1992b**). ''Evidence for two discrete sources of $2f1-f2$ distortion-product otoacoustic emission in rabbit: II. Differential physiological vulnerability,'' J. Acoust. Soc. Am. **92**, 2662–2682.

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.

Whitehead, M. L., McCoy, M. J., Martin, G. K., and Lonsbury-Martin, B. L. (**1993**). ''Click-evoked and distortion-product otoacoustic emissions in adults: Normative data,'' Assoc. Res. Otolaryngol. Abstr. **16**, 99.

Whitehead, M. L., Stagner, B. B., Martin, G. K., and Lonsbury-Martin, B. L. (**1996**). ''Visualization of the onset of distortion-product otoacoustic emissions, and measurement of their latency,'' J. Acoust. Soc. Am. **100**, 1663–1679.

Basilar-membrane responses to clicks at the base of the chinchilla cochlea

Alberto Recio,[a] Nola C. Rich,[b] S. Shyamla Narayan, and Mario A. Ruggero[c]
The Hugh Knowles Center, Department of Communication Sciences and Disorders, and Institute for Neuroscience, Northwestern University, 2299 North Campus Drive, Evanston, Illinois 60208-3550

(Received 25 September 1997; revised 19 December 1997; accepted 31 December 1997)

Basilar-membrane responses to clicks were measured, using laser velocimetry, at a site of the chinchilla cochlea located about 3.5 mm from the oval window (characteristic frequency or CF: typically 8–10 kHz). They consisted of relatively undamped oscillations with instantaneous frequency that increased rapidly (time constant: 200 μs) from a few kHz to CF. Such frequency modulation was evident regardless of stimulus level and was also present post-mortem. Responses grew linearly at low stimulus levels, but exhibited a compressive nonlinearity at higher levels. Velocity-intensity functions were almost linear near response onset but became nonlinear within 100 μs. Slopes could be as low as 0.1–0.2 dB/dB at later times. Hence, the response envelopes became increasingly skewed at higher stimulus levels, with their center of gravity shifting to earlier times. The phases of near-CF response components changed by nearly 180 degrees as a function of time. At high stimulus levels, this generated cancellation notches and phase jumps in the frequency spectra. With increases in click level, sharpness of tuning deteriorated and the spectral maximum shifted to lower frequencies. Response phases also changed as a function of increasing stimulus intensity, exhibiting relative lags and leads at frequencies somewhat lower and higher than CF, respectively. In most respects, the magnitude and phase frequency spectra of responses to clicks closely resembled those of responses to tones. Post-mortem responses were similar to *in vivo* responses to very intense clicks. [S0001-4966(98)02504-1]

PACS numbers: 43.64.Kc, 43.64.Ha [RDF]

INTRODUCTION

The basilar membrane of the mammalian cochlea responds nonlinearly to stimulation with tones (e.g., Rhode, 1971; Sellick *et al.*, 1982; Robles *et al.*, 1986; Cooper and Rhode, 1992; Nuttall and Dolan, 1996; Russell and Nilsen, 1997; reviewed by Ruggero, 1992b; and Patuzzi, 1996). Since, for nonlinear systems, the responses to tones cannot generally be used to predict the responses to arbitrary stimuli, a thorough understanding of basilar-membrane behavior requires the use of other stimuli, such as tone complexes, noise and clicks. Clicks are especially useful because, being punctate and wide-band in nature, they permit precise timing of a system's responses while simultaneously testing it over a wide range of frequencies. In the case of linear systems, the Fourier transform of the unit impulse response is identical to the system's transfer function. Departures from such identity in the case of nonlinear systems may provide clues about the nature of the nonlinearities.

The first *in vivo* study of basilar-membrane responses to clicks was carried out in the 8-kHz region of the squirrel monkey cochlea (Rhode and Robles, 1974; Robles *et al.*, 1976). Although hampered by substantial waveform distortion introduced by the Mössbauer technique, this pioneering investigation established that basilar-membrane responses to clicks consisted of a brief and low-frequency initial segment, which grew linearly with stimulus intensity, and a longer-lasting segment with periodicity corresponding to the characteristic frequency (CF) measured using single tones, which grew at compressive rates with stimulus level. Mildly nonlinear growth of basilar-membrane responses to clicks was also measured with a capacitive probe at the basal region of guinea pig cochleae that had been severely traumatized by the experimental procedures (LePage and Johnstone, 1980). More recently, laser methodology and improved surgical techniques have made it possible to obtain undistorted recordings of responses to clicks in the basal region of relatively healthy cochleae of chinchilla (Ruggero and Rich, 1990, 1991a, b; Ruggero *et al.*, 1991, 1992a, b, 1993, 1996) and guinea pig (Nuttall and Dolan, 1993; de Boer and Nuttall, 1997), as well as near the apex of guinea pig and chinchilla cochleae (Cooper and Rhode, 1996). To date, however, reports of these recordings (including those from our laboratory) have been very limited in scope and detail.

The present paper provides an extensive description of basilar-membrane responses to clicks for the 8–10 kHz region of the chinchilla cochlea. At this site, responses to clicks are frequency modulated, exhibiting a low-to-high-frequency glide both *in vivo* and post-mortem. *In vivo*, the responses to clicks grow at compressive rates within 100 μs of response onset, and accurately predict the main features of responses to tones. Taking the frequency glide into account, nonlinear feedback appears to accompany, nearly instantaneously, the CF spectral components of basilar-membrane vibration.

[a]Current address: Department of Physiology, University of Wisconsin, 1300 University Ave., Madison, WI 53706.
[b]Current address: 1193 Liberty Church Road, Mocksville, NC 27028.
[c]Electronic mail: mruggero@nwu.edu

I. METHODS

A. Animal preparation

Male chinchillas, weighing about 0.5 kg, were anesthetized with an initial dose of Ketamine (100 mg/kg, S.C.) and supplementary doses of sodium pentobarbital (I.P.), or with Ketamine (20 mg/kg, I.M.) and Dial (50 mg/kg) in urethane (200 mg/kg, I.P.) and supplementary doses of Dial in urethane. They were tracheotomized and intubated but forced ventilation was rarely used. Core body temperature, measured using a rectal probe, was maintained at 38 °C using a servo-controlled electrical heating pad. The left pinna was resected, the bulla was widely opened, the tensor tympani was cut and the stapedius was detached from its anchoring. A silver-wire electrode was placed on the round window to record compound action potentials (CAPs) evoked by tone bursts. A small hole was made in the basal turn of the otic capsule by first thinning and drilling the bone using a dental bur and then chipping away bone fragments with a metal pick. The hole allowed direct visualization of the basilar membrane and placement on it of a few glass microbeads (10–30 μm in diameter), which served as reflecting targets for the laser beam. In most experiments, the otic-capsule hole was left open. In six experiments, basilar-membrane recordings were made after the hole was covered with a window fashioned from slide coverslip glass (to minimize motion of the perilymph meniscus overlying the recording site; see section A of Discussion). In these experiments, the basilar-membrane recordings were complemented by vibration measurements from the stapes or the incus, near the incudostapedial joint (without using reflecting beads).

B. Acoustic stimulation

Acoustic stimuli were produced by exciting a Beyer DT-48 earphone with electrical signals from a custom-built digital waveform generator (Ruggero and Rich, 1983) or from a commercial system (Tucker-Davis Technologies). Electrical clicks, which had durations of 50-μs or 10-μs (up to and following experiment L131, respectively), were presented with repetition periods of 20–53 ms (usually 25 ms). Single-tone stimuli were modulated at onset and offset by 1/2 period of a raised cosine waveform (1.16 ms rise/fall time). The tone bursts had durations of 5, 10, 25 or, exceptionally, 3 ms and repetition periods of 25, 50, 100 or 15 ms, respectively. At the beginning of each experiment, the probe tip of a calibrated miniature microphone (Knowles 1842 or 1785) was placed within 2 mm of the tympanic membrane. Using this microphone, the transfer function of the electroacoustic stimulus system was measured for 100 Hz–24 kHz tones, with 100-Hz resolution. The amplitude and phase spectra of the transfer function were stored digitally and were later used to compute the acoustic-click waveform by Fourier synthesis. A simulated electrical click was convolved with the transfer function in the frequency domain to obtain the spectrum of the acoustic click. Inverse Fourier transformation of this spectrum yielded the time waveform of the acoustic click. Throughout this paper click levels are given as the peak pressures of the synthesized clicks, expressed relative to 20 μPa.

C. Laser velocimetry

Basilar-membrane vibrations were recorded using a laser velocimeter, which measures the velocity of a vibrating object by detecting the Doppler frequency shift of light reflected from it. In our application, the laser beam was reflected from glass microbeads placed on the basilar membrane. The velocimeter consisted of a 20 mW He–Ne laser (Spectra Physics 106-1), a fiber vibrometer (Dantec 41X60) and a Doppler frequency tracker (Dantec 55n20). The velocimeter was coupled to a compound microscope (Olympus BHMJ) equipped with 5X and 20X ultralong working-distance objectives (Mitutoyo Plan Apo 5X, N.A. 0.14, and 20X, N.A. 0.42). The electrical output of the Doppler frequency tracker, a voltage proportional to the velocity, was filtered with a bandpass frequency response (1–15 000 Hz). The output of the filter was sampled by a computer at a rate of 40 kHz (up to L131) or 100 kHz (after L131).

D. Data analysis

Responses to clicks or tones were averaged over 512, 1024 or 2048 stimulus repetitions. The magnitude and phase spectra of the average-response waveforms were routinely calculated by Fourier transformation using time windows of 12.8 or 10.24 ms (i.e., 512 25-μs bins or 1024 10-μs bins). In addition, short-term Fourier transforms (STFTs; see Fig. 5(C) were computed using 0.8-ms Hanning windows centered at each delay and zero padded to an overall duration of 3.2 ms. Consecutive STFTs were computed at 0.4-ms intervals. The envelope and instantaneous frequency of the responses to clicks were estimated using their analytic signal representation (Bennett, 1970). The analytic signal of a waveform is a complex quantity whose real part is the wave-

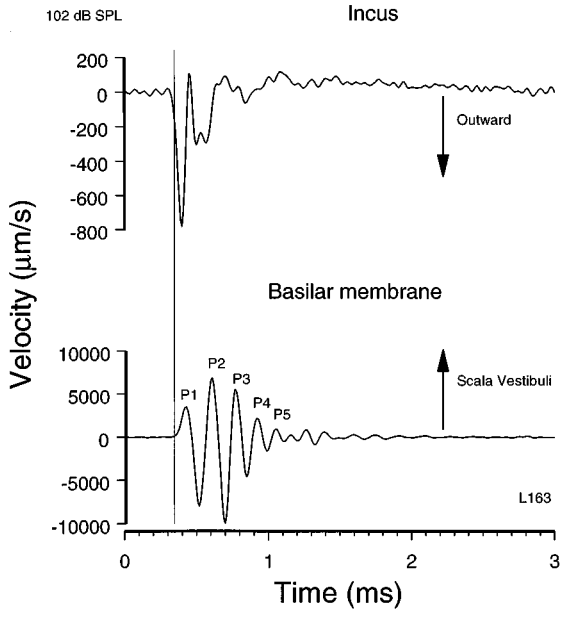


FIG. 1. Responses to intense rarefaction clicks recorded from the incus (top) and a basal site of the basilar membrane (bottom) in a single chinchilla ear. The thin vertical line intersects the incus response at the time it reaches 20% of its maximum value and may be taken as marking the onset of the input to the cochlea. Click peak pressure was 102 dB *re*: 20 μPa.

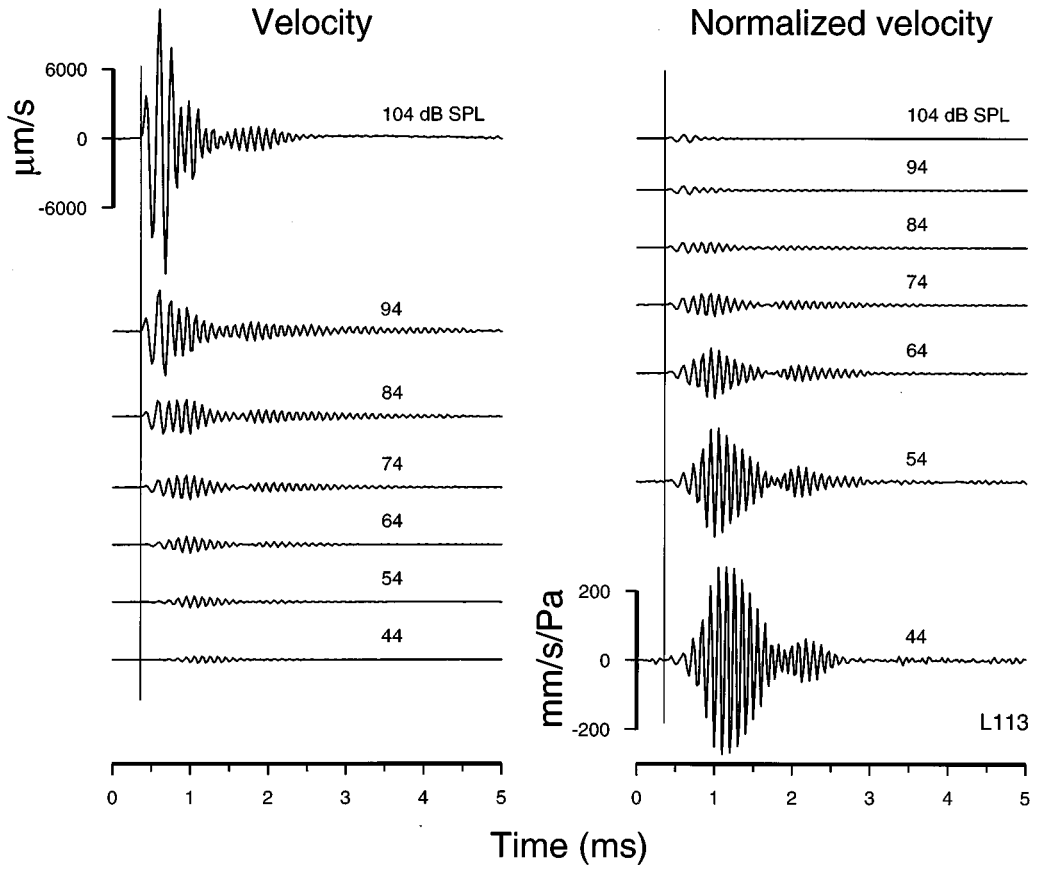


FIG. 2. Basilar-membrane responses to rarefaction clicks presented at several intensities. The responses are displayed with uniform scales of velocity (left column) and normalized velocity (velocity divided by peak click pressure, right column). The thin vertical lines indicate the onset of middle ear ossicular vibration. Response polarity as in Fig. 1. The parameter next to each trace indicates the peak click pressure, expressed relative to 20 μPa.

form itself and whose imaginary part is the Hilbert transform of the real part. The envelope of the waveform is equal to the magnitude of the analytic signal and the instantaneous frequency corresponds to the derivative of the phase of the analytic signal.

II. RESULTS

Basilar-membrane responses to clicks were recorded at a region of the chinchilla cochlea located about 3.5 mm from the oval window. Data from two representative cochleae (L13 and L113) are highlighted throughout the paper. Responses from these cochleae were selected for presentation because: (1) they were exceptionally stable (remaining invariant over several hours of recording); (2) they were collected in near-normal ears (judging from surgically induced CAP threshold elevations at CF of 6–12 dB); (3) they were especially sensitive; and (4) extensive samples of responses to tones in the same cochleae are available for comparison (Ruggero *et al.*, 1997).

A. Main features of basilar-membrane responses to clicks

Figure 1 allows a comparison of velocity responses to identical intense acoustic clicks measured in a single chinchilla ear from the incus (top panel) and from a basilar-membrane site located about 3.5 mm from the oval window (bottom panel). The response of the incus consisted of a short oscillation, consistent with the untuned, wide-band nature of middle-ear vibrations. The basilar-membrane response, much larger (note ordinate scales) and longer lasting than the incus response, was relatively undamped, displaying the characteristic ''ringing'' of a well-tuned bandpass system. In the overwhelming majority of experiments, rarefaction clicks evoked basilar-membrane responses whose first peak (''P1'' in Fig. 1) was in the direction of scala vestibuli. Complementary measurements of incus and basilar-membrane motion, using intense clicks and 100-kHz sampling frequency, were performed in 7 ears. In 6 of these, the basilar-membrane recordings were carried out after covering the otic capsule hole with a glass window. Defining response onset as the time at which oscillations first reached 20% of their maximum value, the cochlear delay of the basilar-membrane response was computed as the interval between the onset of motion at the incus and at the basilar membrane. The cochlear delay measured in these 6 ears was 29.8 ± 12.4 μs (mean$\pm$standard deviation).

Figure 2 presents basilar-membrane responses to clicks plotted with uniform scales of velocity (left column) and velocity normalized to stimulus pressure (right column). If the basilar membrane vibrated linearly, neither the response wave shapes nor their normalized magnitudes would vary as a function of stimulus level. In fact, increases in click level

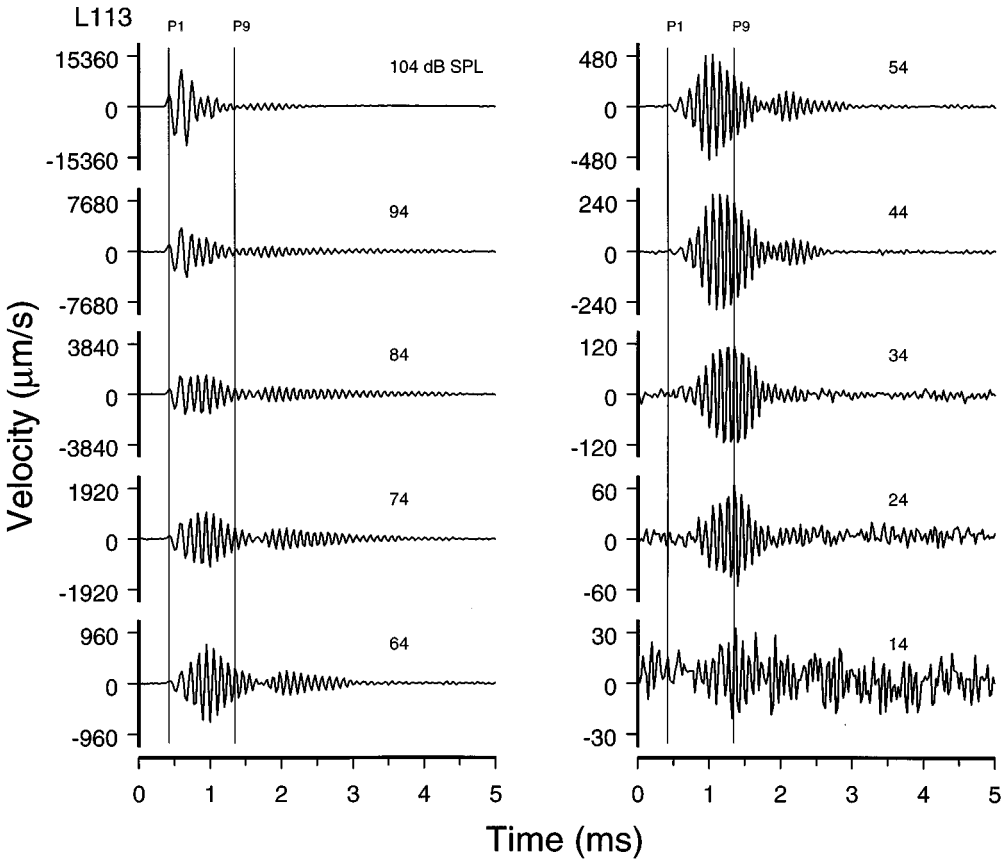


FIG. 3. Basilar-membrane responses to rarefaction clicks. The responses are displayed with scales that are systematically compressed by a ratio of 2 (i.e., 6 dB) for every increment of 10 dB in stimulus intensity. Such scaling, which de-emphasizes intensity-dependent changes in response magnitude, allows for easier comparison of the response wave shapes.

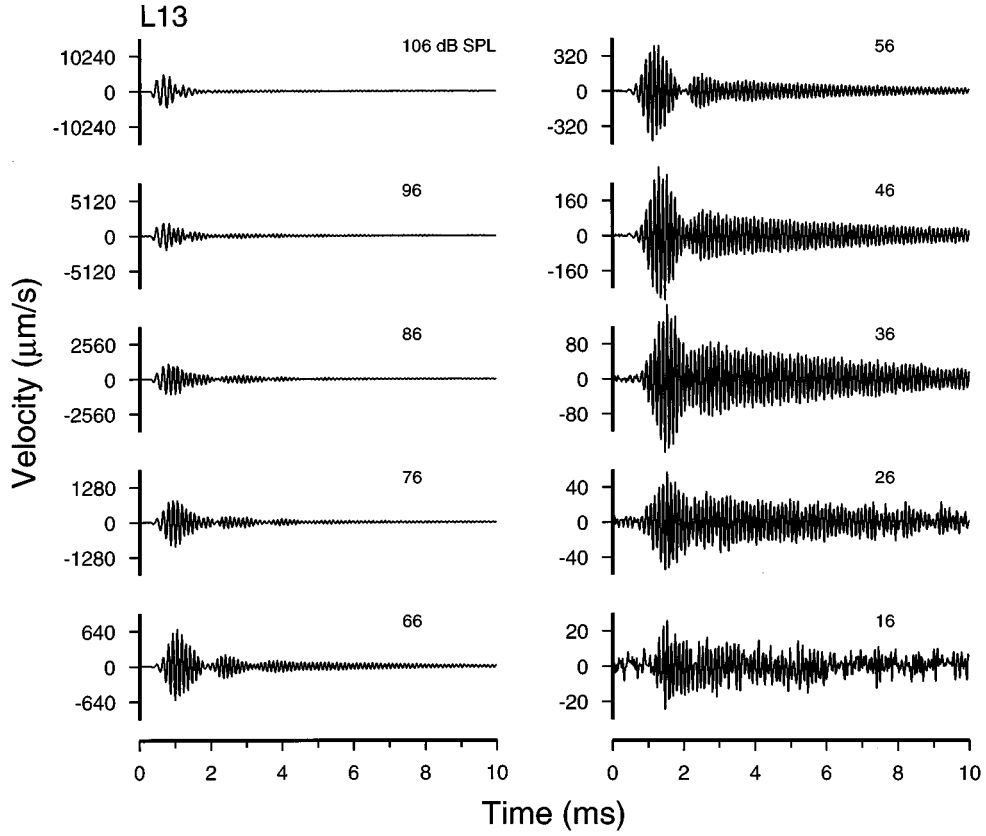


FIG. 4. Basilar-membrane responses to condensation clicks. All details as for Fig. 3.

were accompanied both by changes in response wave shape and by decreases in normalized velocity (right column). Compressive growth was obvious at all stimulus levels but it became less evident at the very highest levels.

Figure 3 shows a more complete series of basilar-membrane responses to rarefaction clicks (from the same cochlea of Fig. 2) with scales chosen to facilitate comparison of the wave shapes without regard to their magnitudes: the scales are compressed by a factor of 2 (6 dB) for every 10-dB increase of stimulus intensity. For most click intensities, the envelopes of the responses consisted of two adjacent spindle-shaped lobes. This ''two-lobe'' wave shape was characteristic of cochleae that were in good physiological state, it was absent in cochleae that were surgically traumatized and disappeared after death (Fig. 13) or acoustic overstimulation (Figs. 6 and 8 of Ruggero *et al.*, 1996). The early response peaks grew with stimulus intensity at faster rates than later ones, giving rise to a systematic skewing of the envelope toward earlier times. For example, at low stimulus levels (14–24 dB SPL), the response maximum corresponded to the ninth positive peak (P9) but the maximum shifted to earlier peaks with increases of stimulus level: P8 at 34 dB, P6 at 54 dB, and P2 at 94 and 104 dB. The time elapsed between the response onset (0.38 ms) and the maximum of the response envelope decreased from 0.97 ms for 24-dB clicks to 0.25 ms for 104-dB clicks.

Figure 4 shows responses to condensation clicks recorded in an exceptionally sensitive preparation. Apart from the long-lasting ringing, with duration longer than in any other cochlea in our sample, these responses share the same characteristics observed in Figs. 1–3, including initial response polarity (i.e., responses to condensation and rarefaction clicks, respectively, begin with motion toward scala tympani and scala vestibuli), compressive growth, and skewing of the envelope toward earlier times as a function of increasing stimulus level.

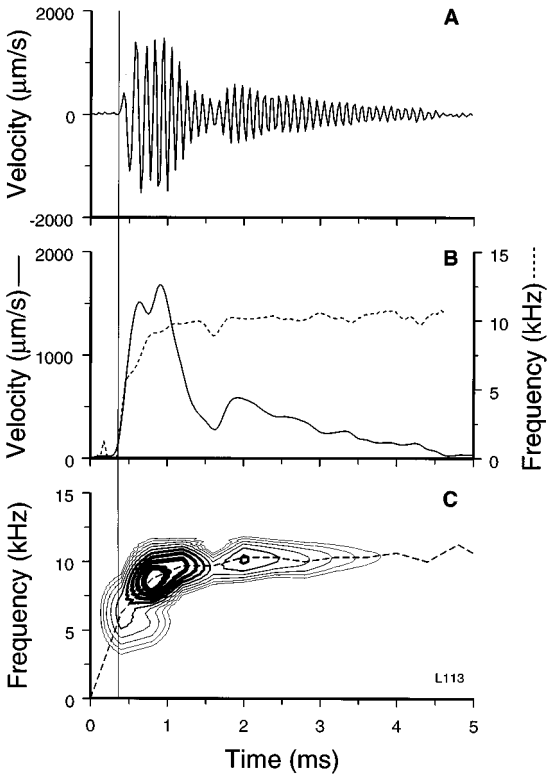


FIG. 5. Instantaneous frequency and envelope of basilar-membrane responses to 84-dB clicks. (A) Original response waveform. (B) Envelope (continuous line) and instantaneous frequency (dashed line) of the waveform depicted in A. (C) Magnitude of short-time Fourier transforms (STFTs; see Methods) of the same waveform, displayed as a function of time (abscissa) and frequency (ordinate). The thickness of the contour lines indicates the STFT amplitude, expressed relative to the ensemble maximum. The thickest line indicates a relative amplitude of −2 dB and the thinner lines correspond to lower levels, in steps of −2 dB. The frequency of the STFT maximum is also shown as a function of time (dashed line). The thin vertical line indicates the onset of middle-ear ossicular vibration.

B. Frequency modulation in basilar-membrane responses to clicks

Close inspection of the waveforms of responses to clicks (e.g., Fig. 2, 104 dB SPL) reveals that their periodicity or instantaneous frequency changed as a function of time. To quantify this frequency modulation, we computed the instantaneous frequency and the envelope of the response waveforms using the ''analytic signal'' representation (Bennett, 1970; see Methods). Figure 5 shows the response to a click presented at 84 dB SPL (panel A), as well as its envelope (continuous line, panel B) and its instantaneous frequency (dashed line, panel B), plotted against time. The instantaneous frequency increased rapidly within several hundreds of microseconds: it was initially about 1 kHz, surpassed 5 kHz by the time of the first positive peak and saturated at about 10 kHz. An exponential fit had a time constant of 0.2 ms.

The variation of frequency with time was also measured using short-term Fourier transforms (STFTs; see Methods). Panel C of Fig. 5 displays the magnitude part of the STFTs for the response of panel A as a family of contour lines spaced at 2-dB intervals relative to the ensemble peak, with the thicker contour lines indicating higher spectral magnitudes. Also indicated is the frequency at which the peak magnitude of the STFT occurs as a function of time (dashed line). Both the STFT contours and the peak instantaneous frequency show that the frequency content glides from low to high ($\approx$CF). The increase in response frequency was accompanied by a narrowing of the bandwidth. This may be ascertained from the STFT contours of Fig. 5C by measuring the frequency range encompassed by a fixed number of contour lines along any given time ''slice.'' The equivalent rectangular bandwidth and the 10- and 20-dB bandwidths were 3.9, 4.0 and 8.3 kHz, respectively, at 0.8 ms and decreased to 2.8, 3.3, and 4.4 kHz at 2 ms.

Instantaneous frequency was influenced by stimulus level but its time trajectory retained its main features even at the highest levels and post-mortem (Fig. 6A). Thus, it is clear that the frequency modulation is not a byproduct of nonlinear or ''active'' cochlear processing. In contrast, the response envelopes were highly dependent on stimulus intensity. Figure 6B shows the envelopes of responses to clicks with peak pressure of 64-, 84- and 104-dB, normalized to 104-dB. Had responses grown linearly, the normalized envelopes would be identical throughout their extent. In fact, due

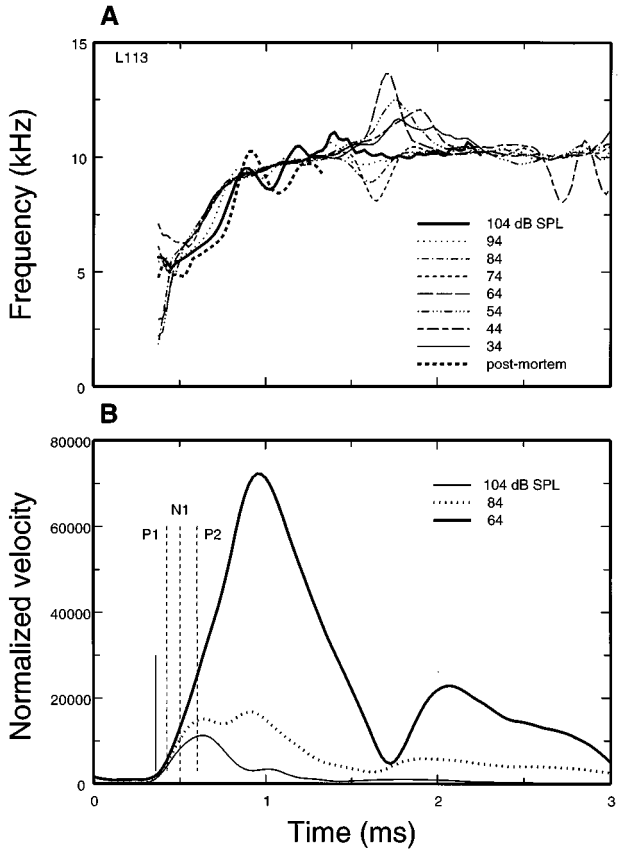


FIG. 6. Instantaneous frequency (A) and envelope (B) of basilar-membrane responses to clicks as a function of time and stimulus level. The thick dashed line in panel A corresponds to responses recorded post-mortem (94-dB clicks). The vertical continuous line in panel B indicates the onset time of middle-ear vibration. The vertical dashed lines indicate the times of the first and second positive peaks and the first negative peak. The ordinate in panel B has units of micrometer/s and are normalized to 104-dB (i.e., the velocities of responses to 84- and 64-dB clicks are magnified 10 and 100 times, respectively). Data are those represented in Figs. 2 and 3.

to the compressive growth of responses, the envelopes coincide only at the onset (where responses grow linearly).

C. Growth of responses as a function of click intensity

The initial peaks of responses to clicks grew with intensity at faster rates than later peaks (Figs. 2–4). To quantify these rates, peak times were determined in the responses to high-intensity clicks and the corresponding magnitudes (measured in the response envelopes) were plotted against click intensity (Fig. 7). With the exception of animal L13, the magnitudes measured were those corresponding to velocity maxima toward scala vestibuli (i.e., positive peaks in responses to rarefaction clicks). At low levels of stimulation (peak pressure: <30–40 dB *re*: 20 μPa), all peaks grew linearly or nearly so. The initial oscillation, P1, grew in an almost linear fashion at all levels. Peaks 3–7 grew compressively at intermediate levels but tended to become linear again at the highest levels of stimulation. Later peaks grew nonlinearly (at rates as low as 0–0.2 dB/dB) throughout the range of moderate and high stimulus levels.

To further explore the development of compressive nonlinearity as a function of time, we computed the slopes of the velocity-intensity functions for the positive peaks and averaged the slopes over 50-dB ranges encompassing the highest stimulus levels. Figure 8 displays such averages for the four cochleae represented in Fig. 7. Nonlinear growth was evident within 100 μs (i.e., 1 CF period) of response onset (vertical line). In a sample of 8 sensitive cochlea (including the 4 of Fig. 8) the first positive and negative peaks, respectively, grew at rates of 0.950 ± 0.035 and 0.883 ± 0.053 dB/dB (over 40-dB ranges). Such slopes differed significantly from 1 dB/dB ($p<0.005$, one-tailed t test).

Slopes diminished rapidly as a function of time, from values approaching linearity immediately after response onset, to minima as low as 0.1–0.2 dB/dB at 1.3–1.9 ms. At later times, growth slopes waxed and waned and exhibited local maxima as high as 0.5–0.6 dB/dB and minima as low as 0.1 dB/dB. Inspection of the curves for cochleae L13 and L113 in Fig. 8 indicates that the slope maxima (i.e., compression minima) were approximately synchronous with the constrictions that demarcate adjacent lobes of the time-domain response waveforms (Figs. 3 and 4). This was also true for the other cochleae represented in Fig. 8.

D. The magnitude-frequency spectra of basilar-membrane responses to clicks

Figure 9 shows the magnitude of the Fourier transforms of the waveforms depicted in Figs. 3 and 4 (computed using 12.8-ms windows). At low spectral frequencies, the responses grew linearly: for 10-dB increments in click level, the responses grew by a factor of 3.1 (10 dB). At frequencies around CF, response growth was linear at the lowest stimulus levels but quite compressive at high intensities. The response bandwidth changed systematically with click intensity: for low stimulus levels, responses were largely confined to frequencies near CF, whereas at high stimulus levels responses encompassed a broad range of frequencies. The change in bandwidth was accompanied by an overall frequency shift toward lower frequencies. At the highest click intensities, the frequency of the spectral maximum was about -0.5 octave relative to CF. The Q_{10} (peak frequency divided by 10-dB bandwidth) decreased substantially as a function of click intensity. In cochlea L113, for example, Q_{10} was 4.06 for 44-dB clicks and 1.08 for 104-dB clicks. At frequencies near CF, some of the frequency spectra included sharp notches. These were seen only in the most sensitive and nonlinear preparations, and disappeared as the state of the preparation deteriorated with the passage of time, after death [note notches at 8.48 kHz in *in vivo* responses of L13 (Figs. 9 and 10) and their absence post-mortem (Fig. 10)] or after acoustic overstimulation (Fig. 9 of Ruggero *et al.*, 1996).

The intensity-dependent nonlinear growth of responses is best visualized by normalizing the spectra of Fig. 9, frequency by frequency, to stimulus pressure (using the acoustic calibration tables; see Methods), thus producing estimates of basilar-membrane gain (Fig. 10). [In a linear system, such a procedure eliminates entirely the effects of irregularities in the stimulus spectrum. The usefulness of the correction is less certain in the case of a nonlinear system.] For clarity, the gains have been smoothed using a 3-point running average. At frequencies lower than 1/2 octave below CF, the curves coincided, indicating linear growth. Nonlinear behavior was

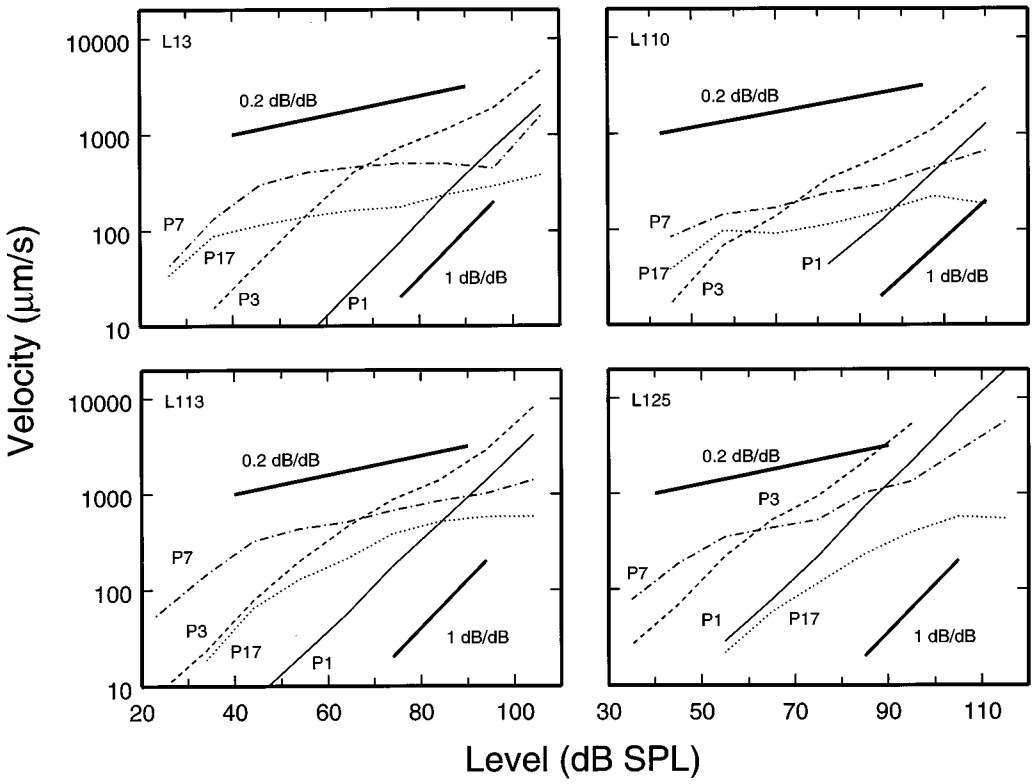


FIG. 7. The magnitudes of selected positive peaks of the responses to clicks, plotted against stimulus level. Also indicated are lines with slopes of 1 dB/dB and 0.2 dB/dB. Each panel represents data from one sensitive cochlea.

evident at higher frequencies (>6 kHz), specially near CF (9 or 10 kHz), where gains became smaller with increases of stimulus level. Compressive growth persisted even at the highest levels of stimulation (compare the curves for 104 or 106 dB with those for 94 or 96 dB).

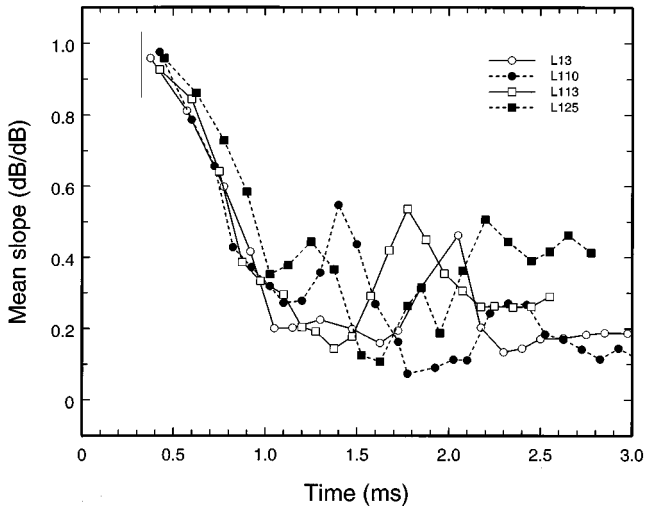


FIG. 8. Average slopes of the magnitude-intensity functions for the responses to clicks of 4 sensitive cochleae. The average slopes of curves such as those depicted in Fig. 7 were computed over 50-dB ranges (54–65 to 103–115 dB SPL). The thin vertical line indicates the onset time of basilar-membrane vibration (325 μs, averaged across the 4 cochleae).

E. The phase-frequency spectra of basilar-membrane responses to clicks

Figure 11A displays the phases of responses to clicks in one cochlea, obtained by Fourier transformation of some of the waveforms of Fig. 3. The curves represent the phases of basilar-membrane displacement toward scala tympani relative to maximum condensation at the eardrum. The curves show a monotonically increasing phase lag, interrupted in some cases by abrupt shifts at frequencies corresponding to notches in the magnitude spectrum (Fig. 9A). Slopes were shallower at frequencies lower than 7 kHz than at higher frequencies and were steepest near 11 kHz, a frequency somewhat higher than the CF (estimated from responses to low-level clicks). [Slopes were steepest at CF only in the cases in which there were abrupt phase changes near this frequency (64–94 dB; see Fig. 13B).] In the range 3–7 kHz, the average slope was -250 μs. In the 8–10 kHz region, the slopes varied between about -560 μs for the responses to clicks presented at 104 dB SPL (both *in vivo* and post-mortem) and -826 μs for 24-dB clicks (not illustrated).

Figure 11B shows a representative sample of phase-vs-frequency curves, normalized to inward stapes displacement, for responses to clicks recorded in several cochleae. The basilar-membrane sites had CFs in the range 8–12 kHz and the clicks were presented at comparable levels (82–88 dB SPL). The low-frequency segments (<6–7 kHz) had slopes confined to a relatively narrow range, -137 to -269 μs, irrespective of CF. The high-frequency segments diverged

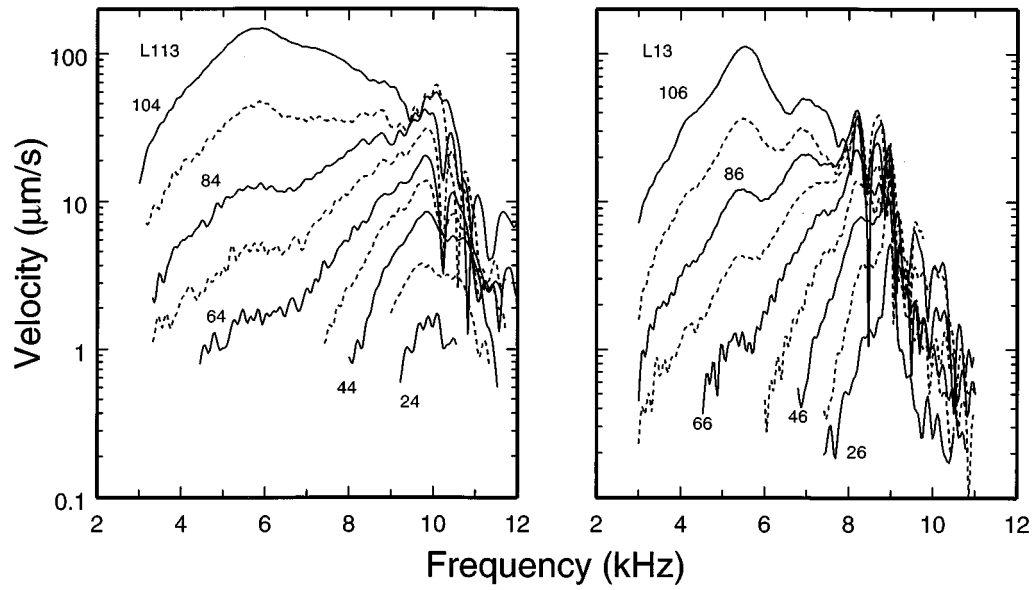


FIG. 9. Magnitude-frequency spectra for the velocity responses to clicks in two sensitive cochleae. Clicks were presented at levels of 24 or 26 dB up to 104 or 106 dB, in steps of 10 dB. Data are those represented in Figs. 2 and 3.

increasingly at frequencies higher than 1/2 octave below CF. In general, the phase-frequency slopes were steepest near CF. At CF, the phase lags amounted to 1.5–2 cycles. Average group delays were measured over 1-kHz ranges centered at CF in responses to low-level clicks. [Phase-vs-frequency curves for responses to more intense clicks (such as those of Fig. 11B) included irregularities at frequencies near CF which rendered the measurement of group delays unreliable.] The cochleae represented in Fig. 11B had near-CF group delays averaging 779±147 μs (mean±s.d.). Group delays measured between 7 kHz and 1 kHz below CF (or somewhat smaller ranges if limited by the noise floor) averaged 410±109 μs.

The dependence of response phases on stimulus intensity (shown in Fig. 15B) re-plots the data of Fig. 11A after normalization to the phases of responses to 64-dB clicks. For low frequencies (<5–6 kHz) the phases varied little with stimulus intensity, consistent with the linearity of response growth in that frequency region. For frequencies (6–9 kHz) just below CF, the phases of responses exhibited increasing lags as a function of increasing click level. The lags were small except at levels higher than 74 dB. For frequencies just

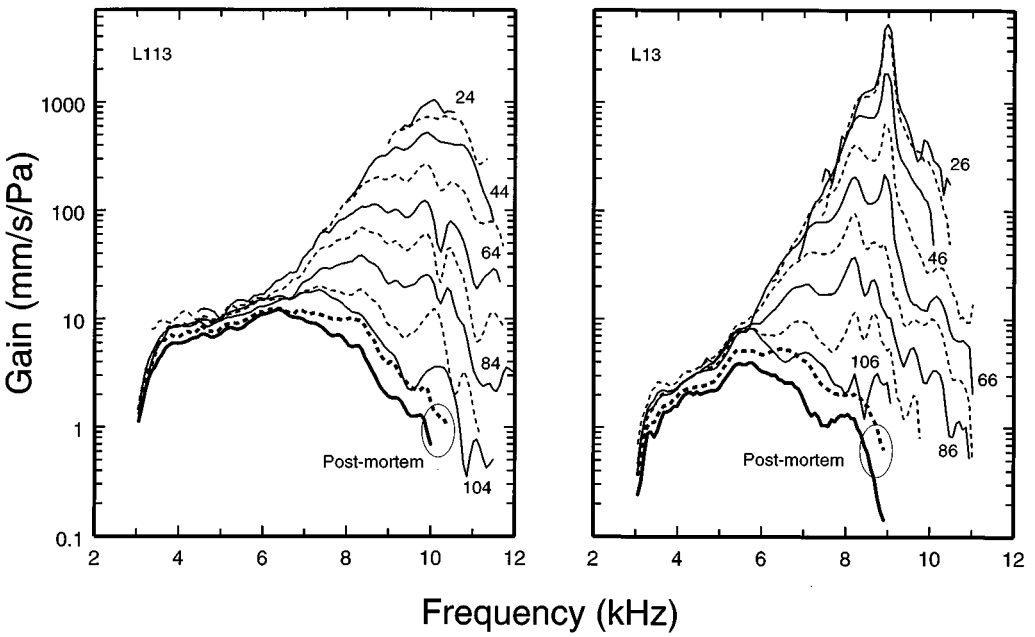


FIG. 10. Gain-frequency spectra for the responses to clicks in two cochleae. Gains were computed by dividing, frequency by frequency, velocity magnitudes (shown in Fig. 9) by the peak click pressure (parameter). The thick lines indicate the gains of responses recorded 10–20 minutes postmortem; dashed lines: 94 dB (L113) and 96 dB (L13); continuous lines: 104 dB (L113) and 106 dB (L13). Curves were smoothed using a 3-point running average.

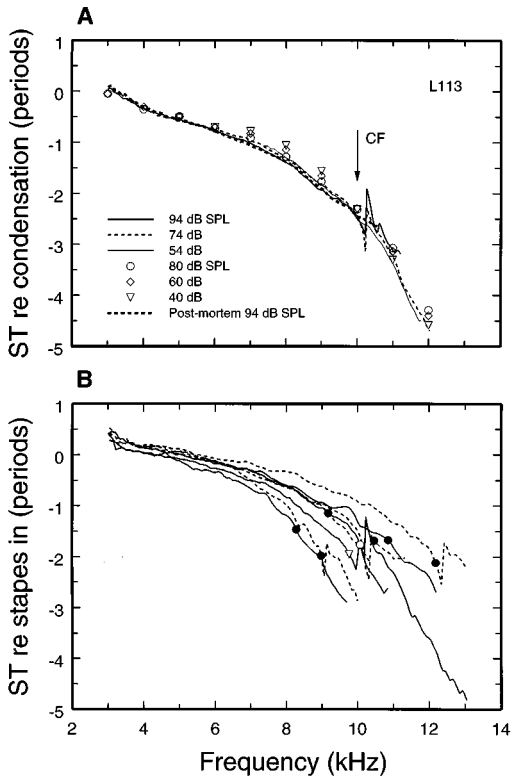


FIG. 11. Phase-vs-frequency curves for basilar-membrane responses to clicks. (A) The ordinate indicates the phases of displacement toward scala tympani (ST) referred to peak positive pressure (condensation) at the eardrum. Phases were obtained by Fourier transformation of the data depicted in Figs. 2 and 3 (cochlea L113). The thick dashed line indicates postmortem responses to intense clicks (94-dB SPL). The symbols indicate the phases of responses to tones. (B) Phase-vs-frequency curves for basilar-membrane responses to clicks in 8 cochleae, including L113 (identified by an open circle). The symbols indicate phases at CF (measured in responses to low-level clicks). The phase curves are presented relative to inward stapes motion. In one case (L163, indicated by a triangle), basilar-membrane and stapes response phases were measured in the same ear; in the other cases, basilar-membrane phases were normalized to previously published average stapes data for chinchilla (Fig. 9 of Ruggero *et al.*, 1990). Clicks had peak pressures of 82–88 dB.

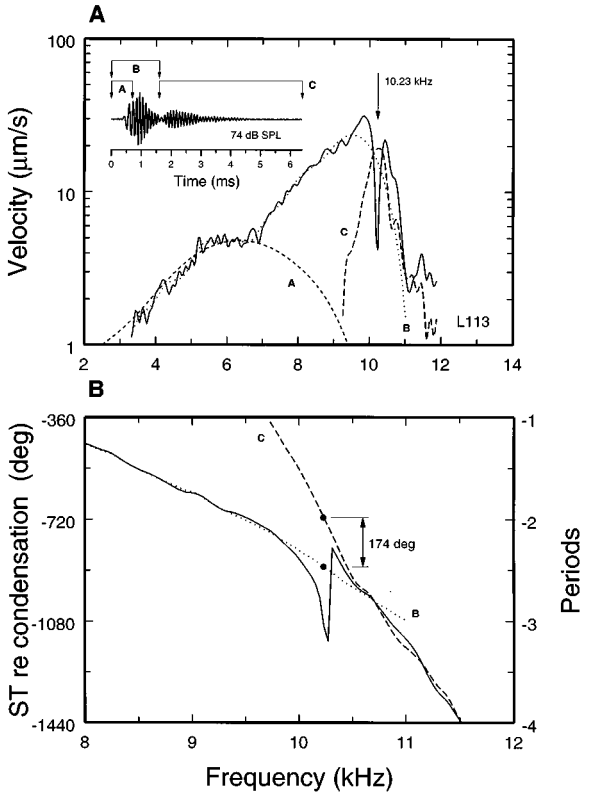


FIG. 12. Time-dependent variation of the magnitude- and phase-frequency spectra of responses to clicks. Top panel: magnitude spectra. The continuous line shows the variation of magnitude as a function of frequency for the entire waveform (6.4 ms) of averaged responses to 74-dB clicks in cochlea L113 (inset). The other curves depict the magnitude-frequency spectra of the waveform segments indicated in the inset (A: 0–0.7 ms, short-dash line; B: 0–1.7 ms, dotted-line; C: 1.7–6.4 ms, long-dash line). Bottom panel: phase spectra. The continuous line indicates the variation of phase with frequency for the entire waveform. The dotted and long-dash lines depicts phases for segments B and C, respectively.

higher than CF, the response phases showed increasing leads as a function of increasing intensity. A similar intensity dependence of phase (lags and leads, respectively, for frequencies lower and higher than CF, as a function of increasing intensity) could be usually demonstrated in the responses of all sensitive cochleae.

F. Relation between time-domain lobes and magnitude notches and phase jumps in the frequency domain

To explore the spectral notches observed in the CF region (Fig. 9), as well as the associated phase jumps (Fig. 11), the responses to clicks were studied using Fourier analysis of selected time windows. Figure 12 shows such an analysis for the response waveform shown in the inset. The waveform was divided into two nonoverlapping segments (B and C, inset in Fig. 12A). Segment B spans the first 1.6 ms of the response and extends to the transition between the two response lobes. Segment C, encompassing the second lobe, starts at 1.6 ms and ends at 6.4 ms. The magnitude spectra for these two segments are shown in Fig. 12A (B: dotted line; C: long-dash line), together with the spectrum of the entire original waveform (solid line). The magnitude spectrum of segment C (long-dash line) is sharply tuned to frequencies centered at 10.234 kHz, the frequency of the spectral notch of the original waveform. The notch is absent from segments B and C, which suggests that its presence in the original waveform resulted from destructive interference between early and late out-of-phase spectral components. Figures 12B and 13 illustrate this.

Figure 12B shows the origin of the 360-degree phase jumps (Fig. 11). At frequencies lower than that of the notch, the phase-vs-frequency curve of the entire response was dominated by the phase spectrum of segment B, whereas at higher frequencies it was dominated by the phase spectrum of segment C. At the notch frequency, the phases of segments B and C differed by 174 degrees (i.e., they were essentially in phase opposition). Figure 13 demonstrates that the response phases at the notch frequency exhibited a monotonically increasing lag in the time interval of 1–2 ms (i.e., straddling the constriction between the two lobes; indicated by the bracket in Fig. 13). Phases became stable within the second lobe, after accumulating a phase lag of nearly 180

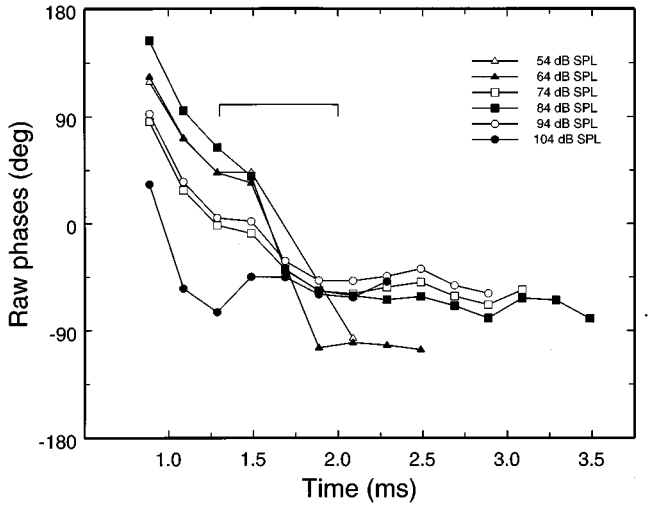


FIG. 13. The variation of response phase at the notch frequency (10.23 kHz) as a function of time. Phases were computed by Fourier transformation of 200-μs segments of the responses to clicks of cochlea L113, including the waveform of the inset of Fig. 12A. The bracket indicates the times of the transitions between the time-domain lobes (see Fig. 3).

degrees. Thus, it is clear that the notch in the magnitude spectrum was generated by a phase cancellation at the notch frequency between response components in the two lobes.

Figure 12A also shows the amplitude spectrum of segment A (short-dash line), which indicates that most of the low-frequency (<7 kHz) response components of the original waveform are confined to the response onset (consistent with the analytic-signal representation and the short-term Fourier transform: Fig. 5).

G. Effects of death on basilar-membrane responses to clicks

Figure 14 illustrates the effects of death on responses to clicks presented at 84 and 104 dB SPL. [The scale for the responses to 104-dB clicks is compressed by a factor of 10 (i.e., 20 dB) relative to the scale for responses to 84-dB clicks. Therefore, response features that grew linearly have identical magnitudes in the 84- and 104-dB traces.] When the cochlea was healthy (left column), all but the earliest peaks grew at highly compressive rates, so that wave shapes were skewed toward earlier times and gains became smaller as a function of increasing stimulus intensity. Post-mortem, the amplitude of all but the earliest response peaks were drastically reduced relative to the *in vivo* responses, with the responses to the weaker stimuli being more strongly affected (right column of Fig. 14; see also Figs. 7 and 8 of Ruggero and Rich, 1991a, and Figs. 1 and 3 of Ruggero *et al.*, 1992a). After death, responses grew almost linearly, so that responses at all stimulus levels resembled scaled versions of a single wave shape, similar to that of *in vivo* responses to intense clicks. However, post-mortem responses (recorded within 10–20 minutes after death) were not completely stable and exhibited vestiges of nonlinearity (note the late oscillations in the traces of Fig. 14 and the post-mortem magnitude spectra of Fig. 10).

Post-mortem, the response gains (thick lines in Fig. 10)

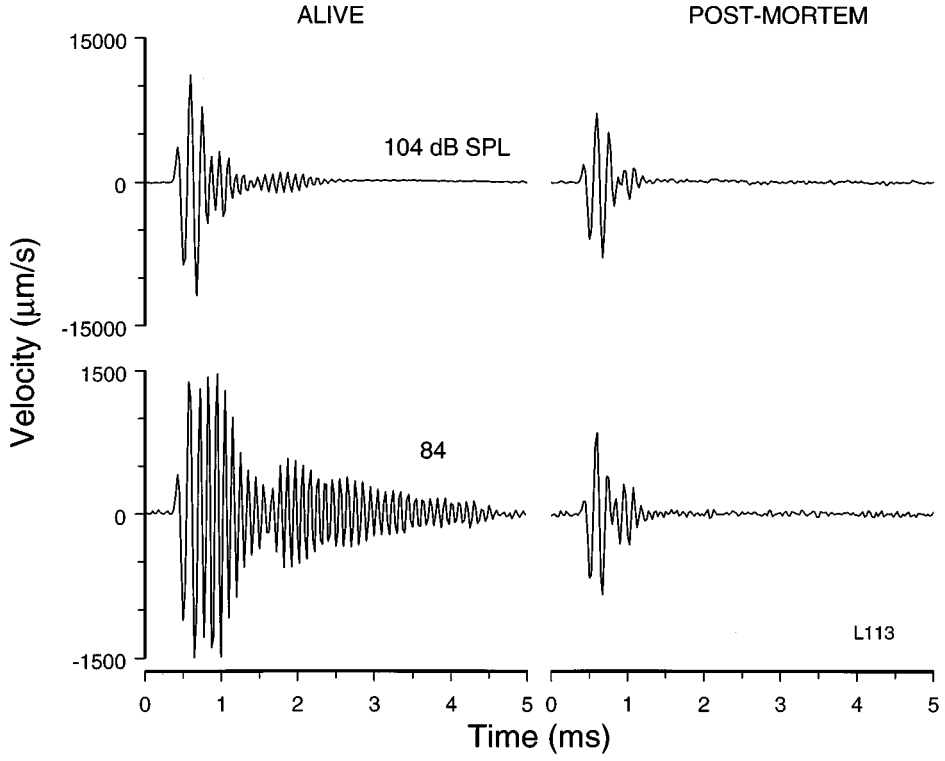


FIG. 14. The effect of death on basilar-membrane responses to clicks. Responses to clicks, presented at 84 and 104 dB SPL, were recorded *in vivo* (left column) and about 10 minutes post-mortem (right column).

were reduced by 3–4 orders of magnitude at CF, relative to *in vivo* responses to low-level clicks (by 58–64 dB in L113 and 78–91 dB in L13), the bandwidths were broadened and the peaks of the gain spectra shifted down to frequencies about 0.5 octave lower than CF. Post-mortem, the Q_{10}'s (peak frequency divided by bandwidth) were reduced to 1.12–1.21 (L113) and 1.42–1.53 (L13). Overall, the post-mortem gain spectra were very similar to (but somewhat lower than) the spectra of *in vivo* responses to intense (104- or 106-dB) clicks.

In relation to *in vivo* responses to low-level clicks, post-mortem responses exhibited phase lags and leads, respectively, at frequencies lower and higher than CF (thick-dash lines in Figs. 11 and 15B; see also Fig. 2 of Ruggero, 1994). Death abolished the phase jumps evident *in vivo* at near-CF frequencies (compare *in vivo* and post-mortem responses to 94-dB clicks in Figs. 11A and 15B). The phase effects of death could not be assessed with any precision at frequencies higher than CF, at which responses were buried in the background noise.

H. Comparison of basilar-membrane responses to clicks and tones

As described in the preceding sections, the features of basilar-membrane responses to clicks varied systematically as a function of time: (1) the initial oscillation was nearly linear, whereas later oscillation peaks grew at compressive rates that waxed and waned as a function of time (Figs. 7 and 8); (2) the instantaneous frequency increased (Figs. 5–6 and 12A); and (3) the phases of near-CF components shifted by nearly 180 degrees (Figs. 12B and 13). We investigated how these time-varying features of basilar-membrane vibration relate to steady-state behavior by recording responses to tones and clicks in the same cochleae in close temporal proximity.

Responses to clicks and tones were compared in the frequency domain by plotting their gains and phases as a function of frequency and stimulus level (Fig. 15). There was an excellent match between the gains (Fig. 15A) of responses to clicks (replotted from Fig. 10) and to tones (from Fig. 9 of Ruggero *et al.*, 1997) not only at low frequencies (below 6–7 kHz), where responses were linear, but also at frequencies around CF, where responses were strongly compressive. Figure 15A is representative of many other comparisons (not illustrated) in that they demonstrate a close correspondence between the gains of basilar-membrane responses to tones and clicks at the 3.5-mm site of the chinchilla cochlea (see also Fig. 3 of Ruggero *et al.*, 1992a and Fig. 6 of Ruggero *et al.*, 1992b). Such close correspondence was evident regardless of the health of the cochlea and/or the strength of the basilar-membrane nonlinearities and could be demonstrated whenever responses to both types of stimulus (click or tones) were recorded over a wide range of stimulus intensity. For the data of Fig. 15A, for example, the magnitude spectrum of responses to clicks with peak pressure of 104 dB matched accurately the magnitude of responses to 90-dB tones; responses to 94-dB clicks matched responses to 80-dB tones; and so on down to 24- and 34-dB clicks, which matched responses to 10- and 20-dB tones. In other words, over wide ranges of frequency and level, the spectrum of responses to clicks presented at a given level predicted accurately the responses to tones presented at a constant level. Furthermore, once the appropriate correspondence was ascertained for one click level (for example: in the case of cochlea L113, 60-dB tones corresponded to 74-dB clicks), all other levels were simultaneously determined.

Figures 11A and 15B allow comparison of the phases of responses to clicks (lines) and tones (symbols) recorded in the same cochlea. As in the case of the gain functions (Fig. 15A), there is a good match between the response phases for the two types of stimulus for frequencies lower than CF. The response phases for click and tone stimuli (lines and symbols, respectively, in Figs. 11A and 15B) vary as a function of increasing stimulus intensity in a qualitatively similar fashion. Phases tend to lag for frequencies (e.g., 8.5–9 kHz) lower than CF and to lead for frequencies (e.g., 10.3–10.8 kHz) higher than CF. There is little phase variation at CF or at frequencies lower than 5–6 kHz. Comparable matches between the phases of responses to tones and clicks were also obtained for other cochleae (not illustrated here; see Fig. 4 of Ruggero *et al.*, 1992a).

III. DISCUSSION

A. Methodological considerations

In analyzing basilar-membrane recordings using laser velocimetry, it is important to take into account an artifact due to stapes-driven motion of the fluid meniscus overlying the recording site (Cooper and Rhode, 1992). Such motion changes the effective path length of the laser beam and mimics the Doppler shifts due to basilar-membrane motion. The artifact principally affects low-frequency measurements, when ossicular motion is largest and basilar-membrane motion is small (Cooper and Rhode, 1992; Ruggero *et al.*, 1997).

We were especially concerned that motion of the fluid meniscus could introduce a spurious low-frequency oscillation, mimicking an early-onset basilar-membrane response to clicks. To control for this possibility, basilar-membrane responses were measured in 6 cochleae before and after placement of a small window (made up of slide coverslip glass) on the otic-capsule hole overlying the basilar-membrane recording site, a procedure that diminishes the motion of the fluid meniscus (Cooper and Rhode, 1996). [For illustration purposes, we present here responses to intense clicks, for which near-CF basilar-membrane response gains are lowest (e.g., Fig. 10) and therefore the effects of fluid-meniscus motion (which grow linearly with stimulus intensity) are relatively large. In the case of low-level stimulation, the effects of fluid-meniscus motion are negligible.] The window altered the recorded waveforms consistently (Figs. 16–17). In 5 of the 6 cochleae, the initial oscillation (as usual, toward scala vestibuli for rarefaction clicks) grew larger after placement of the glass window (e.g., Fig. 16, left panel). In one exceptional cochlea (Fig. 16, right panel) the initial oscillation was originally negative and became positive after the glass window was installed. Significantly, the onset delays were only minimally changed. Placement of the glass win-

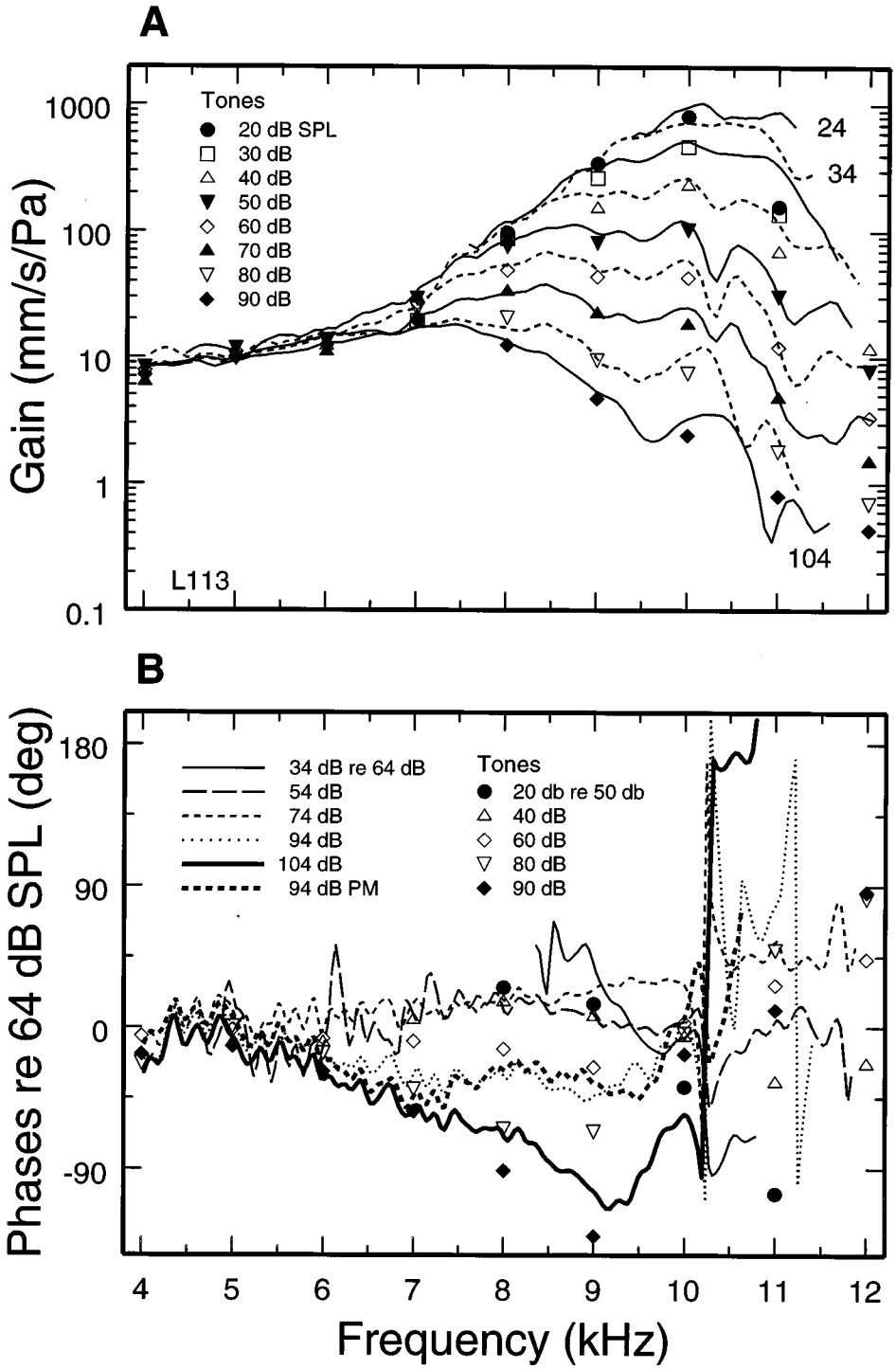


FIG. 15. Comparison of basilar-membrane responses to clicks (curves) and tones (symbols) recorded in the same cochlea. (A) Gain-vs-frequency spectra as a function of stimulus intensity. (B) Phase-vs-frequency spectra as a function of stimulus intensity. At each level, the phases responses to clicks are expressed relative to the phases of responses to 64-dB clicks. The phases of responses to tones are expressed relative to the phases of responses to 50-dB tones, whose magnitudes closely matched the spectral responses to 64-dB clicks (panel A). Positive values indicate relative phase leads

dow reduced the cochlear delays (relative to the onset of incus motion) from 43.0±21.1 μs to 29.8±12.4 μs (mean ±s.d., $N=6$). Cochlea L163 was exceptional in that the delay *increased*, from 14 to 36 μs.

The effects of covering the otic-capsule hole were especially obvious in the frequency domain (Fig. 17). The upper panels of Fig. 17, depicting the magnitude spectra of the waveforms of Fig. 16, show that the response components above 2–3 kHz were largely unaltered but those at lower frequencies decreased in magnitude by as much as 20 dB. Correspondingly, after the placement of the window response phases (lower panel of Fig. 17) increasingly lagged (by as much as 90 degrees) those measured with the open otic capsule as the frequency decreased below 3–4 kHz. These results indicate that fluid-meniscus motion typically does not alter drastically the polarity, magnitude or onset

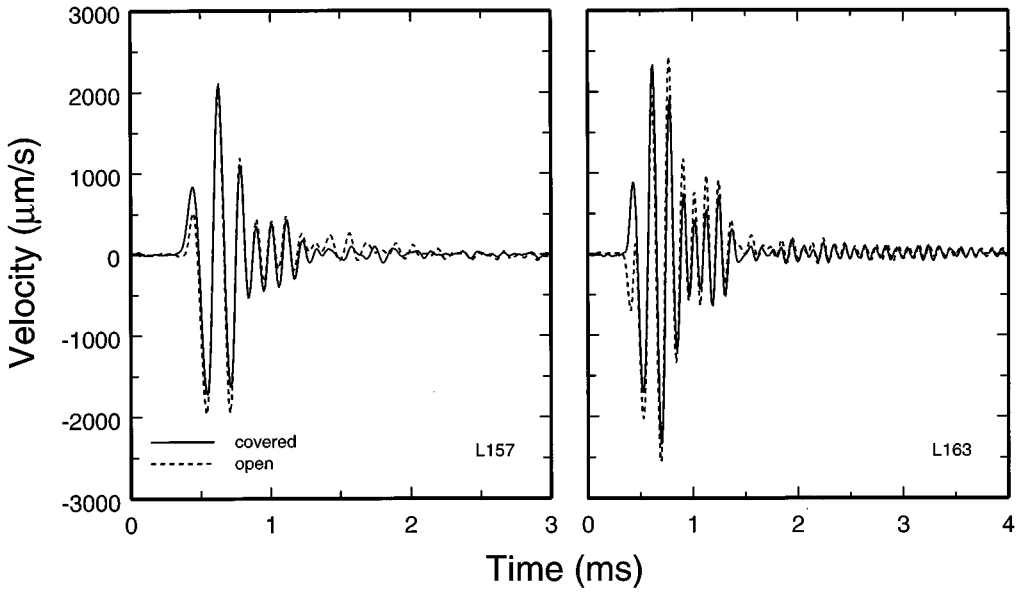


FIG. 16. Basilar-membrane responses to clicks recorded in open and closed cochleae. The dashed and continuous lines, respectively, identify responses recorded before and after covering the otic-capsule hole with a glass window. The clicks had peak pressures of 78 dB (L157) and 92 dB (L163).

time of the initial oscillation of responses to clicks, whose dominant spectral components have frequencies higher than 2–3 kHz (Figs. 5–6 and 12A).

In the course of measuring responses to clicks at the apex of the cochlea, Cooper and Rhode (1996) found evidence for another artifact, also related to the opening of the otic capsule but still present after placement of a glass window. This artifact, which produced an initial oscillation lasting 1–1.5 ms and a spectral magnitude notch, was abolished by restoring the hydraulic seal of the cochlea. [In our preparations, placement of a glass window on the otic capsule hole also failed to abolish the spectral notches (not illustrated) but *increased* the magnitude of the initial oscillation (Fig. 16).] We did not attempt to fully restore the hydraulic seal and, therefore, we cannot rule out that our measurements were affected by a similar artifact. Nevertheless, we deem this an unlikely possibility. The apical preparation of Cooper and Rhode (1996) involved opening scala vestibuli. In contrast, in our preparation the cochlea is opened near its basal end, overlying scala tympani. Although the round window does influence pressure at nearby scala tympani sites (Nedzelnitsky, 1980), the differential pressure across the cochlear partition is dominated by pressure in scala vestibuli, which substantially exceeds that in scala tympani (Dancer and Franke, 1980; Nedzelnitsky, 1980). The implication is that opening a hole in the otic capsule near the round window should not have significantly altered the differential pressure, which is the driving force to basilar-membrane motion (Voss *et al.*, 1996).

B. Basilar-membrane responses to clicks at the base of the chinchilla cochlea compared with measurements in other species and other cochlear locations

The present results are generally consistent with the main findings of the pioneering Mössbauer study of Robles *et al.* (1976; Rhode and Robles, 1974) at a site of the squirrel monkey cochlea with CF of 6–7.8 kHz. At that site, basilar-membrane responses to clicks consisted of two segments. The initial segment was of short duration, exhibited a relatively low frequency of oscillation and grew linearly with stimulus intensity. The second segment was a relatively undamped oscillation with frequency corresponding to CF, which grew at compressive rates. Nonlinear growth of basilar-membrane responses to clicks has also been demonstrated at the base of the cochleae of chinchilla (Ruggero *et al.*, 1992a,b) and guinea pig (LePage and Johnstone, 1980; de Boer and Nuttall, 1997). Frequency modulation has been described in basilar-membrane responses to clicks at the base of the guinea-pig cochlea (de Boer and Nuttall, 1997). At the base of the chinchilla cochlea, both the frequency of oscillation and the extent of the nonlinearity increase continuously over several hundreds of microseconds (Figs. 5–8). The frequency modulation is not abolished by death (Fig. 6A) and, therefore, must be viewed as a ''passive'' feature of both linear and nonlinear basilar-membrane responses. Thus, it is not surprising that frequency modulation can be mimicked by some linear models of basilar-membrane vibration (Nilsson and Møller, 1977; Møller and Nilsson, 1979; Lyon, 1997; de Boer and Nuttall, 1997).

Mechanical responses to clicks at the apex of guinea pig cochleae grow linearly with stimulus intensity (Cooper and Rhode, 1996). It is likely that the linearity of such responses reflects surgically induced cochlear damage. Responses to clicks at the apex of chinchilla cochleae do exhibit a mild compressive nonlinearity, with output/input slopes no lower than 0.76 dB/dB (Cooper and Rhode, 1996), which contrast with slopes as low as 0.2 dB/dB at the cochlear base (e.g., Figs. 7–8). It is not clear whether this quantitative discrepancy reflects greater surgical damage in apical preparations or a genuine difference in the extent of nonlinearity between the apex and the base of the cochlea. The latter possibility

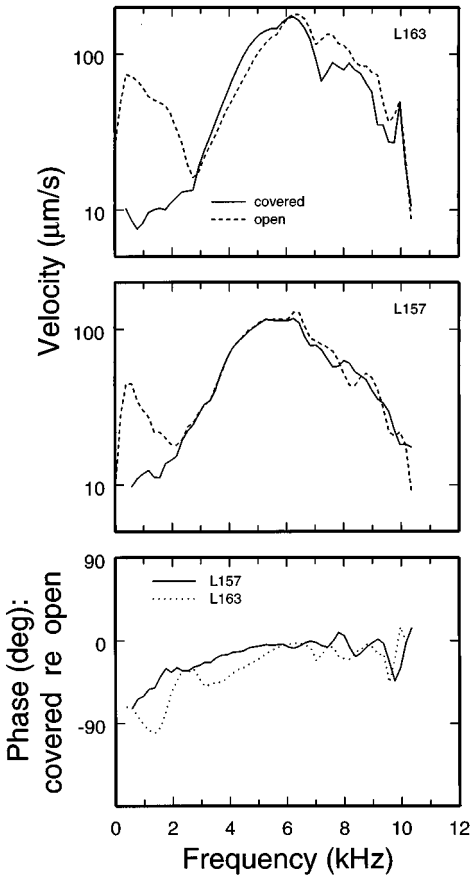


FIG. 17. The effects on basilar-membrane responses to clicks of covering the otic capsule hole with a glass window. Upper panels: magnitude-frequency spectra of the responses of Fig. 16. Lower panel: phase effects. Positive phases indicate leads relative to responses before placement of the glass window.

merits consideration because some properties of auditory-nerve responses suggest, in fact, that compressive nonlinearity is less salient in low-CF than in high-CF cochlear regions (e.g., Cooper and Yates, 1994; Sewell, 1984; Temchin *et al.*, 1997).

Responses to clicks at the apex of the chinchilla cochlea also differ from those at the base in that they apparently lack a (nearly) linear initial segment (Cooper and Rhode, 1996). We suppose that this difference reflects the distinct manner in which nonlinearities are distributed at the two recording sites as a function of time and in the corresponding frequency spectra. At basal locations, the initial (almost linear) oscillation has a spectral content heavily weighted by frequencies well below CF (Figs. 5–6), at which responses grow linearly, whereas nonlinear compressive growth is confined to later oscillations (Figs. 7–8) and to frequencies near CF (the ''tip'' of tuning curves; Fig. 11). At apical locations, compressive growth occurs over a wide frequency region (so that there is no distinction between nonlinear ''tip'' and linear ''tail'') and the spectral content of the initial oscillation of responses to clicks falls within the frequency range of nonlinear growth, which spans more than 3 octaves (Rhode and Cooper, 1996).

C. Polarity and latency of the onset of basilar-membrane responses to clicks

At the base of the chinchilla cochlea, the initial basilar-membrane response to rarefaction clicks consists of motion toward scala vestibuli (Figs. 1–2, 4, and 16) and, appropriately, responses to condensation clicks begin with motion toward scala tympani (Fig. 3). Other investigations of basilar-membrane responses to clicks at the cochlear base (Rhode and Robles, 1974; Robles *et al.*, 1976; LePage and Johnstone, 1980; Nuttall and Dolan, 1993; de Boer and Nuttall, 1997) did not report the polarity of responses. At the apex of chinchilla and guinea pig cochleae (at locations approximately 14 and 16.5 mm, respectively, from the stapes), the responses to clicks (recorded from the scala-media surfaces of Claudius cells, Hensen's cells or the tectorial membrane) contain a ''fast'' (short-latency) component, which is presumably artifactual (see Discussion Section III A), and a ''slow'' component, which begins as late as 1.5 ms after the onset of ossicular motion (Cooper and Rhode, 1996). The polarity of the onset oscillation of this slow component (i.e., motion toward scala vestibuli for rarefaction clicks) is the same as we found at the cochlear base. This suggests that the same onset polarity holds throughout the cochlea.

At the 3.5-mm site of the chinchilla cochlea, basilar-membrane responses to clicks begin approximately 30 μs after the onset of ossicular motion. This delay is smaller than a previous preliminary estimate from our laboratory (90 μs; Ruggero *et al.*, 1992a). The latter were not based on middle-ear and basilar-membrane responses to clicks recorded in the same ears and probably underestimated delays in the stimulus system.

Reports of responses to clicks at basal locations of guinea pig cochleae (LePage and Johnstone, 1980; Nuttall and Dolan, 1993; de Boer and Nuttall, 1997) did not provide estimates of onset delay. A cochlear delay of 300–390 μs was measured at a squirrel monkey basilar-membrane site with CF of 6–7.8 kHz (Robles *et al.*, 1976). According to a CF-distance map proposed by Greenwood (1990), such CFs correspond to a distance of 8.3–9.5 mm from the oval window. The difference between the delays at the 3.5-mm basilar-membrane site of chinchilla and the 8.3–9.5-mm site of squirrel monkey probably reflects the slowing down of the basilar-membrane wave as it travels from base to apex.

D. Is there a traveling wave on the basilar membrane?

The mechanical responses to clicks at the base of the chinchilla cochlea (the present work), at apical sites of the guinea pig and chinchilla cochleae (Cooper and Rhode, 1996), as well as those at an intermediate cochlear location in squirrel monkey (Robles *et al.*, 1976), imply that there is a progressive phase accumulation or increasing delay as a function of distance from the oval window. Responses to identical stimuli at closely spaced basilar-membrane locations in individual cochleae also show phase accumulation or increasing delay as a function of distance from the oval window (Rhode, 1971; Kohllöffel, 1972; Cooper and Rhode, 1996; Russell and Nilsen, 1997). Thus, the bulk of the basilar-membrane data on cochlear delays provide strong

support for the classical view of the traveling wave (von Békésy, 1960). Indirect estimates of traveling-wave delays based on cochlear microphonic responses to low-frequency tones (Tasaki *et al.*, 1952; Dallos and Cheatham, 1971; Dancer *et al.*, 1997) do not differ greatly from the latencies of responses to clicks at the basilar membrane of squirrel monkeys, chinchillas and guinea pigs (Robles *et al.*, 1976; Cooper and Rhode, 1996, and the present work). At sites located about 4, 9 and 14 mm from the oval window in the guinea pig cochlea, such delays amount to 25, 250 and 1000 μs, respectively (Dancer *et al.*, 1997).

The latency of responses to clicks of auditory-nerve fibers increases systematically as a function of decreasing CF (Kiang *et al.*, 1965; Kim and Molnar, 1979; Siegel *et al.*, 1982; Ruggero and Rich, 1983, 1987; Ruggero, 1992a). In the chinchilla, latencies for rarefaction clicks range from about 1 ms for fibers with CFs higher than 3–4 kHz to about 2.7 ms for fibers with CF of 320 Hz (Ruggero and Rich, 1983, 1987; Salvi *et al.*, 1979). If synaptic and neural delays account for a 1-ms delay, regardless of CF (discussed by Ruggero and Rich, 1987), the neural data imply that mechanical travel time is almost nil at basal cochlear locations and 1.7 ms near the apex. These delays are fairly consistent with those measured on the basilar membrane or the organ of Corti at basal (e.g., Fig. 1) and apical (Cooper and Rhode, 1996) sites of the chinchilla cochlea.

Dancer has suggested that the basilar membrane does not sustain a traveling wave (Dancer, 1992; see also Dancer *et al.*, 1997). He offered two main arguments. The first was that the onset latencies of cochlear responses are shorter than those predicted by Zwislocki's (1948) model of basilar-membrane vibration. We do not view this quantitative discrepancy between measurements and one particular mathematical model as evidence against the very *existence* of a traveling wave. The onset latencies of basilar-membrane and auditory-nerve responses to clicks increase systematically as a function of distance from the oval window, as expected for a mechanical wave that propagates from base to apex (Ruggero, 1994). Further, Zwislocki (1974) has shown that his model can be adjusted (by increasing stiffness) to yield latencies that closely match *in vivo* empirical data for the guinea pig cochlea (see his Fig. 5). Dancer's second argument was that the substantial phase accumulations exhibited by basilar-membrane responses at near-CF frequencies (equivalent to 1–2 CF periods at both basal and apical sites of the chinchilla cochlea: the present work and Cooper and Rhode, 1996) are due mostly to ''active'' cochlear processes and should not be taken as evidence for the existence of a (''passive'') traveling wave. Our data on the effects of death on the phases of basilar-membrane responses clearly contradict this argument: the phases of post-mortem responses are fully within the range of those observed *in vivo* (Figs. 11A and 15B).

E. Nonlinear phase shifts in responses to clicks and their relation to gain spectra

A striking aspect of the phases of post-mortem responses to clicks, as well as *in vivo* responses to intense clicks, is that they did not differ substantially from *in vivo* responses to low-level clicks. Although death, as well as increases in click level *in vivo*, did cause phase changes (and abolished the phase jumps), the changes (Fig. 15B) were small in comparison to the accumulated phase lags exhibited by *in vivo* responses at CF, regardless of stimulus level. This is somewhat paradoxical, in view of the substantial broadening of the bandwidth of response gain that accompanies either death or increases in stimulus level *in vivo* (Fig. 10). Since, in general, sharper frequency tuning requires longer response delays (Bode, 1946; Goldstein *et al.*, 1971; Geisler and Rhode, 1982), the large increases in bandwidth observed post-mortem (or at intense click levels *in vivo*) might have been expected to result in correspondingly large decreases of phase lag or group delay at frequencies near CF. In fact, the phases of post-mortem and *in vivo* responses to intense clicks did not change at CF and group delays decreased by only small amounts or not at all (Figs. 11A and 15B).

Basilar-membrane responses to clicks in healthy chinchilla cochleae often have two- (or even multiple-) lobed waveforms (e.g., Figs. 2–4; also see Fig. 6 of Ruggero and Rich, 1991a, and Fig. 6 of Ruggero *et al.*, 1996) and their spectra may include a notch (e.g., Figs. 9, 10, and 12; also see Fig. 9 of Ruggero *et al.*, 1996). Two-lobed mechanical responses to clicks and/or spectral notches have also been observed at a basal cochlear site in guinea pig (Nuttall and Dolan, 1993; de Boer and Nuttall, 1997) and at the apex of guinea pig and chinchilla cochleae (Cooper, 1997).

At the base of the chinchilla the spectral notches arise from cancellations due to phase opposition between same-frequency components in the (time-domain) lobes of the responses to clicks (Figs. 12 and 13). The phase modulation of basilar-membrane responses to clicks may be related to the asymmetry of their magnitude spectra, given that the impulse responses of asymmetrical filters generally exhibit such modulation (Papoulis, 1977, p. 123). However, the origin of the notches and the underlying phase modulation is obscure and we are uncertain about their significance. We doubt that they are solely artifacts (e.g., due to opening the otic capsule: Cooper, 1997; see Section III A of Discussion), since such artifacts should be relatively invulnerable to physiological deterioration and should exhibit linear properties. In contrast, notches and phase shifts were present only in sensitive cochleae, at or near CF at certain stimulus levels (i.e., their presence was CF-specific and intensity dependent) and disappeared postmortem. All these features indicate that the phase modulation and associated notches are linked to nonlinear and physiologically vulnerable processes. Most importantly, the notches occurred at levels at which response growth was highly compressive (see Fig. 7, L113), thus ruling out such linear artifacts as motion of the fluid-meniscus (see Section A of Discussion). However, we cannot rule out that these phenomena arise from an interaction between cochlear nonlinearities and stimulus artifacts (such as ringing in the acoustic click waveform).

F. Interpretation of the close resemblance between basilar-membrane responses to tones and clicks: Linearlike correspondences in spite of nonlinear growth

Robles *et al.* (1976) noted that at the basilar membrane of the squirrel monkey, responses to clicks could be mimicked to a fair extent by a Fourier synthesis based on magnitude-vs-frequency spectra measured with tones and derived phase-vs-frequency functions (computed from the magnitude data on the assumption of minimum-phase behavior). Our present findings show that, indeed, basilar-membrane responses to clicks and tones are generally well matched (Fig. 15; see also Figs. 3 and 4 of Ruggero *et al.*, 1992a, and Fig. 6 of Ruggero *et al.*, 1992b), so that responses to tones can be predicted, with good accuracy, from responses to clicks, and vice versa.

It seems remarkable that reasonably accurate cross-stimulus predictions are possible for such a highly nonlinear system as the one underlying basilar-membrane vibration at the base of the cochlea. We surmise that the system properties that make such predictions possible are the same that severely restrict the prominence of basilar-membrane harmonic distortion and other nonlinearities, especially when stimulation is confined to a single spectral level (e.g., Evans, 1989; de Boer and Kruidenier, 1990; Zwislocki *et al.*, 1997). (1) The magnitude- and phase-frequency spectra of responses to a given type of stimulus (click or tone) presented at any one level accurately predicts responses to the other type of stimulus only at a single, specific level. In the case of Fig. 15A, for example, responses to 54-dB clicks predict accurately the magnitudes of responses to 40-dB tones but grossly overestimate the responses to 70-dB tones. (2) Basilar-membrane responses to single tones at the base of the cochlea exhibit no dc components (Cooper and Rhode, 1992) and contain little (perhaps negligible) harmonic distortion (Cooper and Rhode, 1992; Ruggero *et al.*, 1992c, 1997). (3) Responses to white noise can be accurately predicted by first-order Wiener kernels (Recio *et al.*, 1996, 1997). (4) The magnitude of odd-order distortion products elicited by pairs of equal-level tones rarely exceeds 10% of the responses to the primary tones (Robles *et al.*, 1991, 1997; Nuttall *et al.*, 1990; Rhode and Cooper, 1993).

G. Frequency modulation, the extent of nonlinearity as a function of time, and the speed of the ''cochlear amplifier''

Nonlinear compressive growth at the 3.5-mm site of the chinchilla basilar membrane is evident within 100 μs states of response onset (Figs. 6 and 8). By implication, the organ of Corti feedback that boosts basilar membrane vibrations (''the cochlear amplifier'') starts its operation within 1 CF period following the arrival of the traveling wave at the recording site. The strength of the feedback (as indicated by the magnitude of compressive nonlinearity) varies significantly as a function of time, increasing rapidly from a minimum at onset to a maximum 1–1.6 ms later but waxing and waning thereafter (Fig. 8). One plausible interpretation of the initial variation is that the feedback is tuned to CF and thus merely follows the ''passive'' frequency glide, which is present both *in vivo* and post-mortem (Figs. 5 and 6). Taking the frequency glide into account, the onset of nonlinearity appears to be an almost instantaneous response to the CF spectral components of the ''passive'' vibrations.

However, the time course of the frequency glide cannot fully account for the time course of the compressive nonlinearity. The frequency glide consists of a roughly monotonic saturating-exponential increase, whereas the magnitude of the compressive nonlinearity undergoes waxing and waning, with compression minima (i.e., slope maxima) corresponding approximately to the wave shape constrictions. The nearly synchronous occurrence of the constrictions and the compression minima suggests that response sensitivity and nonlinear growth are tightly linked, which is consistent with the operation of a (nearly) instantaneous nonlinear ''cochlear amplifier.''

H. Comparison between responses to clicks at the basilar membrane and auditory-nerve responses to clicks and noise

The responses to clicks of the basilar-membrane site with CF of 9–10 kHz cannot be compared in any detail with those of auditory-nerve fibers with similar CFs because such neurons do not phase lock at near-CF frequencies. Except for rectification (presumably due to synaptic processes), the responses to clicks of low-CF neurons resemble qualitatively the high-CF basilar-membrane responses in several respects. First, low-CF auditory-nerve fibers respond to rarefaction clicks with shorter latencies [by about 1/(2*CF)] than to condensation clicks (Kiang *et al.*, 1965; Goblick and Pfeiffer, 1969; Pfeiffer and Kim, 1972; Kim and Molnar, 1979; Siegel *et al.*, 1982; Ruggero and Rich, 1983, 1987). This is consistent with the onset polarity of basilar membrane responses (the present work and Cooper and Rhode, 1996) on the assumption that, in apical cochlear regions, auditory-nerve excitation occurs when the basilar membrane is deflected or in motion toward scala vestibuli (Ruggero and Rich, 1983, 1987). Second, neural responses exhibit nonlinear properties reminiscent of those at the basilar membrane, including a skew of their envelope toward earlier times as a function of increasing stimulus intensity (e.g., Kiang *et al.*, 1965; Pfeiffer and Kim, 1972) and an amplitude nonlinearity in responses to paired clicks (Goblick and Pfeiffer, 1969). This amplitude nonlinearity may reflect directly the nonlinear growth of responses at the basilar membrane. The envelope skew, on the other hand, probably reflects mixed contributions from both basilar-membrane nonlinearities (Figs. 2–4, 7 and 8) and neural and synaptic processes (e.g., Gray, 1967; Schoonhoven *et al.*, 1994). Third, the responses to clicks of 7% of low-CF fibers exhibit lobes reminiscent of those (Figs. 3 and 4) characterizing basilar-membrane responses at the cochlear base (Pfeiffer and Kim, 1972). However, it is puzzling that such lobes should be only present in a small minority of fiber responses, if in fact they reflect basilar-membrane vibrations.

Revcors (de Boer, 1967; de Boer and de Jongh, 1978) for responses to noise of low-CF auditory-nerve fibers resemble basilar-membrane responses to clicks even more closely than do auditory-nerve fiber responses to clicks. This

is because the first-order cross-correlation between the spike train and the noise stimulus removes rectification and other even-order nonlinearities, which are largely (but not entirely: Rhode and Cooper, 1996) of synaptic and/or neural origin (Recio *et al.*, 1996, 1997). Revcors of neural responses to noise (Møller and Nilsson, 1979) exhibit a frequency modulation very much like that seen in basilar-membrane responses (Figs. 5–6). They also show a deterioration of frequency tuning and a spectral shift toward lower frequencies with increasing stimulus intensity (Evans, 1977; de Boer and de Jongh, 1978; Møller, 1978). Again, it is likely that these nonlinearities reflect both basilar-membrane and more central (synaptic and neural) contributions.

ACKNOWLEDGMENTS

Many thanks to Andrei Temchin for help in the preparation of figures and to Mary Ann Cheatham, Peter Dallos, Luis Robles and especially Nigel Cooper for their comments on previous drafts of this paper. We were supported by Grants Nos. 5-P01-DC-00110-23 and 2-R01-DC-00419-10 from the National Institute on Deafness and Other Communication Disorders.

Békésy, G. von (**1960**). *Experiments in Hearing* (McGraw-Hill, New York).

Bennett, W. R. (**1970**). *Introduction to Signal Transmission* (McGraw-Hill, New York).

Bode, H. W. (**1946**). *Network Analysis and Feedback Amplifier Design* (Van Nostrand, New York).

Cooper, N. P. (**1997**). ''Mid-band sensitivity notches in apical cochlear mechanics,'' in *Diversity in Auditory Mechanics*, edited by E. R. Lewis, G. R. Long, R. F. Lyon, P. M. Narins, C. R. Steele, and E. Hecht-Poinar (World Scientific, Singapore), pp. 298–304.

Cooper, N. P., and Rhode, W. S. (**1992**). ''Basilar membrane mechanics in the hook region of cat and guinea-pig cochleae: Sharp tuning and nonlinearity in the absence of baseline position shifts,'' Hearing Res. **63**, 163–190.

Cooper, N. P., and Rhode, W. S. (**1996**). ''Fast travelling waves, slow travelling waves and their interactions in experimental studies of apical cochlear mechanics,'' Aud. Neurosci. **2**, 289–299.

Cooper, N. P., and Yates, G. K. (**1994**). ''Nonlinear impact-output functions derived from the responses of guinea-pig cochlear nerve fibres: Variations with characteristic frequency,'' Hearing Res. **78**, 221–234.

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.

Dancer, A. (**1992**). ''Experimental look at cochlear mechanics,'' Audiology **31**, 301–312.

Dancer, A., and Franke, R. (**1980**). ''Intracochlear sound pressure measurements in guinea pigs,'' Hearing Res. **2**, 191–205.

Dancer, A., Avan, P., and Magnan, P. (**1997**). ''Can the travelling wave be challenged by direct intracochlear pressure measurements?'' in *Diversity in Auditory Mechanics*, edited by E. R. Lewis, G. R. Long, R. F. Lyon, P. M. Narins, C. R. Steele, and E. Hecht-Poinar (World Scientific, Singapore), pp. 340–346.

de Boer, E. (**1967**). ''Correlation studies applied to the frequency resolution of the cochlea,'' J. Aud. Res. **7**, 209–217.

de Boer, E., and de Jongh, H. R. (**1978**). ''On cochlear encoding: Potentialities and limitations of the reverse-correlation technique,'' J. Acoust. Soc. Am. **63**, 115–135.

de Boer, E., and Kruidenier, C. (**1990**). ''On ringing limits of the auditory periphery,'' Biol. Cybern. **63**, 433–442.

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.

Evans, E. F. (**1977**). ''Frequency selectivity at high signal levels of single units in cochlear nerve and nucleus,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 185–192.

Evans, E. F. (**1989**). ''Cochlear filtering: A view seen through the temporal discharge patterns of single cochlear nerve fibers,'' in *Cochlear Mechanisms: Structure, Function and Models*, edited by J. P. Wilson and D. T. Kemp (Plenum, New York), pp. 241–250.

Geisler, C. D., and Rhode, W. S. (**1982**). ''The phases of basilar-membrane vibrations,'' J. Acoust. Soc. Am. **71**, 1201–1203.

Goblick, Jr., T. J., and Pfeiffer, R. R. (**1969**). ''Time-domain measurements of cochlear nonlinearities using combination click stimuli,'' J. Acoust. Soc. Am. **46**, 924–938.

Goldstein, J. L., Baer, T., and Kiang N. Y.-S (**1971**). ''A theoretical treatment of latency, group delay and tuning characteristics for auditory nerve responses to clicks and tones,'' in *The Physiology of the Auditory System*, edited by M. B. Sachs (National Educational Consultants, Baltimore, MD), pp. 133–141.

Gray, P. R. (**1967**). ''Conditional probability analyses of the spike activity of single neurons,'' Biophys. J. **7**, 759–777.

Greenwood, D. D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.

Kiang, N. Y.-S., Watanabe, T., Thomas, C., and Clark, L. F. (**1965**). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (MIT, Cambridge, MA).

Kim, D. O., and Molnar, C. E. (**1979**). ''A population study of cochlear nerve fibers: comparison of spatial distributions of average-rate and phase-locking measures of responses to single tones,'' J. Neurophysiol. **42**, 16–30.

Kohllöffel, L. U. E. (**1972**). ''A study of basilar membrane vibrations. II. The vibratory amplitude and phase pattern along the basilar membrane (post-mortem),'' Acustica **27**, 66–81.

LePage, E. L., and Johnstone, B. M. (**1980**). ''Nonlinear mechanical behaviour of the basilar membrane in the basal turn of the guinea pig cochlea,'' Hearing Res. **2**, 183–189.

Lyon, R. F. (**1997**). ''All-pole auditory filter models,'' in *Diversity in Auditory Mechanics*, edited by E. R. Lewis, G. R. Long, R. F. Lyon, P. M. Narins, C. R. Steele, and E. Hecht-Poinar (World Scientific, Singapore), pp. 205–211.

Møller, A. R. (**1978**). ''Responses of auditory nerve fibres to noise stimuli show cochlear nonlinearities,'' Acta Oto-Laryngol. **86**, 1–8.

Møller, A. R., and Nilsson, H. G. (**1979**). ''Inner ear impulse response and basilar membrane modelling,'' Acustica **41**, 258–262.

Nedzelnitsky, V. (**1980**). ''Sound pressures in the basal turn of the cat cochlea,'' J. Acoust. Soc. Am. **68**, 1676–1689.

Nilsson, H. G., and Møller, A. R. (**1977**). ''Linear and nonlinear models of the basilar membrane motion,'' Biol. Cybernet. **27**, 107.

Nuttall, A. L., and Dolan, D. F. (**1993**). ''Basilar membrane velocity responses to acoustic and intracochlear electric stimuli,'' 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. 288–294.

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.

Nuttall, A. L., Dolan, A. F., and Avinash, G. (**1990**). ''Measurements of basilar membrane tuning and distortion with laser Dopler velocimetry,'' in *The Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele (Springer-Verlag, Berlin), pp. 288–295.

Papoulis, A. (**1977**). *Signal Analysis* (McGraw-Hill, New York).

Patuzzi, R. (**1996**). ''Cochlear micromechanics and macro mechanics,'' in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 186–257.

Pfeiffer, R. R., and Kim, D. O. (**1972**). ''Response patterns of single cochlear nerve fibers to clock stimuli: Descriptions for cat,'' J. Acoust. Soc. Am. **52**, 1669–1677.

Recio, A., Narayan, S. S., and Ruggero, M. A. (**1996**). ''Wiener-kernel analysis of basilar membrane responses to noise,'' ARO Mid-Winter Meeting Abstract 19, 55.

Recio, A., Narayan, S. S., and Ruggero, M. A. (**1997**). ''Wiener-kernel analysis of basilar-membrane responses to white noise,'' in *Diversity in Auditory Mechanics*, edited by E. R. Lewis, G. R. Long, R. F. Lyon, P. M. Narins, C. R. Steele, and E. Hecht-Poinar (World Scientific, Singapore), pp. 325–331.

Rhode, W. S. (**1971**). ''Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique,'' J. Acoust. Soc. Am. **49**, 1218–1231.
Rhode, W. S., and Cooper, N. P. (**1993**). ''Two-tone suppression and distortion production on the basilar membrane in the hook region of cat and guinea pig cochlea,'' Hearing Res. **66**, 31–45.
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 Robles, L. (**1974**). ''Evidence from Mössbauer 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 monkey using the Mössbauer effect,'' J. Acoust. Soc. Am. **59**, 926–939.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). ''Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input–output functions, tuning curves and response phases,'' J. Acoust. Soc. Am. **80**, 1364–1374.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1991**). ''Two-tone distortion in the basilar membrane of the cochlea,'' Nature (London) **349**, 413–414.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1997**). ''Two-tone distortion on the basilar membrane of the chinchilla cochlea,'' J. Neurophysiol. **77**, 2385–2399.
Ruggero, M. A. (**1992a**). ''Physiology and coding of sound in the auditory nerve,'' in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A. N. Popper and R. R. Fay (Springer-Verlag, New York), pp. 34–93.
Ruggero, M. A. (**1992b**). ''Responses to sound of the basilar membrane of the mammalian cochlea,'' Current Opinion in Neurobiology **2**, 449–456.
Ruggero, M. A. (**1994**). ''Cochlear delays and traveling waves: Comments on 'Experimental look at cochlear mechanics' (A. Dancer, Audiology 1992, 31: 301–312),'' Audiology **33**, 131–142.
Ruggero, M. A., and Rich, N. C. (**1983**). ''Chinchilla auditory nerve responses to low-frequency tones,'' J. Acoust. Soc. Am. **73**, 2096–2108.
Ruggero, M. A., and Rich, N. C. (**1987**). ''Timing of spike initiation in cochlear afferents: Dependence on site of innervation,'' J. Neurophysiol. **58**, 379–403.
Ruggero, M. A., and Rich, N. C. (**1990**). ''Systematic injection of furosemide alters the mechanical response to sound of the basilar membrane,'' in *The Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele (Springer-Verlag, Berlin), pp. 314–321.
Ruggero, M. A., and Rich, N. C. (**1991a**). ''Application of a commercially-manufactured Doppler-shift laser velocimeter to the measurement of basilar-membrane vibrations,'' Hearing Res. **51**, 215–230.
Ruggero, M. A., and Rich, N. C. (**1991b**). ''Furosemide alters organ of Corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane,'' J. Neurosci. **11**, 1057–1067.
Ruggero, M. A., Rich, N. C., Robles, L., and Shivapuja, B. G. (**1990**). ''Middle ear response in the chinchilla and its relationship to mechanics at the base of the cochlea,'' J. Acoust. Soc. Am. **87**, 1612–1629.
Ruggero, M. A., Rich, N. C., and Recio, A. (**1991**). ''Responses to clicks of the chinchilla basilar membrane,'' Soc. Neurosci. Abst. **17**, 1106.
Ruggero, M. A., Rich, N. C., and Recio, A. (**1992a**). ''Basilar membrane responses to clicks,'' in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, London), pp. 85–91.
Ruggero, M. A., Robles, L., Rich, N. C., and Recio, A. (**1992b**). ''Basilar membrane responses to two-tone and broadband stimuli,'' Philos. Trans. R. Soc. London, Ser. B **336**, 307–315.
Ruggero, M. A., Robles, L., and Rich, N. C. (**1992c**). ''Two-tone suppression in the basilar membrane of the cochlea: Mechanical basis of auditory-nerve rate suppression,'' J. Neurophysiol. **68**, 1087–1099.
Ruggero, M. A., Rich, N. C., and Recio, A. (**1993**). ''Alteration of basilar membrane responses to sound by acoustic overstimulation,'' in *Biophysics of Hair Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 258–264.
Ruggero, M. A., Rich, N. C., and Recio, A. (**1996**). ''The effect of intense acoustic stimulation on basilar-membrane vibrations,'' Aud. Neurosci. **2**, 329–345.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the ase of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
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. USA **94**, 2660–2664.
Salvi, R. J., Henderson, D., and Hamernik, R. P. (**1979**). ''Single auditory-nerve fiber and action potential latencies in normal and noise-treated chinchillas,'' Hearing Res. **1**, 237–251.
Schoonhoven, R., Keijzer, J., Versnel, H., and Prijs, V. F. (**1994**). ''A dual filter model describing single-fiber responses to clicks in the normal and noise-damaged cochlea,'' J. Acoust. Soc. Am. **95**, 2104–2121.
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.
Sewell, W. F. (**1984**). ''The effects of furosemide on the endocochlear potential and auditory-nerve fiber tuning,'' Hearing Res. **15**, 69–72.
Siegel, J. H., Kim, D. O., and Molnar, C. E. (**1982**). ''Effects of altering organ of Corti on cochlear distortion products $f2\text{-}f1$ and $2f1\text{-}f2$,'' J. Neurophysiol. **47**, 303–328.
Tasaki, I., Davis, H., and Legouix, J.-P. (**1952**). ''The space-time pattern of the cochlear microphonics (guinea pig) as recorded by differential electrodes,'' J. Acoust. Soc. Am. **24**, 502–519.
Temchin, A. N., Rich, N. C., and Ruggero, M. A. (**1997**). ''Low-frequency suppression of auditory-nerve responses to characteristic-frequency tones,'' Hearing Res. **113**, 29–56.
Voss, S. E., Rosowski, J. J., and Peake, W. T. (**1996**). ''Is the pressure difference between the oval and round windows the effective acoustic stimulus for the cochlea?'' J. Acoust. Soc. Am. **100**, 1602–1616.
Zwislocki, J. J. (**1948**). ''Theorie der Schneckenmechanik: Qualitative and Quantitative Analyse,'' Acta Oto-Laryngol. Suppl. **72**, 1–76.
Zwislocki, J. J. (**1974**). ''Cochlear waves: Interaction between theory and experiments,'' J. Acoust. Soc. Am. **55**, 578–583.
Zwislocki, J. J., Szymko, Y., and Hertig, L. Y. (**1997**). ''The cochlea is an automatic gain control system after all,'' in *Diversity in Auditory Mechanics*, edited by E. R. Lewis, G. R. Long, R. F. Lyon, P. M. Narins, C. R. Steele, and E. Hecht-Poinar (World Scientific, Singapore), pp. 354–360.

Responses of auditory nerve fibers to trains of clicks

Robert E. Wickesberg and Hanna E. Stevens
Department of Psychology, 603 East Daniel Street, University of Illinois at Urbana-Champaign, Champaign, Illinois 61820

(Received 17 April 1997; revised 24 November 1997; accepted 25 November 1997)

Responses of auditory nerve fibers to trains of clicks were recorded in ketamine anesthetized chinchillas. By varying the number of clicks and the interclick interval, this study examined whether ''post-onset adaptation,'' described in psychoacoustic experiments on localization, occurred in auditory nerve fibers. The results showed that the number of action potentials recorded from a nerve fiber in response to a train of clicks was a power function of the number of clicks. For interclick intervals of 2 ms or greater the exponent of the power function was 0.5, and this exponent did not change over a 20-dB range of intensities. The timing of action potentials relative to the click stimuli was measured using synchronization coefficients. The coefficients increased with interclick interval, decreased with increasing intensity, and were greater for fibers with low rates of spontaneous activity than for high spontaneous fibers. Recovery functions showed that for interclick intervals of 2 ms or more, the responses to the second click were at least 70% of the response to the initial click. The recovery depended upon the number of clicks in the train. These findings indicate that auditory nerve fibers respond to high rates of stimulus presentation and do not display the adaptation observed in localization studies. [S0001-4966(98)04803-6]

PACS numbers: 43.64.Pg, 43.64.Ri, 43.66.Mk [RDF]

INTRODUCTION

During sound localization tasks, the auditory system displays a decrease in sensitivity with an increase in either the duration of the acoustic signal or the rate of information presentation. In psychoacoustic experiments on lateralization (Hafter and Dye, 1983; Hafter *et al.*, 1983; Hafter and Wenzel, 1983; Dye and Hafter, 1984; Hafter *et al.*, 1988), each click in a train reduced the amount of information provided by subsequent clicks, and the shorter the interclick interval, the greater the reduction. The decrease in the information conveyed by the trailing clicks for binaural estimates of sound source location was attributed to a process of ''post-onset'' or ''binaural'' adaptation by Hafter and co-workers, and this adaptation was hypothesized to occur in the cochlear nuclear complex prior to the comparison of binaural information for localization (Hafter *et al.*, 1988). The goal of the experiments reported here was to determine whether the post-onset adaptation is present in the responses of auditory nerve fibers, which provide the input from the cochlea to the cochlear nucleus.

To describe the effectiveness of successive clicks in conveying information for lateralization estimates, Hafter and co-workers (e.g., Hafter *et al.*, 1988) used a signal detection model. This approach assumed that an error in lateralization was the result of a normally distributed, temporal noise in the auditory system. According to this model, increasing the number of stimuli should reduce the error by an amount proportional to the square-root of the original number divided by the increased number. Ideally, therefore, the interaural time difference (ITD) detectable using N clicks would be proportional to the time difference detectable with one click divided by the square-root of N. Plotting this relationship on a graph with logarithmic axes yields a straight line with a slope of -0.5. The experiments demonstrated that for rates of click presentation with interclick intervals between 5 and 10 ms, the slopes of the log–log functions were indeed -0.5. With higher rates of click presentation, the slopes became shallower. For interclick intervals of 2 ms and less the slope approached the limiting case of 0.0 and almost no information was conveyed after the initial click. The cause of the change in slope was postulated by Hafter and co-workers to be a process of adaptation.

A reduction in the information available for binaural judgments with high rates of information presentation has been described in several other lateralization experiments. Harris *et al.* (1963) described a decrease in information conveyed by the second click in a pair, when the interclick interval was between 1 and 8 ms. For shorter or longer intervals, however, the trailing click provided cues for localization. Similar results have been reported by Toole and Sayers (1965), Yost and Soderquist (1984) and by Saberi and Perrott (1990) in investigations of the precedence effect (Wallach *et al.*, 1949; Haas, 1951; Gaskell, 1983). The experiments demonstrating a reduced ability to make binaural discriminations with short interclick intervals have generally used pairs of clicks presented with a low repetition rate (100 ms between pairs) as the stimuli. Sayers and Toole (1964) used a higher presentation rate with only 6 ms between pairs with 3 ms between clicks, which effectively produced a train of clicks with an interclick interval of 3 ms. Both clicks produced an image that was lateralized but the images were located nearer to the midline (Sayers and Toole, 1964). These observations indicate that for all repetition rates, the initial clicks changed how the subsequent ones were processed.

Although the amount of information provided by the trailing click(s) appears to be reduced in localization tasks, estimates of timbre, pitch or intensity make use of this infor-

mation (Guttman, 1965; Harris *et al.*, 1963; Hafter and Richards, 1988). For example, the discrimination of click rate with trains of impulses, which produces periodicity pitch, increases as the number of clicks in a train increases (Hafter and Richards, 1988). These observations indicate that separate channels convey information for localization and pitch discriminations. Results from *in vitro* studies (Wickesberg and Oertel, 1988, 1990; Wickesberg *et al.*, 1991) and from across species comparisons (Takahashi *et al.*, 1984; Covey, 1993) imply that these channels are not just functionally distinct but different neural pathways. The process responsible for the binaural adaptation would not be expected to occur in any common pathway located prior to the separation into parallel channels.

The psychoacoustic results have indicated that adaptation occurs prior to the binaural comparison of information at the level of the superior olive (Harris *et al.*, 1963; Toole and Sayers, 1965; Hafter *et al.*, 1988). The cochlear nuclear complex, where information from auditory nerve fibers is divided into multiple parallel ascending channels, has been proposed as the specific site of the adaptation (Harris *et al.*, 1963; Hafter *et al.*, 1988). The results on pitch discriminations indicate that the adaptation should not be present in the responses of the auditory nerve fibers, the common input to the different channels. The purpose of the experiments reported here was to determine whether auditory nerve fibers can transmit the information contained in a click train, as either the number of clicks in a train was increased or the intervals between clicks are decreased. The results indicate that auditory nerve fibers are able to convey information both with increasing numbers of clicks and at short interclick intervals.

I. METHODS

Responses to trains of clicks were recorded extracellularly from auditory nerve fibers in chinchillas anesthetized with ketamine (40 mg/kg) supplemented with acepromazine (4 mg/kg) and xylazine (3 mg/kg). Supplemental doses of anesthetic were given as needed to maintain an areflexive state. After anesthetizing the animal, the internal temperature was maintained at 35 °C by a feedback-controlled heating pad. Following a tracheotomy, the external auditory meatus was opened by removal of the bony lateral wall, which allowed direct visualization of the tympanic membrane. An earpiece fitted with a Grayson-Stadler eartip was inserted in the opened canal to within several millimeters of the tympanic membrane and petroleum jelly was used to seal the sound delivery system. The bulla was vented by drilling a small hole and inserting a 19 gauge needle. Sounds were delivered using a pair of shielded Etymotic Research ER-1 Tubephones and the sound delivery system was calibrated at the start of the experiment using an ER-7c probe microphone.

Stimuli consisted of tone bursts and trains of clicks. As shown in Fig. 1, click trains contained 2, 4, 8, 12, or 16 clicks with interclick intervals between 1 and 5 ms. A train with a set number of clicks and a fixed interclick interval was presented once every 140 ms with a total of 100 repetitions. The data collection program began with two clicks and a 1-ms interclick interval, incremented the interclick interval

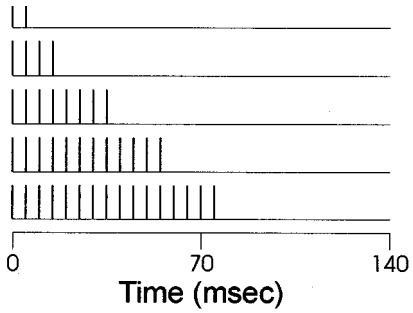


FIG. 1. Diagram depicting the click trains used as stimuli for the auditory nerve fibers. The click trains contained either 2 (top), 4, 8, 12 or 16 (bottom) clicks and a train was presented once every 140 ms. Each train of clicks containing a fixed number of clicks at a fixed interclick interval was presented 100 times, then the interclick interval was incremented. The intervals between clicks were varied from 1 ms to 5 ms (shown) in 1 ms steps. After the maximum interclick interval, the number of clicks was increased and the interclick interval was set to 1 ms.

in 1-ms steps to 5 ms, and then the number of clicks was increased with a return to a 1-ms interclick interval. The clicks were 100 μs, rectangular pulses. Click intensity was calculated from the peak amplitude of the click measured in the ear canal with the probe microphone using the amplitude of the fixed 94 dB SPL, 1000 Hz calibration tone as the reference. Since the intensity is based on the equivalent of the peak, it will be referred to as the peak equivalent SPL or ''pe SPL.'' The intensity of the unattenuated clicks was 99 dB pe SPL, and the levels presented to the fibers were 59, 69 and 79 dB pe SPL. The lowest intensity clicks, 59 dB pe SPL, were 10 to 20 dB above the thresholds of most units.

The auditory nerve was approached dorsally, following the ablation of the overlying cerebellum. The recording electrode was positioned visually either in the auditory nerve or in the lateral edge of the ventral cochlear nucleus where the auditory nerve enters. Auditory nerve fibers were identified physiologically by their primary-like peristimulus time histograms, and short latency responses to clicks. Characteristic frequencies and thresholds were determined using an automated tuning curve procedure. The auditory nerve fibers were classified on the basis of their spontaneous activity either as high ($\geqslant$20 spikes/s) or low ($<$20 spikes/s) spontaneous. Spontaneous rate was measured both by the tuning curve program (three samples of 50 ms each) and during the last 70 ms of the 140-ms repetition interval in the responses to trains of two clicks with an interclick interval of 1 ms. The extracellular recordings were made from single units using high impedance (15–35 Mohms) glass microelectrodes filled with 2M KCl and mounted in a Trent-Wells hydraulic microdrive. A stable recording environment was provided by filling the cerebellar cavity with Type I agarose (4% solution). A Dagan Cornerstone intracellular preamplifier followed by a Tucker-Davis Spike Conditioner amplified and filtered the recorded waveshapes. Spike times were discriminated using the windowed spike discriminator in the Tucker-Davis system and then measured relative to the onset of a stimulus using an event timer. These spike times were used to generate peristimulus time (PST) histograms, dot raster displays, and spike counts both on-line and during the later

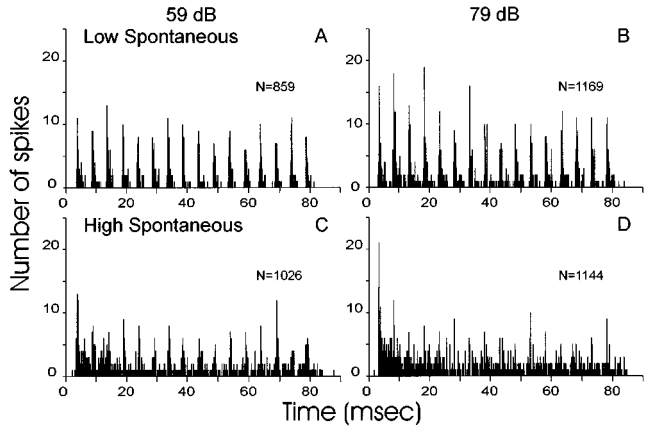


FIG. 2. Peristimulus time (PST) histograms of the responses to trains of 16 clicks for two auditory nerve fibers with (A,B) low (1.2 spikes/s) and (C,D) high (36 spikes/s) spontaneous activity. The interval between clicks was 5 ms and the trains were presented 100 times at levels of 59 (A,C) and 79 (B,D) dB pe SPL. Both auditory nerve fibers had characteristic frequencies of 2500 Hz. Increasing the stimulus intensity increased the numbers of spikes elicited for both units, but reduced the synchrony of the responses for the high spontaneous unit and individual peaks in the histogram were no longer visible. N=total number of action potentials.

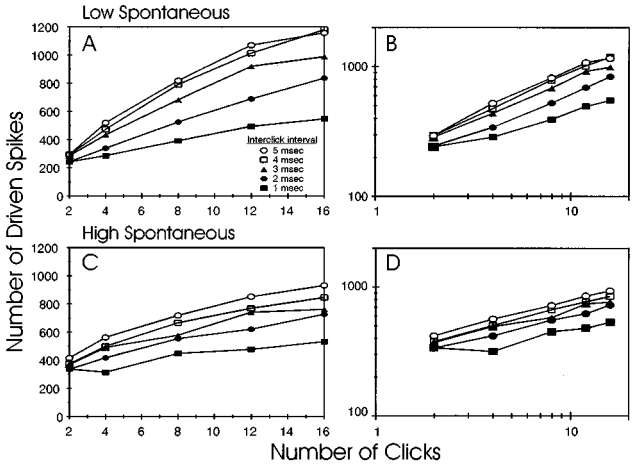


FIG. 3. The number of driven spikes in the responses to trains of 79 dB pe SPL clicks recorded from the same low (A,B) and high (C,D) spontaneous auditory nerve fibers whose PST histograms were shown in Fig. 2 as the number of clicks in a train and interclick intervals were varied. The driven spike counts were calculated by subtracting the expected number of spontaneous spikes from the total number of spikes. These counts were plotted either on linear (A,C) or log–log axes (B,D). On linear axes there was a decrease in slope as the number of clicks was increased, while on log–log axes the slope was constant. The intervals between clicks were 1 ms (■), 2 ms (●), 3 ms (▲), 4 ms (□), or 5 ms (○).

off-line analyses. The data collection, the generation of the stimuli, and the Tucker-Davis system were controlled by a Microsoft Windows™-based program on a PC-compatible, 486/33 MHZ computer. Off-line data analysis was accomplished using a PC version of the Response Analysis Package (RAP) from the Department of Neurophysiology at the University of Wisconsin–Madison. The results of these analyses were exported to Quattro Pro™ for statistical tests and summary plots were made with CorelDraw™.

II. RESULTS

Responses to trains of clicks (Fig. 1) were recorded from 67 auditory nerve fibers in 12 chinchillas. There were 33 high spontaneous and 34 low spontaneous fibers. Several characteristics of the responses of auditory nerve units are illustrated by the peristimulus time (PST) histograms shown in Fig. 2. The histograms were calculated from the responses to trains of 16 clicks presented with a 5-ms interclick interval at two intensities for a low spontaneous and a high spontaneous unit with identical characteristic frequencies (CF =2500 Hz). In response to the lower level click train, both units displayed a distinct peak in the histogram for each click that was presented. As the level of the stimulus increased, both the high and the low spontaneous rate units produced more action potentials, but individual peaks were no longer apparent in the PST for the high spontaneous unit. These observations indicate a differential responsiveness to trains of clicks for high and low spontaneous units. The decrease in the salience of the individual peaks in the PST suggests a loss of information on the times of occurrence of individual clicks in the train. The ability of auditory nerve fibers to transmit timing information is clearly important for the detection of interaural time differences, while an increase in the number of spikes suggests that more information is available for detecting the differences. The responses of a unit, therefore, must be analyzed using the timing of the action potentials as well the number of spikes elicited by the stimulus. The analyses based on the spike counts will be presented first.

A. Spike counts

The numbers of spikes elicited by click trains with increasing numbers of clicks presented at different interclick intervals are shown in Fig. 3 for the same high and low spontaneous units whose responses are shown in Fig. 2. The data were plotted on both linear and logarithmic axes. The counts of elicited or driven spikes were calculated as the increase in the number of spikes above the spontaneous rate from the onset of a train of clicks until 5 ms after the last click in the train. There were two reasons for computing the driven spike count, which subtracts the expected number of spontaneous spikes from the total count. The first was that the number reflects the counts of the spikes produced in response to the click train and represents the increase in the information available for making a judgement of source location. The second reason was that subtracting the spontaneous rate eliminated an artificial correlation of slope with the spontaneous rate when the results were plotted on logarithmic axes. Without the subtraction, units with higher spontaneous rate always had lower slopes, since the change in the number of spikes was smaller relative to the spontaneous spike count.

On linear axes, the slopes of these plots represent the change in the number of action potentials elicited as the number of clicks was increased, and the units for these slopes is spikes per click. These functions (e.g., Fig. 3(A) and (C)) show a decrease in slope with increasing numbers of clicks for both the high and the low spontaneous units. Using a paired-comparison t-test, the slopes were signifi-

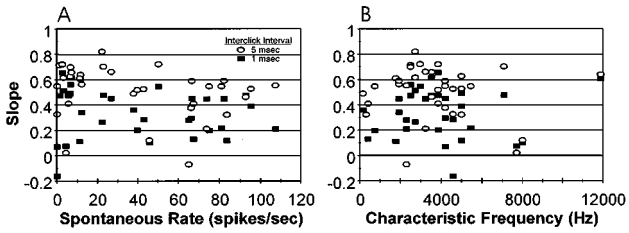


FIG. 4. The slopes from the responses to 16 clicks presented at 69 dB pe SPL with interclick intervals of 1 ms (■) or 5 ms (○) plotted as a function of the spontaneous rate of the nerve fiber (A) or the characteristic frequency (B).

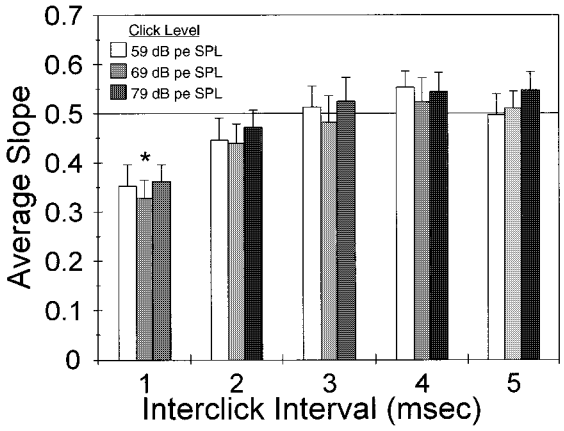


FIG. 5. The average slope as a function of the interval between clicks for both low and high spontaneous units for trains of clicks presented at levels of 59 (white), 69 (light gray) and 79 (dark gray) dB pe SPL. Error bars represent the standard error of the mean. Over a 20-dB range there was little change in slope for a fixed interclick interval. Only the slope at a 1-ms interclick interval was significantly different from a slope of 0.5 ($*p<0.01$). The value of 0.5 is consistent with Hafter's power-law model for the case of no binaural adaption.

cantly smaller for the increase from 12 to 16 clicks than for the increase from 2 to 4 clicks ($p<0.001$). As the number of clicks in a train increased, therefore, the number of spikes elicited by each additional click decreased. Plotting the same numbers on logarithmic axes (e.g., Fig. 3(B) and (D)) yielded, with two exceptions, linear plots with no significant differences in slopes as the number of clicks increased.

The linear functions on logarithmic axes are consistent with power function approximations of the responses, as proposed by Hafter *et al.* (1988) for the psychoacoustic data. The slopes on logarithmic axes represent the values of the exponents of the power functions. The values of the slopes were calculated as the slopes of the linear regression lines through the data on the log–log plots. With increasing interclick interval, the slopes of these plots for the low spontaneous fiber (Fig. 3(B)) are 0.41 (1 ms), 0.60 (2 ms), 0.62 (3 ms), 0.68 (4 ms) and 0.67 (5 ms); for the high spontaneous fiber (Fig. 3(D)) the slopes are 0.25 (1 ms), 0.37 (2 ms), 0.35 (3 ms), 0.39 (4 ms) and 0.39 (5 ms). The slopes for the low spontaneous fiber were slightly higher than average while those for the high spontaneous fiber were slightly lower (cf. Fig. 5).

Two deviations from linearity are apparent for the low spontaneous unit in Fig. 3. There was a decrease in slope for the increase from 2 to 4 clicks at interclick intervals of 1 and 2 ms, and a decrease in slope also occurs for the increase from 12 to 16 clicks for interclick intervals of 3, 4 and 5 ms. Similar changes were not generally observed for high spontaneous units, although in Fig. 3 there appears to be a reduced slope for a 1 ms interclick interval as the number of clicks was increased from 2 to 4. The decreased slope for the larger number of clicks and longer interclick intervals is likely to be a result of the overall repetition rate of the click train, as described below. The decrease for the shortest click trains at the shortest interclick intervals is consistent with refractory properties of the nerve fibers (Hafter *et al.*, 1988).

The slopes of the number of action potentials elicited by increasing numbers of clicks were not correlated with either spontaneous rate or characteristic frequency. For trains of clicks presented at a level of 69 dB pe SPL, the slopes on log–log axes were plotted as a function of either the spontaneous rate of the fiber (Fig. 4(A)) or its characteristic frequency (Fig. 4(B)). In neither case was the correlation coefficient significant. With an interclick interval of 1 ms the low spontaneous rate units had an average slope of 0.324, while the average slope for high spontaneous rate units was 0.329. At an interclick interval of 5 ms the average slopes were 0.535 (low spontaneous) and 0.493 (high spontaneous). In terms of the numbers of spikes elicited by trains of clicks presented at a specific interclick interval, therefore, the average slopes of the responses, i.e., the mean exponents of a power function approximations to the responses, were the same across characteristic frequency for both low and high spontaneous units.

The slopes of the spike counts varied as a function of interclick interval, but not as a function of intensity over a 20-dB range. The average slopes as a function of the interclick interval for three different intensities of clicks are shown in Fig. 5. Since at each intensity the slopes of the high and the low spontaneous units were not significantly different (e.g., Fig. 4), the unit types were combined for calculating the average slopes. For all units, increasing the intensity of the clicks resulted in an increased number of spikes. However, for each interclick interval, the slope of the functions were almost identical over a 20-dB intensity range. At interclick intervals between 3 and 5 ms, the average slope was approximately 0.5. With a decrease to a 2 ms interclick interval, there was a decrease in the average slope to about 0.45, which was not significant. A reduction in interclick interval to 1 ms produced a decrease in slope to 0.35 for all three click intensities, and this decrease was significant ($p<0.01$).

B. Timing

The times of occurrence of action potentials relative to the presentation of the clicks in the train were examined by folding the PST histograms on the interclick interval. This process is illustrated in Fig. 6 for a low spontaneous unit (7.2 spikes/s, CF=3535 Hz) and a high spontaneous unit (73 spikes/s, CF=3242 Hz) using the responses to 69 dB pe SPL clicks. The result is a period or cycle histogram determined for the frequency whose period is equal to the interclick interval. The cycle histograms (Fig. 6(B), (D), (F), (H)) can be

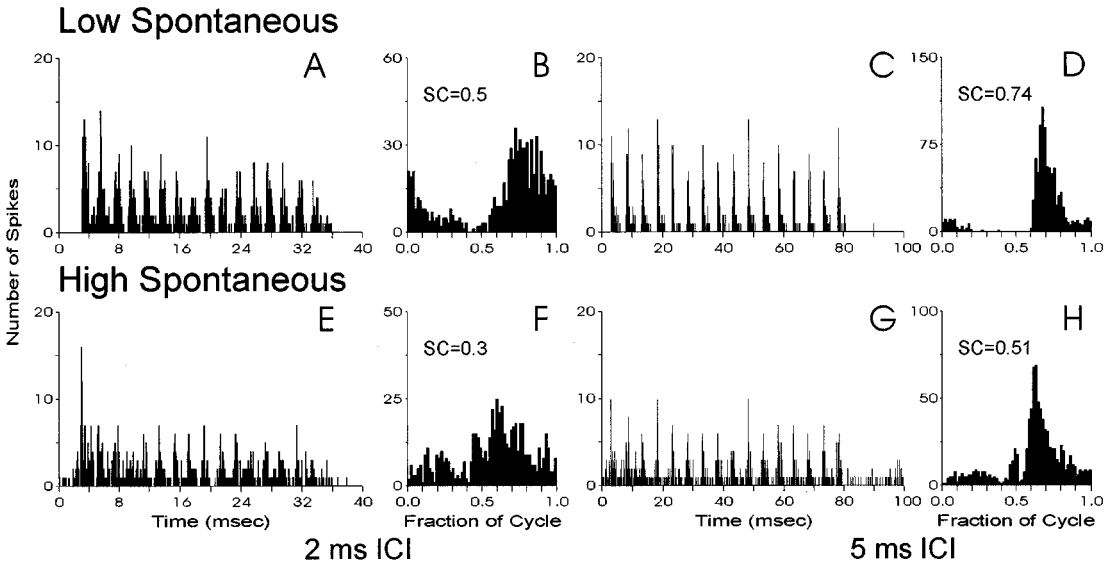


FIG. 6. PST (A,C,E,G) and period (B,D,F,H) histograms for the responses of low and high spontaneous auditory nerve fibers to trains of 16 clicks. The low spontaneous unit (A–D) had a characteristic frequency of 3535 Hz and the high spontaneous fiber (E–H) had a characteristic frequency of 3242 Hz. The click trains were presented with interclick intervals of 2 ms (A,B,E,F) or 5 ms (C,D,G,H) at a level of 69 dB pe SPL. The synchronization coefficients (SC) are displayed for each period histogram.

used to calculate synchronization (SC) coefficients (Goldberg and Brown, 1968). This measure of the timing of the action potentials relative to the click stimuli was chosen because the value reflects the relative numbers of precisely and poorly timed action potentials. It has a value of 1.0 when all spikes are precisely timed relative to the clicks and fall into one bin in the period histogram. To the extent that the spikes are not precisely timed, the coefficient decreases. With a complete loss of timing information, there would be an equal number of spikes in all bins and a coefficient of zero. For the two auditory nerve fibers shown in Fig. 6, the unit with the lower spontaneous activity was better able to transmit timing information, as indicated by the higher synchronization coefficients, but as the interclick interval decreased so did the ability of both fibers to transmit timing information.

The average synchronization coefficients for the low and high spontaneous units calculated in response to trains containing 16 clicks and presented at three different levels are shown in Fig. 7. As the interval between clicks increased, the synchronization coefficients increased. For the high spontaneous units (Fig. 7(B)), there was a smaller increase as interclick interval increased from 1 to 5 ms than for the low spontaneous units (Fig. 7(A)). However, the increases between interclick intervals of 2 and 5 ms were almost identical for both types of units. For the low spontaneous units the average coefficient at 59 dB pe SPL increased by 0.22 from 0.53 (2 ms ICI) to 0.75 (5 ms ICI), while for the high spontaneous units the values increased by 0.23 from 0.36 to 0.59. At all interclick intervals, the coefficients for the low spontaneous auditory nerve fibers were significantly ($p<0.05$) greater than the values for the high spontaneous fibers.

As the intensity of the clicks increased, the coefficients for both high and low spontaneous units decreased. This decrease can also be observed in Fig. 2, where the synchronization coefficient for the low spontaneous unit was 0.79 for the response to the 59-dB click train and 0.65 for the 79-dB train. For the high spontaneous unit in Fig. 2, the coefficients decreased from 0.36 to 0.22 with the 20-dB increase in click level. Comparing these values with Fig. 7, the low spontaneous unit was almost average, while the coefficients for the

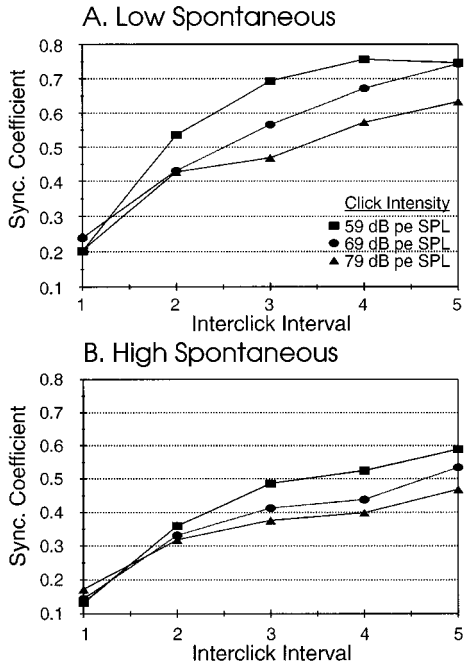


FIG. 7. Average synchronization coefficients determined at the frequency whose period equaled the interclick interval, and calculated from the responses to 16 clicks presented at levels of 59 (■), 69 (●) and 79 (▲) pe SPL for low (A) and high (B) spontaneous units. Both types of units showed a gradual decrease in the synchronization coefficient as the interval between clicks was reduced from 5 to 2 ms, and then a much larger reduction as the interval was reduced to 1 ms. There was also a decrease in synchronization with increasing intensity for all interclick intervals.

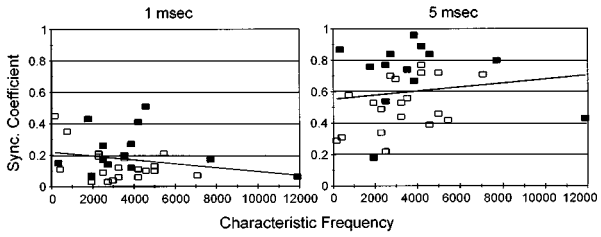


FIG. 8. The synchronization coefficients plotted as a function of the characteristic frequency of the auditory nerve fiber. The coefficients were calculated from the response to trains of 16 clicks presented at 69 dB pe SPL with interclick intervals of 1 ms (left) or 5 ms (right). Values from low spontaneous rate fibers are shown with filled symbols (■) and open symbols (□) indicate data from low spontaneous rate fibers. The regression lines have slopes of -1.2×10^{-5} (1 ms) and 4.9×10^{-6} (5 ms). There was no significant correlation of synchronization coefficient with characteristic frequency at either interclick interval.

high spontaneous unit were lower than average. For both units, the cause of the decrease was an increase in the number of spikes in the intervals between the peaks of the PST. This was also a significant cause of the difference in synchronization between the low and high spontaneous units (cf. Fig. 6(D) and (H)). While synchronization depended on the level of spontaneous activity and the intensity of the clicks, the coefficients were not significantly correlated with the characteristic frequency of the fiber as shown in Fig. 8. Figure 8 also shows the increase in synchronization with increasing interclick interval for all characteristic frequencies.

C. Recovery functions

A recovery function is a measure of the extent to which the response to a trailing click is equal to the response to a single click, and reflects the ability of the auditory nerve to respond throughout the stimulus. Two methods for calculating recovery functions were used. Both methods for calculating the recovery function counted the number of spikes during a window of time equal to the interclick interval for each click in the train, with the first window beginning at the onset of the response to the initial click. For example, if a fiber had a latency of 2.3 ms in response to a single click, the response to two clicks presented at an interclick interval of 2 ms would be calculated using time windows of 2.3 to 4.3 ms and 4.3 to 6.3 ms. The first method then divided these counts by the number of spikes elicited from that unit in response to a single click over a 4 ms analysis window (2.3 to 6.3 ms in the example), and expressed the result as a percentage (% single click). The single click responses were averages of four counts obtained from the responses to the initial click of a train of two or four clicks with interclick intervals of 4 and 5 ms. This method yielded values (Fig. 9(A)) that are similar to previously published results (Parham *et al.*, 1996).

For the response to the second click in the train, the recovery functions calculated using the first method were different for high and the low spontaneous units (Fig. 9(A)). In response to the presentation of just two clicks, the trailing click had a significantly better recovery for the low spontaneous units than for the high spontaneous units at interclick intervals of 1 ($p<0.05$), 2 ($p<0.02$), 4 ($p<0.02$) and 5 ms ($p<0.01$). This finding is similar to the results reported in decerebrate cats (Parham *et al.*, 1996). Doubling the number of clicks in the train to four decreased the rate of recovery for the low spontaneous units, while the recovery of the high spontaneous units was unchanged. With four clicks the recoveries of the low and high spontaneous units were not significantly different. An increase to eight clicks in the train further reduced the recovery, and both the low and high spontaneous units had similar recovery functions. For all click trains, the recovery relative to a single click response was gradual with increasing interclick interval, as reported

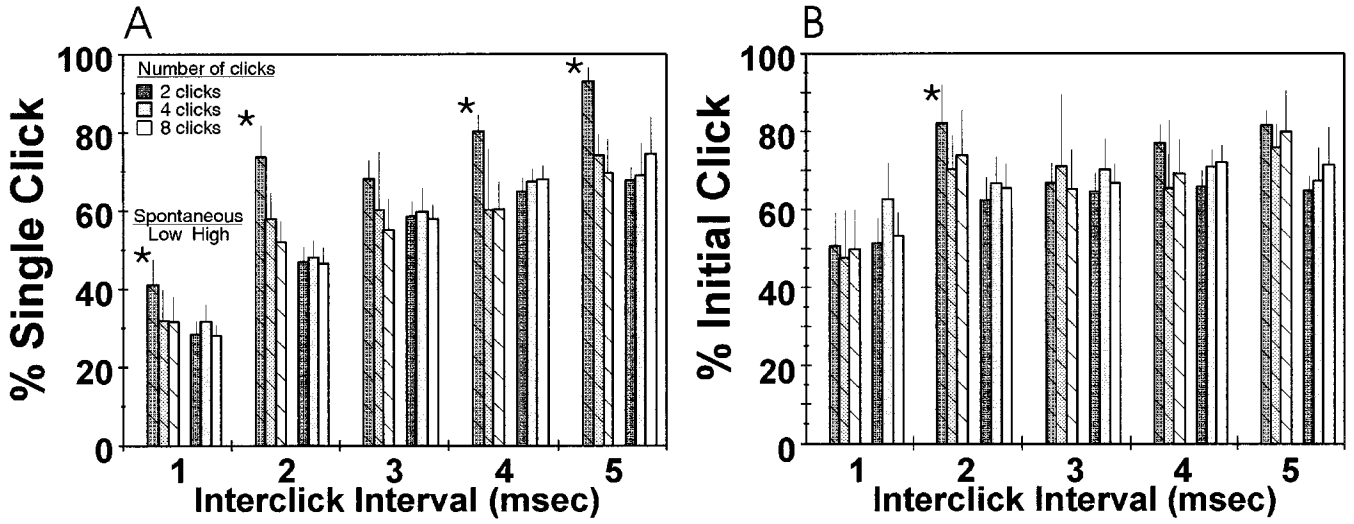


FIG. 9. The average response to the second click in a train of 2 (dark gray), 4 (light gray), or 8 (white) clicks for low (hatched) and high (unhatched) spontaneous rate auditory nerve fibers as a function of interclick interval. A. The response is expressed as a recovery function with the number of spikes elicited to the second click divided by the number evoked by a single click and expressed as a percentage. With only two clicks in the train, the percentages for the low spontaneous units were significantly (*) higher than for the high spontaneous units at interclick intervals of 1, 2, 4, and 5 ms. Increasing numbers of clicks in the train eliminated this faster recovery. B. The response is expressed as a percentage of the initial click. With this measure there is little difference between low and high spontaneous units for low and high spontaneous units. In response to two clicks with an interval of 2 ms, the low spontaneous units again outperform the high spontaneous fibers.

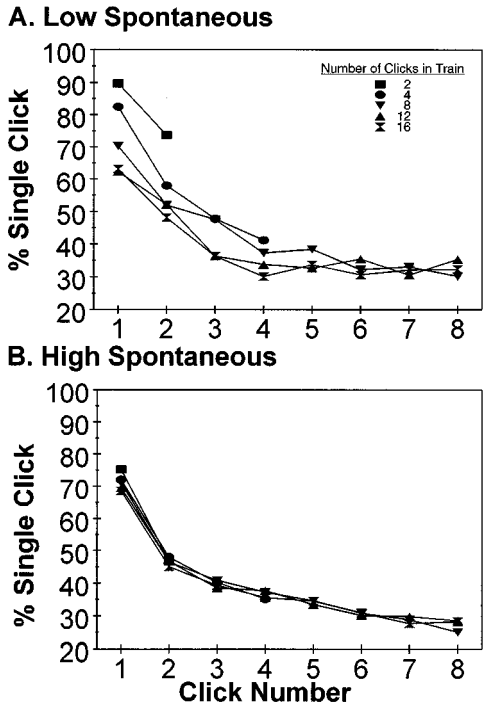


FIG. 10. The average responses to the initial clicks in a train expressed as a percentage of the response to a single click for low (A) and high (B) spontaneous units. Details of the calculation are described in the text. Click trains contained 2 (■), 4 (●), 8 (▼), 12 (▲) or 16 (⧗) clicks. The interval between clicks was 2 ms and each train was presented once every 140 ms.

previously for responses to pairs of clicks (Parham *et al.*, 1996).

This first method of calculating the recovery function compares the response with an arbitrary template and can underestimate how much information is available at any given time. An alternative method for computing the recovery function is to represent the amount of information available relative to the amount at the onset of a stimulus. For trains of clicks, this alternative method calculates the size of the trailing peak as a percentage of the number of spikes in the initial peak using equal length analysis windows. In the example given earlier, the number of spikes occurring between 4.3 and 6.3 ms would be divided by the number occurring between 2.3 and 4.3 ms. The averages of the responses to the second click, expressed as a percentage of the number of spikes elicited by the first click (% initial click), were plotted as a function of interclick interval in Fig. 9(B) for high and low spontaneous units. This measure of response was relatively constant with a value of approximately 70% for interclick intervals between 2 and 5 ms. For an interclick interval of 1 ms, the size of the second click was approximately 50% that of the initial click. For a train of two clicks with an interclick interval of 2 ms, the response of the low spontaneous units was again significantly ($p<0.05$) higher than the response for the high spontaneous fibers.

The decrease in the recovery with an increasing number of clicks in the stimulus was also observed in responses to individual clicks in a train. Figure 10 shows average percentages of the response to a single click calculated, using the first method described above, for the initial clicks (up to eight) in trains of 2, 4, 8, 12 and 16 clicks and plotted as a function of the number of the click. The click trains were all presented with an interclick interval of 2 ms. For the low spontaneous units, as the number of clicks increased from 2 to 12 the sizes of the first and second peaks decreased. The greater recovery of the response to the second click for the low spontaneous units demonstrated in Fig. 9(A) was coupled with a larger response to the initial click. With 8, 12 or 16 clicks, little change in the size of the initial peaks was observable. The responses to the later clicks, 5 through 8, did not change as the number of clicks was increased. For the high spontaneous units, there was little change in the size of any of the peaks with increasing numbers of clicks.

III. DISCUSSION

The firing patterns of chinchilla auditory nerve fibers in response to trains of clicks indicate that information was preserved even at high rates of stimulus presentation. Both high and low spontaneous units displayed similar increases in their responsiveness, as measured by the number of action potentials, with increases in the number of clicks in the train (Figs. 2 and 3). However, the high spontaneous units had a much larger loss of timing information as the click intensity was increased (Fig. 7). With changes in intensity, the low spontaneous units were better able to preserve the temporal acuity at high stimulus rates.

The increases in spike counts with increasing numbers of clicks in a train were consistent with a power-law description of the response. On linear axes, the slopes of the functions relating spike count to click number decreased significantly with increasing numbers of clicks. This decrease was eliminated by plotting the results on log–log axes (Fig. 3). For the log–log plots the functions were linear with an average slope of 0.5 at interclick intervals of 3, 4 and 5 ms, with a slightly lower value for an interval of 2 ms. Only for an interclick interval of 1 ms was the average slope significantly below 0.5 (Fig. 5). The slopes were not correlated with either spontaneous rate or characteristic frequency (Fig. 4). These findings are consistent with an interpretation that for interclick intervals equal to or longer than 2 ms, the number of spikes elicited by a train of clicks was proportional to the number of clicks in the train raised to the power 0.5.

There are two implications of the power-law description of the responses of auditory nerve fibers to trains of clicks. First, a power-law was proposed by Hafter and co-workers to express the increase in detectability of interaural time differences (ITDs) as a function of the number of clicks in a train. The increase in detectability of the ITD was measured as a decrease in the minimum detectable time difference. For interclick intervals of 4 and 5 ms, the rate of decrease was proportional to the number of clicks in the train raised to the power -0.5. A decrease in the minimum detectable ITD implies an increase in the information available for detection. The information available for judgments of ITD, therefore, should be increasing as a power function with an exponent of 0.5. While a direct comparison between the detectability of ITDs and the number of spikes elicited by a train of clicks may or may not be appropriate, the values of the exponents suggests that for interclick intervals of 4 and 5 ms the audi-

tory nerve is supplying enough information, in terms of spike counts, to account for the decreases in the detectable ITDs observed in the psychoacoustic experiments. The second implication follows from the finding that for the auditory nerve fibers, the exponent of the power function (the slope) was the same for interclick intervals between two and five milliseconds. This contrasts with the decrease in the exponent that was observed in the psychoacoustic experiments as the interclick intervals decreased. For shorter interclick intervals, the responses of the auditory nerve appear to be more robust than the psychoacoustic responses. A decrease in the exponent (or slope) for the auditory nerve responses occurred only at the shortest interclick interval (1 ms). In terms of spike counts, because auditory nerve fibers did not display the large reductions in responsiveness at short interclick intervals that were inferred from the localization tasks, the ''post-onset'' adaptation process does not occur at or prior to the auditory nerve.

These observations suggest that there may be two adaptation processes involved. The first adaptation process is observable in responses of auditory nerve fibers as a reduction in the number of spikes elicited by successive clicks in a train. This is likely the short-term adaptation process described by Smith and colleagues (Smith and Zwislocki, 1975; Westerman and Smith, 1984). This process limits the amount of information available for both localization and timbre judgments. Since the increase in the information (the number of spikes) available for localization estimates increases as a power function with an exponent of 0.5, this can account for the observed increase in detectability of the interaural time differences. The second adaptation process is the ''binaural'' process proposed by Hafter and colleagues and it occurs at a subsequent stage in the auditory system and may take place in the cochlear nuclear complex, as proposed by Hafter *et al.* (1988).

There are three minor differences between the stimuli used by Hafter and co-workers and the clicks trains presented in this study. The psychoacoustic studies used interclick intervals of 1, 2, 5, and 10 ms (Hafter and Dye, 1983; Hafter *et al.*, 1983), while the intervals between the clicks presented to the auditory nerve fibers varied from 1 to 5 ms. A 10-ms interval was not used in order to reduce the data collection time and because the psychoacoustic results for 5 and 10 ms were similar. The second difference was that Hafter and colleagues presented narrow-band clicks filtered at a center frequency of 4000 Hz, and the wideband clicks presented in this study were 100 μs, rectangular pulses. Presumably the filtering action of the cochlea narrows the effective bandwidth of the stimulus delivered to the auditory nerve fibers. The third difference was in the intensity levels of the clicks. Dye and Hafter (1984) presented 55, 75, and 95 dB clicks normalized using the root-mean-square sound pressure levels for tones with amplitudes equal to the peaks of the clicks. The clicks presented here were 59, 69 and 79 dB normalized to the sound pressure level of a 1000-Hz calibration tone with an amplitude equal to the peaks of the clicks. The lowest level clicks used in this study were 10 to 20 dB above the thresholds of the fibers. While it is difficult to directly compare levels of clicks, for neither the psychoacoustic (Dye and Hafter, 1984) nor the physiological data (Fig. 5) was there an effect of intensity on slope.

The time of occurrence at each ear of an acoustic event provides a major cue for the localization of a sound source. The values of the synchronization coefficients show that auditory nerve fibers can continue to encode these times throughout the duration of the acoustic signal, even for trains of clicks presented with short interclick intervals. At an interclick interval of 2 ms, the average synchronization coefficient for the high spontaneous units was greater than 0.3 (Fig. 7), which indicates the presence of a significant amount of timing information (cf. Fig. 6(E) and (F)). The averaging of the synchronization coefficients assumes that each nerve fiber contributes equally to the available information on the times of occurrence of acoustic events. It is not known currently what aspects of the information from the auditory nerve are used by some central processor to accomplish the lateralization task in the psychoacoustic experiments. The results from Dye and Hafter (1984) show that the detectable ITD gets smaller as intensity is increased, which reflects an increased ability to detect timing differences. This contrasts with the decrease in the synchronization coefficients with increasing intensity for both high and low spontaneous units. This difference implies that the synchronization coefficients alone are not good measures of the timing information used by the central processor, although they do demonstrate that the timing information was present. A better measure may be some combination of spike counts and synchrony, but this measure should be defined using responses from the cochlear nucleus which provides the direct inputs to the olivary complex where the signals from the two ears are compared.

The synchronization of the auditory nerve response recorded in the chinchilla is similar to the results obtained in the cat (Kiang *et al.*, 1965). While the ability of low frequency auditory nerve fibers to transmit timing information has long been known (e.g., Kiang *et al.*, 1965; Rose *et al.*, 1967), the findings presented here demonstrate that this ability also exists for high frequency fibers in response to rapid, transient stimuli such as trains of clicks. Only with the decrease in the interclick interval to 1 ms was there a decrease in the ability of the nerve fibers to transmit the timing information (Fig. 7). Further, the low spontaneous fibers continued locking their responses to the individual clicks in the train as the intensity of the clicks was increased. The continued locking demonstrates the ability of low spontaneous fibers to encode fine timing information. This temporal acuity of low spontaneous fibers contrasts with the relative loss of timing information in the high spontaneous fibers.

The ability of auditory nerve fibers to continue to increase the number of action potentials with increasing numbers of clicks in a train combined with their continued signaling of the times of acoustic events demonstrates the capacity of auditory nerve fibers to transmit information accurately. The pattern of responses of the auditory nerve was distinctly different from the changes in detectability observed in psychoacoustic experiments on lateralization. In those experiments a large reduction in detectability was found at short interclick intervals, indicating a decrease in the information available for judgments of sound source lo-

cation. The decrease was attributed by Hafter and co-workers to an adaptation process that occurs in the cochlear nucleus. The continued responsiveness of auditory nerve fibers at short interclick intervals is consistent with the location of the adaptation proposed by Hafter and colleagues. The continued responsiveness is also consistent with results on judgments of pitch and timbre (Harris *et al.*, 1963; Guttman, 1965; Hafter and Richards, 1988), which demonstrated that significant information was provided by late arriving clicks.

The recovery functions calculated using the first method for the auditory nerve in the chinchilla for pairs of clicks are comparable to those reported in the auditory nerve of the decerebrate cat (Parham *et al.*, 1996) and in the cochlear nucleus of the chinchilla (Wickesberg, 1996), but they are not as complex as those reported in the inferior colliculus (Yin, 1994; Fitzpatrick *et al.*, 1995). The first method for calculating recovery divided the number of spikes in the response to the trailing click by the response to a single click, which is similar to the technique used by Parham *et al.* (1996). The major difference between the two techniques is that the method used here does not subtract the percentage of spikes in the tail, i.e., the proportion of the response to a single click that would overlap the response to a trailing click. If the tail is small relative to the non-overlapping portion, this percentage will be small. The results from the two different techniques are very similar. The findings reported here replicate the differences in recovery rates for low and high spontaneous auditory nerve fibers observed by Parham *et al.* (1996). However, this difference was only found in the responses to pairs of clicks. Increasing the number of clicks in the train eliminates the different recovery rates. While increasing the number of clicks reduces the recovery time available to the auditory nerve fibers, psychoacoustic data (Toole and Sayers, 1965) demonstrates that lateralization judgements can be made using pairs of clicks presented as often as once every 6 ms.

An interesting characteristic of the responses to click trains was the change with increasing numbers of clicks in a train. Even the relatively small change from two to four clicks with an interclick interval of 2 ms reduced the response to both the first and second clicks. Since the increase in the number of clicks resulted in only a 4 ms (or 3%) reduction in recovery time, the change in the initial peaks was likely the result of the increase in the number of stimuli rather than a decrease in recovery time. Further, the increase from two to four clicks produced a bigger reduction in the response to the second click than in the response to the initial click. Since the responsiveness to the initial click is reduced, the change in the response to the trailing click is not likely to result from the refractory period of the fiber. One explanation for the change in responsiveness is a reduction in the available neurotransmitter.

An alternative definition of recovery was a calculation of peak size as the percentage of spikes in the initial peak (Fig. 7(B)). This definition implies that the auditory system is not comparing responses to an arbitrary template, but looking at the information available at any instant in time. The alternative definition produced very different ''recovery'' functions with only a 50% reduction in the response to the second click for a 1-ms interclick interval, and for longer interclick intervals a relatively constant 30% reduction in responsiveness. For a pair of clicks with a 2 ms interclick interval the size of the second peak averaged more than 80% of the initial peak. This alternative definition demonstrated that auditory nerve fibers were capable of effectively signaling the time of occurrence of a trailing click, even if not all spikes in the trailing peak were elicited solely by the trailing click. This conclusion is consistent with the findings of Siegel and Relkin (1987), the data of Parham *et al.* (1996), and the psychoacoustic results of Hafter and Richards (1988) and Harris *et al.* (1963).

The findings presented here demonstrate that auditory nerve fibers are able to maintain their responsiveness with rapid presentation of acoustic stimuli. The low spontaneous auditory nerve fibers preserve that temporal acuity better with increasing intensity levels than do high spontaneous fibers, although both high and low spontaneous fibers increase their firing rate as intensity is increased. This observation suggests a parallel with the sensitivity versus spatial acuity problem that must be solved in visual processing. In the fovea, there is a low convergence of receptor cells onto the retinal ganglion cells, which produces very high spatial acuity but low sensitivity. In contrast, there is a high convergence in the periphery of receptors onto the ganglion cells, which results in high sensitivity but low spatial acuity. In the auditory system, the lower threshold, high spontaneous auditory nerve fibers maintain the sensitivity of the system to low level stimuli, while the low spontaneous fibers function to maintain temporal acuity.

The responses of auditory nerve fibers to trains of clicks indicate that the adaptation process described by Hafter and colleagues is not apparent at the level of the auditory nerve and is, therefore, likely to occur in the cochlear nucleus or higher. The information is transmitted by the auditory nerve fibers to the cochlear nuclear complex, where it is sent to parallel ascending pathways. The differences in both the projection patterns of high and low spontaneous fibers (Liberman, 1991; Snyder *et al.*, 1997) and their abilities to transmit timing information may be reflected in the ability of different cochlear nucleus neurons to respond to temporal cues. Further, the ''post-onset'' adaptation process may occur in the cochlear nucleus as proposed by Hafter and co-workers (1988) in the pathway associated with sound localization, while in other parallel pathways, there may be no adaptation. These ''non-adapting'' pathways could convey the information provided by the later clicks in the train to account for the findings on pitch judgments (Hafter and Richards, 1988).

ACKNOWLEDGMENTS

The authors would like to thank Ekta Shah, Sandhya Vollala, Jon Cantey and Sujal Shah for their assistance in collection of the data and critically reading the manuscript. We would also like to express our appreciation to the two anonymous reviewers for their helpful critiques. This work was supported by the National Science Foundation (IBN-9396231).

Covey, E. (**1993**). ''The monaural nuclei of the lateral lemniscus: parallel pathways from cochlear nucleus to midbrain,'' in *The Mammalian Cochlear Nuclei—Organization and Function*, edited by M. A. Merchán, J. M. Juiz, D. A. Godfrey, and E. Mugnaini (Plenum, New York), pp. 321–334.

Dye, Jr., R. H., and Hafter, E. R. (**1984**). ''The effects of intensity on the detection of interaural differences of time in high-frequency trains of clicks,'' J. Acoust. Soc. Am. **75**, 1593–1598.

Fitzpatrick, D. C., Kuwada, S., Batra, R., and Trahiotis, C. (**1995**). ''Neural responses to simple simulated echoes in the auditory brain stem of the unanesthetized rabbit,'' J. Neurophysiol. **74**, 2469–2486.

Gaskell, H. (**1983**). ''The precedence effect,'' Hearing Res. **11**, 277–303.

Goldberg, J. M., and Brown, P. B. (**1969**). ''Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: Some physiological mechanisms of sound localization,'' J. Neurophysiol. **32**, 613–636.

Guttman, N. (**1965**). ''Binaural interactions of three clicks,'' J. Acoust. Soc. Am. **37**, 145–150.

Haas, H. (**1951**). ''On the influence of a single echo on the intelligibility of speech,'' Acustica **1**, 49–58.

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. Gall, and W. M. Cowan (Wiley, New York), pp. 647–676.

Hafter, E. R., and Dye, Jr., 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**, 644–651.

Hafter, E. R., Dye, Jr., R. H., and Wenzel, E. (**1983**). ''Detection of interaural differences of intensity of high-frequency clicks as a function of interclick interval and number,'' J. Acoust. Soc. Am. **73**, 1708–1713.

Hafter, E. R., and Richards, V. M. (**1988**). ''Discrimination of the rate of filtered impulses,'' Percept. Psychophys. **43**, 405–414.

Hafter, E. R., and Wenzel, E. M. (**1983**). ''Lateralization of transients presented at high rates: Site of the saturation effect, in *Hearing—Physiological Basis and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer-Verlag, Berlin), pp. 202–208.

Harris, G. G., Flanagan, J. L., and Watson, B. J. (**1963**). ''Binaural interaction of a click with a click pair,'' J. Acoust. Soc. Am. **35**, 672–678.

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*, Res. Monograph **35** (MIT, Cambridge, MA).

Liberman, M. C. (**1991**). ''Central projections of auditory-nerve fibers of differing spontaneous rate. I. Anteroventral cochlear nucleus,'' J. Comp. Neurol. **313**, 240–258.

Parham, K., Zhao, H. E., and Kim, D. O. (**1996**). ''Responses of auditory nerve fibers of the unanesthetized decerebrate cat to click pairs as simulated echoes,'' J. Neurophysiol. **76**, 17–29.

Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (**1967**). ''Phase-locked responses to low-frequency tones in single auditory nerve fibers of the squirrel monkey,'' J. Neurophysiol. **30**, 769–793.

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.

Sayers, B. McA., and Toole, F. E. (**1964**). ''Acoustic-image lateralization judgements with binaural transients,'' J. Acoust. Soc. Am. **36**, 1199–1205.

Siegel, J. H., and Relkin, E. M. (**1987**). ''Evidence for preseynaptic facilitation in primary cochlear afferent neurons,'' Hearing Res. **29**, 169–177.

Smith, R. L., and Zwislocki, J. J. (**1975**). ''Short-term adaptation and incremental responses of single auditory-nerve fibers,'' Biol. Cybern. **17**, 169–182.

Snyder, R. L., Leake, P. A., and Hradek, G. T. (**1977**). ''Quantitative analysis of spiral ganglion projections to the cat cochlear nucleus,'' J. Comp. Neurol. **379**, 133–149.

Takahashi, T., Moiseff, A., and Konishi, M. (**1984**). ''Time and intensity cues are processed independently in the auditory system of the owl,'' J. Neurosci. **4**, 1781–1786.

Toole, F. E., and Sayers, B. McA. (**1965**). ''Inferences of neutral activity associated with binaural acoustic images,'' J. Acoust. Soc. Am. **38**, 769–779.

Wallach, H., Newman, E. B., and Rosenzweig, M. R. (**1949**). ''The precedence effect in sound localization,'' Am. J. Psychol. **52**, 315–336.

Westerman, L. A., and Smith, R. L. (**1984**). ''Rapid and short-term adaptation in auditory nerve responses,'' Hearing Res. **15**, 249–260.

Wickesberg, R. E. (**1996**). ''Rapid inhibition in the cochlear nuclear complex of the chinchilla,'' J. Acoust. Soc. Am. **100**, 1691–1702.

Wickesberg, R. E., and Oertel, D. (**1988**). ''Tonotopic projection from the dorsal to the anteroventral cochlear nucleus of mice,'' J. Comp. Neurol. **268**, 389–399.

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.

Wickesberg, R. E., Whitlon, D., and Oertel, D. (**1991**). ''Tuberculoventral neurons project to the multipolar cell area but not the octopus cell area of the posteroventral cochlear nucleus,'' J. Comp. Neurol. **313**, 457–468.

Yin, T. C. T. (**1994**). ''Physiological correlates of the precedence effect and summing localization in the inferior colliculus of the cat,'' J. Neurosci. **14**, 5170–5186.

Yost, W. A., and Soderquist, D. R. (**1984**). ''The precedence effect: Revisited,'' J. Acoust. Soc. Am. **76**, 1377–1383.

Effects of wide band inhibitors in the dorsal cochlear nucleus. II. Model calculations of the responses to complex sounds

Jacob J. Blum
Division of Physiology, Department of Cell Biology, Duke University Medical Center, Durham, North Carolina 27710

Michael C. Reed
Department of Mathematics, Duke University, Durham, North Carolina 27708-0320

(Received 9 May 1997; revised 9 January 1998; accepted 12 January 1998)

In a previous modeling study of signal processing in the dorsal cochlear nucleus [Reed and Blum, J. Acoust. Soc. Am **96**, 1442–1453 (1997)] it was shown that inclusion of a wideband inhibitor (WBI) greatly improved the fit between model response maps and the experimental response maps of type IV units to pure tones. In this study we examine the effect of the WBI on the responses to complex sound stimuli such as broadband noise (BBN), notch noise, noise bands, and band/notch combinations. A new and more realistic model for auditory nerve (AN) response in the presence of different levels of noise is used. It is shown that one can explain and understand the qualitative features of virtually all the published data on type II and type IV unit responses to BBN, notch noise and noise bands. The monotone decreasing response of the maximum firing rate of type II units to noise bands of increasing width that is observed experimentally occurs in the model due to the increasing inhibition of type II cells by the WBI. Similarly, the various nonmonotone patterns of maximum firing rate of type IV units to noise bands of increasing width is shown to arise from the complex and highly nonlinear effects of inhibition from the type II to type IV and the WBI to type IV cells and the nonlinear direct excitation from the AN to the type IV cells. A number of experiments using double notches, double noise bands, or notch-noise band pairs are suggested which, by comparison with model results, would allow one to infer probable connectional patterns between type II and type IV units and between the WBI units and the type IV units. [S0001-4966(98)04104-6]

PACS numbers: 43.64.Qh, 43.64.Bt [RDF]

INTRODUCTION

It has been known for some time that type II units inhibit type IV units in the dorsal cochlear nucleus (DCN); see Voigt and Young (1990) and references therein. It was initially suggested by Young *et al.* (1988) that there are other inhibitory inputs to type IV units, since massive stimulation of the auditory nerve inhibits both type II and type IV units. This question was investigated in detail by Nelken and Young (1994) who studied the responses of type IV units to complex sounds such as notch noise and noise bands. On the basis of these studies they argued for the existence of units which strongly inhibit type II units and weakly inhibit type IV units. Since this inhibition occurred over a wide range of sound frequencies, Nelken and Young termed these units wide-band inhibitors (WBI). Reed and Blum (1995) and Blum *et al.* (1995) showed that the inclusion of a WBI cell which inhibits model type II cells (T2) markedly improves the fits to much of the experimental data. In those two modeling studies, no direct connections from the WBI inhibitors to the model type IV cells (T4) were included. In those papers we tentatively identified these inhibitory cells as stellate cells before the term WBI was introduced. For the purposes of the model, it does not matter what the unit is called, so for consistency we will continue to use the term stellate cell as well as WBI. Following the publication of Nelken and Young (1994), it was of great interest to ascertain whether adding to the model such direct connections would further improve the fits to experimental data. Indeed, it was found that including the WBI connection to model type IV cells improves the similarity between model and experimental response maps to pure tones (Reed and Blum, 1997). In particular, the model responses now had upper inhibitory sidebands—the main feature lacking in our 1994 and 1995 studies—and the fit to data at lower frequencies was markedly improved.

In this study we examine the effect of including direct WBI connections to model type IV cells when the system is stimulated by complex sounds such as broadband noise (BBN), notch noises of various depths centered on best frequency (BF), and noise bands centered on BF. We show that the model rate functions capture the essential features of the data of Nelken and Young (1994). The model allows one to investigate why the rate-intensity curves exhibit the various complex patterns observed, and also to predict how the rate-intensity curves would look if more complicated sound inputs such as double notches, double noisebands, or notch-noise band pairs are given. Experimental tests of these predictions would provide additional insight into the strengths and connectional features of the DCN units being investigated. Thus the theoretical model can be used to design experiments to extract information about particular features of the anatomy of the DCN connections.

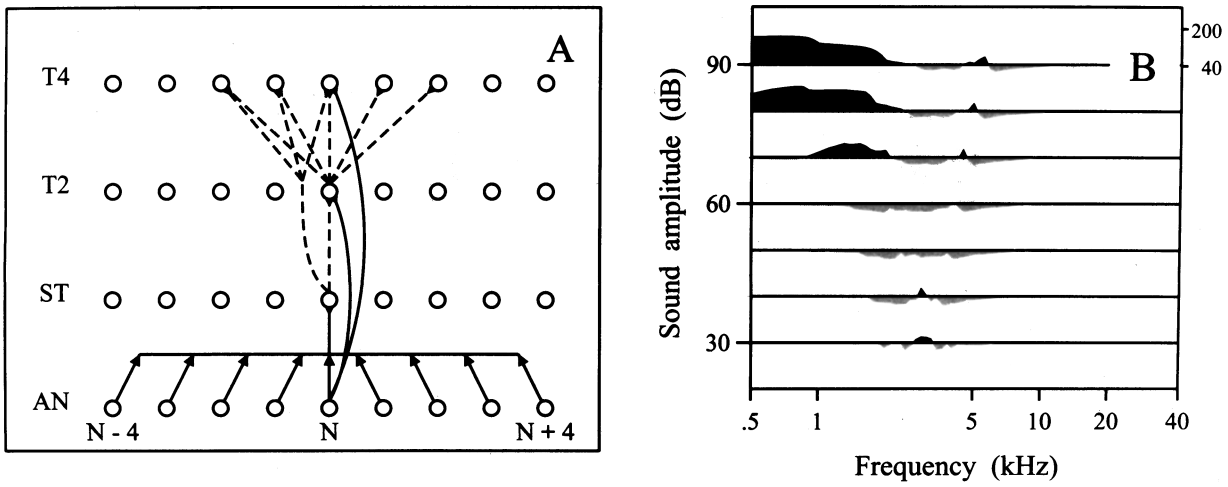


FIG. 1. Panel A shows connectional scheme #1. Solid lines indicate excitatory connections and the dashed lines inhibitory connections. Each row corresponds to a different type of cell, 100 to each row, with best frequency increasing from left to right. We give first the connections and strengths (in parentheses) which remain unchanged for all simulations shown unless otherwise specified. Nine AN fibers converge on each ST cell (0.11). In addition, each AN fiber sends a single excitatory projection to the T2 (1.0) and T4 (1.0) cells at the same BF. Each ST cell inhibits the T2 cell at the same BF (0.62). These connections and parameter values are as in our previous studies. For this connectional scheme, each ST cell also inhibits T4 cells N, $N-1$, and $N-2$ (''three to the left'') with strengths, σ to be specified. The Nth T2 cell inhibits the $N-2$, $N-1$, N, $N+1$, and $N+2$ T4 cells. The connectional pattern is repeated for all values of N. Panel B shows the response map of T4 cell #6 when each ST-to-T4 connection has strength $\sigma=0.1$, and the T2-to-T4 connectional strengths are 0.166, 0.332, 0.664, 0.332, 0.166 for T4 cells $N-2$, $N-1$, N, $N+1$, and $N+2$, respectively, as in RB95 and BRD95. The response map shows the spike output in response to pure tones at different frequencies at seven different sound amplitudes. The scale for spike output in spikes/second is shown at the upper right of panel B.

I. METHODS

A. Connectional patterns

The computational model introduced in Reed and Blum (1995), hereafter referred to as RB95, consists of auditory nerve (AN) fibers, stellate cells (ST), type II cells (T2), and type IV cells (T4). The input–output relations of the ST, T2, and T4 cells are as discussed in RB95. For this study, which deals with complex sound inputs, we have utilized a more realistic AN input–output relationship than used in Blum *et al.* (1995), hereafter referred to as BRD95, as will be described in detail below. There are three major patterns in the pure-tone response maps recorded by Spirou and Young (1991) for 35 ''standard'' and 14 ''nonstandard'' type IV cells. In our model calculations, these three patterns arise from the directional bias of the T2-to-T4 connections as discussed in detail in Reed and Blum (1997). Therefore, in this paper we restrict our attention to these three connectional patterns. Figures 1, 2, and 3 show both the connectional patterns and the general features of the corresponding response maps. The points on the abscissas in Figs. 1A, 2A, and 3A correspond to the BF of the cells located above, with lower BF to the left and higher BF to the right. All three patterns have the following features in common. Each ST cell receives excitatory input from nine AN fibers as shown. Each AN fiber, in addition, makes a direct excitatory connection to the T2 and T4 cells at the same BF. Each ST cell makes a

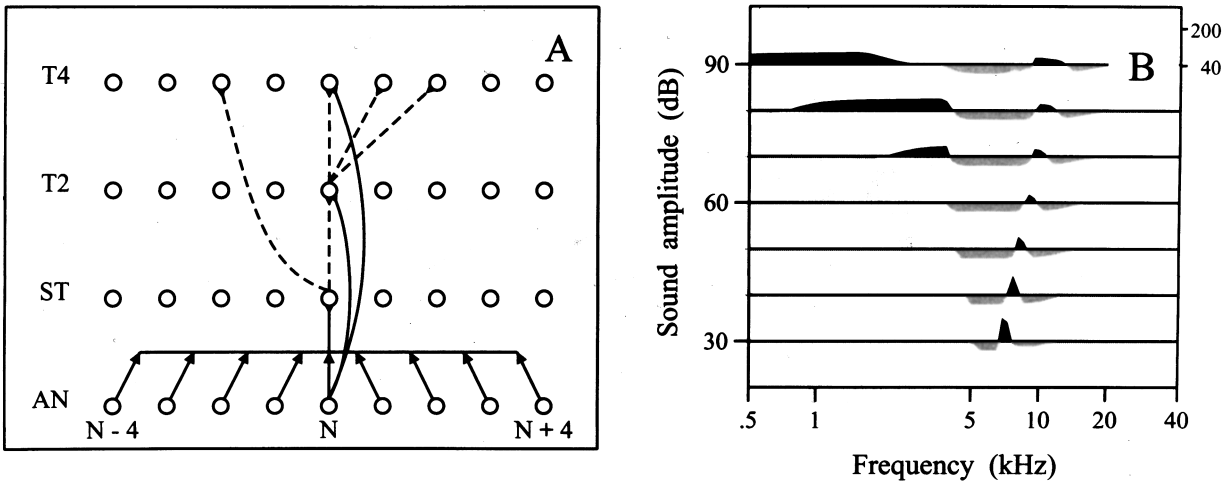


FIG. 2. Panel A shows connectional scheme #2. This differs from connectional scheme #1 in that ST cell N connects to T4 cell $N-2$ and T2 cell N connects to T4 cells N, $N+1$, and $N+2$. Panel B shows the response map of T4 cell 14 when the ST-to-T4 connection has strength $\sigma=0.7$ and each T2-to-T4 connection has strength 1.66.

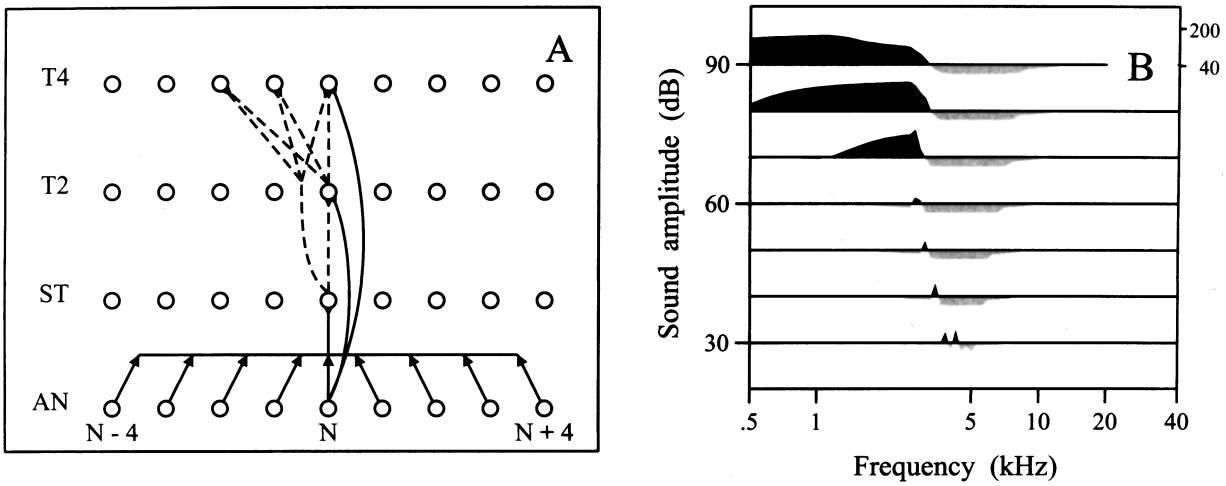


FIG. 3. Panel A shows connectional scheme #3. This differs from connectional scheme #1 in that ST cell #N connects to T4 cells $N-2$, $N-1$, and N and the Nth T2 cell also connects to the $N-2$, $N-1$, and N T4 cells. Panel B shows the response map of T4 neuron 8 when each ST-to-T4 connection has strength $\sigma=0.1$ and each T2-to-T4 connection has strength 1.66.

direct inhibitory connection to the T2 cell at the same BF. The ST-to-T4 and the T2-to-T4 connections differ in the three patterns.

In the first connectional pattern, Fig. 1A, the ST-to-T4 connections are three to the left and the T2-to-T4 connections are five straight up. As discussed in our previous papers, the response map to pure tones depends on the BF of and the connectional pattern to the particular T4 neuron being examined. This connectional pattern results in response maps which have an isolated BF excitatory region at low amplitudes and both low- and high-frequency excitatory regions at high amplitudes. The response map for T4 neuron 6 for this connectional pattern is shown in Fig. 1B. This response map pattern is similar to that of the ''standard cell C'' in Fig. 2 of Spirou and Young (1991), Fig. 4b of Young *et al.* (1992), and Fig. 2 of Nelken and Young (1994).

In the second connectional pattern, Fig. 2A, each ST cell makes a single connection to the T4 cell two the left. Each T2 cell projects to three T4 cells with a right bias. Figure 2B shows the pure-tone response map for T4 neuron 14. This connectional pattern results in response maps in which the BF excitatory region is no longer isolated but extends from BF at low amplitudes to higher frequencies at higher amplitudes. This band of excitation is separated from the low-frequency, high amplitude excitatory region. This response map pattern is similar to that of the ''standard'' cell A in Fig. 2 and ''nonstandard'' cell B in Fig. 3 of Spirou and Young (1991), as well as Fig. 2b of Young *et al.* (1992) and Fig. 4B of Voigt and Young (1990).

In the third connectional pattern, Fig. 3A, both the ST-to-T4 and the T2-to-T4 connections project three to the left. Figure 3B shows the pure-tone response map for T4 neuron 8. This connectional pattern results in response maps in which the BF excitatory region is no longer isolated but extends from BF at low amplitudes to lower frequencies at higher amplitudes. There is no high-frequency, high amplitude excitatory region. This response map pattern is similar to that of the ''standard'' cell B in Fig. 2 and ''nonstandard'' cell A in Fig. 3 of Spirou and Young (1991), as well as the type IV cells in Fig. 3 of Young and Brownell (1976).

Each of the connections in the model has a ''strength'' which is a scale factor that multiplies the output of the presynaptic cell before it is used as input for the post-synaptic cell. In this paper and in our previous papers, the scale factor for AN-to-ST is 0.11 for each of the nine connections, is 1.0 for AN-to-T2 and AN-to-T4, and is 0.62 for ST-to-T2. For connectional pattern #1 the scale factors for the T2-to-T4 connections are 1.66 times 0.1, 0.2, 0.5, 0.2, and 0.1, respectively, proceeding from left to right. The ST-to-T4 connections are each 0.1. For pattern #2 the scale factors for the T2-to-T4 connections are each 1.66 and that for the ST-to-T4 connection is 0.7. For pattern #3 the scale factors for the T2-to-T4 connections are each 1.66 and each of the ST-to-T4 connections is 0.1. The strengths of the T2-to-T4 connections were chosen as a result of the investigations in RB95 and BRD95. The ST-to-T4 scale factors were chosen as a result of the investigations in Reed and Blum (1997); we examine the effect of varying these strengths below. In what follows we denote the strength of the ST-to-T4 inhibitory connections by σ.

B. Auditory nerve fiber properties

The steady state spike output of AN fibers in response to pure tones at various frequencies and amplitudes is relatively simple to characterize. Each AN cell has a tuning curve (which depends on BF) and fires at a rate above its spontaneous rate only if the amplitude-frequency pair is above the tuning curve. The firing rate increases with amplitude above the curve. Although there is considerable variation between fibers, we chose in RB95 a simple sigmoidal response which satisfactorily represents many AN fibers (see, e.g., Kiang, 1984). This AN model was also used in Reed and Blum (1997).

In contrast, modeling the steady state spike output of AN fibers in response to complex sounds is fraught with great difficulties. This is because of the highly nonlinear properties of the cochlea which lead to phenomena such as

two-tone suppression and to reduced responses to pure tones in the presence of noise backgrounds. In general, AN fibers respond less strongly if fibers at neighboring frequencies are also stimulated. Furthermore, these properties vary even among fibers at the same BF depending, for example, on their spontaneous rates. In our previous study on broadband and notch noise (BRD95), these difficulties were ignored. Given a complex sound with different amplitudes at different frequencies, we assumed that the output of each AN fiber was simply the output that it would produce if the sound were a pure tone with that amplitude at its BF. Thus a spectral notch in the sound input was reproduced as a notch in the outputs of the family of AN fibers in the model. Studies on AN responses to spectral notches (Poon and Brugge, 1993) show that this is approximately true. In our previous paper, however, we did not take into account that the individual AN response functions are quite different in the presence of background noise (Gibson *et al.*, 1985). In particular, with increasing noise amplitude, the AN response often has a higher spontaneous rate, a higher threshold, a more gradual slope, and a lower saturation level (Costalupes *et al.*, 1984; Ruggero, 1973; Schalk and Sachs, 1980; Geisler and Sinex, 1980). The extent of these changes depends on many factors including the spontaneous rate of the AN fiber (Smith, 1988; Winter *et al.*, 1990; Winter and Palmer, 1995) and the frequency range and amplitude of the noise. Since it is not known which type of AN fibers connect to any particular type II or type IV cells, we have chosen to model the AN output in the presence of noise by a Michaelis–Menten-type formula with a single parameter λ given in units of dB. This parameter depends on the sound levels at the AN fiber's BF and on the sound levels at neighboring frequencies. The algorithm for computing λ is described below. Once λ is computed, the response in spikes/s of the AN fiber under consideration is given in terms of the sound amplitude, I, at its BF, and λ by the formula:

$$AN(I,\lambda)=R(\lambda)+\frac{A(\lambda)(I-T(\lambda))^{\beta(\lambda)}}{40+(I-T(\lambda))^{\beta(\lambda)}},\quad I-T(\lambda)\geqslant 0, \tag{1}$$

$$AN(I,\lambda)=R(\lambda),\quad I-T(\lambda)<0. \tag{2}$$

We assume that the spontaneous rate, $R(\lambda)$, the threshold $T(\lambda)$, the saturation level $A(\lambda)$, and the sigmoidal exponent $\beta(\lambda)$ depend linearly on λ in the range 0–100 dB:

$$R(\lambda)=10+\frac{40}{100}\lambda, \tag{3}$$

$$T(\lambda)=20+\frac{30}{100}\lambda, \tag{4}$$

$$A(\lambda)=200-\frac{60}{100}\lambda, \tag{5}$$

$$\beta(\lambda)=2-\frac{\lambda}{100}. \tag{6}$$

The slopes of the linear dependencies on λ were chosen as a rough approximation to effect of noise on these four characteristics of AN response functions as shown in the experimental papers referred to above.

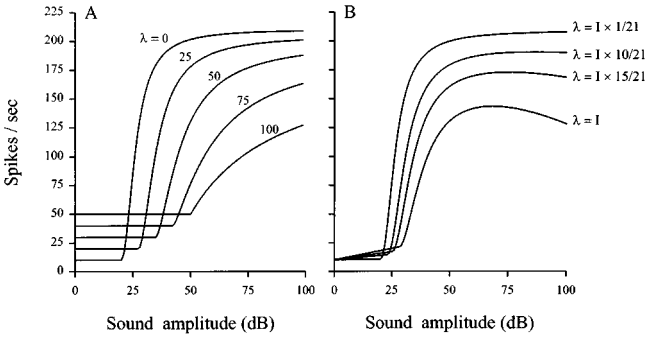


FIG. 4. Panel A shows the response of AN fibers as I is increased for different fixed levels, λ, of background noise. The analytical formulae are given in Eqs. (1)–(6). Panel B shows the output of an AN fiber to noise bands centered at its BF. The widths of the bands are 1, 9, 15, and 21 units.

For every complex sound the noise parameter, $\lambda(N)$, is chosen differently for each of the 100 AN fibers, as follows. For the Nth AN fiber $\lambda(N)$ is the average sound input to the 21 fibers closest in frequency to the Nth fiber, i.e., the average of the inputs to AN fibers $N-10,...,N,\ldots,N+10$. Recall that the BF of the AN fibers in the model increases as N goes from 1 to 100. Thus $\lambda(N)$ averages (on a scale of 0–100 dB) the sound input to fibers with best frequencies near to the BF of fiber N. For example, suppose that AN fibers 20 and 50–70 are stimulated by an input sound. Since fibers 10–19 and 21–30 are not being stimulated, the parameter $\lambda(20)$ will be very small and the response of AN fiber 20 will be close to that for a pure tone, as depicted in Fig. 4A for $\lambda=0$. On the other hand, the sound levels at fibers 50–70 will affect the parameter $\lambda(60)$ and will therefore change the output behavior of AN fiber #60 as the sound level increases.

Figure 4A shows graphs of the AN response function for five different values of λ. The value $\lambda=0$ corresponds to the AN response to a pure tone at BF in the absence of noise; this is the simplistic response function which was used in BRD95. It is important to remember that parameter λ for a given AN fiber depends on the sound levels at the neighboring fibers. To see how this works in a simple case, consider noise bands of different widths centered at the BF of the AN fiber under consideration. The sound level at BF, I, is the same as the sound level in the rest of the noise band. If the band has width 1, then $\lambda=\frac{1}{21}I$. If the band has width 9, then $\lambda=\frac{9}{21}I$. As the band widens, λ becomes a larger and larger fraction of I and the response of the AN fiber changes as shown in Fig. 4B. For a noise band wider than 21 there is no additional effect. Thus as the noise band widens the threshold of the AN fiber increases, the response slope decreases, the saturation level decreases. Note that below threshold, there is small increase in firing rate as I is raised because of the effect of neighboring AN fibers on the spontaneous rate of the fiber under consideration. Also, for very wide noise bands ($\lambda=I$) there is a slight decrease in AN output at very high sound amplitudes. This occurs in the model because the AN fiber is already saturated and the only change with increasing amplitude is the decrease in saturation level [see Eq. (5)]. This effect has been seen in only a few BBN experi-

ments; see, for example, Figs. 13 and 14 of Schalk and Sachs (1980) and Fig. 13 of Ruggero (1973). In any case, this small effect, which occurs only at very high amplitudes, has little or no effect on the results described below.

C. Definition of terms and units

In our model, each of the 100 AN fibers (see Figs. 1 and 2) has a different BF with BF increasing as N increases. In BRD95 and in Reed and Blum (1997) where we study model response maps to pure tones, a specific correspondence between the AN fiber number and BF was specified. However, in BRD95 and in this paper there is no specific correspondence between neuron number and BF. Thus each AN fiber has the the same input–output relation which depends only on the inputs to fibers $N-10$ to $N+10$. Thus ''frequency'' is specified by AN fiber number. We note that we use the term dB to specify the strength of the sound input to a particular model AN fiber. If there were a specific correspondence between the AN fiber number and frequency, this ''dB level'' would correspond to the ''spectrum level'' at that frequency, that is, the energy in a fixed narrow width band about that frequency.

II. RESULTS

It is useful to summarize the general features of type IV responses to broadband noise, notch noise, and noise bands.

(1) Experimental studies show a variety of type IV response curves to increasing BBN levels. In almost every case there is an initial moderate increase in spike output with increasing sound level. With further increases in sound level, there may be: continued increase to a moderate saturation level (cf. Figs. 9 and 10, cell 74-11-2, of Young and Brownell, 1976; Fig. 11D of Spirou and Young, 1991; Figs. 6B and 6D of Nelken and Young, 1994); a gradual decrease (Fig. 10, cells 74-11-3 and 74-13-3, of Young and Brownell, 1976); or a gradual decrease followed by a gradual increase (Figs. 2A, 3A, and 6F of Nelken and Young, 1994).

(2) The general response to a notch centered at BF is inhibition at moderate intensities followed by excitation at higher sound intensities. As the notch gets wider the inhibitory dip initially gets deeper and then gets shallower until at sufficiently wide notch widths the response curve approaches that of BBN. These properties are the main features of the experimental data seen, for example, in Figs. 12D and 12E of Spirou and Young (1991), and Figs. 2B and 6 of Nelken and Young (1994).

(3) The general response to noise bands centered at BF is an increase of spike output at low to moderate sound amplitudes. This increase is followed by complete inhibition (cf. Figs. 6A and 6C of Nelken and Young, 1994), or inhibition followed by excitation at higher sound amplitudes (Fig. 6E of Nelken and Young, 1994, and Fig. 13B of Young *et al.*, 1988).

The purpose of this modeling study is not just to capture these features but to provide an in-depth explanation of the complex, nonlinear interactions which give rise to the variety of experimental patterns observed. Furthermore, we explain how other complex sounds (such as double notches or double noise bands) can be used to obtain insight into the connectional patterns from the type II and WBI neurons to the type IV neuron being observed.

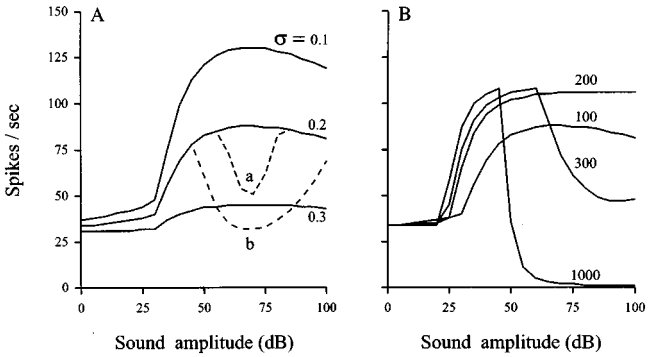


FIG. 5. Panel A shows the response to BBN of a T4 cell in connectional scheme #1. The solid lines have the parameter values specified in the legend to Fig. 1 except that $\sigma=0.1$, 0.2, and 0.3 as indicated. If the strength, σ, of the ST-to-T4 inhibition is kept at 0.2 but the T2-to-T4 connectional strengths are each reduced by a factor of 0.48 and the ST-to-T2 connection is reduced in strength from 0.62 to 0.47, the T4 response curve is the same except for a large dip labeled ''a.'' If σ is kept at 0.2 but the T2-to-T4 connectional strengths are each reduced by a factor of 0.12 and the ST-to-T2 connection is reduced in strength from 0.62 to 0.40, the T4 response now has a larger dip labeled ''b.'' For $\sigma=0.2$, panel B shows the effect of changing the coefficients of λ in Eqs. (3)–(6) as discussed in the text.

A. Response to broadband noise

Figure 5A shows the response to BBN of a T4 cell in the model with connectional pattern #1 for three different values of the strength, σ, of the ST-to-T4 wide-band inhibitor connections. The three curves have the same general shape, a gradual increase after threshold followed by slight decrease in spike output at high amplitudes. The overall response decreases markedly as σ increases. The slight decrease at high sound levels is due to the decrease in AN saturation level, $A(\lambda)$ [see Eqs. (1) and (4)] at high noise levels. In each of these cases there is no output from the T2 cells at any sound level because they are inhibited by the ST cells. If the ST-to-T2 inhibition is weakened, the T2 cells will fire and partially inhibit the T4 cells over a range of BBN sound levels. Two examples of this phenomenon are shown by the curves labeled a and b in Fig. 5A. At sufficiently high sound amplitudes, the T2 cells are again inhibited by the ST cells and therefore the T4 output rises again. The depth of the dip in T4 output depends on the strength of the inhibitory T2-to-T4 connections. Similar results are obtained for connectional patterns #2 and #3 (simulations not shown). Thus the details of the connectional pattern should have relatively little effect on the response of a type IV unit to BBN as compared to the effect of the strength of the inhibitory connections to that unit. It is not surprising that in a network with two different sources of inhibition, one can obtain different complicated response functions to even so ''simple'' a sound as BBN. In our model, the AN response functions and the parameters and connection patterns are the same across the entire frequency range. Local and global variations of these properties would cause even more variation in the T4 response curves to BBN, as would the introduction of different connectional patterns between, for example, the AN and ST cells and the

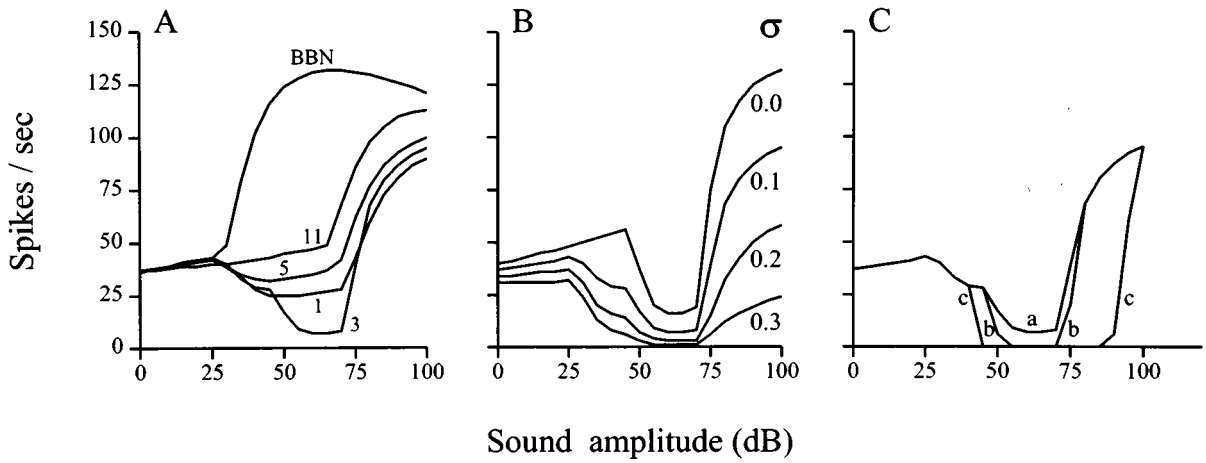


FIG. 6. Response of T4 cells to notch noise. Panel A shows the response of a T4 cell with connectional scheme #3 to BBN and to a series of notches, centered at BF, of depth 30 dB and widths 1, 3, 5, 11, as shown. In panel B the width and depth of the notch are kept constant at 3 and 30 dB, respectively, and the strength, σ, of the ST-to-T4 connections is varied from 0 to 0.3 as shown. Simulation ''a'' in panel C is the same as simulation ''3'' in panel A. In simulation ''b'' the strength of the T2-to-T4 connections has been increased by a factor of 3. In simulation ''c'' the strength of the T2-to-T4 connections has been increased by a factor of 3 and the strength of the ST-to-T2 connection has been lowered from 0.62 to 0.50.

ST and T2 cells. Nevertheless, the model experiments show how the qualitative features of the BBN response curves of type IV cells depend on the interplay between the type II inhibition of type IV cells and the WBI inhibition of both type II and type IV cells.

In the computations shown in Fig. 5A we used the algorithm for computing the effects of noise on AN output described in Sec. I. In particular, once λ is computed for a particular AN fiber from the sound at all frequencies, the effect of λ on the AN parameters is given by Eqs. (3)–(6). The magnitude of the coefficients of λ in these equations determines how strongly the sound at surrounding frequencies affects the output of the AN fiber under consideration. We chose these coefficients, which are $\frac{40}{100}$, $\frac{30}{100}$, $\frac{60}{100}$, and $\frac{1}{100}$, respectively, to give a fairly strong effect. Figure 5B shows the effect on the BBN response (with $\sigma=0.2$) of reducing the magnitude of these coefficients. The curve labeled 100 is the same as the $\sigma=0.2$ curve in Fig. 5A. The curves labeled 200, 300, and 1000, were obtained, respectively, by replacing 100 by these these values in the coefficients of λ. Reducing the effect of λ by $\frac{1}{2}$, i.e., the case 200, results in an increase in slope and a higher plateau level. Note that the decrease in output at high amplitudes is no longer present. Reducing the coefficient of λ by $\frac{1}{3}$, i.e., the case 300, causes an increase in initial slope, followed by a marked decline at high sound levels. When λ is reduced by $\frac{1}{10}$ the T4 response is similar to that for a pure tone at BF, which is reasonable since we have eliminated virtually all of the effects of sounds at the surrounding frequencies. In the rest of this paper we use the values shown in Eqs. (3)–(6) that give a reasonably strong effect of the surrounding sound.

B. Response to notch noise

Figure 6A shows the effect of changing notch width for a fixed notch depth of 30 dB for a T4 cell with connectional pattern #3. Similar results were obtained for connectional patterns #1 and #2 (simulations not shown). The behavior observed—inhibition at moderate intensities followed by excitation at higher sound intensities—is the main feature of the experimental data referred to above. Furthermore, the progression of curves as the notch width increases is similar to the progression seen in the experimental data, i.e., the inhibitory dip initially gets deeper and then gets shallower until at sufficiently wide notch widths the response curve approaches that of BBN. In Fig. 6 (and in Fig. 10), dB refers to the spectrum level in the noise bands to the right and left of the notch, as in the graphs shown by Young and coworkers.

There are potentially two different sources of inhibitory input to T4 cells in the model—the inhibitory connections from the T2 cells and the (weak) inhibitory connections from the ST cells. Both kinds of inhibition are evident in the response curve to the notch of width 3 in Fig. 6A. The initial decrease in T4 output (up to approximately 45 dB) results from ST inhibition. By examining the output of the ST and T2 cells (simulations not shown), we ascertained that the sharp dip between 45 and 65 dB is caused by T2 inhibition. At higher sound levels, the ST cells turn off the T2 inhibition and the weak ST inhibition of T4 cells is then overwhelmed by direct excitation from AN fibers. The T2 inhibition is a ''shoulder effect'' of the notch as discussed in detail in BRD95. Whether this T2 effect occurs depends on notch width. For example, it does not occur for this cell for notch widths of 1 and 5. When it does occur, it causes an inhibitory dip over a relatively narrow dB range; this occurs in the response to notch width 3 in Fig. 6A and in the simulations shown in BRD95. In that paper, in which wide-band inhibition of the T4 cells was not included (i.e., there were no ST-to-T4 connections), we attributed the experimental inhibitory dip in response to notch noise entirely to the T2 effect. It is now clear that the inhibitory dip can arise from both sources, T2-to-T4 and ST-to-T4, but that it is the ST inhibition to the T4 cells which is responsible for the breadth of the dip.

In Fig. 6A, each of the three ST-to-T4 connections has strength $\sigma=0.1$. The effect of changing this strength is

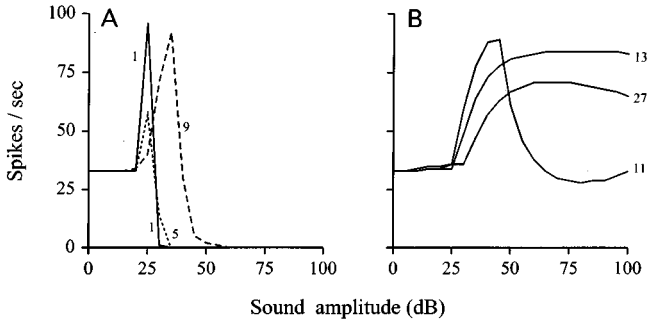


FIG. 7. Response of T4 cells to noise bands. Panel A shows the response of a T4 cell with connectional scheme #2 to a series of noise bands with widths 1 (solid line), 5 (dotted), and 9 (dashed). Panel B shows the response of the same T4 cell for noise bands widths of 11, 13, and 27.

shown in Fig. 6B for a notch width of 3. If $\sigma = 0$ then there is no ST inhibition of the T4 cell and only the narrow inhibition caused by the T2 cell remains. With increasing σ the breadth as well as the depth of the inhibitory dip increases.

It was also of interest to examine the effect of changing both the strengths of the T2-to-T4 inhibition and of the ST-to-T2 inhibition. Increasing the strength of each of the T2-to-T4 connections from 1.66 to 5.0 causes an appreciable deepening of the dip without affecting its width (Fig. 6C, curve b). If now one also decreases the strength of the direct ST-to-T2 inhibition, the inhibitory dip in T4 response becomes much wider because the T2 cells are less inhibited (Fig. 6C, curve c).

C. Response to noise bands

In our previous study (BRD95), the response of T4 cells to noise bands was examined. In that paper, the noise band, centered at BF, was accompanied by broadband noise (at a lower amplitude) at all other frequencies and there were no ST-to-T4 connections. In this paper, the noise bands have no energy outside the band itself (except for Fig. 10), as in the experimental studies of Young *et al.* (1988) and Nelken and Young (1994), and the ST-to-T4 connections are included. Figures 7A and 7B show the responses of a T4 cell in connectional pattern #2 to noise bands centered at BF of widths 1, 5, 9, and 11, 13, and 27, respectively. For band width 1, i.e., essentially a pure tone, there is a sharp increase in spike output at threshold, followed by complete inhibition. At width 3 the result is similar to width 1 (not shown). At width 5, the increase is much smaller. At width 9, the excitatory response increases and the peak is much broader. At width 11, the unit is no longer completely inhibited at high sound amplitudes. At higher widths the response becomes similar to the response to BBN. The reason for these progessive changes in the T4 response is that the ST cells are turned on more and more as the noise band gets wider, thus inhibiting the T2 cells and allowing the T4 cells to fire at high sound amplitudes.

These simulations capture the essential features seen in the experimental data referred to above. Furthermore, the progression of curves as the noise band widens is very similar to that seen in Fig. 9 of Young and Brownell (1976), Fig. 13B of Young *et al.* (1988), and Fig. 12A of Nelken and Young (1994).

The interplay between ST and T2 inhibition of T4 cells can be seen in Fig. 8, which shows the outputs of all model neurons near BF at a sound amplitude of 50 dB. In column A the noise band has width 1. At this amplitude the AN fiber excites the T2 neuron at the same BF, which in turn inhibits the corresponding T4 neuron. The stellates are hardly turned on because only one AN fiber is firing. For a noise band of width 5 (column B) at the same sound amplitude, the stellate cells close to BF are fully activated, thereby inhibiting the T2 cells close to BF. In fact, at all amplitudes the T2 output is a monotone decreasing function of the width of the noise band as found experimentally by Nelken and Young (1994). The inhibition of the T2 cells permits increased firing by the T4 cells near BF. This effect can be seen even more clearly at width 21 (column C) where the T2 cells are inhibited over a wide region near BF. We remark that for these simulations the ST-to-T4 connections skip two to the left and the T2-to-T4 connections are three to the right (see Fig. 2). The consequences of this particular connectional scheme can be seen in the details of the pattern of T4 responses of the cells adjacent to T4 neuron #50. Note that although the sound level input is 50 dB for each AN fiber within the noise band, the AN responses are not uniform. This is because the AN fibers near the middle of the band sense a noisier environment and therefore their noise parameter, λ, is set higher (and their output is lower) as explained in Sec. I.

Nelken and Young (1994) examined the dependence of maximal firing rate for type II and type IV cells on noise bandwidth. For type II cells the maximal firing rate was a monotone decreasing function of bandwidth for all cells examined (see their Fig. 14, Panel A). In our model, the maximal firing rate of the T2 cells also decreases monotonically with bandwidth as shown in panel A of Fig. 9. Curve a shows the maximal firing rate of a T2 cell in connectional pattern #1 with all the standard parameters. Curve b shows that reducing the strength of the ST-to-T2 inhibitory connections from 0.62 to 0.5 causes a more gradual decline in T2 output. Similar results (not shown) are obtained for the other connectional patterns.

In contrast, the maximal firing rates of type IV cells showed a wide range of behavior as bandwidth was increased (Nelken and Young, 1994, Fig. 14, panel A). Some of the curves increased monotonically, some decreased monotonically, but the majority exhibited complex nonmonotonic behavior. Curves a and b in Fig. 9, panel B, show the maximal firing rates of T4 cells with connectional patterns #1 and #2, respectively, as a function of bandwidth. In both cases one sees the same kind of nonmonotone behavior as observed by Nelken and Young. This behavior results from the complex interplay of the nonlinear excitation (from AN) with the two sources of nonlinear inhibition (from ST and T2 cells). Monotone behavior can also be obtained in the model. If one increases σ from 0.1 to 0.4 in connectional scheme #1, then the ST-to-T4 inhibition dominates and the maximal T4 output decreases monotonically with increasing bandwidth as shown in curve c of panel B. If one allows the AN-to-T4 connections to spread across neighboring isofre-

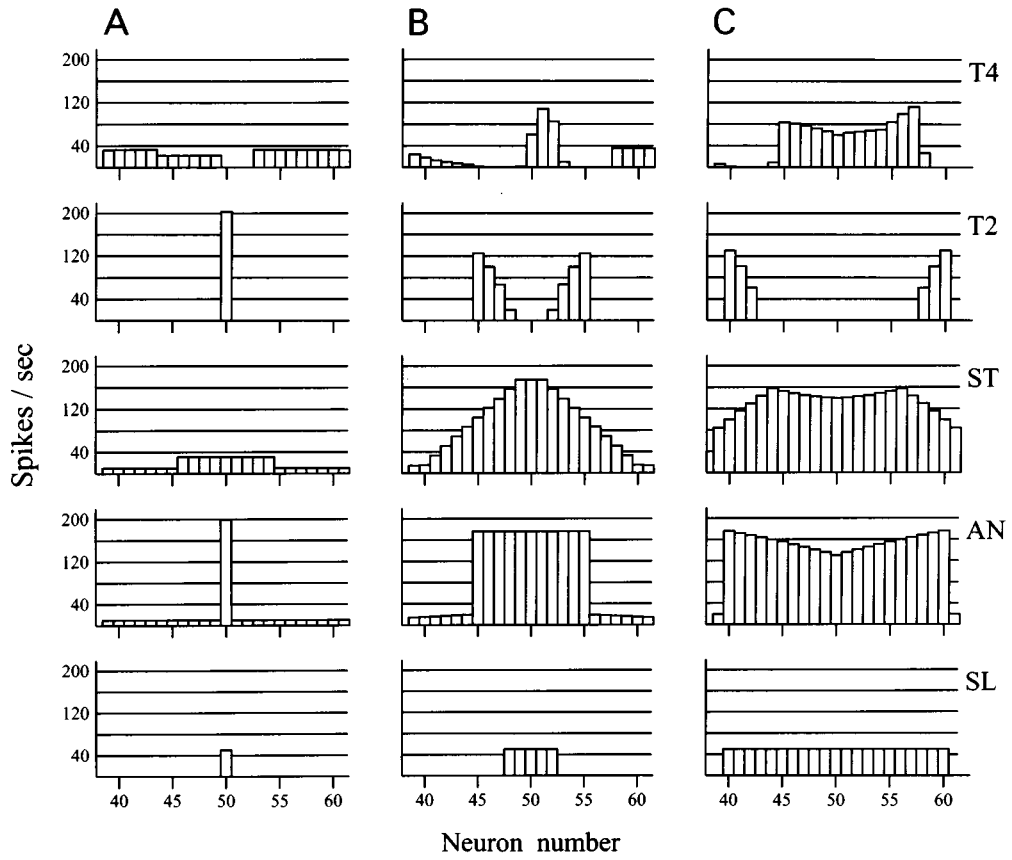


FIG. 8. Population response to noise bands. Connectional scheme #2 is used. The bottom row shows (for noise band widths 1, 5, and 21) the dB input to each AN fiber for *N* ranging from 39 to 61. All other AN fibers have zero input. The spike outputs of the AN fibers, ST cells, T2 cells, and T4 cells are shown for each of the three noise bandwidths.

quency sheets, then one obtains a curve which is monotone increasing except for a small decrease at high bandwidth as shown in curve d of panel B. The maximal firing rate increases in this case because the excitatory drive on the T4 cell increases with increasing bandwidth. Further increase in bandwidth beyond neighboring frequencies does not increase the excitatory drive and, in fact, AN output falls somewhat because the noise parameter λ is higher, thereby decreasing maximal T4 firing rate at high bandwidths. Similar behavior is seen for some of the type IV cells examined by Nelken and Young (1994, Fig. 14, panel A).

D. Predicted responses to complex sounds

Figure 10 shows the response of T4 neuron 50 to BBN, to two notches of depth 30 dB, width 3, located at AN neurons 46, 47, 48, and 52, 53, 54, and to two noise bands at the same locations. The double notch causes a large inhibitory dip at moderate sound amplitudes due entirely to the inhibitory effect of T2 neurons. For notches of lower depth, the inhibitory dip is narrower and for notches of greater depth the inhibitory dip is wider (simulations not shown). For double noise bands, the T4 response is flat followed by complete inhibition at higher sound amplitudes. Although Fig. 10 shows only the responses for connectional pattern #1, the responses for the other patterns are very similar (simulations not shown). This is because of the symmetry around BF of the stimulating sounds. For noise bands, if the T2-to-T4 pattern is three to the right (connectional pattern #2), then T4 neuron 50 is inhibited by T2 neuron 48, which is turned on by the lower noise band. If the T2-to-T4 pattern is three to the left (connectional pattern #3), then T4 neuron 50 is inhibited by T2 neuron 52, which is turned on by the upper noise band. In the case of connectional pattern #1, where the T2 inhibition spreads both to the right and to the left, T4 neuron 50 is inhibited by both T2 neurons 48 and 52. Note that the presence of an ST-to-T4 connection of strength 0.2 causes a marked decrease of the response to a double notch at high sound amplitudes, but has little effect on the response to a double noise band.

If the symmetry of the stimulating sound is broken, then one can detect substantial differences in the T4 responses to sounds off BF due to different connectional patterns. For example, suppose that one presents the following complex sound and then increases the sound amplitude equally at all frequencies. All AN fibers except 46–48 and 52–54 are stimulated by 15-dB sound input, while fibers 46–48 are stimulated with 30-dB input and fibers 52–54 are have 0-dB input. Thus there is a noise band of width 3 and height 15 dB on the left of neuron 50 and a notch of width 3 and depth 15 dB on its right. The response of T4 neuron 50 to increasing sound level is shown in Fig. 11A for each of the three connectional patterns. Panel B shows the responses of T4 neuron

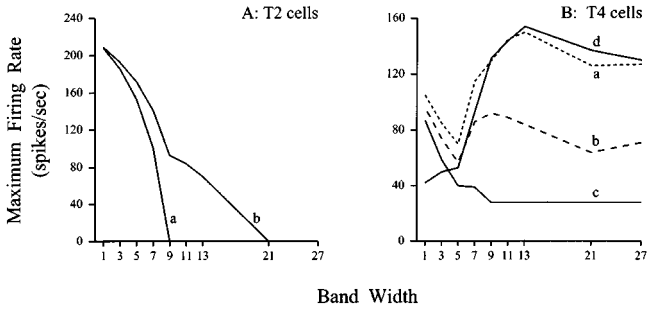


FIG. 9. Dependence of maximum firing rate on bandwidth. Panel A shows the maximal firing rate (over the range 0–100 dB) of two model T2 cells. Curve a: connectional pattern #1 with the standard parameters. Curve b: the strength of the inhibition from ST units to T2 units has been reduced from 0.62 to 0.50. Panel B shows the maximal firing rate (over the range 0–100 dB) of four model T4 cells as a function of the bandwidth (i.e., number of AN fibers stimulated) of noise centered at BF. Curves a and b: connectional patterns #1 and #2, respectively, with the standard parameters. Curve c: connectional pattern #1 and the strength of the inhibitory ST-to-T4 connections has been increased from 0.1 to 0.4. Curve d: connectional pattern #1 except that the single connection of strength 1 from the Nth AN fiber to the Nth T4 unit at the same BF has been changed to nine connections projecting from the Nth AN fiber to nine T4 cells, $N-4,N-3,...,N+4$, each with strength 0.11.

50 when the positions of the noise band and notch are reversed. For connectional pattern #1, where the T2-to-T4 connections are symmetric, there is no difference in the T4 response for these two opposite situations. For connectional pattern #2 (T2-to-T4 connections three to the right), if the noise band is on the right of BF a mild inhibition is followed by an increase to a plateau. If, however, the noise band is on the left of BF, a marked inhibition preceeds the rise to the plateau. Exactly the opposite occurs for connectional pattern #3, i.e., the marked inhibition occurs when the noise band is on the right. Thus experiments with paired notch noise and noise bands off BF have the potential to reveal the T2-to-T4 connectional pattern to that T4 unit.

III. DISCUSSION

Experimental studies have shown that several cell types exist in the DCN and that the physiological responses of DCN units to sound inputs such as pure tones, broadband

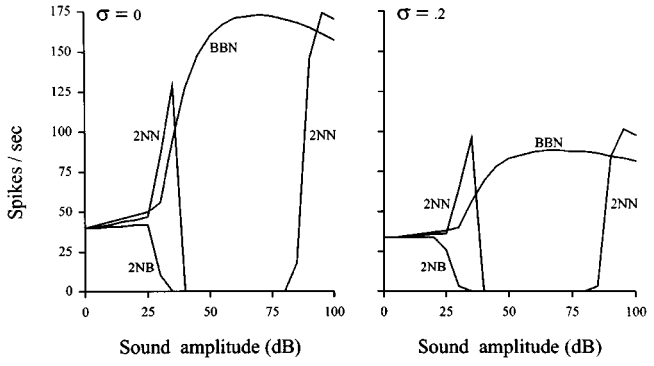


FIG. 10. Response to double notches and to double noise bands: effect of the strength of the ST-to-T4 connections. Each panel shows the response of T4 cell 50 with connectional pattern #1 to broadband noise (BBN), double notches of width 3 and depth 30 centered at $N=47$ and $N=53$ (2NN), and double noise bands of width 3 also centered at numbers 47 and 53 (2NB). The strength of the ST-to-T4 inhibition, σ, is shown in each panel.

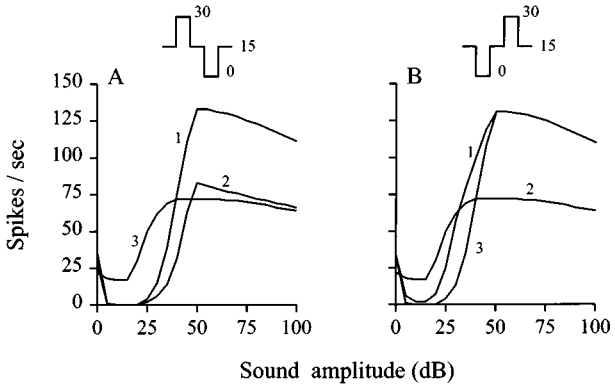


FIG. 11. Responses to noise band notch pairs. Each panel shows the response of T4 cell 50 with the indicated connectional pattern (#1, #2, or #3) to a noise band notch pair centered at neuron numbers $N=47$ and $N=53$ as indicated. The notch and noise band depths are 15 dB as shown and the widths are three neurons. In panel A the noise band is to the left of neuron #50 and the notch is to the right. In panel B the locations are reversed.

noise, notch noise, and noise bands are varied and complex. Certain features of response maps to pure tones, such as upper, lower, and BF excitatory regions, central inhibitory areas, and upper inhibitory sidebands, were observed in many type IV cells. However, each particular cell exhibited some or all of these features to a larger or smaller extent. Similarly, the responses to more complex sounds vary considerably from cell to cell but have some common features. These studies led experimentalists, such as Young and coworkers, to propose connectional schemes between type II, type IV, and wide-band inhibitors, that could account for these features.

The present modeling study lends strong support to the existence of a wide-band inhibitor in the DCN. As explained in Reed and Blum (1997), inclusion of ST-to-T4 connections produces an upper inhibitory sideband in response maps, the main feature that was missing in our earlier study of response maps (Reed and Blum, 1995) and significantly improves the quantitative fit to data in lower-frequency regions of the response map. The present study shows that the inclusion of the ST-to-T4 connections allows one to account for the wide variety of responses observed by Young and collaborators in response to notches, noise bands, and broadband noise.

In our previous study of responses to complex sounds (Blum *et al.*, 1995), we attributed the inhibitory dip in T4 response to notches as sound amplitude is raised to the inhibitory effect of T2 cells. It is now apparent that both T2 inhibition and ST inhibition is required to explain the width of the dip, since T2 inhibition alone gives a very narrow dip. This was pointed out by Nelken and Young (1997) and is confirmed by our present study as shown in Fig. 6 and the associated discussion.

In the present study we show that the model T4 cells show rate-level responses to noise bands that are similar to those in the experimental literature. Furthermore, the progression of rate-level response curves as band width increases (see Fig. 7) follows the same progression as the experimental curves observed by Young and Brownell (1976, Fig. 9) and Young *et al.* (1988, Fig. 13). Nelken and Young

emphasize in their 1994 study that the maximal firing rate of type II units is monotone decreasing with increasing bandwidth, whereas type IV units show a great variety of different behaviors. We have shown that our model cells have these behaviors for different choices of connectional patterns and strengths and we have given explanations for each type of behavior observed.

One of the powers of modelling based on specific connectional patterns is that it enables one to understand the likely reasons for specific features in the experimental data. Young and co-workers have repeatedly emphasized that narrow noisebands and narrow notches (centered on BF) by themselves result in inhibition of type IV units, while the sum of these stimuli produces excitation. These nonlinear effects can be completely understood on the basis of our model. A narrow noise band turns on T2 cells which inhibit the T4 cell. A ''shoulder effect'' at the edge of a narrow notch also turns on the T2 cells and inhibits the T4 cell (for a detailed explanation, see BRD95). The sum of these two stimuli, i.e., BBN, excites the T4 cell because there is no shoulder effect and the wide-band nature of the stimulus produces strong stellate firing which inhibits the T2 cells thus preventing inhibition of the T4 cell.

Another important use of models is to suggest experiments which can provide insight into the connectional patterns of any specific unit being examined experimentally. In our discussion of paired notches and noise bands, we showed how such stimuli can be used to show the probable direction of connections (e.g., connectional patterns #1, #2, or #3) from the type II units leading to the particular type IV unit being studied. To determine the probable direction of the WBI connection to a particular type IV unit one can proceed as follows. First, determine the amplitude of a pure tone at BF which excites the type IV unit but which has low enough amplitude that type II inhibition is not activated. Then add to the pure tone a noise band well above BF. The noise band will excite the WBI units at frequencies near the noise band. As one moves the noise band closer to the BF of the type IV unit under study, one will see some inhibition of the type IV response when the WBI's connected to that type IV unit are activated. This experiment can also be done by moving the noise band toward BF from lower frequencies. By comparing the two results one will obtain a strong indication of the direction of the WBI to type IV connections on this particular unit.

There are, of course, many simplifying assumptions in the present model. Each isofrequency sheet has only one AN fiber, ST cell, T2 cell, and T4 cell. Only three connectional patterns are examined and each of them has only a single connection from the *N*th AN fiber to the *N*th T2 cell and the *N*th T4 cell. We have made frequency independent assumptions about the effect of noise on AN response (see Sec. I). Finally, the connectional pattern is identical across the frequency axis. Despite these limitations and the simple input–output relations for the cells in the model, the present model not only accounts for but gives explanations of virtually all of the observed steady state experimental data on type IV unit responses, both to pure tones and to complex sounds.

ACKNOWLEDGMENTS

This research was supported by NSF Grant No. DMS9501404. The authors are grateful to John Davies for the preparation of software and figures. We are also grateful to I. Nelken and E. Young for providing us with preprints of unpublished work.

Blum, J. J., Reed, M., and Davies, J. M. (**1995**). ''A computational model for signal processing by the dorsal cochlear nucleus. II. Responses to broadband and notch noise,'' J. Acoust. Soc. Am. **98**, 181–191.

Costalupes, J. A., Young, E. D., and Gibson, D. J. (**1984**). ''Effects of continuous noise backgrounds on rate response of auditory nerve fibers in cat,'' J. Neurophysiol. **51**, 1326–1344.

Geisler, D., and Sinex, D. G. (**1980**). ''Responses of primary auditory fibers to combined noise and tonal stimuli,'' Hearing Res. **3**, 317–334.

Gibson, D. J., Young, E. D., and Costalupes, J. A. (**1985**). ''Similarity of dynamic range adjustment in auditory nerve and cochlear nuclei,'' J. Neurophysiol. **53**, 940–958.

Kiang, N. Y. S. (**1984**). ''Peripheral neural processing of auditory information,'' in *Handbook of Physiology, Section 1. The Nervous System, Vol. III*, edited by I. Darian-Smith Sensory Processes (American Physiological, Society, Bethesda), Chap. 15, pp. 639–674.

Nelken, I., and Young, E. D. (**1994**). ''Two separate inhibitory mechanisms shape the responses of dorsal cochlear nucleus Type IV units to narrowband and wide-band stimuli,'' J. Neurophysiol. **71**, 2446–2462.

Nelken, I., and Young, E. D. (**1997**). ''Linear and non-linear spectral integration in Type IV neurons of the dorsal cochlear nucleus: I. Regions of linear interaction,'' pre-print.

Poon, P. W. F., and Brugge, J. F. (**1993**). ''Sensitivities of auditory nerve fibers to spectral notches,'' J. Neurophysiol. **70**, 655–666.

Reed, M. C., and Blum, J. J. (**1995**). ''A computational model for signal processing by the dorsal cochlear nucleus. I. responses to pure tones,'' J. Acoust. Soc. Am. **96**, 1442–1453.

Reed, M. C., and Blum, J. J. (**1997**). ''Envelope coding in the auditory brainstem,'' in *Simulation in the Medical Sciences*, edited by J. G. Anderson and M. Katzper (Society for Computer Simulation International, San Diego), pp. 182–190.

Ruggero, M. A. (**1973**). ''Response to noise of auditory nerve fibers in the squirrel monkey,'' J. Neurophysiol. **36**, 569–587.

Schalk, T. B., and Sachs, M. B. (**1980**). ''Non linearities in auditory nerve responses to band limited noise,'' J. Acoust. Soc. Am. **67**, 903–913.

Smith, R. L. (**1988**). ''Encoding of sound intensity by auditory neurons,'' in *Auditory Function, Neurobiological Bases of Hearing*, edited by G. Edelman, W. Gall, and W. Cowan (Wiley, New York), pp. 243–276.

Spirou, G., and Young, E. (**1991**). ''Organization of dorsal cochlear nucleus type IV unit response maps and their relationship to activation by band-limited noise,'' J. Neurophysiol. **66**, 1750–1768.

Voigt, H. F., and Young, E. D. (**1990**). ''Cross-correlation analysis of inhibitory interactions in dorsal cochlear nucleus,'' J. Neurophysiol. **64**, 1590–1610.

Winter, I. M., Robertson, D., and Graeme, K. Y. (**1990**). ''Diversity of characteristic frequency rate-intensity functions in guinea pig auditory nerve fibers,'' Hearing Res. **45**, 191–202.

Winter, I. M., and Palmer, A. R. (**1995**). ''Level dependence of cochlear nucleus onset unit responses and facilitation by second tones or broadband noise,'' J. Neurophysiol. **73**, 141–159.

Young, E. D., and Brownell, W. E. (**1976**). ''Responses to tones and noise of single cells in dorsal cochlear nucleus of unanesthetized cats,'' J. Neurophysiol. **39**, 282–300.

Young, E., Shofner, W., White, J., Robert, J.-M., and Voigt, H. (**1988**). ''Response properties of cochlear nucleus neurons in relationship to physiological mechanisms,'' in *Auditory Function: Neurological Bases of Hearing*, edited by G. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 277–312.

Young, E., Spirou, G., Rice, J., and Voigt, H. (**1992**). ''Neural organization and responses to complex stimuli in the dorsal cochlear nucleus,'' Philos. Trans. R. Soc. London, Ser. B **336**, 407–419.

Basilar membrane nonlinearity and loudness[a)]

Robert S. Schlauch,[b)] Jeffrey J. DiGiovanni, and Dennis T. Ries
Department of Communication Disorders, University of Minnesota, Minneapolis, Minnesota 55455

(Received 26 August 1997; revised 17 November 1997; accepted 7 January 1998)

Loudness matching functions for tones for persons with one shifted-threshold ear (hearing loss and noise-shifted thresholds) and one ear within normal limits were used to derive the presumed basilar membrane (BM) input–output (I/O) function in a normal ear. The comparison was made by assuming that the BM I/O function for the ear with the cochlear threshold shift has a slope of one (a linearized cochlea). The function for the normal ear was derived from the loudness matching function based on this assumption. Comparisons were made for archival basilar membrane data [M. A. Ruggero, N. C. Rich, A. Recio, S. S. Narayan, and L. Robles, J. Acoust. Soc. Am. **101**, 2151–2163 (1997)] for chinchilla and archival loudness matches for long-duration tones for persons with various degrees of cochlear hearing loss [F. Miskolczy-Fodor, J. Acoust. Soc. Am. **32**, 486–492 (1960)]. Comparisons were made also between BM I/O functions and ones derived from loudness matches for persons with unilateral hearing loss simulated by broadband noise. The results show a close resemblance between the basilar membrane I/O function and the function derived from loudness matches for long-duration tones, even though the comparison was between human and chinchilla data. As the degree of threshold shift increases from 40 to 80 dB, the derived BM I/O functions become shallower, with slopes for losses of 60 dB or more falling in the range of values reported for physiological data. Additional measures with short-duration tones in noise show that the slope of the loudness function and the slope of the derived basilar membrane I/O function are associated with the behavioral threshold for the tone. The results for long-duration tones suggest a correspondence between BM displacement and loudness perception in cases of recruitment, but the relation between the degree of loss and the amount of BM compression and the relation between signal duration and compression suggests that other factors, such as the neural population response, may play a role. [S0001-4966(98)05104-2]

PACS numbers: 43.66.Ba, 43.66.Cb, 43.66.Dc, 43.64.Bt [RVS]

INTRODUCTION

The basilar membrane input–output function in a healthy mammal (e.g., chinchilla and guinea pig) is nonlinear and can be described as a compressive function with a slope of 0.13–0.45 for most of its dynamic range when basilar membrane velocity (dB) or displacement (dB) is plotted as a function of sound pressure level (SPL) (Cooper and Yates, 1994; Ruggero *et al.*, 1997).

Yates *et al.* (1990) presented evidence that they argue supports the notion that the basilar membrane determines the form of the loudness function, which is usually described as a compressive function of intensity for persons with normal hearing sensitivity (Stevens, 1970). Yates *et al.* (1990) compared a BM I/O function (derived from single unit data) to loudness magnitude estimation (ME) data for a 1.0-kHz tone from a study by Viemeister and Bacon (1988). Viemeister and Bacon's (1988) loudness slopes (re: sound pressure) for their three subjects were 0.182, 0.192, and 0.11, which bear a striking resemblance to the derived basilar membrane curves with a slope of about 0.2. They concluded that ''loudness perception may be based on a simple coding of basilar membrane displacement at CF'' (characteristic frequency) (Yates *et al.*, 1990; pg. 217).

The striking resemblance between the BM I/O functions in Yates *et al.* (1990) and the loudness ME functions from Viemeister and Bacon's study is interesting, but the generality of this result is debatable when the loudness ME data are considered in the context of other studies. In an analysis of 78 different studies of loudness ME, Hellman (1991) found that the mean value of the slope is 0.6,[1] the same value adopted for an international standard (ISO/R 131-1959). By contrast, the slopes obtained by Viemeister and Bacon (1988) are more than three standard deviations below the average value found in Hellman's retrospective analysis. This discrepancy in the loudness data presents a problem for comparing the *absolute* values of BM I/O measures, a peripheral process, and loudness scaling, a process that is possibly influenced by more central factors (Zeng *et al.*, 1988). This criticism does not imply that such a comparison should not be made, but rather that *relative* measures (loudness matching) might provide a method of examining this idea in greater depth. A good candidate for such a comparison is a cochlear threshold shift, a result known to affect loudness and the BM I/O function. If the loudness function is determined by the BM I/O function, it follows that cochlear hearing loss, which alters the shape of the BM I/O function, should alter the shape of the loudness function in a predictable manner. Recent experiments demonstrate that the basilar membrane I/O function becomes steeper with certain

[a)]Portions of this paper were presented at the 19th Midwinter Research Meeting of the Association for Research in Otolaryngology, February 1996, St. Petersburg.

[b)]Electronic mail: Schla001@maroon.tc.umn.edu

types of cochlear hearing losses, chiefly ones that affect the outer hair cells (Ruggero, 1992). It is well known that cochlear hearing loss can cause loudness recruitment, an abnormally rapid growth of loudness. In loudness recruitment, the loudness function (log loudness versus dB SPL) is steeper for much of its range than it is in someone with hearing thresholds within normal limits. This comparison suggests that the form of the BM I/O function and loudness are in qualitative agreement in cochlear hearing loss; that is, both become steeper with cochlear threshold shifts. The intent of our study is to compare these two phenomena quantitatively.

I. GENERAL METHOD

Our method is based on loudness matches between an ear with hearing within normal limits and a threshold-shifted ear. One cannot derive unique functions from matching data if neither function contributing to the matching function is known. However, if one of the functions contributing to the matching function is known, the other function is defined uniquely. For our analysis, we assumed that the slope of the BM I/O function for the threshold shifted ear is 1.0 as is reported when the active process is not operative (Ruggero, 1992). As mentioned above, the BM I/O function is a compressive function of intensity in a healthy ear. When the active process is inoperative, the system shows a loss in gain of between 40 and 80 dB and the I/O function becomes linear (Ruggero *et al.*, 1997). By deriving the BM I/O function for a healthy ear from the loudness-matching function assuming that the BM I/O function for the threshold-shifted ear has a slope of 1.0, we can compare BM I/O functions and loudness functions directly. If the form of the loudness function is determined by BM displacement, our derived functions for normal ears should resemble the BM I/O functions for healthy ears in cases where the hearing loss ear contributing to the loudness matching function had a loss that equals or exceeds the gain of the active process (i.e., in situations presumably yielding a linear cochlear response).

II. ANALYSIS OF COCHLEAR HEARING LOSS: LONG-DURATION TONES

Figure 1 illustrates loudness matching data for tones from a study by Miskolczy-Fodor (1960) for listeners with unilateral cochlear hearing losses of 40, 50, 60, or 80 dB SPL. These data represent loudness matches from 200 published studies and from 100 subjects from Miskolczy-Fodor's laboratory. The test frequency was not specified, but studies show that the form of the loudness function does not change much for persons with normal hearing at frequencies between 500 and 8000 Hz (Scharf, 1978), the likely range of frequencies selected for these measurements.

Lines were fitted to the group data in Fig. 1 to quantify the slope of the matching function for different degrees of hearing loss.[2] The lines were anchored at threshold in both ears and the best-fitting line was found using an iterative process that found the intercept and slope that minimized the sum of squared deviations of the data from the predicted line. This procedure was used: (1) for the entire range of tested levels; and (2) for levels between threshold and 30-dB sensation level (SL) in the impaired ear. The second analysis, with the range of levels restricted to 30 dB above threshold, was performed to examine the slope of the matching function for levels below where complete recruitment usually occurs (some authors fit the data in the region of recruitment and the region above this point with separate lines). Our analysis of Miskolczy-Fodor's data, shown in the upper portion of Table I, demonstrates that the same trend holds for both methods; the slope of the loudness-matching function becomes steeper with increased hearing loss. This finding is consistent with

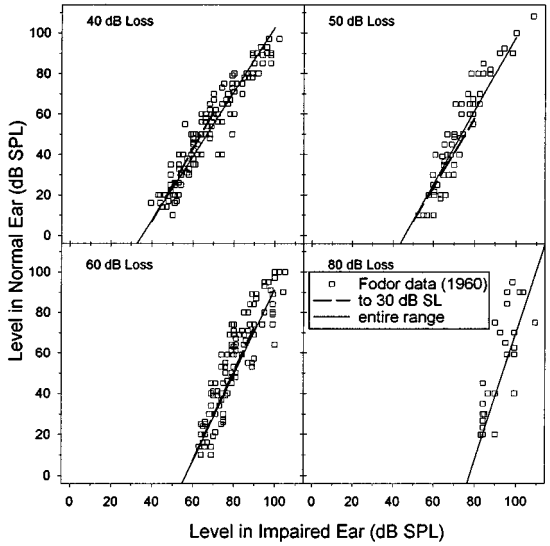


FIG. 1. Loudness matches for persons with unilateral losses from a study by Miskolczy-Fodor (1960) are denoted by squares. The upper-left and upper-right panels show results for persons with 40 dB SPL losses and 50 dB SPL losses, respectively. The lower-left and lower-right panels show results for persons with 60 dB SPL and 80 dB SPL losses, respectively. Regression lines were fitted to the entire data set in each panel (solid lines) and to the data between threshold and 30-dB sensation level (dashed lines). In the lower right panel, the regression lines for both fits are nearly identical. The slope of these regression lines increases with the amount of hearing loss.

TABLE I. The slope of the loudness-matching function categorized by degree of cochlear threshold shift. Data are from a study by Miskolczy-Fodor (1960) for persons with unilateral cochlear hearing loss and from the present study for persons with simulated unilateral hearing loss. Slopes were calculated using two methods. In one method, the entire range of data was used. In the other method, data between threshold and 30-dB sensation level in the poor ear were fitted. For both methods, the line was anchored at threshold in both ears and slope was varied in small steps to minimize the sum of squared deviations of the data from the prediction.

Detection threshold (poor ear, dB SPL)	Slope of the matching function	
	Threshold to 30 dB SL	Entire range of data
Miskolczy-Fodor (1960)		
40	1.79	1.59
50	1.68	1.81
60	2.17	2.10
80	3.12	3.12
Present study		
38 (250 ms)	1.18	1.38
60 (250 ms)	2.17	2.17
62 (2 ms)	1.77	1.82
83 (2 ms)	5.43	5.43

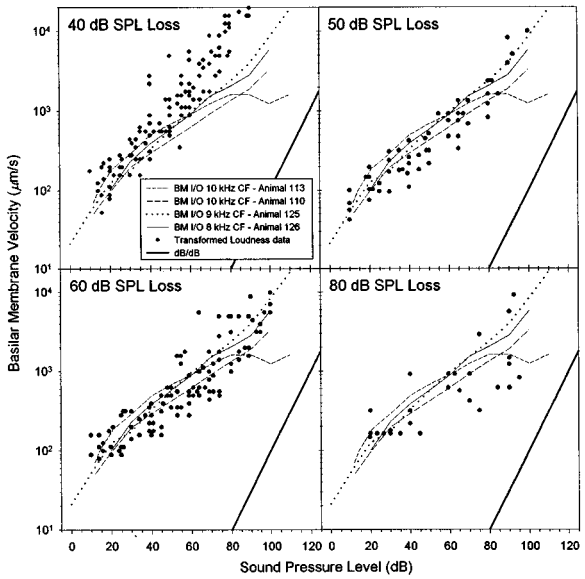


FIG. 2. A comparison of basilar-membrane (BM) input–output (I/O) functions from a study by Ruggero *et al.* (1997) and BM I/O functions derived from Miskolczy-Fodor's (1960) loudness-matching data (Fig. 1). Lines in each panel represent BM I/O functions for the same four chinchillas from the study by Ruggero *et al.* (1997). Filled circles represent BM I/O data for normal ears derived from loudness matching data from the study by Miskolczy-Fodor (1960). As in Fig. 1, Miskolczy-Fodor's (1960) data are categorized by the degree of hearing loss.

studies of loudness matching (Hellman, 1993), magnitude scaling (Hellman and Meiselman, 1990) and cross-modality matching of line length and loudness (Hellman and Meiselman, 1993).

Figure 2 illustrates BM I/O functions for normal preparations at CF[3] and BM functions for normal ears derived from Miskolczy-Fodor's (1960) loudness-matching data between impaired ears and a normal ears. The four BM I/O functions from healthy chinchilla ears are from a study by Ruggero *et al.* (1997). Ruggero *et al.* (1997) measured BM responses from 129 chinchillas and obtained useful data from 43 of their preparations. The four shown in our figure represent the four most sensitive BM responses to CF tones for which measurements were made between threshold and at least 90 dB SPL. The slopes of these four functions range from 0.2 to 0.42 (mean = 0.32) for stimulus levels between 50 dB SPL and 90 dB SPL.

The derived BM I/O functions are based on the simple assumption that the hearing loss ear is linear. For this analysis, the slope of the derived function is simply the inverse of the matching function (i.e., the ordinate in Fig. 1 is the abscissa in Fig. 2). To compare quantities for the derived functions (loudness) to the physiological data (velocities), ordinal values on the derived functions in Fig. 2 were scaled so that values for 20 dB SPL were lined up for both data sets. For the 40 dB SPL condition, the transformed loudness data are, on average, steeper than the BM I/O curves, especially for high levels. This finding is consistent with the idea that the hearing loss ear was not linear and the active process was not eliminated completely. The derived BM data for the remaining conditions resemble the physiological data.

III. EXPERIMENT 1. NOISE-SHIFTED THRESHOLDS: SHORT AND LONG DURATION TONES

Thermal noise, either broadband or shaped, is often used to simulate cochlear hearing loss or the effect of a cochlear threshold shift (Steinberg and Gardner, 1937; Schlauch and Wier, 1987; Florentine *et al.*, 1988; Schlauch *et al.*, 1994; De Gennaro and Braida, 1997). The effect of a broadband noise on the response of tones at the level of the basilar membrane has not been investigated; however, studies of two-tone suppression in the basilar membrane show that a suppressor tone decreases the sensitivity of a second tone and linearizes its response [e.g., Fig. 3 in Ruggero *et al.* (1992b), p. 1090]. If broadband noise acts to suppress the response to a tone in noise, the growth of the response of the tone in noise may be linear as it is in the case of two-tone suppression. The idea that either suppression or cochlear damage linearizes cochlear responses to tones is an intriguing one, but it is likely that this presumed linearization of responses as a result of suppression by broadband noise is not identical to linearization due to cochlear damage. For instance, the errors in speech recognition by listeners with simulated, flat hearing losses are only qualitatively similar to those observed in a listener with a flat hearing loss of the same magnitude (De Gennaro and Braida, 1997).

For the loudness of long-duration tones, noise simulated hearing loss using broadband noise produces results nearly identical to those observed for cochlear hearing loss (Hellman, 1988). Given this similarity in results, one would expect that a noise simulated hearing loss greater than 60 dB SPI, would yield loudness results consistent with known changes in the basilar membrane I/O function, as was demonstrated in the previous section for cochlear hearing loss. But what happens when the duration of the tone is only 2 ms? Ruggero *et al.* (1992a) examined basilar membrane responses to clicks and tones and found that the responses were similar. A temporal analysis showed that initial responses to clicks were linear, but later responses grew nonlinearly, as they do for long-duration tones. Cooper and Rhode (1996) showed that two-tone suppression, which is strongly believed to be related to the active process, appears in basilar membrane responses within 1 ms of the stimulus onset. The significance of these findings is that the cochlear amplifier is operative after only a minimal delay.

Assuming that the loudness of a short-duration tone (i.e., duration greater than 1 ms but less than 10 ms) for an ear with thresholds within normal limits grows in a compressive manner, does a 60-dB threshold shift yield results consistent with a linearized cochlea? The detection threshold for a short-duration tone in noise is elevated relative to the threshold for a long duration tone by an amount approximated by differences in energy in the two stimuli (for duration up to 100–300 ms) (Florentine *et al.*, 1988). Thus a lower noise spectrum level is required to shift the threshold of the short-duration tone to 60 dB SPL than is required to shift the threshold for a long-duration tone to the same level. It is unknown whether the system behaves in a linear manner when the detection threshold for a particular stimulus is elevated to a certain point (e.g., 60 dB SPL) or when a fixed

TABLE II. Detection thresholds (dB SPL) for 4.0-kHz tones in quiet and in noise for each subject. The abbreviation ''DNT'' means *did not test.*

	Quiet		Low-level noise		High-level noise	
Subject	2 ms	250 ms	2 ms	250 ms	2 ms	250 ms
S1	25.3	6.0	63.9	37.5	82.4	61.2
S2	26.4	6.8	59.8	DNT	84.1	60.7
S3	25.3	7.6	DNT	DNT	81.7	59.3
S4	29.4	15.3	61.0	38.5	82.0	59.6
Average	26.6	8.9	61.6	38.0	82.6	60.2

noise spectrum level is placed into an ear. To examine this idea, loudness matches between tones in quiet and in noise were made for (1) short-duration tones and (2) long-duration tones.

A. Subjects

Four young adults (age 22–24) with hearing sensitivity within normal limits participated. Subjects were selected based on the following criteria. First, thresholds were 10 dB HL or better (less) at audiometric frequencies (octave intervals between 250 and 8000 Hz). Second, bilateral thresholds for a 4.0-kHz, 250-ms tone differed by 5 dB or less. These thresholds at the test frequency (4.0 kHz) were measured using an adaptive procedure that targeted 79% correct detections and were based on four 50-trial blocks.

B. Stimuli

A stimulus frequency of 4.0 kHz was selected to minimize the effect of splatter on judgments of loudness. At this frequency, the critical bandwidth is 700 Hz (Scharf, 1970), which would encompass nearly all of the energy of a 2-ms tone with a gradual rise and fall time.

Pure tones at 4.0 kHz were digitally generated at a sampling rate of 20 kHz by a custom-designed, 16-bit digital-to-analog converter. Tones were either 2 ms or 250 ms with $\cos^2$ rise and fall times of 1 ms. Two channels were employed. One channel-generated the standard tone. The other channel generated the comparison tone. The level in each channel was adjusted by separate, computer controlled attenuators.

One of the pure-tone channels was mixed with broadband noise that was low-pass filtered at 10 kHz (Kemo model VBF/25). The noise was produced by a custom-designed generator. The spectrum level of the noise, which was left on continuously during a block of trials, was either 18 dB/Hz or 38 dB/Hz. Noise levels were selected based on pilot data which showed that detection thresholds for a 250-ms tone were shifted to 40 dB SPL (18 dB/Hz) or 60 dB SPL (38 dB/Hz). In subsequent sections of the paper the 18- and 38-dB/Hz conditions will be referred to as the low-level and high-level noise conditions, respectively.

C. Procedure

Prior to data collection, tonal thresholds were measured in continuous broadband noise and in quiet. Thresholds were measured using a 2IFC adaptive procedure that targeted 79% correct detections (Levitt, 1971). Correct answer feedback was provided after each trial. The results for 2-ms and 250-ms tones are shown in Table II. As expected based on temporal integration (Watson and Gengel, 1969; Florentine *et al.*, 1988), thresholds for the 2-ms condition are elevated by roughly 20 dB more than are the thresholds for the 250-ms tone in quiet or in conditions with the same level of noise.

Loudness matches were measured using a 2IFC adaptive procedure for subjective judgments (Jesteadt, 1980). Each trial contained two observation intervals marked by lights and separated by 500 ms. The standard tone was presented to one ear and the comparison tone to the other. Two performance levels were tracked concurrently by interleaving trials controlled by separate decision rules. One track converged on the stimulus level judged louder 21% of the time; the other track converged on the stimulus level judged louder 79% of the time (Levitt, 1971). The subject's task was to indicate which interval contained the louder tone by depressing a response button. Listeners were not given feedback regarding performance because loudness is a subjective measure.

Starting levels for each track were begun 10 dB above (79% track) or 10 dB below (21% track) the expected equal loudness level. If the initial starting levels were not close to the final stopping point of the run, that initial track was discarded and subsequent runs were based on a revised starting level that allowed subjects to bracket the point of subjective equality. The step size was 3 dB at the beginning of a block of trials and was reduced to 1.5 dB after two reversals in level. Thresholds for each track were calculated based on the mean of reversals in stimulus level direction excluding the first two reversals, which were discarded. Data points plotted for each subject represent the mean thresholds of between four and eight 100-trial blocks.

To control for potential biases, measurements were made for conditions with the standard tone in the quiet ear and for the standard tone in the noise-shifted ear. For both noise conditions (high and low level) and a 2-ms tone, standard levels in quiet ranged from 30 dB SPL to 90 dB SPL in 10-dB steps. Standard levels for the 2-ms tone in the low-level noise condition (comparison tone in quiet) ranged from 65 to 90 in 5-dB steps whereas standard levels for the 2-ms tone in the high-level noise condition ranged from 85 dB SPL to 103 dB SPL in 3-dB steps. For the 250-ms tone, standard levels in quiet were 15, 25, 35, 45, 55, 65, and 80 dB SPL. For the high-level noise condition, standard levels

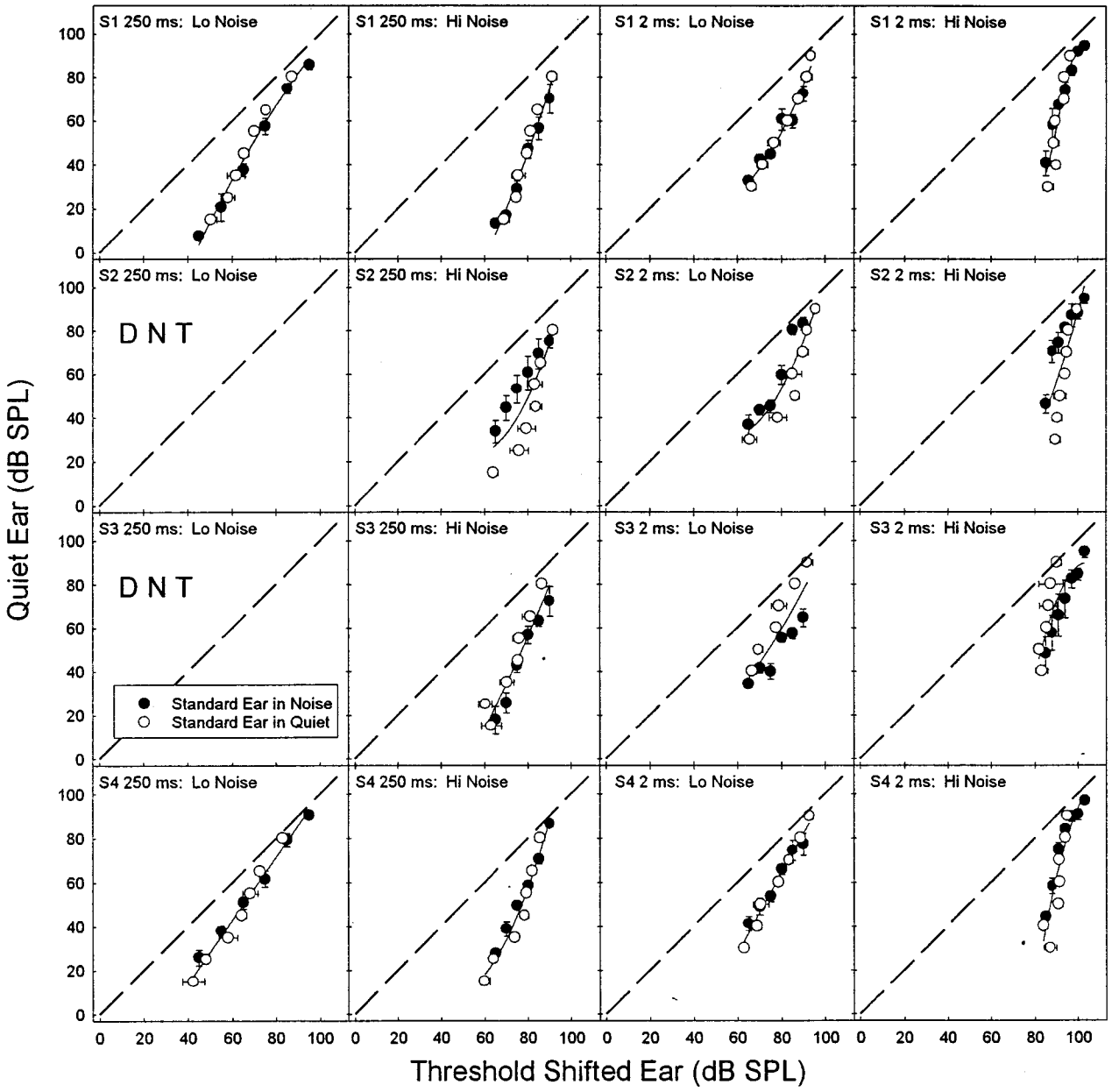


FIG. 3. Loudness matches for individual subjects for 2-ms tones and 250-ms tones for high and low noise levels. Open symbols represent conditions where the standard tone was in the quiet ear. Filled symbols represent conditions where the standard tone was presented in noise. The lines fitted to each data set are best-fitting second-order polynomial regressions. Error bars represent plus and minus one standard deviation. Error bars are omitted for levels with a standard deviation less than 1 dB.

for the 250-ms tone in noise ranged from 65 dB SPL to 90 dB SPL in 5-dB steps. Due to attrition, only two of the four subjects completed the low-level noise condition for the 250-ms tone. Standard levels for the low-level noise condition ranged from 45 dB SPL to 95 dB SPL in 10-dB steps.

Data for the 2-ms tone for both noise levels and for the 250-ms tone in the high-level noise condition were collected in semi-random order. Data for the 250-ms low-level noise condition were collected at the end of the study. Subjects were run in 1- or 2-h blocks until data collection was completed.

D. Results

Figure 3 illustrates loudness matches for four listeners in two levels of noise. The matching functions are very similar across listeners for identical conditions. Group-mean data for each of the conditions are shown in Fig. 4. As reported in other studies, the slope of the loudness function is steeper in a high-level noise condition than it is in a low-level noise condition (Stevens and Guirao, 1967). We also found that short-duration tones can produce a steeper matching function than long-duration tones in the same level of background noise. Richards (1977) reported a minor increase (a few degrees) in the slope of the matching function with a decrease in signal duration, but comparisons are difficult with his study because the stimuli were different. Richard's (1977) shortest-duration tone was 10 ms and the tones being matched were of different durations.

The lower portion of Table I shows an analysis of matching-function slopes for each condition for the average

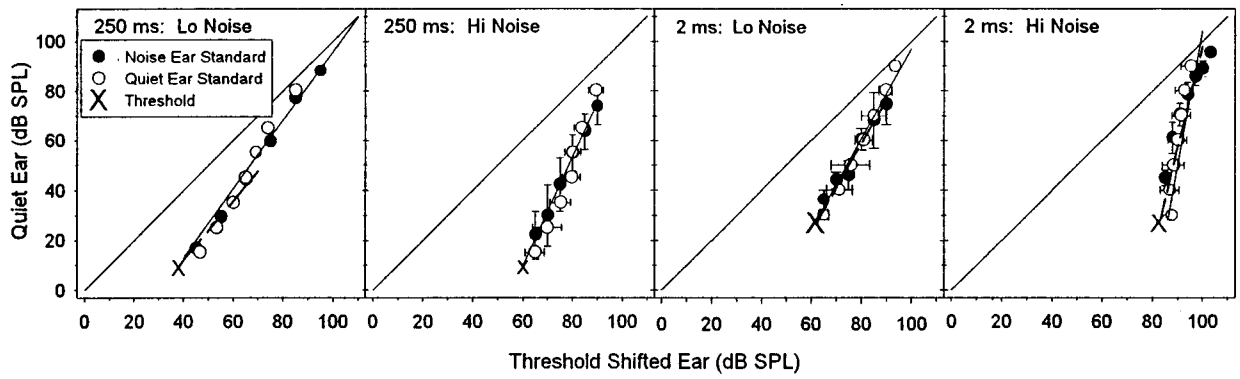


FIG. 4. Group-mean loudness matches for 2-ms tones and for 250-ms tones for high and low noise levels. The symbols are identical to those of Fig. 3. Regression lines were fitted to the entire data set in each panel (solid lines) and to the data between threshold and 30-dB sensation level (dashed lines), as in Fig. 1. Standard deviations were not calculated for the 250-ms, low-noise condition given that only two subjects participated.

data. The simulated thresholds show the same trend as the cochlear losses in the upper portion of the table; the matching-function slope increases with the amount of threshold shift in the poorer ear.

The variability of data obtained with the standard tone in the noise-shifted ear is greater than the variability of data obtained with the standard tone in quiet. This finding has been reported by others (Hellman and Zwislocki, 1964; Hellman *et al.*, 1987; Rankovic *et al.*, 1988). Average within subject standard deviations are 3.5 dB and 2.2 dB for conditions with the standard tone in the noise-shifted ear and the quiet ear, respectively. Average standard deviations are nearly identical for 2-ms tones (s.d.=2.8 dB) and for the 250-ms tones (s.d.=2.9 dB).

Figure 5 shows a comparison of Ruggero *et al.*'s (1997) BM I/O functions and BM I/O functions derived from loudness matches for group-mean data. The method for this comparison is identical to that used to transform loudness data shown in Fig. 2. The group-mean data are shown for conditions with the standard tone in noise (filled symbols) and for the standard tone in quiet (open symbols). The derived functions are similar in appearance to the BM I/O functions of Ruggero *et al.* (1997).

Figure 6 shows the quantitative relation between the derived basilar membrane slope (50–90 dB SPL) and detection threshold in the continuous noise for the various conditions for each subject. These values are shown for conditions with

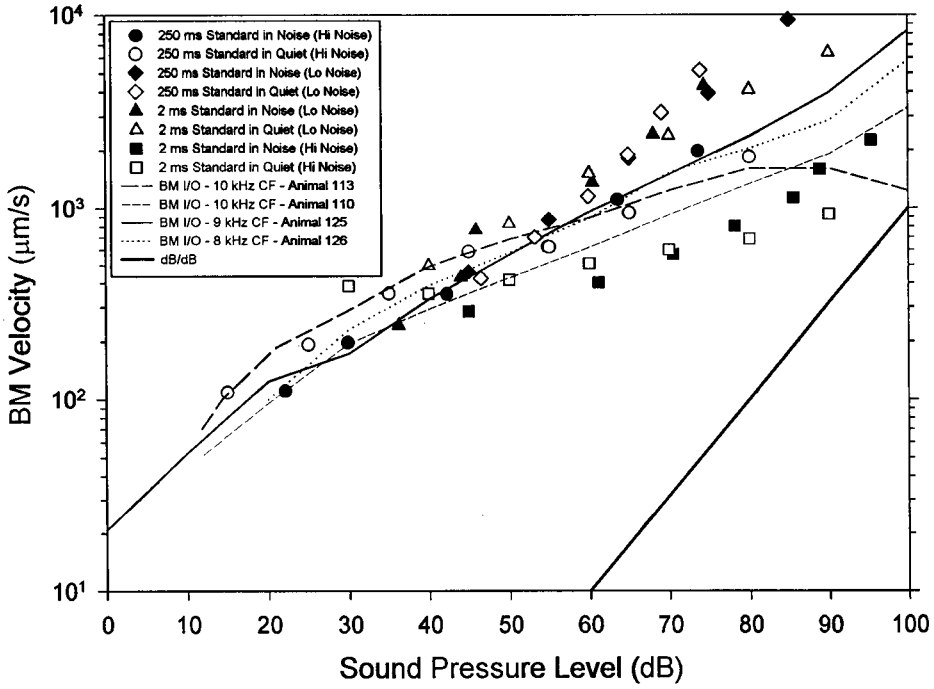


FIG. 5. A comparison of basilar-membrane (BM) input–output (I/O) functions from a study by Ruggero *et al.* (1997) and BM I/O functions derived from group-mean loudness-matching data for conditions of simulated unilateral hearing loss (Fig. 4). Lines represent BM I/O functions for four chinchillas from the study by Ruggero *et al.* (1997). Filled circles represent data derived from conditions with the standard presented in the threshold-shifted ear. Open symbols represent conditions with the standard in presented in quiet. The lowest levels of the derived functions for standards in quiet and in noise were scaled so that the midpoint of those two measures matched the BM velocities for Ruggero *et al.*'s (1997) functions for those levels.

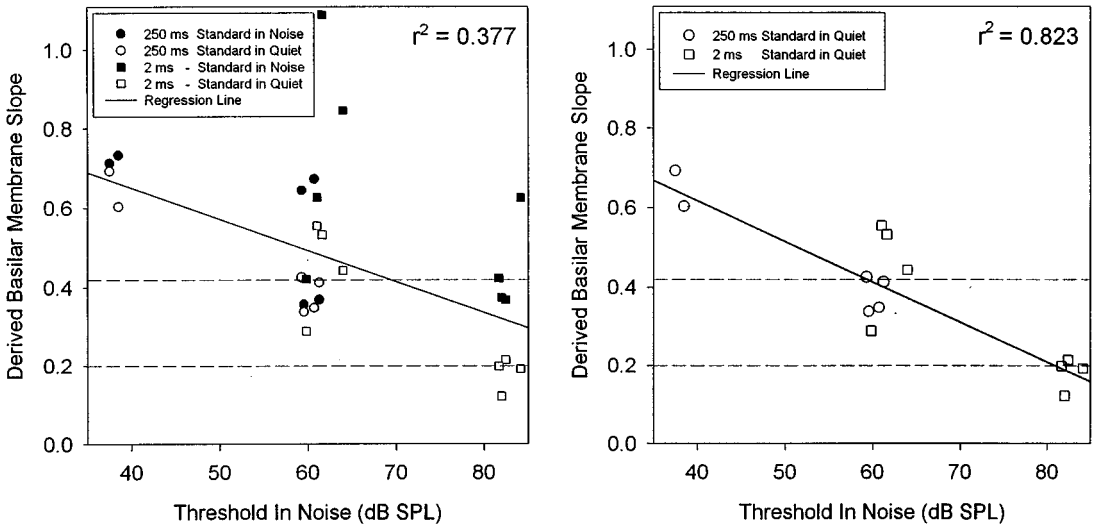


FIG. 6. The relation between detection threshold in noise (dB SPL) and the slope of the derived BM I/O function between 50 and 90 dB SPL. Squares represent conditions with 2-ms tones. Circles represent conditions with 250-ms tones. Open symbols are for conditions with the standard tone in the quiet ear. Filled symbols are for conditions with the standard tone in the threshold-shifted ear. The solid line represents a regression line fitted to data for all of the conditions (left-hand panel) or for conditions with the standard tone presented to the ear in quiet (right-hand panel). Dotted and dashed horizontal lines represent the range of BM I/O slopes for four healthy chinchilla cochleas from Ruggero *et al.* (1997).

the standard tone in quiet (open symbols) and the standard tone in noise (filled symbols). Both methods show the same general trend; the slope of the derived BM I/O function becomes shallower as threshold increases. The standard tone in the noise-shifted ear produced more variable data, as noted earlier, and for all but one condition for one subject (13 out of 14 comparisons), the slope of the derived BM I/O function was steeper when the standard tone was presented in noise than when the standard tone was presented in quiet. When all of the data are evaluated (standard tone in quiet and standard tone in noise), the regression line relating derived basilar membrane slope and detection threshold in noise shows a functional relation. The regression line for this analysis, shown in the left-hand panel of Fig. 6, accounts for 38% of the variance ($r=0.615$; $df=24$; $p=0.001$).[4] The right-hand panel of Fig. 6 shows that when only the less variable conditions with the standard tone presented to the ear in quiet are evaluated, the fit improves. The regression line in the right-hand panel accounts for 82% of the variance ($r=0.91$; $df=11$; $p<0.001$). By comparison, when threshold was shifted to 60 dB SPL (2-ms low-level noise and 250-ms high-level noise), the slope of the derived BM I/O function showed no apparent difference due to changes in spectrum level. The regression line fitted to derived BM I/O function slope and spectrum level for conditions with threshold shifted to roughly 60 dB SPL accounted for 7% of the variance for conditions with the standard presented to either ear (all data with threshold shifted to 60 dB SPL) ($r=0.269$; $df=12$; $p=0.35$). When only the conditions with the standard tone presented to the quiet ear are fitted, the regression line accounts for only 9% of the variance ($r=0.295$; $df=7$; $p=0.52$).

Horizontal dotted lines in Fig. 6 delimit the range of slopes for BM I/O functions for the four functions selected from Ruggero *et al.* (1997). All of the data for our derived slopes for the low-level noise and long-duration tone fall outside this range. By contrast, for conditions with threshold shifted to 60 dB SPL many of the derived slopes are within the range of BM I/O slopes. For threshold shifts of 80 dB SPL, the slopes for the derived functions and actual functions are more nearly comparable.

IV. GENERAL DISCUSSION

The nonlinear BM I/O function is presumably produced by an active process or ''cochlear amplifier'' that is physiologically vulnerable (Ruggero, 1992). The cochlear amplifier is described well by a compressive function with a shallow slope. When the active process is removed temporarily by a drug, such as furosemide, the BM I/O function steepens until the system becomes linear and there is a corresponding loss in sensitivity (Ruggero, 1992). Our BM I/O functions derived from loudness matching data between shifted threshold ears and ears with normal hearing sensitivity resemble actual BM I/O functions, but the correspondence between loudness and BM I/O functions is not perfect. The notable differences are related to the degree of hearing loss and the effect of tonal duration.

Physiological data show that elimination of the active process results in a roughly 40–80 dB loss in gain and that this loss is closely related to the health of the preparation (Ruggero, 1992; Ruggero *et al.*, 1997). The reason for the large range in estimates of the gain of the cochlear amplifier results from differences in methods commonly used to specify the gain. One method compares the difference in sensitivity to tones at CF in a healthy specimen and in the same animal shortly after death (Ruggero *et al.*, 1997). This method yields gain estimates of between 60 and 81 dB (Ruggero *et al.*, 1997). Another method compares the difference in gain between low-level tones at CF and high-level tones at the place yielding the largest response. This second method yields gains of between 39 and 60 dB (Ruggero *et al.*, 1997).

Moore and Glasberg (1997) modeled loudness growth in cochlear hearing loss and assumed that the gain of the cochlear amplifier has an upper limit consistent with estimates made using this second approach (55 dB–65 dB).

In their model of loudness in cochlear hearing loss, Moore and Glasberg (1997) assume that the loss of inner hair cells (IHC) results in a simple attenuation of gain whereas the loss of outer hair cells (OHC), which are linked with the active process, results in a steepening of the loudness function as well as a loss of gain. Moore and Glasberg (1997) assumed that complete damage to OHC would result in a loss of 55 dB for frequencies below 2.0 kHz; this assumption implies that the loudness function reaches its maximum slope for complete OHC loss. Moore and Glasberg (1997) compared the predictions of their model to Miskolczy-Fodor's (1960) data and found close agreement with this idea for losses between 40 and 60 dB SPL. For Miskolczy-Fodor's group with 80 dB SPL losses, however, many data points fell outside the range of slopes predicted by their model.[5] This result is probably due to the loudness function steepening for losses greater than 60 dB. Although Miskolczy-Fodor's (1960) data for 80 dB SPL losses are somewhat variable and not as extensive as his data for losses of 40, 50, and 60 dB SPL, Hellman and Meiselman (1993) found that, on average, the slope of the loudness function continues to steepen for hearing losses greater than 60 dB SPL. Our data for conditions of noise-simulated hearing losses support this idea as well.

There are several possible explanations for the loudness function steepening for losses as great as 80 dB SPL. First, the effective gain of the cochlear amplifier could be as great as 80 dB, as estimated by the technique that compares sensitivity in healthy, live preparations to that of fresh, dead ones. Another possibility is that cochlear gain is more than simply ''active'' gain. Ruggero *et al.* (1997) found 66 dB–76 dB of gain between the BM and the stapes. This estimate includes passive and active gain, both of which may be relevant for loudness coding in shifted thresholds. Finally, the neural population response that contributes to loudness may differ in an important way from BM I/O functions. BM I/O functions represent the velocity (or displacement) at a single place whereas the neural response to an intense sound represents excitation along a major extent of the BM. Although loudness measures in cochlear hearing loss (Hellman, 1994; Moore *et al.*, 1985) and in noise simulated hearing loss (Schlauch, 1994) are influenced to a small extent by spreading excitation, the effect may be large enough to result in slope changes across conditions as noted in our study.

The results from our study for conditions with noise-simulated hearing loss and short-duration tones are also difficult to explain based on known properties of the BM. In our study, the loudness function steepened and the derived BM I/O function became more compressive as tonal duration was reduced from 250 ms to 2 ms. This result suggests that the slope of the loudness function in noise is coupled tightly with threshold, which changes with tonal duration, and not the spectrum level of the noise.[6] Given that the active process is believed to be operative within about 1 ms of stimulation (Ruggero *et al.*, 1992a; Cooper and Rhode, 1996), elimination of the active process by cochlear hearing loss should result in a parallel shift of the BM I/O function with changes in tonal duration. That is, the slope of the BM I/O function is predicted to remain the same for changes in tonal duration. Oxenham and Plack (1997), who examined the relation between BM I/O functions and psychophysical forward masking in persons with cochlear hearing loss, report such a finding, but their range of durations was much smaller (4 ms–14 ms) than ours (2 ms–250 ms).

A possible explanation for the finding that the slope of the derived BM I/O function follows threshold rather than the spectrum level of the noise is that noise-shifted thresholds do not mimic cochlear hearing loss in every way. Although broadband noise and cochlear hearing loss result in nearly identical loudness functions for long-duration tones (Hellman, 1988), the same may not be true for short-duration tones and this difference may be related to changes in temporal integration for detection. The amount of temporal integration for detection thresholds for tones in noise follows signal energy (Florentine *et al.*, 1988) as it does in quiet (Plomp and Bouman, 1959; Watson and Gengel, 1969), but in cochlear hearing loss the amount of integration is reduced significantly (Florentine *et al.*, 1988). Why is the amount of temporal integration different in noise-shifted ears than it is in ears with cochlear hearing loss? It is possible that noise and cochlear hearing loss both linearize the cochlea but the cochlear hearing loss damages the system in an additional manner not simulated by the noise. Indirect evidence for this finding is seen in the slope of the psychometric functions for detection. In noise shifted thresholds, the psychometric function slopes are identical to those obtained in quiet from persons with hearing thresholds within normal limits (comparison between results in Green and Swets, 1966, p. 192, and those of Watson, 1972).[7] In cochlear hearing loss, the slope of the psychometric function for detection is sometimes steeper than the slope for persons with normal hearing (Arehart *et al.*, 1990; Carlyon *et al.*, 1990).

The possibility that mechanisms responsible for loudness perception may be different in cochlear hearing loss and noise-shifted thresholds has implications for modeling loudness. Many investigators assume that loudness can be modeled as a power function of intensity and that hearing loss represents a subtractive component (Steinberg and Gardner, 1937; Zwislocki, 1965, 1970; Humes *et al.*, 1992), perhaps due to a reduction in the number of neurons (Steinberg and Gardner, 1937) or a reduction in the overall firing rate of neurons that are able to convey information about signal intensity. For this class of models, the exponent relating intensity to loudness does not change with the degree of hearing loss. For this genre of models, the equation formulated by Zwislocki (1965) provides an excellent description of loudness functions for tones for persons with sensorineural hearing loss (Hellman and Meiselman, 1990) and hearing loss simulated by broadband noise (Schlauch *et al.*, 1995). An alternative method would be to describe loudness based on a power function in which the exponent varies to reflect a steepening of the loudness function with cochlear hearing loss (e.g., Stevens, 1966; Launer, 1995). To model changes in BM function, the exponent would be a low value (less

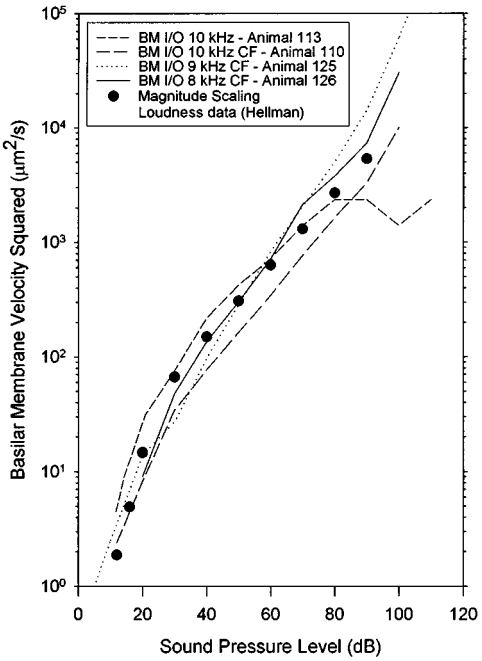


FIG. 7. A comparison of the square of the BM I/O function and the loudness function for a 3.0-kHz tone. The BM I/O functions are from Ruggero *et al.* (1997). The loudness function is from Hellman (1976). The loudness function was arbitrarily scaled so that the magnitude of the datum for 40 dB SPL was centered among squared velocities for the four BM I/O functions.

than one) in normal hearing and increase to a value of one when the active process is eliminated completely. Further studies need to be conducted to examine the appropriateness of either of these models for conditions of hearing loss and simulated hearing loss.

Finally, it is of interest to note that the average slope of the BM I/O functions in Ruggero *et al.*'s (1997) study is 0.32. This value is about half the value reported typically for magnitude scaling procedures (average=0.6) (Scharf, 1978). Assuming that loudness is determined in large part by the BM I/O function, the slope difference may be due to the finding that IHCs square the output of the BM (Goodman *et al.*, 1982). To examine this idea, the square of basilar membrane velocity for Ruggero *et al.*'s (1997) BM data was plotted as a function of level and compared with an archival loudness function for a 3.0-kHz tone obtained by Hellman (1976) using magnitude estimation and magnitude production. The functions in Fig. 7 plotted in this manner show a striking resemblance. Thus if the link between the BM I/O function and loudness is real and interspecies differences are minimal, the slope of the loudness magnitude scaling function may be accounted for at the level of the cochlea. This is in contrast to the report by Zeng *et al.* (1998) who argue that central factors play a large role in determining the shape of the loudness function. They assume that loudness compression is a peripheral process that occurs at the level of the BM and is followed by more central (i.e., at the level of the brain) loudness exponentiation expansion which recovers the original 100-dB dynamic range of the input stimulus. For the comparison in Fig. 7, we found it unnecessary to assume a central exponentiation-expansive process to account for the discrepancy between the loudness function and the BM I/O function.

V. CONCLUSIONS

(1) There is a striking resemblance between BM I/O functions and ones derived from loudness-matching functions for long-duration tones between an ear with normal hearing and one with threshold shift of 60 dB or more.

(2) Despite the similarity between BM I/O functions and ones derived from matching functions, there are some notable differences. Namely: (a) the loudness-matching function continues to steepen for hearing losses as great as 80 dB which results in a derived BM I/O function that becomes more compressive as the hearing loss increases. According to some measurements, the gain of the active process is thought to be 60–65 dB and a tight coupling between loudness and BM I/O functions would predict that the loudness function would not become steeper for losses exceeding the active gain of the system (Moore and Glasberg, 1997). (b) The loudness-matching function became steeper as the duration of a tone in noise was shortened resulting in the derived BM I/O function becoming more compressive. If broadband noise linearizes the cochlear response, tonal duration, in the range of values selected for this study, is not predicted to affect the slope of the derived BM I/O function.

(3) On average, the slope of the BM I/O function is one-half the slope of loudness functions obtained using magnitude scaling procedures for long-duration tones. Given that IHCs square the output of the BM (Goodman *et al.*, 1982), the slope of the cochlear output function in log units would double and be consistent with loudness scaling data. Thus the slope of the loudness magnitude scaling function follows the output of the cochlea for long-duration tones.

ACKNOWLEDGMENTS

We thank Rhona and Bill Hellman for meaningful discussions about the issues in this paper as well as for comments on an earlier draft of our manuscript. Brian Glasberg simplified our analysis by providing the coordinates for Miskolczy-Fodor's (1960) data. Thanks are also extended to Craig Champlin and an anonymous reviewer for helpful suggestions during the review process. This research was supported by NIH-NIDCD R29 DC01542.

[1]Loudness between roughly 30 and 90 dB SPL is described by a simple power law:

$$L = kP^{0.6}, \quad (1)$$

where L is loudness (in sones), P is pressure (N/m^2), and k is a scaling constant (Scharf, 1978). If units of power are used, the exponent is 0.3.

[2]A modified power function (e.g., Zwislocki, 1965) provides a better fit to loudness-matching data than a linear function; however, the slope obtained from a linear regression summarizes the main effect of threshold shift and provides a method for summarizing and comparing different conditions in a manner that is interpreted easily (e.g., Stevens, 1966).

[3]Ruggero *et al.* (1997) measured BM I/O functions in chinchilla for CFs between 9.0 and 10.0 kHz. Chinchilla hearing sensitivity is excellent at these frequencies. The mean threshold for frequencies of 8.0 kHz and 11.0 kHz is 7 dB SPL for six studies (Fay, 1988). Hearing sensitivity for humans at 4.0 kHz (Fay, 1988), the test frequency in the present study, and the one used for comparison with the chinchilla results, is also excellent.

[4]Detection thresholds for subject 3 (S3) were not measured for the 2-ms,

low-level noise condition. Her derived slopes for this condition are plotted in Fig. 6 at the group-mean threshold for this condition, but these points were not used in calculating the proportion of variance accounted for by the regression lines.

[5]Moore and Glasberg (1997) assumed for their model predictions for comparison with Miskolczy-Fodor's (1960) data that complete OHC loss results in a 55-dB loss of sensitivity. They state that assuming a 65-dB loss of sensitivity for complete OHC loss would yield results more consistent with Miskolczy-Fodor's (1960) data for 80 dB SPL losses. A study of BM I/O functions derived from single-unit data by Cooper and Yates (1994) showed that frequencies above 2.0 kHz show more gain than lower frequencies, which is the basis for Moore and Glasberg's (1997) assumption of two gain values corresponding to the loss of OHC. However, even this greater amount of gain associated with OHC loss would not account for loudness slope increases that occur for losses between 65 and 80 dB SPL as reported in Hellman and Meiselman's (1993) study.

[6]The slope of the loudness function for tones in noise is also dependent on noise bandwidth. For long-duration tones, Hellman (1970, 1972) showed that given the same noise spectrum level and the same thresholds, the slope of the loudness-matching function, and hence the slope of the loudness function, changed with variations in the noise bandwidth. The noise bandwidth was held constant in the present study and the use of a broadband noise produces loudness recruitment functions nearly identical to those seen in cochlear hearing loss (Hellman, 1988).

[7]A single psychometric function slope describes tonal detection in quiet and in noise between roughly 0.5 and 4.0 kHz. Green and Swets (1966) show psychometric functions for tonal detection for a wide range of durations in noise (10 ms, 100 ms, and 1000 ms) for various frequencies (0.25 Hz–6.0 Hz). Watson *et al.* (1972) measured psychometric functions for detection of a 150-ms tone in quiet for various frequencies (0.125–4.0 kHz). Viemeister and Schlauch (unpublished data) measured psychometric functions for detection for tones in quiet and in noise in the same listeners. The slope of the psychometric function, within the resolution of our confidence limits, did not vary with duration (10 ms, 20 ms, 40 ms, 80 ms, and 160 ms) and frequency (0.5 Hz, 2.0 Hz, and 4.0 kHz). Viemeister and Schlauch's result corroborates Green and Swets' (1966) result and extends Watson *et al.*'s (1972) result to shorter tonal durations.

Arehart, K., Burns, E. M., and Schlauch, R. S. (**1990**). ''A comparison of psychometric functions for detection in normal hearing and hearing-impaired listeners,'' J. Speech Hear. Dis. **54**, 429–438.

Carlyon, R. P., Buus, S., and Florentine, M. (**1990**). ''Temporal integration of trains of tone pulses by normal and by cochlearly impaired listeners,'' J. Acoust. Soc. Am. **87**, 260–268.

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 fibres: Variations with characteristic frequency,'' Hearing Res. **78**, 221–234.

De Gennaro, S. V., and Braida, L. D. (**1997**). ''Lippmann *et al.* revisited: A study of speech reception by listeners with hearing loss simulated by masking noise. Results for words in sentences,'' in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Mahwah, NJ), pp. 131–148.

Fay, R. (**1988**). *Hearing in Vertebrates* (Heffernan, Worcester).

Florentine, M., Fastl, H., and Buus, S. (**1988**). ''Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking,'' J. Acoust. Soc. Am. **84**, 195–203.

Goodman, D. A., Smith, R. L., and Chamberlain, S. C. (**1982**). ''Intracellular and extracellular responses in the organ of corti of the gerbil,'' Hearing Res. **7**, 161–179.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

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. (**1976**). ''Growth of loudness at 1000 and 3000 Hz,'' J. Acoust. Soc. Am. **60**, 672–679.

Hellman, R. P. (**1988**). ''Loudness functions in noise-induced and noise-simulated hearing losses,'' in *Noise as a Public Health Problem: Vol. 2*, edited by B. Berglund, U. Berglund, J. Karlsson, and T. Lindvall (Swedish Council for Building Research, Stockholm).

Hellman, R. P. (**1991**). ''Loudness measurement by magnitude scaling: implications for intensity coding,'' in *Ratio Scaling of Psychological Magnitude*, edited by S. J. Bolanowski, Jr. and G. A. Gescheider (Erlbaum, London), pp. 215–228.

Hellman, R. P. (**1993**). ''Can magnitude scaling reveal the growth of loudness in cochlear impairment?,'' in *Sensory Research Multimodal Perspectives*, edited by R. T. Verrillio (Erlbaum, London), pp. 1–18.

Hellman, R. P. (**1994**). ''Relation between the growth of loudness and high frequency excitation,'' J. Acoust. Soc. Am. **96**, 2655–2663.

Hellman, R. P., and Meiselman, C. H. (**1990**). ''Loudness relations for individuals and groups in normal and impaired hearing,'' J. Acoust. Soc. Am. **88**, 2596–2606.

Hellman, R. P., and Meiselman, C. H. (**1993**). ''Rate of loudness growth for pure tones in normal and impaired hearing,'' J. Acoust. Soc. Am. **93**, 966–975.

Hellman, R. P., and Zwislocki, J. J. (**1964**). ''Loudness function for 1000-cps tone in the presence of masking noise,'' J. Acoust. Soc. Am. **36**, 1618–1624.

Hellman, R. P., Scharf, B., Teghtsoonian, M., and Teghtsoonian, R. (**1987**). ''On the relation between growth of loudness and the discrimination of intensity for pure tones,'' J. Acoust. Soc. Am. **82**, 448–453.

Humes, L. E., Jesteadt, W., and Lee, L. W. (**1992**). ''Modeling the effects of sensorineural hearing loss on auditory perception,'' in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 617–623.

International Organization for Standardization (**1959**). *Expression of the Physical and Subjective Magnitudes of Sound* [ISO/R-131-1959 (E)] (International Organization for Standardization, Geneva).

Jesteadt, W. (**1980**). ''An adaptive procedure for subjective measurements,'' Percept. Psychophys. **28**, 85–88.

Launer, S. (**1990**). ''Loudness perception in listeners with sensorineural hearing impairment,'' Ph.D. thesis, Oldenburg, Germany.

Levitt, H. (**1971**). ''Transformed up-down methods in psychophysics,'' J. Acoust. Soc. Am. **49**, 467–447.

Miskolczy-Fodor, F. (**1960**). ''Relation between loudness and duration of tonal pulses. III. Response in cases of abnormal loudness function,'' J. Acoust. Soc. Am. **32**, 486–492.

Moore, B. C. J., and Glasberg, B. R. (**1997**). ''A model of loudness perception applied to cochlear hearing loss,'' Aud. Neurosci. **3**, 289–311.

Moore, B. C. J., Glasberg, B. R., Hess, R., and Birchall, J. P. (**1985**). ''Effects of flanking noise bands on the rate of growth of loudness of tones in normal and recruiting ears,'' J. Acoust. Soc. Am. **77**, 1505–1513.

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.

Plomp, R., and Bouman, M. A. (**1959**). ''Relation between hearing threshold and duration for tone pulses,'' J. Acoust. Soc. Am. **31**, 749–758.

Rankovic, C. M., Viemeister, N. F., Fantini, D. A., Cheesman, M. F., and Uchiyama, C. L. (**1988**). ''The relation between loudness and intensity difference limens for tones in quiet and noise backgrounds,'' J. Acoust. Soc. Am. **84**, 150–155.

Richards, A. M. (**1977**). ''Loudness perception for short-duration tones in masking noise,'' J. Speech Hear. Res. **20**, 684–693.

Ruggero, M. A. (**1992**). ''Responses to sound of the basilar membrane of the mammalian cochlea,'' Curr. Opin. Neurobiol. **2**, 449–456.

Ruggero, M. A., Rich, N. C., and Recio, A. (**1992a**). ''Basilar membrane responses to clicks,'' in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, London), pp. 85–91.

Ruggero, M. A., Robles, L., and Rich, N. C. (**1992b**). ''Two-tone suppression in the basilar membrane of the cochlea: Mechanical basis of auditory-nerve rate suppression,'' J. Neurophysiol. **68**, 1087–1099.

Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S., and Robles, L. (**1997**). ''Basilar membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.

Scharf, B. (**1970**). ''Critical Bands,'' in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. 1, pp. 157–202.

Scharf, B. (**1978**). ''Loudness,'' in *Handbook of Perception, Vol. IV Hearing*, edited by E. C. Caterette and M. P. Friedman (Academic, New York), pp. 187–242.

Schlauch, R. S. (**1994**). ''Intensity resolution and loudness in high-pass noise,'' J. Acoust. Soc. Am. **95**, 2171–2179.

Schlauch, R. S., Harvey, S., and Lanthier, N. (**1995**). ''Intensity resolution and loudness in broadband noise,'' J. Acoust. Soc. Am. **98**, 1895–1902.

Schlauch, R. S., and Wier, C. W. (**1987**). ''A method for relating loudness matching and intensity discrimination in impaired ears,'' J. Speech Hear. Res. **30**, 13–20.
Steinberg, J. C., and Gardner, M. B. (**1937**). ''The dependency of hearing impairment on sound intensity,'' J. Acoust. Soc. Am. **9**, 11–23.
Stevens, S. S. (**1966**). ''Power transformations under glare, masking and recruitment,'' J. Acoust. Soc. Am. **39**, 725–735.
Stevens, S. S. (**1970**). ''Neural events and psychophysical law,'' Science **170**, 1043–1050.
Stevens, S. S., and Guirao, M. (**1967**). ''Loudness functions under inhibition,'' Percept. Psychophys. **2**, 459–465.
Viemeister, N. F., and Bacon, S. P. (**1988**). ''Intensity discrimination, increment detection and magnitude estimation for 1.0 kHz tones,'' J. Acoust. Soc. Am. **84**, 172–178.
Watson, C. S., Franks, J. R., and Hood, D. C. (**1972**). ''Detection of tones in the absence of external masking noise. I. Effects of signal intensity and signal frequency,'' J. Acoust. Soc. Am. **52**, 633–643.
Watson, C. S., and Gengel, R. W. (**1969**). ''Signal duration and signal frequency in relation to auditory sensitivity,'' J. Acoust. Soc. Am. **46**, 989–997.
Yates, G. K., Winter, I. M., and Robertson, D. (**1990**). ''Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range,'' Hearing Res. **45**, 203–220.
Zeng, F.-G., Shannon, R. V., and Hellman, W. (**1998**). ''Physiological processes underlying psychophysical laws,'' in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr Publishers, London, in press).
Zwicker, E. (**1970**). ''Masking and psychological excitation as consequences of the ear's frequency analysis,'' in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden), pp. 376–394.
Zwislocki, J. J. (**1965**). ''Analysis of some auditory characteristics,'' in *Handbook of Mathematical Psychology, Vol. III*, edited by R. D. Luce, R. R. Bush, and E. Galanter (Wiley, New York), pp. 1–97.

Interactions of forward and simultaneous masking in intensity discrimination[a)]

Fan-Gang Zeng[b)]
Auditory Perception Laboratory, Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

(Received 3 September 1997; accepted for publication 22 December 1997)

Intensity coding mechanisms are explored in a paradigm involving both forward and simultaneous masking. For intensity discrimination of 1000-Hz pure tone in quiet, a near-miss to Weber's law is observed. However, as more stimulus components are added to this relatively simple experiment, interactions among components produce a more complex pattern of results. An intense forward masker, while not causing any threshold shift for the test tone, produces a nonmonotonic intensity discrimination function [''the midlevel hump,'' Zeng *et al.*, Hearing Res. **55**, 223–230 (1991)]. The midlevel hump can be removed by the presence of additional notched noise [Plack and Viemeister, J. Acoust. Soc. Am. **92**, 1902–1910 (1992)] or narrow-band noise whose level is increased along with the test tone's standard level. The same midlevel hump can also be enhanced by a fixed-low-level notched noise or a high-level, high-pass noise which causes minimal masking at the test frequency. Interactions of forward masking and simultaneous masking present a serious problem for a clear interpretation of these results. For example, the notched noise was originally intended to restrict off-frequency listening, but on-frequency masking compromised this original purpose and confounded the interpretation of the notched noise effects. By measuring systematically the growth-of-masking functions, the present study identified various interactions of forward and simultaneous masking and clarified the role of off-frequency listening in forward-masked intensity discrimination. Both peripheral and central mechanisms may have contributed to the occurrence, reduction and enhancement of the midlevel hump under these masking conditions. [S0001-4966(98)01404-0]

PACS numbers: 43.66.Ba, 43.66.Cb, 43.66.Dc, 43.66.Fe [WJ]

INTRODUCTION

The peripheral auditory system is often modeled as a bank of tuned filters that perform frequency analysis on sounds. The high-frequency selectivity of the auditory filter is accomplished by a passive resonance of the basilar membrane coupled with an active process linked to the nonlinear motor activity of outer hair cells (e.g., Allen and Neely, 1992). Because of this nonlinearity, the output of the array of auditory filters or ''the excitation pattern'' generated by an acoustic stimulus is also highly nonlinear as a function of stimulus level. Both psychophysical and physiological data indicate that for sinusoidal stimuli at low stimulus levels, localized activity occurs over a small range of filters tuned around the stimulus frequency, but as the level is increased, the excitation spreads towards filters with higher and lower center frequencies, particularly on the high-frequency side of the excitation pattern (e.g., Egan and Hake, 1950; Kim and Molnar, 1979).

Auditory discrimination of intensity and frequency has been modeled using both a single-filter model in which the filter is located at the frequency where the largest difference in excitation occurs (Zwicker, 1970) and a multi-filter model in which information from all excited filters is combined either in an optimal fashion (Florentine and Buus, 1981) or in a nonoptimal, unweighted fashion (Moore and Sek, 1994). The ability to use information in different frequency regions to improve auditory performance is often termed ''off-frequency listening.'' Off-frequency listening has been shown to be involved in a wide range of psychophysical measures including intensity discrimination (Viemeister, 1972; Moore and Raab, 1974), the slope of loudness growth functions (Hellman, 1978), differences in temporal integration between normal and impaired listeners (Hall and Fernandes, 1983), and differences in tuning curves obtained in simultaneous and forward masking (O'Loughlin and Moore, 1981).

Attempts to restrict off-frequency listening have used a notched or band-stop noise masker (e.g., Viemeister, 1974; Moore *et al.*, 1985; Schneider and Parker, 1987; Plack and Viemeister, 1992a,b). In these experiments, as the signal level is increased, the level of the notched noise is also increased to restrict off-frequency listening. This increase in the notched-noise level elevates the signal threshold so that intensity discrimination is measured at low sensation levels for all standard levels. To reduce this masking problem, alternative approaches have been used to restrict off-frequency listening. Hellman (1978) used a high-pass noise to mask a 100-Hz tone and reasoned that the high-pass noise would limit off-frequency listening at high frequencies while the apical edge of the basilar membrane would limit naturally

[a)]Portions of this paper were presented at the 125th meeting of Acoustical Society of America, 1993, Ottawa and the 3rd Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan, 1996, Hawaii.

[b)]Electronic mail: zeng@hei.org

off-frequency listening at low frequencies. Viemeister (1972) measured intensity discrimination for a tone in high-pass noise and suggested that the ''near-miss'' to Weber's law reflects the observer's use of information of aural harmonics; on the other hand, Moore and Raab (1974) compared intensity discrimination results from relatively low-level, wideband, high-pass, and notched noises, and found that information from both low- and high-frequency sides of the test tone, not just the aural harmonics, contributes to intensity discrimination at high levels. Florentine (1983) controlled the audibility of very high frequencies (6–19 kHz) using either a high-pass noise or the high threshold naturally present in normal-hearing listeners and found that intensity discrimination for a 1000-Hz tone at 80–90 dB SPL was related to the ability to hear at these very high frequencies. Additional studies found similar results using cochlear-impaired listeners with normal hearing at the test frequency and significant hearing loss in regions of excitation spread (e.g., Penner *et al.*, 1974; Zeng and Turner, 1991).

Because the excitation spreads more toward high frequencies, the low-pass band in a notched noise may play a smaller role than the high-pass band in restricting off-frequency listening. On the other hand, the low-pass noise spreads to the signal frequency and actually interferes with the excitation pattern at the signal frequency. Schlauch (1994) measured intensity discrimination using only a high-pass noise presented at a fixed high level for all standard levels. This high-pass noise masker produced a minimal threshold elevation at the signal frequency but significantly elevated thresholds for higher frequencies in the noise pass band. Under these conditions, Schlauch showed that intensity discrimination and loudness are unaffected for standard levels below 40 dB SPL. However, for standard levels above 40 dB SPL, intensity discrimination is degraded and loudness is reduced, indicating a significant role for off-frequency listening at these higher levels.

The present study evaluates the role of off-frequency listening in intensity discrimination under forward masking. Zeng *et al.* (1991, 1992, 1995) showed that, for a brief pure tone following a high-level masker, intensity discrimination is degraded significantly at medium levels, but by little or not at all at low and high levels, a pattern termed the ''midlevel hump.'' Plack and Viemeister (1992a) used a notched noise in an attempt to restrict off-frequency listening and suggested that ''limiting the intensity information to neurons tuned close to the pedestal frequency should, if anything, increase the magnitude of the effect.'' Instead, Plack and Viemeister found that the notched-noise masker actually reduced the midlevel hump. Although the original motivation for the Plack and Viemeister study was to use the notched noise to restrict off-frequency listening, the apparent contradiction to their original prediction of the notched-noise effects led them to conclude that ''the notched noise obviously does more than limit off-frequency listening, the role of this later process in the original Zeng *et al.* experiment is still unclear (Plack and Viemeister, 1992a, p. 1902).''

In particular, the present study addresses the issue of on-frequency masking caused by the notched noise. Indeed, threshold data from a single subject in the Plack and Viemeister study indicated that the notched noise produced significant masking at the signal frequency so that intensity discrimination was measured at sensation levels between 5 and 15 dB for all standard levels. This on-frequency masking problem, also noted by Plack and Viemeister (1992a), compromised their original purpose of using the notched noise to restrict off-frequency listening and possibly confounded their interpretation of the data. In this paper, three experiments were conducted in an attempt to separate the effects of on-frequency masking and off-frequency listening on forward-masked intensity discrimination. The first experiment used a notched noise to replicate the Plack and Viemeister study; in addition, this experiment extended their study by systematically measuring the growth-of-masking function for a tone in the notched noise in both the presence and absence of forward masking. The second experiment used a narrow-band noise to match threshold shifts caused by the notched noise to study whether the on-frequency masking at the threshold level by itself can remove the midlevel hump at the suprathreshold level. The third experiment used a high-pass noise to reduce the on-frequency masking while limiting effectively off-frequency listening on the high-frequency side of the excitation pattern.

I. GENERAL METHOD

A. Subjects

Six experienced listeners, two females and four males, participated in the present study. Three subjects participated in experiment 1 and the other three participated in experiments 2 and 3. They were between 25 and 30 years old at the time of the experiment. All had normal hearing with less than 10 dB HL thresholds at frequencies of 125, 250, 500, 1000, 2000, and 5000 Hz. For all measurements, subjects were seated in a double-walled, sound-treated booth and tested individually.

B. Stimuli

Figure 1 shows the temporal and spectral configurations of the three experiments. In all three experiments, the forward masker was a 1000-Hz sinusoid of 100 ms in duration. The test tone was a 1000-Hz sinusoid of 25 ms in duration. There was a 100-ms delay between the offset of the masker and the onset of the tone, at which the average threshold shift due to the forward masker was negligible. In noise conditions, the 25-ms test tone was presented simultaneously at the temporal center of a noise. Unless otherwise noted, the noise had a duration of 125 ms. The noise was a notched noise (experiment 1), a narrow-band noise (experiment 2), or a high-pass noise (experiment 3). All durations included 2.5-ms cosine-squared rise and fall ramps.

All tonal stimuli were digitally generated by an IBM-PC computer using a 16-bit D/A converter at a sampling rate of 20 kHz (TDT model QDA2, Tucker-Davis Technologies). The stimuli were smoothed by an anti-aliasing filter (TDT model FLT3) with a cutoff frequency of 8 kHz. The forward masker was always presented at 90 dB SPL. The level of the standard tone varied from 30 to 90 dB SPL in steps of 10 dB.

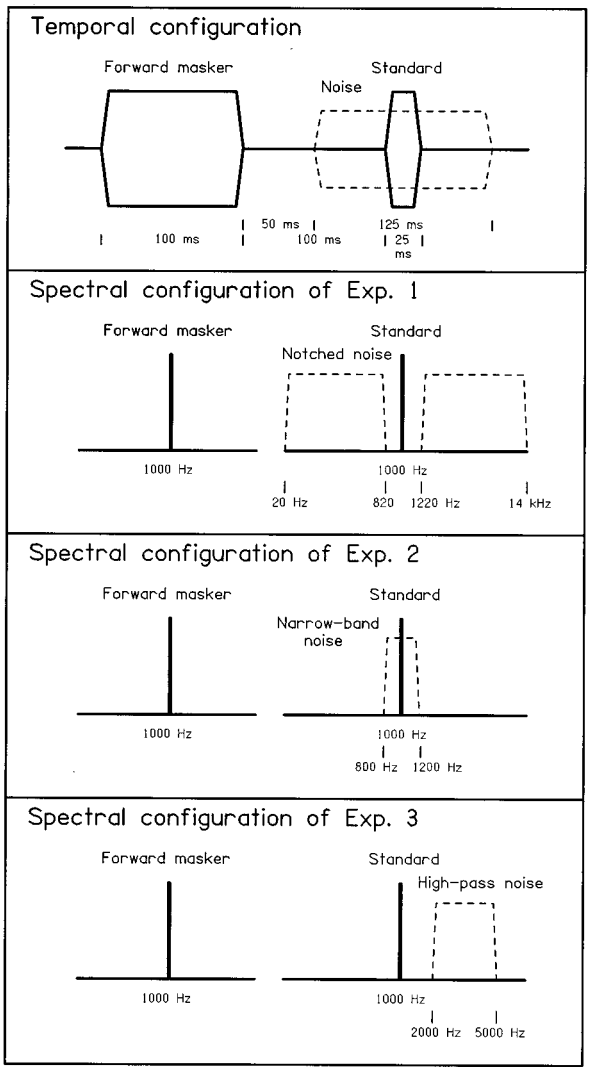


FIG. 1. The top panel shows the temporal configuration of stimuli in experiments 1, 2, and 3. A notched noise was used in experiment 1, a narrow-band noise was used in experiment 2, and a high-pass noise was used in experiment 3. All stimuli were turned on and off with 2.5-ms cosine-squared ramps. The second, third, and bottom panel shows spectral configurations for the stimuli in experiment 1, 2, and 3, respectively.

The levels of the masker and the test tone were separately controlled by two programmable attenuators (TDT model PA3).

All noise stimuli were produced by passing a white noise (TDT model WG1) through a dual-channel filter with a nominal attenuation slope of 135 dB/oct (Stewart VBF-10M). The timing of the noise was controlled by an electronic switch (TDT model SW1). In experiment 1 (the second panel in Fig. 1), the notched noise was similar to the one in Plack and Viemeister's study (1992a) and consisted of a low-pass noise (3-dB cutoff frequencies: 20–820 Hz) and a high-pass noise (1220–14 000 Hz). In experiment 2 (the third panel of Fig. 1), the bandpass noise had a width of 400 Hz with 3-dB down points at 800 and 1200 Hz, respectively. In experiment 3 (the bottom panel of Fig. 1), the high-pass noise (relative to the 1000-Hz signal frequency) had a width of 3000 Hz with 3-dB down points at 2000 and 5000 Hz, respectively. The stimuli in experiment 1 were delivered through an ER-2 insert earphone and calibrated periodically by a B&K sound level meter using a Zwislocki coupler. The ER-2 earphone had a relatively flat frequency response up to 16 kHz. The stimuli in experiments 2 and 3 were delivered through TDH-49 headphones mounted in an MX41/AR cushion and calibrated by the B&K sound level meter using an NBS-9A coupler. Monaural stimulation of the right ear was used for all subjects.

C. Procedure

A two-interval, two-alternative, forced-choice, adaptive procedure was employed to measure the increment in intensity that produces a 79.4% correct response level (3-down, 1-up rule, Levitt, 1971). The two observation intervals were indicated visually on a computer monitor and separated by 650 ms. The silent interval from trial to trial was about 2 s plus the subject's response time. During each trial, the observer had to indicate which one of two intervals contained the signal in the detection experiment or the louder sound in the discrimination experiment. The signal was randomly presented in either one of the two intervals. Visual feedback was provided indicating the correct response. The starting level for the test tone was 20 dB above the estimated threshold in the detection experiment and 10 or 20 dB higher than the standard tone in the discrimination experiment. An initial 5-dB step size was reduced to 1 dB after the first four reversals. Testing continued until 12 reversals had occurred or 60 trials were completed. At the end of each run, intensity discrimination in units of $10\log(1+\Delta I/I)$ was taken as the mean of the values for the last eight reversals. The result from a run was discarded if less than 10 reversals were obtained or the standard deviation was greater than 5 dB in the run. Three-to-eight runs were obtained for each data point. Following Buus and Florentine's (1991) suggestion, intensity discrimination data in terms of level difference in dB will be plotted on a logarithmic scale.

II. THE NOTCHED-NOISE EXPERIMENT

A. Rationale and method

To measure systematically the degree of on-frequency masking caused by the notched noise, the growth-of-masking function was measured for a 25-ms, 1000-Hz tone as a function of the notched-noise level from -10 to 50 dB in 10 steps. The growth-of-masking function was obtained in both the presence and absence of forward masking in all three subjects. In addition, two control masking functions were obtained in one subject. The ''gated'' control represents a condition where the notched noise had a duration of 25 ms and was gated simultaneously with the 25-ms tone (experiment 2C in Plack and Viemeister, 1992a). Threshold elevations caused by the ''gated'' notched noise were presumably due to suppression (Costalupes *et al.*, 1984). In contrast to the ''gated'' control, the ''fringe-only'' control represents a condition where the notched noise was turned off in the presence of the test tone but was on for 50 ms before the onset of the tone and for 50 ms after the offset of the tone. Threshold

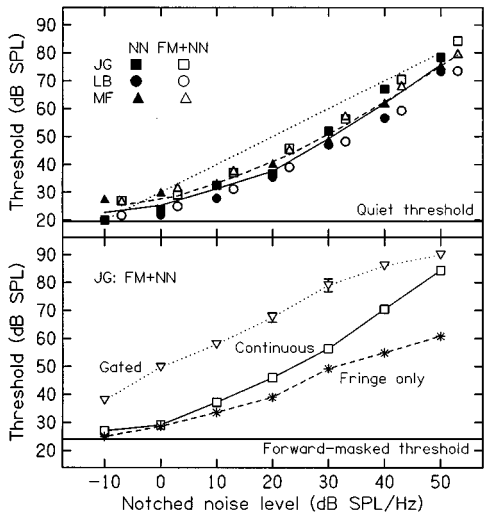


FIG. 2. Growth-of-masking function for a 1000-Hz tone in a notched noise with and without forward masking (upper panel). Individual data are represented by symbols and average data are represented by lines. The dotted line represents a hypothetical linear growth-of-masking function. Growth-of-masking functions for an individual subject obtained with three notched-noise conditions in the presence of forward masking (lower panel). The ''Continuous'' data were the same as in the upper panel. The ''Gated'' data were obtained when the notched noise was gated simultaneously with the test tone. The ''Fringe only'' data were obtained under a condition opposite to the ''Gated'' condition.

elevations caused by the ''fringe-only'' condition were mainly due to adaptation (Smith, 1979; Costalupes *et al.*, 1984).

Forward-masked intensity discrimination was measured first using a paradigm identical to Plack and Viemeister (1992a). The standard level for the tone varied from 30 to 80 dB SPL in 10-dB steps. The notched noise level was increased by 10 dB whenever the standard level was increased by the same amount, in an attempt to keep a constant 30-dB ratio between the tonal standard level and the spectrum level of the notched noise. Data at the 90 dB SPL standard level were not collected because all subjects reported that the 60-dB notched noise was uncomfortably loud. In addition, the forward-masked intensity discrimination function was measured for a fixed notched-noise level at -10, 10, 30, or 50 dB.

B. Results and discussion

The upper panel of Fig. 2 displays the growth-of-masking functions obtained in the continuous notched-noise condition. The filled symbols represent individual data and the solid line represents the averaged data for the notched-noise alone condition; the open symbols represent individual data and the dashed line represents the averaged data for the notched noise plus forward-masking condition. The error bar represents the size of one standard deviation. To reduce overlap between the two data sets in this display, the data in the notched-noise plus forward-masking condition are shifted by 3 dB on the x axis. The horizontal solid line represents the averaged tonal threshold in quiet (19 dB SPL). The diagonal dotted line represents a hypothetical linear growth-of-masking function, in which a 1-dB increment in the notched-noise level would result in a 1-dB threshold elevation for the test tone.

A nonlinear, increasingly steep, growth-of-masking function is clearly noted for the detection of the 1000-Hz test tone in both the notched-noise alone and the notched-noise plus forward-masking conditions. The forward masker appears to produce an additional 2–3 dB threshold elevation, mostly at low and medium notched-noise levels, compared with the masking function in the notched-noise alone condition. It is also noted that the single subject's data in the Plack and Viemeister (1992a) study were very similar to subject LB's threshold data (open circles) in the present study. To demonstrate quantitatively the nonlinear growth-of-masking function, the slope of the masking function was calculated using a linear regression analysis at low notched-noise levels (-10, 0, and 10 dB) and at high noise levels (30, 40, and 50 dB), respectively. For the notched-noise alone condition, the slope averaged across three subjects was 0.43 ($r=0.97$) at low notched-noise levels and 1.32 ($r=0.99$) at high levels; whereas for the notched-noise plus forward-masking condition, the slope was 0.51 ($r=0.99$) at low notched-noise levels and 1.26 ($r=1.00$) at high levels. This increasing steep masking function in notched noise is similar to the ''upward spread of masking'' typically observed for masker frequencies well below the signal (e.g., Egan and Hake, 1950; Oxennam and Moore, 1995) and likely reflects the growth of suppression of the auditory nerve activity (Delgutte, 1990a, b).

The lower panel of Fig. 2 displays three growth-of-masking functions obtained from subject JG only, and representing the ''Continuous'' notched-noise condition (open squares, same as in the upper panel), the ''Gated'' condition (inverted triangles), and the ''Fringe only'' condition (asterisks). The increasingly steep ''Continuous'' masking function is sandwiched between the more linear ''Gated'' and ''Fringe only'' masking functions. A linear regression analysis across all notched-noise levels reveals a slope of 0.89 ($r=0.99$) for the ''Gated'' condition and 0.67 ($r=0.99$) for the ''Fringe only'' conditions. In other words, a 60-dB increment in the notched-noise level results in an approximately 54-dB and 42-dB overall threshold elevation for the ''Gated'' and the ''Fringe only'' conditions, respectively. The threshold difference between the ''Gated'' and the ''Continuous'' conditions shows the classic level-dependent overshoot effect, in which the gated (or onset) threshold is greater than the continuous (or steady-state) threshold at low and medium levels than at high levels (Bacon, 1990). To summarize, the present data show that the notched-noise produced considerable on-frequency masking, even under conditions where the notched noise did not overlap with the test tone either spectrally or temporally (the ''Fringe only'' condition).

Figure 3 shows both individual (JG, LB, and MF) and average data that were obtained using the Plack and Viemeister paradigm (experiment 1, 1992a). The intensity discrimination function in the quiet condition is plotted as a shaded area (mean plus one standard deviation) and the forward-masked intensity discrimination function is plotted

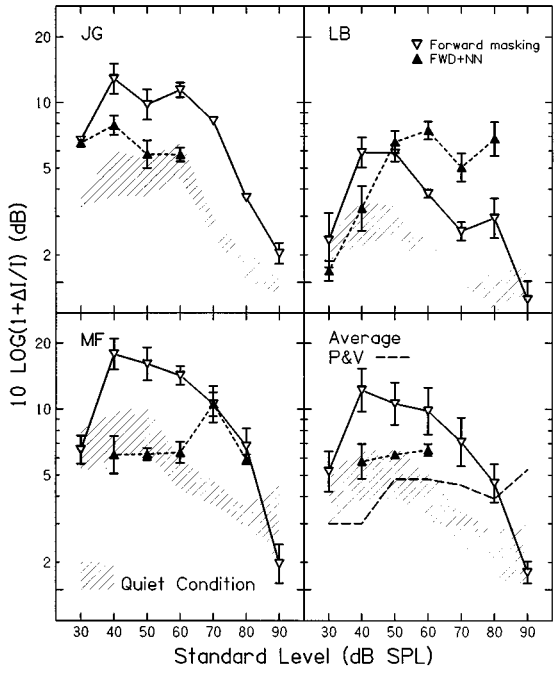


FIG. 3. Replication of the Plack and Viemeister study (1992a), in which the notched-noise level was increased along with the tonal standard level.

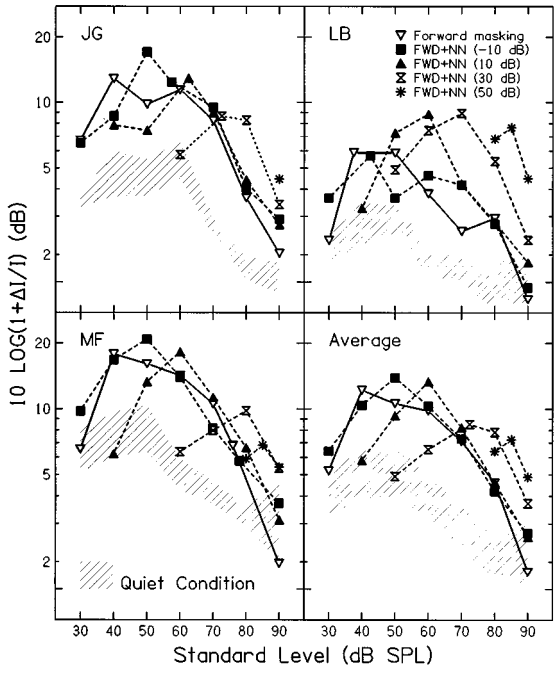


FIG. 4. Forward-masked intensity discrimination functions as a function of the fixed noise level. Note both consistent shift for the hump position and the change for the hump size as a function of the notched-noise level. Some data points are shifted by −2 and 2 dB along the *x* axis to avoid overlap.

as inverted triangles connected by solid lines. The forward-masked intensity discrimination function in the notched-noise condition is plotted as filled triangles connected by dotted lines. Note the missing data points at 70 and 80 dB SPL for subject JG and at 30 dB SPL for subject MF, because the combination of the forward masker and the notched noise elevated the tonal threshold above these standard levels (see Fig. 2, lower panel). Plack and Viemeister's results are converted from Weber fractions in dB to level differences in dB and are plotted as a dashed line on the ''Average'' panel.

Figure 3 shows that intensity discrimination in quiet followed the ''near-miss'' to Weber's law (e.g., McGill and Goldberg, 1968) and that the forward masker produced a nonmonotonic intensity discrimination function (Zeng *et al.*, 1991). The addition of the notched noise essentially removed the midlevel hump, with the exception of subject LB, who had the smallest hump in the forward masking alone condition and appeared to shift the peak of the ''hump'' from 40 to 60 dB SPL. Although only three data points (at 40, 50, and 60 dB SPL) were averaged from all three subjects, the removal of the midlevel hump could be observed in the average data between present study and the Plack and Viemeister (1992a) study. The 1–2 dB overall elevation in the average data for the present study could be due to a procedural difference (2IFC vs 3IFC) between the two studies.

Figure 4 shows individual forward-masked intensity discrimination functions measured at fixed noise levels of −10, 10, 30, and 50 dB SPL, which is represented by squares, triangles, ''hour glasses,'' and asterisks, respectively. For subject JG, only the 90 dB SPL data point was collected at the 50-dB notched-noise level because of his high threshold under this condition (86 dB SPL, see Fig. 2). Shown again in Fig. 4 are the quiet and forward-masked intensity discrimination functions without the notched noise. Despite individual variability, both individual and average data show a consistent peak shift as a function of the notched-noise level: the peak of the ''midlevel hump'' shifted from 40 dB SPL in the forward-masking alone condition to about 50, 60, 70, and 85 dB SPL at the notched-noise level of −10, to 10, 30, and 50 dB, respectively. The size of the ''hump'' was increased usually at low notched-noise levels (−10 dB for JG and MF, and 10 and 30 dB for LB), and decreased at high noise levels (30 dB for JG, and 30 and 50 dB for MF). The average data show a similar trend in the peak shift and the dependence of the hump size on the notched-noise level. Note also in the average data that the reduced hump at 30-dB notched-noise level is consistent with Plack and Viemeister's observation (1992a) at the same fixed notched-noise level.

In terms of the average data, the present experiment (Fig. 3 and the 30-dB notched-noise condition in Fig. 4) replicated the Plack and Viemeister result (1992a) that the notched noise removed or reduced the midlevel hump. In addition, the present experiment also measured forward-masked intensity discrimination functions at other fixed notched-noise levels from −10 to 50 dB in 20-dB steps. The results indicate that the notched noise did not always remove the midlevel hump, but rather produced level-dependent dual effects on forward-masked intensity discrimination. At low notched-noise levels where on-frequency masking was negligible, the notched noise actually increased the size of the midlevel hump in all three listeners. At high notched-noise levels where significant on-frequency masking occurred, the notched noise decreased the size of the midlevel hump. These new data (Figs. 2 and 4) suggest a close association

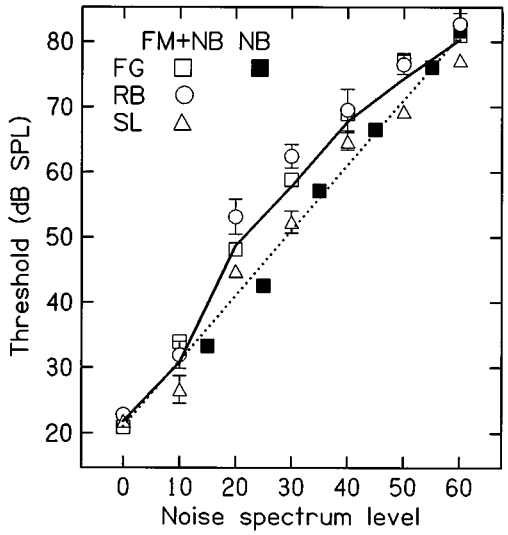


FIG. 5. Growth-of-masking function for a 1000-Hz tone in a narrow-band noise with forward masking: individual data (symbols) and the average data (solid line). The filled squares represent the measured growth-of-masking in noise without forward masking from a single listener and the dotted line represents predicted linear masking function. Bars on each data point represent the size of one standard deviation.

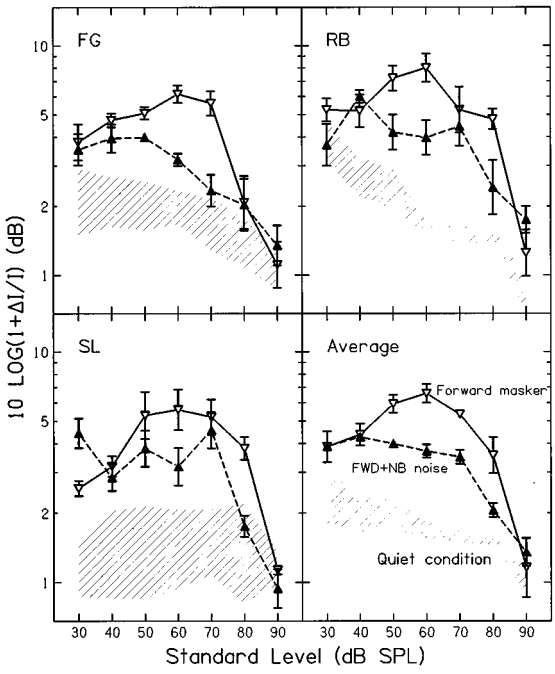


FIG. 6. Effects of the narrow-band noise on intensity discrimination under forward masking (solid triangles). Also shown are the intensity discrimination functions under the quiet control condition (shaded areas) and the forward-masking control condition (open, inverted triangles).

between the removal of the midlevel hump and the on-frequency masking caused by the notched noise.

III. THE NARROW-BAND NOISE EXPERIMENT

A. Rationale and method

A direct test of the relation between on-frequency masking at the threshold level and reduction in the midlevel hump by the notched noise would be to use a narrow-band noise centered on the test frequency. The narrow-band noise would produce direct excitatory masking at the test frequency in contrast to the remote ''suppressive'' masking caused by a notched noise (Delgutte, 1990a, b). Thresholds for the 25-ms, 1000-Hz tone were first measured in both the narrow-band noise only condition and the noise masker plus the forward-masker condition. A 400-Hz bandwidth of the ''narrow-band'' noise was chosen to fill in the ''notch'' in the Plack and Viemeister study. It was noted that 400 Hz was about three times the ''equivalent-rectangular-bandwidth'' at 1000 Hz (Moore and Glasberg, 1983b) and 2.5 times of the critical bandwidth at 1000 Hz (Zwicker and Fastl, 1990). Thus the present narrow-band noise may limit to some degree the off-frequency listening. The noise spectrum levels ranged from 0 to 60 dB in 10-dB steps. Once the growth-of-masking function was obtained, a narrow-band noise level was selected individually for each subject to approximate the threshold shift caused by the equivalent notched noise in the Plack and Viemeister study (1992a).

B. Results and discussion

Figure 5 shows the growth-of-masking function for the 1000-Hz tone in the narrow-band noise condition with and without forward masking. The open symbols represent individual data and the thick solid line represents the average data in forward masking. The solid squares represent the measured growth-of-masking function in the noise alone condition from a single subject (FG). The dotted line represents predicted thresholds (21 dB above the noise spectrum level) based on the 130-Hz auditory bandwidth at 1000 Hz (Moore and Glasberg, 1983b). Note the consistency between the measured and predicted threshold data.

Similar to previous observations made by Carlyon and Beveridge (1993), the forward masker caused additional masking at medium narrow-band noise levels but not much at low and high noise levels. This nonlinear, *decreasingly* steep (except for narrow-band noise levels between 0 and 10 dB) growth-of-masking function in the narrow-band condition contrasts sharply with the *increasingly* steep growth-of-masking function in the notched-noise condition (Fig. 2). Using the same linear regression analysis, the average slope of the growth-of-masking function was estimated to be 1.34 ($r=0.97$) at low narrow-band noise levels (0, 10, and 20 dB) and 0.63 ($r=1.00$) at high noise levels (40, 50, and 60 dB). These slope values were almost reversed from the values at low and high notched-noise levels. These slope differences may reflect two different physiological mechanisms: a suppressive masking caused by the notched noise and an excitatory masking caused by the narrow-band noise (Delgutte, 1990a, b).

Figure 6 shows individual and average intensity discrimination data in the forward masker plus the narrow-band noise condition (solid triangles) as well as data in two additional control conditions. The quiet control condition produced an intensity discrimination function abiding by the near-miss to Weber's law. The forward-masking alone control condition produced a nonmonotonic intensity discrimination with a midlevel hump which was similar in magnitude

to that found in the original Zeng *et al.* study (1991) but smaller than that found in the Plack and Viemeister study (1992a) and the present experiment 1. Similar to the notched noise, the narrow-band noise also removed the midlevel hump. Both noise data suggest that on-frequency masking, independent of its origin, can remove or reduce the midlevel hump.

IV. THE HIGH-PASS NOISE EXPERIMENT

A. Rationale and method

The fixed-low-level, notched-noise data (Fig. 4) indicate that, under conditions where on-frequency masking was minimized, limiting off-frequency listening could increase the midlevel hump as Plack and Viemeister (1992a) originally predicted. To further separate the interactions between off-frequency listening and on-frequency masking, the present experiment used a high-level, high-pass noise to produce minimal on-frequency masking while effectively limiting off-frequency listening on the high-frequency side of the excitation pattern.

First, a forward-masking pattern of the 1000-Hz pure tone was measured to avoid common problems such as combination tones and beats in tone-on-tone simultaneous masking (Egan and Hake, 1950). The forward masker was a 1000-Hz tone and had a duration of 100 ms. Thresholds were measured for a 5-ms test tone with 1-ms signal delay. Test frequencies were at 500, 1000, 1500, 2000, 3000, and 4000 Hz. All stimuli had cosine-squared 2.5-ms ramps. A masking pattern measured under these conditions has been referred as the ''internal representation'' of the forward masker (Bacon and Brandt, 1982; Sidwell and Summerfield, 1985), and can be used to derive the actual ''excitation pattern'' by taking into account the nonlinear characteristics of forward masking (Moore and Glasberg, 1983a).

Second, a simultaneous masking pattern of the high-level, high-pass noise was obtained by measuring thresholds of a 25-ms test tone presented at the temporal center of the high-pass noise. The signal frequency was chosen at 1000, 1250, 1500, 2000, 3000, and 4000 Hz. The forward masker was also included in this experiment, but its effect seemed to be negligible because the obtained masking pattern was very similar to that of Schlauch (1994) in the absence of forward masker. Finally, forward-masked intensity discrimination was measured in the presence of this additional high-level, high-pass noise masker.

B. Results and discussion

Figure 7 shows the average data of the three subjects, except for the 60 dB SPL tonal forward-masking pattern which is based on the data of a single subject (SL). The absolute thresholds of the tonal stimuli are represented by the dotted line and marked as ''Quiet threshold.'' The forward-masking patterns of the 1000-Hz tone (solid lines) show broad excitation areas across frequency, particularly at the high level (90 dB) and toward the high-frequency side. The high-pass noise simultaneous masker caused a 10-dB threshold shift for the 1000-Hz tone, while elevating the threshold to 60 dB SPL for the 1500-Hz tone and to about 80 dB SPL

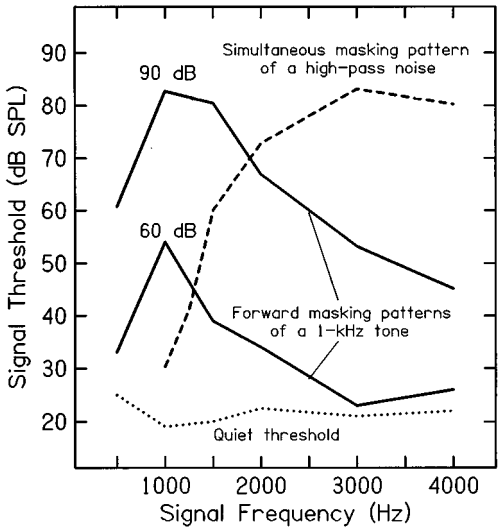


FIG. 7. Forward-masking patterns for the 1000-Hz test tone at 60 and 90 dB SPL (two solid lines). A simultaneous masking pattern for a high-pass noise presented at a 50-dB spectrum level (the dashed line). Absolute thresholds in the quiet condition (the dotted line).

for frequencies higher than 2500 Hz (dashed line). The data suggest that the present high-pass noise can effectively limit off-frequency listening on the high-frequency side while producing minimal on-frequency masking.

Figure 8 shows forward-masked intensity discrimination data in the presence of the high-pass noise (solid triangles). The same two control conditions as in the narrow-band noise experiment are also included. Both individual and average data show degraded intensity discrimination performance at medium and high standard levels in the high-pass noise condition, and the same or slightly better performance at low

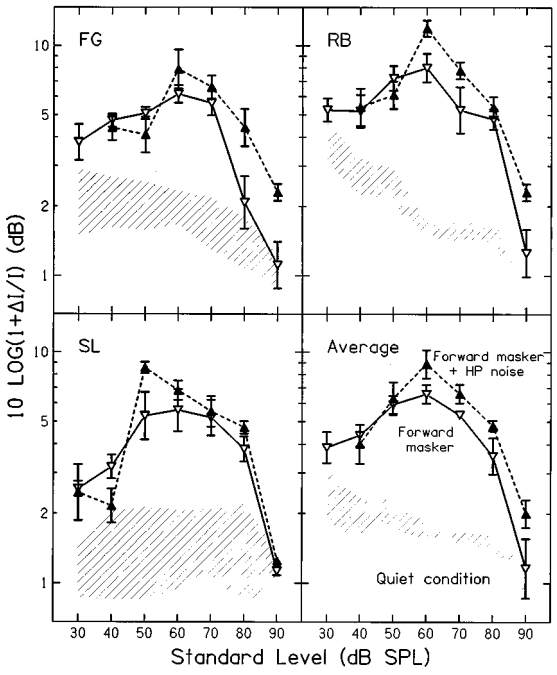


FIG. 8. Effects of the high-pass noise on intensity discrimination under forward masking (solid triangles). Also shown are the same two control conditions as in Fig. 6.

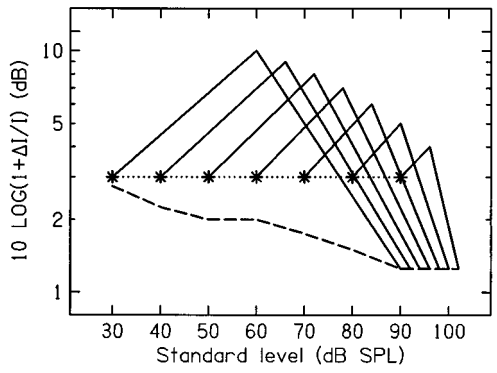


FIG. 9. A schematic model to show how on-frequency masking at the threshold level can remove the midlevel hump in forward-masked intensity discrimination. The dashed line represents the intensity discrimination function in quiet. Solid lines represent a family of forward-masked intensity discrimination functions with successive signal threshold shifts caused by the noise simultaneous masker. Asterisks connected by the dotted line represent the intensity discrimination function under a condition similar to that in the Plack and Viemeister study (1992a).

levels. In contrast to the removal of the midlevel hump by the notched noise and narrow-band noise, a greater midlevel hump is observed in the high-pass noise condition than in the control condition.

The present results indicate a clear role of off-frequency listening in forward-masked intensity discrimination: Under conditions where the on-frequency masking problem is controlled, indeed as Plack and Viemeister (1992a) suggested, limiting the off-frequency listening cue ''should, if anything, increase the magnitude of the effect.'' The increased midlevel hump under limited off-frequency listening conditions adds another piece of evidence to the body of literature demonstrating the role of off-frequency listening in auditory perception. Consistent with previous studies without forward masking (Schlauch, 1994), the present result also shows that the effect of off-frequency listening is greater at moderate and high levels than at low levels. An obvious interpretation for this level effect is that a low-level tone generates localized excitation around the test frequency while moderate to high level tones cause excitation to spread to other frequency regions, which causes minimal effect of the high-pass noise on intensity discrimination at low levels and a significant effect at moderate to high levels.

V. FINAL REMARKS

The most important finding in the present study is the clarification of the role of off-frequency listening in forward-masked intensity discrimination. If the on-frequency masking is minimized by using a low-level notched noise or a high-level high-pass noise, then limiting off-frequency listening does indeed increase the magnitude of the midlevel hump in forward masking. In addition, based on the systematic measurement of the growth-of-masking functions, the present study also shows that the removal of the midlevel hump by the notched noise or the narrow-band noise is always associated with significant on-frequency masking.

Figure 9 presents the data (Fig. 4) in a schematic fashion to demonstrate at a phenomenological level how on-frequency masking could reduce the midlevel hump in forward masking. The dashed line represents the near-miss to Weber's law, or the intensity discrimination function in quiet. The solid lines represent a family of ''idealized'' forward-masked intensity discrimination functions, which from left to right reflect an increasingly elevated threshold at the test frequency (on frequency masking) as a function of the noise level. The asterisk symbol on each function represents intensity discrimination performance measured at the lowest standard level. Figure 9 indicates that the notched noise, when presented at a fixed level, did not actually remove the midlevel hump but rather shifted the hump position to higher levels. However, when the noise level was increased along with the standard level (asterisks connected by the dashed line), the noise appeared to ''remove'' the midlevel hump due to the increased on-frequency masking.

Peripheral mechanisms like adaptation and suppression can account for the removal of the midlevel hump by the notched and narrow-band noises. Based on a differential recovery from prior stimulation by low- and high-threshold auditory neurons (Relkin and Doucet, 1991), Zeng *et al.* (1991) proposed that the midlevel hump is a result of an intensity coding gap at moderate levels in which the high-threshold neurons that normally code intensity at these moderate levels are not recovered whereas the recovered low-threshold neurons are saturated. By shifting the neural dynamic range (Geisler and Sinex, 1980; Palmer and Evans, 1982; Delgutte, 1990a), the notched noise effectively shifts the coding gap and its resulting midlevel hump towards high levels (Fig. 4). The narrow-band noise also shifts the neural dynamic range but not as effectively as the notched noise (Gibson *et al.*, 1985), and accordingly reduces the midlevel hump (Fig. 6) but to a lesser degree than the notched noise (Fig. 3).

Central mechanisms like profile analysis in the frequency domain (Green, 1988) and a similar mechanism in the time domain, namely, ''referential coding'' (Plack *et al.*, 1995; Plack, 1996) can also account for the removal of the midlevel hump by the notched and narrow-band noises. The central hypothesis explains the reduced hump at low sensation levels due to relative comparisons across frequency and time that are more effective when the signal and the noise ''context'' are closer in level (Durlach and Braida, 1969; Green and Kidd, 1983). When the noise level was increased along with the signal, both the temporal and spectral profiles of the signal-noise complex were kept constant at the same favorable level, resulting in a removal of the midlevel hump by the noise.

At present, there is a body of evidence for a central origin of the midlevel hump, for example, the greater midlevel hump in backward masking than in forward masking (Plack and Viemeister, 1992b), the presence of a midlevel hump in cochlear implant users (Zeng and Shannon, 1995), or the greater midlevel hump with a shorter forward masker (Schlauch *et al.*, 1997). On the other hand, there is also a body of evidence for a peripheral origin of the midlevel hump (Turner *et al.*, 1992; Zeng and Shannon, 1995; Schlauch and Clement, 1997; Zeng *et al.*, 1996). In particular, Schlauch and Clement (1997) took advantage of a

binaural hearing phenomenon in which a 80 dB SPL sound was perceptually unnoticeable in the simultaneous presence of a 93 dB SPL sound in the contralateral ear. However, Schlauch found that this unnoticeable 80-dB forward masker in the ipsilateral ear produced a significant midlevel hump even though the contralateral 93-dB forward masker alone did not produce a midlevel hump. It appears that both peripheral and central mechanisms are involved in the midlevel hump in nonsimultaneous masking and the removal of the midlevel hump by the simultaneous noise masker. Quantitative contributions of these peripheral and central mechanisms to the nonmonotonic intensity discrimination function in forward masking remains to be evaluated.

ACKNOWLEDGMENTS

The author thanks Manuel Don for his suggestion on the design of experiment 3 and John Galvin for his technical support in experiment 1. The author also thanks Monita Chatterjee, Bob Shannon, Chris Turner, Alena Wilson, Brian Moore, Chris Plack, and Bert Schlauch for their comments on an earlier version of this paper. This work was supported by the National Institutes of Health (NIDCD-DC02267).

Allen, J. B., and Neely, S. T. (**1992**). ''Micromechanical models of the cochlea,'' Phys. Today **7**, 40–47.

Bacon, S. P. (**1990**). ''Effect of masker level on overshoot,'' J. Acoust. Soc. Am. **88**, 698–702.

Bacon, S. P., and Brandt, J. F. (**1982**). ''Auditory processing of vowels by normal-hearing and hearing-impaired listeners,'' J. Speech Hear. Res. **25**, 339–347.

Buus, S., and Florentine, M. (**1991**). ''Psychometric functions for level discrimination,'' J. Acoust. Soc. Am. **90**, 1371–1380.

Carlyon, R. P., and Beveridge, H. A. (**1993**). ''Effects of forward masking on intensity discrimination, frequency discrimination, and the detection of tones in noise,'' J. Acoust. Soc. Am. **93**, 2886–2895.

Costalupes, J. A., Young, E. D., and Gibson, D. J. (**1984**). ''Effects of continuous noise backgrounds on rate responses of auditory nerve fibers in cat,'' J. Neurophysiol. **51**, 1326–1344.

Delgutte, B. (**1990a**). ''Physiological mechanisms of psychophysical masking: Observation from auditory-nerve fibers,'' J. Acoust. Soc. Am. **87**, 791–809.

Delgutte, B. (**1990b**). ''Two-tone suppression in auditory-nerve fibers: Dependence on suppresser frequency and level'' Hearing Res. **40**, 225–246.

Durlach, N. I., and Braida, L. D. (**1969**). ''Intensity perception. I. Preliminary theory of intensity resolution,'' J. Acoust. Soc. Am. **46**, 372–383.

Egan, J. P., and Hake, H. W. (**1950**). ''On the masking pattern of a simple auditory stimulus,'' J. Acoust. Soc. Am. **22**, 622–630.

Florentine, M. (**1983**). ''Intensity discrimination as a function of level and frequency and its relation to high-frequency listening,'' J. Acoust. Soc. Am. **74**, 1375–1379.

Florentine, M., and Buss, S. (**1981**). ''An excitation-pattern model for intensity discrimination,'' J. Acoust. Soc. Am. **70**, 1626–1654.

Geisler, C. D., and Sinex, D. G. (**1980**). ''Responses of primary auditory fibers to combined noise and tonal stimuli,'' Hearing Res. **3**, 317–224.

Gibson, D. J., Young, E. D., and Costalupes, J. A. (**1985**). ''Similarity of dynamic range adjustment in auditory nerve and cochlear nuclei,'' J. Neurophysiol. **53**, 940–958.

Green, D. M. (**1988**). *Profile analysis: Auditory Intensity Discrimination* (Oxford U.P., Oxford, England).

Green, D. M., and Kidd, G. Hr. (**1983**). ''Further studies of auditory profile analysis,'' J. Acoust. Soc. Am. **73**, 1260–1265.

Hall, J. W., and Fernandes, M. A. (**1983**). ''Temporal integration, frequency resolution, and off-frequency listening in normal-hearing and cochlear-impaired listeners,'' J. Acoust. Soc. Am. **74**, 1172–1177.

Hellman, R. P. (**1978**). ''Dependence of loudness growth on the skirts of excitation patterns,'' J. Acoust. Soc. Am. **63**, 1114–1119.

Kim, D. O., and Molnar, C. E. (**1979**). ''A population study of cochlear nerve fibers: Comparison of spatial distribution of average-rate and phase-locking measures of responses to single tones,'' J. Neurophysiol. **42**, 16–30.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

McGill, W. J., and Goldberg, J. P. (**1968**). ''A study of the near-miss involving Weber's law and pure-tone intensity discrimination,'' Percept. Psychophys. **4**, 105–109.

Moore, B. C. J., and Glasberg, B. R. (**1983a**). ''Masking patterns of synthetic vowels in simultaneous and forward masking,'' J. Acoust. Soc. Am. **73**, 906–917.

Moore, B. C. J., and Glasberg, B. R. (**1983b**). ''Suggested formulae for calculating auditory-filter shapes and excitation patterns,'' J. Acoust. Soc. Am. **74**, 750–753.

Moore, B. C. J., and Raab, D. H. (**1974**). ''Pure-tone intensity discrimination: Some experiments relating to the ''near-miss'' to Weber's law,'' J. Acoust. Soc. Am. **55**, 1049–1054.

Moore, B. C. J., and Sek, A. (**1994**). ''Effects of carrier frequency and background noise on the detection of mixed modulation,'' J. Acoust. Soc. Am. **96**, 741–751.

Moore, B. C. J., Glasberg, B. R., Hess, R. F., and Birchall, J. P. (**1985**). ''Effects of flanking noise bands on the rate of growth of loudness of tones in normal and recruiting ears,'' J. Acoust. Soc. Am. **77**, 1049–1054.

O'Loughlin, B. J., and Moore, B. C. J. (**1981**). ''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 Moore, B. C. J. (**1995**). ''Overshoot and the 'severe departure' from Weber's law,'' J. Acoust. Soc. Am. **97**, 2442–2453.

Palmer, A. R., and Evans, E. F. (**1982**). ''Intensity coding in the auditory periphery of the cat: Responses of cochlear nerve and cochlear nucleus neurons to signals in the presence of bandstop masking noise,'' Hearing Res. **7**, 305–323.

Penner, M. J., Leshowitz, B., Cudahy, E., and Richard, G. (**1974**). ''Intensity discrimination for pulsed sinusoids of various frequencies,'' Percept. Psychophys. **15**, 568–570.

Plack, C. J. (**1996**). ''Temporal factors in referential intensity coding,'' J. Acoust. Soc. Am. **100**, 1031–1042.

Plack, C. J., and Viemeister, N. F. (**1992a**). ''The effects of notched noise on intensity discrimination under forward masking,'' J. Acoust. Soc. Am. **92**, 1902–1910.

Plack, C. J., and Viemeister, N. F. (**1992b**). ''Intensity discrimination under backward masking,'' J. Acoust. Soc. Am. **92**, 3087–3101.

Plack, C. J., Carlyon, R. P., and Viemeister, N. F. (**1995**). ''Intensity discrimination under forward and backward masking: Role of referential coding,'' J. Acoust. Soc. Am. **97**, 1141–1149.

Relkin, E. M., and Doucet, J. R. (**1991**). ''Recovery from prior stimulation. I. Relationship to spontaneous firing rates of primary auditory neurons,'' Hearing Res. **55**, 215–222.

Schlauch, R. S. (**1994**). ''Intensity discrimination and loudness in high-pass noise,'' J. Acoust. Soc. Am. **95**, 2171–2179.

Schlauch, R. S., and Clement, B. R. (**1997**). ''Mechanisms underlying nonmonotonic Weber functions in forward-masked intensity discrimination,'' J. Acoust. Soc. Am. (in revision).

Schlauch, R. S., Lanthier, N., and Neve, J. (**1997**). ''Forward-masked intensity discrimination: Duration effects and spectral effects,'' J. Acoust. Soc. Am. **102**, 461–467.

Schneider, B. A., and Parker, S. (**1987**). ''Intensity discrimination and loudness for tones in notched noise,'' Percept. Psychophys. **41**, 253–261.

Sidwell, A., and Summerfield, Q. (**1985**). ''The effect of enhanced spectral contrast on the internal representation of vowel-shaped noise,'' J. Acoust. Soc. Am. **78**, 495–506.

Smith, R. L. (**1979**). ''Adaptation, saturation, and physiological masking in single auditory-nerve fibers,'' J. Acoust. Soc. Am. **65**, 166–178.

Turner, C. W., Zeng, F.-G., Relkin, E. M., and Horwitz, A. R. (**1992**). ''Frequency discrimination in forward and backward masking,'' J. Acoust. Soc. Am. **92**, 3102–3108.

Viemeister, N. F. (**1972**). ''Intensity discrimination of pulsed sinusoids: The effects of filtered noise,'' J. Acoust. Soc. Am. **51**, 1265–1269.

Viemeister, N. F. (**1974**). ''Intensity discrimination of noise in the presence of band-reject noise,'' J. Acoust. Soc. Am. **56**, 1594–1600.

Zeng, F.-G., and Shannon, R. V. (**1995**). ''Possible origins of the nonmonotonic intensity discrimination function in forward masking,'' Hearing Res. **82**, 216–224.

Zeng, F.-G., and Turner, C. W. (**1991**). ''Binaural loudness matches in unilaterally impaired listeners,'' Q. J. Exp. Psych. **43A**, 565–583.
Zeng, F.-G., and Turner, C. W. (**1992**). ''Intensity discrimination in forward masking,'' J. Acoust. Soc. Am. **92**, 782–787.
Zeng, F.-G., Turner, C. W., and Relkin, E. M. (**1991**). ''Recovery from prior stimulation. II. Contribution to intensity discrimination,'' Hearing Res. **55**, 223–230.
Zeng, F.-G., Galvin, J. J., and Relkin, E. M. (**1996**). ''Relative contributions of two neural populations responses to intensity coding,'' J. Acoust. Soc. Am. **100**, 2719.
Zwicker, E. (**1970**). ''Masking and psychophysical excitation as consequences of the ear's frequency analysis,'' in *Frequency Analysis and Periodicity in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden), pp. 376–396.
Zwicker, E., and Fastl, H. (**1990**). *Psychoacoustics—Facts and Models* (Springer-Verlag, Berlin).

Intensity discrimination for precedence effect stimuli

Richard L. Freyman
Department of Communication Disorders, University of Massachusetts, Amherst, Massachusetts 01003

Daniel D. McCall and Rachel K. Clifton
Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01003

(Received 1 July 1997; revised 13 January 1998; accepted 21 January 1998)

When two identical stimuli are presented from two loudspeakers with a brief delay between them, a single image is heard near the source of the leading sound. The delayed sound or echo appears to be suppressed whereas the preceding sound determines perceived location, hence the name, the precedence effect. This study investigated normal-hearing listeners' sensitivity to changes in the intensity of the lagging sound. Pairs of 2-ms white noise bursts, with a 2-ms delay between the onsets of lead and lag, were presented from two loudspeakers 45° left and right of midline in an anechoic chamber. A 2AFC procedure was used to test discrimination of intensity changes in the lead, lag, and both sounds together. The untreated results showed discrimination to be poorest for changes in the lag stimulus. However, when the intensity differences were transformed into predictions of equivalent monaural level based on KEMAR measurements and binaural loudness summation, discrimination for the lag was equal to the other two conditions. A follow-up experiment found that listeners were highly sensitive to the presence of the lag, more sensitive than would be predicted from loudness changes. It is concluded that the precedence effect does not consist of a general suppression or attenuation of the lagging sound, but rather that suppression may be limited to directionality cues. © *1998 Acoustical Society of America.* [S0001-4966(98)04504-4]

PACS numbers: 43.66.Fe, 43.66.Pn, 43.66.Qp [WJ]

INTRODUCTION

Research on the precedence effect has been concerned with at least two related phenomena. First, if there is only a brief delay between them, two coherent sounds delivered from two loudspeakers produce only one perceptual image. The fact that the delayed sound is not perceived as a separate auditory event has been called echo suppression (Green, 1976, p. 216). The second feature is that as long as the delay exceeds approximately 1 ms, the single image is perceived to come from the vicinity of the lead loudspeaker, suggesting that the directional information in the lag has been suppressed (see Zurek, 1987). Because the image is heard near the location of the lead, we might assume, incorrectly, that only the lead sound is heard. In fact, the lag sound does contribute to perception of the image. The most well known of these contributions is the small but measurable influence the lag sound has on the location of the fused image. Both the classic and more recent data indicate that the lag's contribution to perceived location is approximately $\frac{1}{6}$ to $\frac{1}{10}$ the size of the lead's, as measured by direct localization judgments or discrimination measures of the precedence effect (e.g., Wallach *et al.*, 1949; Yost and Soderquist, 1984; Saberi and Perrott, 1990; Shinn-Cunningham *et al.*, 1993; Litovsky and Macmillan, 1994).

Less well understood is the lag sound's influence on other attributes of the sound image. Although the lead-lag combination sounds clearly different from the lead alone in qualities such as loudness, timbre, and spaciousness of the auditory image (Blauert, 1983), the specific contributions of the lag sound to these qualities have not been fully quantified. From common experience we know that rooms of different sizes, shapes, and absorbency of room surfaces sound quite different from one another, implying that listeners have at least a gross sensitivity to variables such as the intensity, frequency content, and delays of reflections. The quantitative work that has been done in this topic has mostly been concerned with the detectability of delayed sounds in the presence of primary sound (Burgtorf, 1961; Seraphim, 1961; Sommerville *et al.*, 1966; Olive and Toole, 1989; Bech, 1995; Guzman and Yost, 1996). Bech (1995) also investigated the just-noticeable difference (jnd) for intensity of simulated reflections. The results of these studies indicate that the detectability and discriminability of delayed sounds depends, at a minimum, upon the type of stimulus, the delay of the reflection, the spatial separation of the primary and delayed sounds, and the presence or absence of other reflections. What is difficult to determine from these studies is which aspects of the lag sound have been suppressed. Increasing our knowledge in this area is important for a general understanding of how listeners process reflected sounds. Precise descriptions of what is being suppressed, and by how much, are necessary to guide both physiological research and psychophysical models of the precedence effect.

The current experiment attempted to answer the following question: With a lead sound presented from one loudspeaker and a lag sound from a second speaker, how does the discriminability of changes in the level of the lag compare with that for changes in the level of the lead? For simple stimuli delivered monaurally, discrimination is quite good for practiced listeners. When expressed in terms of ΔL, i.e., the difference in decibels between two sounds being discriminated, difference limens are roughly 0.5 dB for broadband noise (Miller, 1947) and pure tones presented at levels well above threshold (Jesteadt *et al.*, 1977). Current concep-

TABLE I. Example of changes in loudspeaker outputs for the three conditions in experiment I where $\Delta L = 6$ dB. Values in the tables are in dBC as measured with a microphone at the position of the listener's head with the listener absent.

	Interval 1		Interval 2		
Condition	L spkr	R spkr	L spkr	R spkr	
Lead discrim.	44	44	50	44	The lead changed by 6 dB
Lag discrim.	50	44	50	50	The lag changed by 6 dB
Control	44	44	50	50	Both changed by 6 dB

tualizations of the precedence effect (e.g., Lindemann, 1986; Zurek, 1987) include an inhibition of the lag sound, but only of the *directional* information in the lag. Thus they do not predict explicitly that changes in the intensity of a lag should be less discriminable than for a lead. The previous empirical research most relevant to this question was conducted by Shinn-Cunningham (1988). Subjects listened monaurally under headphones to pairs of 1-ms noise bursts, with one of the bursts delayed by 1, 5, or 10 ms relative to the other. The leading and trailing bursts were presented at the same level in one interval of a 2AFC trial. In the other interval the level of either the lead or lag burst was reduced, while the other was unchanged. The results showed that discrimination performance did not depend on whether the leading burst or the trailing burst was attenuated, even at the shortest delay of 1 ms where the precedence effect operates strongly.

The purpose of the current study was to determine whether a lead sound suppresses sensitivity to intensity changes of a lag sound when there is spatial separation between lead and lag in a sound field. In the first of two experiments, listeners discriminated changes in the level of a lagging noise burst, in the level of a leading noise burst, and in the level of both bursts changed together. With a conceptualization of the precedence effect as an inhibition of only *directional* information of the later sound, our prediction was that there would be no difference between intensity discrimination of lead and lag. An extension of this prediction is that if the lag sound is not generally suppressed then its detectability should be limited only by its level relative to the lead. The second follow-up experiment examined the detectability of the lag sound compared to when the lead sound was presented with the lag absent.

I. EXPERIMENT I. INTENSITY DISCRIMINATION

A. Method

1. Stimuli and apparatus

Experiments were conducted in an anechoic chamber with the dimensions 4.9 m×4.1 m×3.1 m. The floor, ceiling, and walls of the chamber are lined with 0.72-m foam wedges. Subjects sat in a chair near the center of the room equidistant between two loudspeakers located at 45° left and right at a distance of 1.9 m. The center of the loudspeakers is 1.4 m above the wire mesh floor of the chamber, the approximate height of the average subject's ears while seated in the chair. Both lead and lag stimuli were gaussian noise bursts of 2-ms duration, gated on and off without rise and fall times. Ten such noise burst samples were synthesized digitally on a computer. They were presented at a 20-kHz digital-to-analog conversion rate (TDT QDA2). The signals were low-pass filtered at 8.5 kHz (TTES J1390), attenuated under computer control (TDT PA3), amplified (NAD 2100) and delivered to a pair of Realistic Minimus 7 loudspeakers. The lead stimulus was presented to the left loudspeaker, and the lag stimulus, delayed by 2 ms in all but the Control condition, was delivered to the right loudspeaker. The level of the stimuli was calibrated with the subject not present and a microphone at the position at the center of the listener's head, while the stimuli were presented repeatedly at rate of 4/s. The microphone output was read by a sound level meter (B&K 2104) set to the ''C'' scale and ''fast'' response. The nominal level of the bursts from an individual loudspeaker was 50 dBC.

2. Subjects and procedures

Four college-aged adults with normal hearing confirmed by audiometry served as listeners. They were presented with three conditions, as shown in Table I and displayed schematically in Fig. 1: (1) discrimination of an increase in the level of the lead—''Lead discrimination''; (2) discrimination of a reduction in the level of the lag—''Lag discrimination''; and (3) a ''Control'' condition in which both stimuli changed together and there was no delay between them. A two-interval-forced-choice (2IFC) paradigm was used in which subjects were instructed to report the louder interval. In Lead discrimination both lead and lag were presented ΔL dB lower than the nominal level of 50 dBC in one interval; in the other interval the lead was raised by ΔL to 50 dBC. In Lag discrimination, both lead and lag were presented at 50 dBC in one interval; in the other the lag was reduced by ΔL. In the Control condition both stimuli were presented at 50 dBC in one interval, and both were reduced by ΔL in the other interval. The asymmetry created by reducing the level

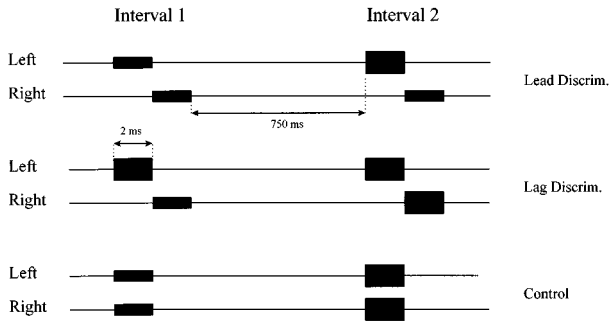


FIG. 1. Schematic diagram of an example of a two-interval trial for each of the three conditions. Amplitude envelopes of the noise bursts are shown as a function of time for the left and right loudspeakers. Note that the timeline is not to scale.

TABLE II. Individual and mean discrimination performance in d' in experiment I.

	ΔL in dB							
	1	2	3	4	5	6	10	20
Lead								
ejc	···	0.40	0.43	1.44	1.05	1.44	···	···
djr	···	−0.25	1.08	0.75	1.99	1.82	···	···
bam	···	0.57	0.71	0.92	1.30	2.20	···	···
hao	···	0.04	0.28	0.45	0.88	1.21	···	···
mean	···	0.19	0.63	0.89	1.30	1.67	···	···
stderr	···	0.18	0.18	0.21	0.24	0.22	···	···
Lag								
ejc	···	0.70	0.90	0.99	1.44	1.31	1.86	2.61
djr	···	0.53	0.66	0.98	0.61	0.9	1.18	1.39
bam	···	0.44	0.53	1.40	1.39	1.33	1.55	1.92
hao	···	0.52	0.80	0.70	1.04	0.83	1.21	1.99
mean	···	0.55	0.72	1.02	1.12	1.09	1.45	1.98
stderr	···	0.05	0.08	0.15	0.19	0.13	0.16	0.25
Control								
ejc	0.37	1.03	1.26	2.40	2.48	3.00	···	···
djr	0.28	0.83	1.31	1.52	1.85	2.83	···	···
bam	−0.30	1.46	1.28	1.82	2.61	3.00	···	···
hao	0.53	0.82	1.82	1.04	1.66	1.66	···	···
mean	0.22	1.03	1.41	1.69	2.15	2.62	···	···
stderr	0.18	0.15	0.13	0.29	0.23	0.33	···	···

for the Lag discrimination conditions and increasing it for the Lead conditions was necessary to avoid the situation where the lag sound would be higher in level than the lead sound. This relationship tends to produce large shifts in the position of the image toward the lag location (see Leakey and Cherry, 1957), and would provide an easy cue for subjects. Table I displays the stimulus levels for both lead and lag in the two intervals for an example where ΔL is 6 dB. In the actual experiment, the order of the stimuli in the intervals was reversed on a random basis on 50% of trials. During a 2IFC trial, the same sample of noise was presented from both loudspeakers in a given interval in each trial, but across the two intervals of a forced-choice trial, and across trials, the samples were chosen randomly from among the ten samples created.

Values of ΔL in the initial testing ranged from 2 to 6 dB in 1-dB steps. Data were collected in blocks of 30 trials each with the condition (Lead, Lag, or Control) and ΔL fixed within a block. Each condition was repeated three times in a random order for a total of 90 trials per data point for each subject. After the analysis of the initial set of data (45 blocks) two more values of ΔL, 10 dB and 20 dB, were added for the Lag condition, and one more value, 1 dB, for the Control condition, with data collection again consisting of three blocks of 30 trials each. Subjects received approximately 15 practice blocks before data collection began. Correct-answer feedback was given on both practice and experimental trials.

B. Results

From the raw percent correct data, values of d' were computed for each condition and level difference, using Macmillan and Creelman (1991). The resulting psychometric functions, listed for individual subjects in Table II and summarized with means and standard errors in Fig. 2, show that the best performance was for the Control condition. Mean performance for the Control condition increased smoothly as the intensity difference was increased, reaching a d' of 1.5 (an arbitrary criterion) at 3.3 dB. The function for the Lead discrimination was roughly parallel to the Control condition, shifted by 1–2 dB. A d' of 1.5 was achieved at approximately 5.5 dB. The lag discrimination data increased more gradually, reaching a mean d' of approximately 1.5 when the lag was reduced by 10 dB, and a d' of about 2.0 at 20 dB. The individual data displayed in Table II indicate that although there was some variability, all four subjects showed the same general trends.

The dashed line in Fig. 2 shows the simple sum of the d's of the Lead and Lag conditions. These sums approximate the d's of the Control condition, indicating that performance for the condition in which lead and lag changed together was close to the sum of the performance when the lead and lag were changed separately. If the perceptual consequences of

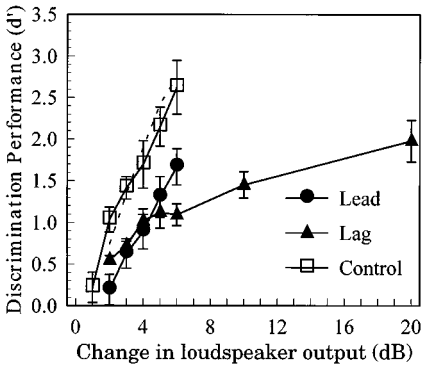


FIG. 2. Discrimination performance in d' as a function of the nominal change in presentation level from the lead, lag, or both loudspeakers.

TABLE III. Loudspeaker outputs, KEMAR measurements, and predicted ΔL's for three transformations of the experimental conditions.

	Speaker output (dBC)				KEMAR levels (dB)							Bionaural loudness		
	Interval 1		Interval 2		Interval 1		Interval 2		Individual ears		Summed energy	L_m (dB SL)		
ΔL	L spkr	R spkr	L spkr	R spkr	L ear	R ear	L ear	R ear	ΔL_L	ΔL_R	ΔL_T	Int. 1	Int. 2	ΔL_m
Lead														
2	48	48	50	48	56.00	56.00	57.75	56.22	1.75	0.22	1.05	65.75	66.95	1.20
3	47	47	50	47	55.00	55.00	57.63	55.44	2.63	0.44	1.67	64.60	66.45	1.85
4	46	46	50	46	54.00	54.00	57.50	54.66	3.5	0.66	2.31	63.40	65.90	2.50
5	45	45	50	45	53.00	53.00	57.47	53.94	4.47	0.94	3.05	62.25	65.50	3.25
6	44	44	50	44	52.00	52.00	57.41	53.16	5.41	1.16	3.79	61.05	65.05	4.00
Lag														
2	50	48	50	50	57.75	56.22	58.00	58.00	0.25	1.78	0.95	66.95	68.10	1.15
3	50	47	50	50	57.63	55.44	58.00	58.00	0.37	2.56	1.33	66.45	68.10	1.65
4	50	46	50	50	57.50	54.66	58.00	58.00	0.5	3.34	1.69	65.90	68.10	2.20
5	50	45	50	50	57.47	53.94	58.00	58.00	0.53	4.06	1.95	65.50	68.10	2.60
6	50	44	50	50	57.41	53.16	58.00	58.00	0.59	4.84	2.21	65.05	68.10	3.05
10	50	40	50	50	57.22	50.72	58.00	58.00	0.78	7.28	2.91	63.75	68.10	4.35
20	50	30	50	50	57.07	47.59	58.00	58.00	0.93	10.41	3.48	62.25	68.10	5.85
Control														
1	49	49	50	50	57.00	57.00	58.00	58.00	1	1	1.00	66.95	68.10	1.15
2	48	48	50	50	56.00	56.00	58.00	58.00	2	2	2.00	65.75	68.10	2.35
3	47	47	50	50	55.00	55.00	58.00	58.00	3	3	3.00	64.60	68.10	3.50
4	46	46	50	50	54.00	54.00	58.00	58.00	4	4	4.00	63.40	68.10	4.70
5	45	45	50	50	53.00	53.00	58.00	58.00	5	5	5.00	62.25	68.10	5.85
6	44	44	50	50	52.00	52.00	58.00	58.00	6	6	6.00	61.05	68.10	7.05
Col. #	1	2	3	4	5	6	7	8	9	10	11	12	13	14

changing lead and lag were independent, then the d' predicted when both were changed would be the square root of the sum of the two individual d's squared (Macmillan and Creelman, 1991, Chapter 10). The linear additivity suggests that the effects of changing the intensity of lead and lag were along a single perceptual dimension (Macmillan and Creelman, 1991), presumably loudness in this case. It is likely that lead and lag, being only 2 ms apart, were integrated within the same temporal window and hence were effectively added together.

To achieve a clearer understanding of the results, the effects of changing levels in the Lead, Lag, and Control conditions on the signals reaching the listener should be evaluated. Changes in lead and lag levels are clearly not symmetrical with respect to energy summation because of the restriction that the lag sound never be higher in level than the lead sound. Consider the Lead conditions shown in Table I, where ΔL is 6 dB. For interval 1, the sum of the energies from the left and right loudspeakers is 47 dBC $(44+44)$; for interval 2 it is 51 dBC $(50+44)$. The difference between the two intervals is 4 dB. For the Lag discrimination conditions, the sums are 51 and 53 and dBC for intervals 1 and 2, respectively. The difference between the intervals is only 2 dB. Thus decreasing the lag by 6 dB has a smaller effect on the total energy than increasing the lead by 6 dB. This suggests that the data in Fig. 2 should be transformed to make a fairer comparison among conditions. However, the energy summation just described is perhaps too simplistic because it does not take into account the differing levels that might occur at the two ears due to head shadow effects or alternative ways in which the two signals might be summed by the auditory system.

To predict how changing the levels of the signals from the loudspeakers affected the levels reaching the listeners' ears, additional acoustical measurements were made using a KEMAR manikin. The measurements were then treated in three different ways to determine which, if any, transformations of ΔL would produce a clustering of the results from the different conditions. A KEMAR manikin fitted with large pinnae and ER-11 $\frac{1}{2}$-in. microphones was placed at the position normally occupied by the listener. The preamplifier outputs were fed to a signal analyzer (HP 3569A). The signal was a continuous white noise source presented at an equal level from both loudspeakers at 60 dBC, or with the right or left loudspeaker reduced by ΔL dB. The signal was split, led through two attenuators, and delivered to the two loudspeakers at 45° left and right of midline with no delay to either loudspeaker. Symmetry was assumed, so all of the actual measurements were made at the left ear of KEMAR. The right ear was evaluated by reversing the outputs of the loudspeakers (as opposed to keeping the loudspeaker outputs the same and switching the measurement to the right ear). Table III displays the results of these measurements. Columns 1–4 in the table display the nominal levels presented from the loudspeakers for the different conditions in the two intervals. Columns 5–8 display the unweighted overall levels measured at the ears of KEMAR. For two of the three types of analyses performed on these data, only the relative levels at the ears were of importance, not the absolute level. For the third analysis, the general range of sensation levels (SLs) of

the stimuli were also important. For this reason the KEMAR measurements shown in columns 5–8 in Table III are expressed in terms of the predicted sensation levels of the bursts as they were presented during the experiment. These sensation levels were estimated from threshold measures described in a later section. After determining that the sensation level for the highest presentation level (50 dBC from each loudspeaker) was estimated to be 58 dB SL, all of the other measurements made through KEMAR were adjusted according to that reference (e.g., if a measurement was 2 dB less than that maximum, it was expressed as 56 dB SL).

The first of the three types of analysis assumed that listeners could base their discrimination on the overall intensity difference at the ear which produced the larger difference across the two intervals of the experiment, and could effectively ignore the other ear. The second analysis assumed that the discrimination was based on differences in the combined energy at the two ears. The third analysis was similar to the second, but rather than a simple summation of energies at the two ears, the two inputs were summed according to the predictions of a binaural loudness summation model. For these analyses, our subjects' data are replotted with appropriate transformations of the abscissa in an effort to determine which model best describes performance on the three conditions.[1]

1. Intensity difference at the most advantageous ear

This treatment of the data assumed that listeners use intensity differences occurring at the left ear if those differences were larger than at the right ear, and the differences at the right ear if those were larger. Table III, column 9 displays ΔL_L, the difference in decibels measured at the left ear between the two intervals of a trial (i.e., column 7 minus column 5). Column 10 displays ΔL_R (column 8 minus column 6). Consider the example of Lag discrimination where the lag was reduced by 20 dB. The difference between the two intervals of the trial was approximately 1 dB in the left ear, but was approximately 10 dB in the right ear. This difference between the ears was due to the effect of head shadow. A reduction of 20 dB from the lag (right) loudspeaker had little effect on the SPL at the left ear because the lead sound dominated the energy at that ear. However, the response of the right ear to the lead sound was attenuated by the head shadow, so changes in the lag affected the overall level. Conversely, as shown in upper rows of the table, changing the level of the lead sound produced larger differences at the left ear. Changing the output of both loudspeakers, as in the Control condition, produced equal differences at the two ears. For this analysis, it was assumed that listeners used the differences from one ear while ignoring the information from the other ear. For simplicity, we call the ear where the larger difference existed the ''better'' ear and the size of the difference ΔL_B. The data shown in Fig. 3 are a replotting of Fig. 2, with the abscissa adjusted to reflect ΔL_B the larger of columns 9 and 10 in Table III. The abscissa is plotted in terms of the Weber fractions as per Jesteadt *et al.* (1977), and because comparisons with masked detection data from experiment II are more intuitive in these terms.[2] The dB values shown on the abscissa $[10 \log \Delta I/I]$ were computed

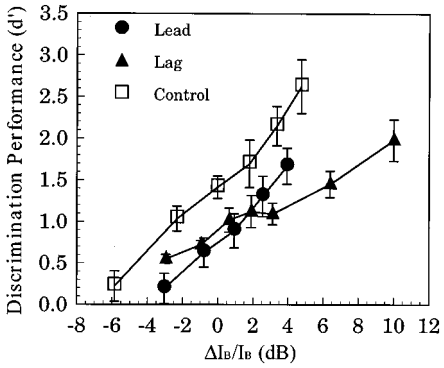


FIG. 3. Data from Fig. 2 replotted with the abscissa transformed to reflect the intensity changes at the ear of KEMAR that showed the larger difference. The abscissa is plotted in terms of the Weber fraction in dB.

by performing $10 \log[10^{(\Delta L_B/10)} - 1]$ for each value of ΔL_B. The plot is similar in form to Fig. 2, and shows large condition differences for equivalent Weber fractions for the better ear. The nature of these differences suggests that subjects did not rely on information available from the better ear while ignoring the other ear. For example, there was approximately a 3-dB difference between the Control condition and the Lead condition for equivalent levels of performance. The simplest explanation for this is that the intensity difference was available in both ears for the Control condition, but mostly in one ear for the Lead condition. Also, the difference between Lead and Lag conditions would be difficult to explain if performance depended upon the better ear, because performance is better for the lag at small Weber fractions, and better for the lead at larger Weber fractions.

2. Intensity difference for the two ears combined

An alternative analysis assumed that intensity discrimination performance was based upon the differences in the combined binaural intensities. For each condition, the intensities at the left and right ears were summed, then the difference in total level, ΔL_T, across the two intervals of the forced choice trial was computed.[3] For example, for the top row in Table III, the result of adding columns 5 and 6 for interval 1 is 59.01 dB; the sum of columns 7 and 8 for interval 2 is 60.06 dB. The difference of 1.05 dB is displayed in column 11. The differences were converted to Weber fractions and substituted for the abscissas used in the earlier figures. The data replotted in Fig. 4 show smaller differences across conditions than revealed in the previous plots. The differences were sufficiently small that we decided not to connect the data points for the different conditions but to fit a single line through all the data and evaluate the quality of the fit. The line is a linear equation in d' vs ΔL coordinates, in agreement with Rabinowitz *et al.* (1976), Buus and Florentine (1991), and Buus *et al.* (1995). The curvature of the line is due to the conversion of the abscissa to Weber fractions before plotting. Although the resulting R^2 of 0.92 suggests that the scatter of the data is reasonably small, the figure seems to show better performance for the Lag condition than for the Lead or Control conditions. The remaining differences led us to consider alternative ways in which the

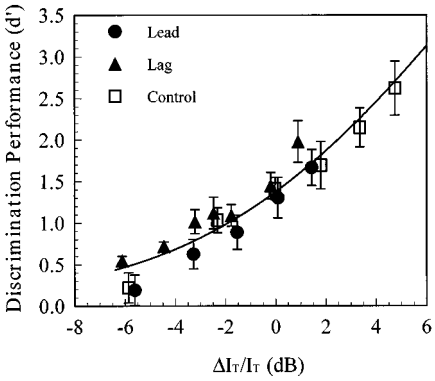


FIG. 4. Data from Fig. 2 replotted with the abscissa transformed to reflect the difference in combined intensity at left and right ears. The abscissa is plotted in terms of the Weber fraction in dB.

sound energies at the two ears might be combined that would be more relevant to intensity discrimination. The combining of loudness is one obvious possibility, as several models assume a relationship between intensity discrimination and loudness (e.g., Riesz, 1933; Lim *et al.*, 1977; Hellman and Hellman, 1990).

3. Binaural loudness summation

A few studies (e.g., Reynolds and Stevens, 1965; Irwin 1965; Treisman and Irwin, 1967) have examined the manner in which loudness is combined across the two ears. For equal inputs to the two ears, the trend is for greater summation as the level of the signal is increased. Reynolds and Stevens showed that the combined loudness produced by 10 dB SPL is approximately equivalent to that produced by monaural presentation of 13 dB SPL. However, binaural presentation of 70 dB SPL produces a loudness approximately equivalent to monaural presentation of 79 dB SPL. Similar data were presented by Irwin (1965), which included cases of asymmetric levels at the two ears. The changes in additivity as a function of level may be due to compression on the basilar membrane (see Oxenham and Moore, 1994). Analysis of the literature by Treisman and Irwin (1967) led to prediction equations for loudness for both symmetric and asymmetric presentation levels. Their general power function equation [Eq. (10)], applicable to both equal and unequal levels at the two ears, is

$$E_R = kI_R^{\log 2/[(1-b)\log I_R + \log a]}$$

and

$$E_L = kI_L^{\log 2/[(1-b)\log I_L + \log a]},$$

where E is the effect produced by the stimulus intensity I for the right and left ears, $a = 1.7$, and $b = 0.96$. E_R and E_L are then added linearly across the two ears to compute a total effect E_T. The power function is compressive because the exponent decreases as I increases.[3] One of the difficulties in applying this equation to the current data was the specification of the value of I. Unlike the previous analyses in this paper, the absolute level is important, not only the relative levels between the two ears. In particular, because the E_R and E_L are nonlinear functions of absolute signal intensity, the relative contribution of the signal at each ear to the overall loudness depends upon the absolute signal intensities at each of the ears, not just their relative intensities (see Reynolds and Stevens, 1960; Treisman and Irwin, 1967). Our stimuli were briefer than the stimuli of Irwin (1965) used to generate the equation, and were presented in a sound field rather than via earphones, so the calibrations are not easily comparable. Because the data from Irwin (1965) and the equations from Treisman and Irwin (1967) were based on the SL, not the SPL of the stimulus, we decided to estimate the sensation level of the bursts by finding their absolute thresholds. Thresholds were measured in two normal-hearing listeners using a Bekesy tracking procedure for the condition where lead and lag were at equal levels. As in the main listening conditions, the lead was to the left and the lag was to the right with a 2-ms delay. Thresholds were determined for four of the ten samples of noise bursts used during the main experiments. The threshold data indicated that during the stimulus intervals where both lead and lag were presented at 50 dBC, the stimulus was approximately 58 dB SL, which was assumed to be the same in both the left and right ears. It was for this reason that the KEMAR measurements were adjusted by a constant value before displaying them in columns 5–8 in Table III, so that the level with lead and lag at 50 dBC is shown as 58 dB SL.

The last three columns of Table III display the results of applying Eq. (10) from Treisman and Irwin (1967) to this analysis for each of the experimental conditions. For intervals 1 and 2, both E_R and E_L were computed and summed. The total, E_T was expressed as L_m, the SL of the monaural sound that would have produced the same value of E_T. Because there is no simple reverse expression for the equation, this was accomplished solving Eq. (10) from Treisman and Irwin (1967) iteratively with I varying in 0.5-dB steps to find the value of I_m (then L_m after converting to decibels) that produced the correct value of E_T. Finally, the difference in equivalent monaural SL between the two intervals (ΔL_m) was computed for each condition and is shown in the last column. Thus the last column in the table represents the change in monaural SL that would have produced the predicted change in binaural loudness.

The data shown in the previous figures were plotted once again in Fig. 5, this time with the abscissa taken from the last column of Table III after converting the values of ΔL_m to the Weber fraction in dB [$10\log(\Delta I_m/I_m)$]. With the abscissa transformed in this manner, the data from the three conditions cluster more closely together than they do with any of the previous transformations, with $R^2 = 0.95$. If the assumption is correct that discrimination performance is related to loudness, then our tentative conclusion must be that information about loudness changes is no less available when the changes are in the lag as when they are in the lead. The line through the data is a linear fit in d' vs ΔL_m. The additional symbol on the graph is taken from Viemeister (1974), who examined intensity discrimination for white noise bursts as a function of burst duration. The data point was interpolated from Viemeister's fitting of the data, which shows a change in the Weber fraction of -3.6 dB per decade

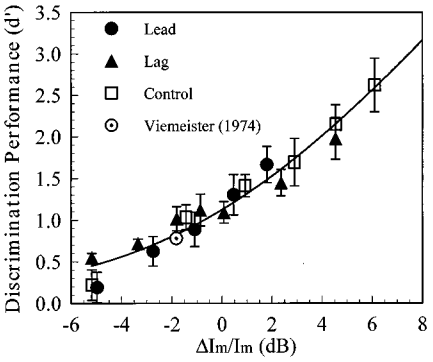


FIG. 5. Data from Fig. 2 replotted with the abscissa transformed to reflect the change in monaural level based on KEMAR measurements and the binaural loudness summation equations from Treisman and Irwin (1967). The abscissa is plotted in terms of the Weber fraction in dB.

increase in duration. We estimated and computed that the fit from Viemeister (1974) passed through a duration of 2 ms (our burst duration) at a Weber fraction of -1.82 dB. Because the adaptive threshold tracking estimated 70.7% correct performance in a 2IFC procedure (Levitt, 1971), the associated d' is 0.78 (Macmillan and Creelman, 1991). The point plotted at -1.82 dB and $d'=0.78$ demonstrates the similarity between Viemeister's subjects and ours despite the differences in methods (earphone versus loudspeaker) and the loudness summation transformation we applied to our data.

Further analysis determined the sensitivity of this basic result to the changes in the inputs to the equations. In acknowledgment of the approximate nature of the sensation level estimates, the analysis was repeated with the sensation level of each stimulus raised or lowered by 10 dB. For example, in the case where SL was raised by 10 dB, the 58 dB SL stimulus was considered instead to be 68 dB SL, and the values for all the other conditions were adjusted upward by 10 dB as well. The results of these two variations were evaluated by fitting the data in the same manner as for the 58 dB SL assumption, and examining the quality of the fit. The equations for the best fitting lines changed slightly, but the clustering of the data remained essentially the same, with R^2 at 0.95 for both variations.

Because we were also concerned about how well the KEMAR measurements represented the stimulus levels at the human ears, the analysis was repeated again with levels predicted from the tables of Shaw and Vaillancourt (1985) in place of the KEMAR measurements. Shaw and Vaillancourt (1985) presented in tabular form the data of Shaw (1974), in which the level of sound at the two ears was measured as a function of azimuth angle for different frequencies. The data were the average measurements from 100 subjects. They represent the difference between measurements made from a microphone at the center of the head (with the subject absent) and with a microphone deep in the ear canal. Our own stimulus calibration was with a microphone at the center of the head, so these differences were added to our measured SPLs. The relevant values for ± 45 degrees were taken from the rows for 45° and 315° as well as the 0° reference row from Tables I to III in Shaw and Vaillancourt (1985). The predictions were converted to overall SPL by integrating the intensity across a band from one table frequency to the next in the range of 200 Hz–8500 Hz. The differences between adjacent values on the table are typically about 1 dB or less, so little error was added by assuming a uniform SPL between the two adjacent frequencies. When these predictions were used in place of the KEMAR measurements as an input to the equation, the computed values of ΔL_m never changed by more than 0.1 dB from those displayed in the last column of Table III. Thus the KEMAR measurements appear to provide acceptable estimates of the overall levels in average human ears for the conditions of this experiment. Within a 20-dB range of SLs and with either KEMAR measurements or predictions from human ears, the analysis suggests a close agreement between discrimination performance and the change in equivalent monaural SL based on binaural loudness summation.

C. Discussion

Although the untreated data displayed in Fig. 2 reveal a difference in discrimination performance between the Lead and Lag conditions, the additional analyses performed on the stimulus conditions, which computed the changes in levels at the two ears individually (Fig. 3), changes in the combined intensities across the two ears (Fig. 4) and changes in the equivalent monaural level (Fig. 5), changed the picture completely to one of virtually no difference between Lead and Lag conditions. Interpretation of the data depends upon how these different analyses are considered. We argue that the analysis based on combined loudness, which resulted in almost no difference among the Lead, Lag, and Control conditions (Fig. 5), is also the most reasonable way of thinking about the discrimination of the level differences. With a short delay of 2 ms, subjects presumably heard one image, the loudness of which would likely depend on the inputs to both ears. Their assigned task was to determine which image was louder between the two intervals of a trial, and it seems sensible to consider, at least initially, that they used a loudness cue as the basis of their discrimination.

Figure 5 demonstrates that performance is approximately the same across the three conditions when the data are related to the change in monaural SL that would have produced the predicted change in loudness. There was equal sensitivity to level differences in lead and lag, with no priority given to the first arriving signal. Because the loudness calculation assumes no suppression of the lag, these data suggest that the precedence effect *does not* involve suppression of the lag's contribution to the loudness changes. This finding is consistent with that of Shinn-Cunningham (1988) for monaural stimuli. It can be reconciled with a conceptualization of the precedence effect where inhibition of the delayed sound is limited to its directional information (e.g., Zurek, 1987), and does not involve a general attenuation of the delayed sound.

The fact that the changes in equivalent monaural SL seem to describe the data reasonably well should not be taken as evidence that other cues such as the spaciousness or quality of the image are unavailable. Indeed, we think that at some level such cues must be available when there are

changes in the relative levels of reflections. For example, it should be easy to distinguish sounds presented in anechoic versus reverberant rooms, even when overall presentation level is allowed to vary. However, it appears that for our subjects with the current task these kinds of cues were not more salient than loudness differences. Otherwise, when plotted as in Fig. 5, discrimination for the Lead and Lag conditions should have been better than for the Control condition, where loudness was almost certainly the basis of discrimination. A less likely possibility is that both lead and lag sounds suppress one another and the agreement among conditions seen in Fig. 5 is a mere coincidence. Even if that were the case, there is still no evidence of greater suppression of the lag than of the lead.

Loudness might also be one of several cues used to detect the presence of the lag when it is added to a single source sound. In the follow-up experiment described below, we examined the detectability of the lag sound as a function of its level relative to the lead, and sought to determine whether such data could also be explained on the basis of loudness changes.

II. EXPERIMENT II. MASKED DETECTION

This experiment examined the detection of the lag sound as a function of its level relative to the lead sound. Criterion performance on this type of task has been referred to as the masked threshold in Blauert (1983), and has been studied by several researchers (e.g., Burgtorf, 1961; Seraphim, 1961; Sommerville *et al.*, 1966; Zurek, 1979; Olive and Toole, 1989; Bech, 1995; Guzman and Yost, 1996). The masked threshold is to be distinguished from the echo threshold in that the former means only that the lag sound is detectable (i.e., discriminable from a single-source sound), while the latter term depends on listeners' subjective reports that the lag sound is audible as a separate event. In the current experiment, masked detection of the lag sound was examined for the same stimuli and loudspeaker configuration used in experiment I. Intuitively we believe that the lag adds more than just loudness, and that subjects should hear a larger, fuller, more spacious image as the level of the lag approaches that of the lead. To the extent that these types of changes provide additional information beyond loudness differences, listeners should perform better than would be predicted by the equation describing the intensity discrimination data in Fig. 5. Experiment II employed a 3AFC procedure, as opposed to the 2AFC procedure used in experiment I. This may have encouraged listeners to attend to a larger set of cues because they could pay attention to the interval that was most different from the other two, instead of focusing only on loudness.

A. Method

Stimuli and procedures were similar to those used in experiment I except that the discrimination was between single-source bursts from the lead loudspeaker and lead-lag pairs of bursts. The paradigm was 3AFC, with a burst presented from the left loudspeaker during all three intervals and the lag added from the right loudspeaker in one of the three intervals. Subjects were instructed to identify the odd interval. Lead and lag were always the same sample of noise, but the samples varied across intervals and across trials. Correct-answer feedback was again provided after each trial. The level of the single-source or lead sound was 50 dBC. Lag attenuations of 0, 5, 10, 15, or 20 dB relative to the lead were fixed within a block of 30 trials. Within a block, the lag was present in each of the three intervals ten times, with the order shuffled. The five different blocks were presented in a random order three times for a total of 90 trials per data point. Listeners were four young normal-hearing individuals who had not participated in experiment I. Subjects practiced until they could perform nearly perfectly on the 0-dB attenuation condition, which usually took only a few blocks.

TABLE IV. Individual subject performance in d' for experiment II.

$\Delta I/I$ (dB)	−20	−15	−10	−5	0
$\Delta Im/Im$ (dB)	−13.26	−7.91	−3.19	0.84	4.81
S1	0.20	0.63	0.91	1.93	3.11
S2	0.20	0.91	1.70	2.90	3.62
S3	0.38	0.59	1.56	2.48	3.62
S4	−0.111	0.70	1.02	1.65	3.43
S5	0.24	0.42	0.52	1.25	2.90
Mean	0.18	0.65	1.14	2.04	3.34
Std. error	0.08	0.08	0.22	0.29	0.14

B. Results

Individual d's obtained in this experiment are shown in Table IV and mean results are displayed in Fig. 6. The abscissa of the figure is plotted in terms of 10 log $\Delta Im/Im$, as computed from the KEMAR measurements and loudness equations used in experiment I. Two KEMAR measurements not needed for experiment I were also required, those for the single-source stimulus and for the condition where the lag was 15 dB down from the lead. The top row of the Table IV shows the nominal presentation level of the lag relative to the lead, whereas the second row displays the transformation based on binaural loudness summation. For example, the addition of a lag sound from the right loudspeaker at a level 20 dB down from the lead is equivalent, according to the measurements and loudness equation, to adding a noise that would be 13 dB down in a monaural presentation. Adding a

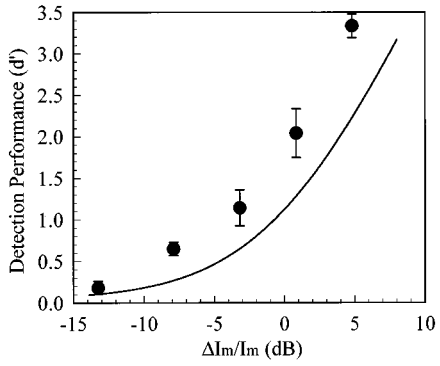


FIG. 6. Detection performance from experiment II, with the abscissa transformed to reflect the relative level of the lag based on KEMAR measurements and the binaural loudness summation equation from Treisman and Irwin (1967). The solid line is the equation used to fit the data in Fig. 5.

lag sound equal in level to the lead (0-dB attenuation) is equivalent to the monaural addition of a noise 4.7 dB higher than the lead. The larger values obtained for the loudness predictions relative to the nominal relative levels of the lag are mostly due to head shadow effects. The lead is attenuated at the right ear, increasing the contribution of the lag to the total.

Performance for all subjects was poor when the nominal level of the lag was 20 dB down, and increased gradually but smoothly as the level of the lag was increased. The lead-lag pair presented at an equal level was almost perfectly discriminable from the lead alone, with the mean d' above 3.0 for the 0-dB attenuation condition. This supports our general understanding of echo suppression: Even when the time delay of the lag sound is brief and well within the region of the precedence effect (see Zurek, 1987), the presence of the lag is nevertheless easily detectable when it is equal in level to the lead.

The solid line in Fig. 6 is the same fitting equation from Fig. 5 that represented the data from the intensity discrimination experiment very well. The line clearly underestimates actual detection performance. Because the line is based on predicted loudness differences only, the underestimation suggests that other differences in addition to loudness may contribute to the discriminability of sounds with echoes from single-source sounds. These cues might include the compactness or diffuseness of the image, and might have also been enhanced by the use of a 3AFC paradigm. As stated earlier, the 3AFC paradigm might have made it easier for subjects to attend to the full range of available cues. In addition, the presentation of the two single-source sounds during each 3AFC trial could have produced a partial breakdown of echo suppression. Freyman *et al.* (1991) showed that the presentation of a train of single source sounds prior to a lead-lag pair decreases echo threshold (i.e., the lag is audible as a separate auditory event at a shorter delay). In the same way, the fact that two-thirds of the presentations were single source sounds could have helped the lag stand out when it was added during one of the intervals.

III. DISCUSSION

The term ''echo suppression,'' frequently used to describe the fact that the lag sound is not heard as a separate event, implies some inhibition or attenuation of the lagging sound. The data from the current experiments do not support this way of thinking about echo suppression. In the intensity discrimination experiment, when predicted loudness differences were considered, listeners were shown to be just as sensitive to changes in the level of the lag as they were to changes in the lead (see Fig. 5). In the lag detection task, listeners were shown to be exquisitely sensitive to the presence of the lag, even more so than could have been predicted on the basis of loudness differences alone.

The most straightforward interpretation for the intensity discrimination data is that listeners relied primarily on loudness differences. In addition to the agreement among conditions when the current data were plotted according to the loudness analysis (Fig. 5), this interpretation is also supported by the consistency of the current results with intensity discrimination data obtained for brief noises when subjects listened monaurally (Viemeister, 1974) and loudness was presumably the cue. One possible explanation for the better performance obtained in the detection experiment (Fig. 6) is that attention to spatial or other sound quality cues was facilitated by the 3AFC procedure, but this remains speculative. The use of spatial cues is likely to be highly dependent on an individual's echo threshold, the delay at which subjects begin to hear two rather than one sound. This is highly variable among listeners and it is not known how different the two subject groups might have been on this type of measure. An additional consideration is that in the current analyses lead and lag were added without regard to the delay. Predicted d's in experiment II would likely have been in better agreement with the data had we employed a temporal window model (e.g., Oxenham and Moore, 1994) in which the lead burst is attenuated by a window centered on the lag burst. However, implementation of this model has other complications when used with our stimuli (see Footnote 3).

The apparent absence of a general ''suppression'' of the lag sound is consistent with a view of the precedence effect in which inhibition is specific to the directional information of the delayed sound (Zurek, 1987). It is also consistent with the idea that reflections, as long as the delay is not excessively long, can actually enhance performance on tasks such as speech perception (Haas, 1972) probably because they increase the acoustic signal energy at the ears. This would not be predicted if delayed sounds were generally suppressed. Finally, it is consistent with the notion that although people are usually not consciously aware of the presence of reflections in a room, they are nevertheless quite sensitive to changes in echoes that carry information relevant to room acoustics (Clifton *et al.*, 1994; Clifton and Freyman, 1997; McCall *et al.*, in press).

Why would it be useful for directionality information to be suppressed but not intensity? To answer this question it is necessary to consider what each type of information tells the listener about the auditory environment. Perception of the lagging sound's directionality as distinct from the leading sound's direction would indicate that there were two sound sources, which in the natural world would be an incorrect conclusion. Echoes should not be confused with true sound sources if the auditory world is to remain a sensible place with reliable information. At delays below echo threshold where lead and lag are fused into a single image, any substantial contribution the lagging sound would make to the perceived direction of the image would lead to a mislocalization of the source. Thus if echoes contributed substantially to perceived location, sound localization in most environments would be inaccurate or even impossible.

While this consideration explains the advantage of suppressing directionality information, there are at least three reasons why it is useful for the auditory system to preserve information about the intensity of the lagging sound. First, as discussed above, reflections can aid speech communication by increasing the signal level reaching the ears (Haas, 1972). As an example, because of reflected acoustic energy, a talker in a normal room need not be facing the listener in order to be understood. In contrast, communication can be difficult

when the talker is not facing the listener in an anechoic chamber. Second, it is well known that reflected sounds enhance the overall quality of complex sounds such as music heard in large rooms (Blauert, 1983, p. 348f). Third, intensity properties of echoes signify aspects of the space one is in, such as the hardness and distance of reflecting surfaces. This type of information may be an important supplement to that provided by vision. Visual information is not always available because of our limited field of view at any moment in time, visual problems, and of course darkness. Under such conditions, information contained in acoustic reflections might help us in a variety of ways such as sensing the presence of an object in a room or avoiding stepping off a precipice. The brain mechanism involved in echo suppression is ''smart'' in its specificity, preventing us from confusing a reflection as a true source, while preserving information contained in the reflection that provides crucial evidence about the auditory environment.

ACKNOWLEDGMENTS

This research was supported by NIDCD Grant No. DC-01625 and a Research Scientist Award from NIMH (MH-00332) to RKC. The authors would like to thank Mary Florentine for her helpful comments concerning the proportionality of d' and ΔL and Neal Viemeister for providing data on the effects of duration on intensity discrimination. We would also like to thank Christopher Plack and an anonymous reviewer for their comments on an earlier version of the manuscript.

[1]Continuous noise, presented simultaneously through two loudspeakers, was used in making the measurements through KEMAR because these measurements would presumably be more stable than with the multiple tokens of brief noise samples. However, as a precaution, because of spectral interactions that might be introduced by the delay in the actual brief stimuli, we repeated the measurements with a sample of noise used in experiment I with the appropriate delay between loudspeakers (4 ms). Highly similar results were obtained. For example, the data in Table III obtained with continuous noise indicate that attenuating the lag by 10 dB produced 0.78-dB and 7.28-dB reductions in the left and right ears respectively (columns 7–5, and 8–6 for the $\Delta L = 10$-dB condition). The repeated measurements with the brief noise sample showed 0.69-dB and 7.10-dB reductions for the same condition. As expected, the effect of using the brief noise sample on the predictions of the analyses was also minimal. For example, the values computed for the final analysis based on binaural loudness summation (column 14 in Table III) were changed only slightly (typically by 0.1 dB with a maximum of 0.25 dB) when the brief noise sample measurements were substituted for the continuous noise measurements.

[2]It is widely understood that the plotting of ΔL in linear units is inappropriate because of compression and differences in the variability of the data at low and high ΔL's (see Jesteadt *et al.*, 1977; Florentine, 1986). In addition to the method of plotting psychometric functions used in the present paper, an alternative appropriate method is $\log d'$ as a function of $\log \Delta L$ (Buus and Florentine, 1991). This gives linear functions with a slope of nearly 1.0 for monaural intensity discrimination for individual subjects. We chose not to use these coordinates because differences among conditions producing low d's are exaggerated by the log scale, and such conditions are part of our data set. Also, as noted in the text, comparisons with experiment II are more straightforward in the coordinates we plotted.

[3]Because the delay was so brief in the Lead and Lag conditions, the intensities were summed without regard to delay for all three conditions. Alternatively, we could have assumed a temporal window (see, for example, Oxenham and Moore, 1994), and summed the window outputs. We chose not to do so because of the added complexity of the bandpass filter stage of the Oxenham and Moore model, which precedes the temporal window. The bandpass filters clearly would produce different degrees of temporal smearing in different frequency regions, and would have necessitated the use of a multichannel version of the model for our broadband stimuli. We thought it would be premature to use a multichannel model with these precedence effect stimuli without first having tested the model with narrow-band stimuli and a single channel.

[4]As noted in the text, the model of Oxenham and Moore (1994) also assumes a compression and places it at a peripheral stage due to physiological data from the basilar membrane. We did not use this model for the reasons cited in Footnote 3.

Bech, S. (**1995**). ''Timbral aspects of reproduced sound in small rooms. I,'' J. Acoust. Soc. Am. **97**, 1717–1726.

Blauert, J. (**1983**). *Spatial Hearing* (MIT, Cambridge, MA).

Burgtorf, W. (**1961**). ''Untersuchungen zur Wahrnehmbarkeit verzogeter Schallsignale [Investigations of the perceptibility of delayed sound signals],'' Acustica **11**, 97–111.

Buus, S., and Florentine, M. (**1991**). ''Psychometric functions for level discrimination,'' J. Acoust. Soc. Am. **90**, 1371–1380.

Buus, S., Florentine, M., and Zwicker, T. (**1995**). ''Psychometric functions for level discrimination in cochlearly impaired and normal listeners with equivalent-threshold masking,'' J. Acoust. Soc. Am. **98**, 853–861.

Clifton, R. K., and Freyman, R. L. (**1997**). ''The precedence effect: beyond echo suppression,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. B. Anderson (Erlbaum, Hillsdale, NJ).

Clifton, R. K., Freyman, R. L., Litovsky, R. Y., and McCall, D. (**1994**). ''Listener expectations about echoes can raise or lower echo threshold,'' J. Acoust. Soc. Am. **95**, 1525–1533.

Florentine, M. (**1986**). ''Level discrimination of tones as a function of duration,'' J. Acoust. Soc. Am. **79**, 792–798.

Freyman, R. L., Clifton, R. K., and Litovsky, R. Y. (**1991**). ''Dynamic processes in the precedence effect,'' J. Acoust. Soc. Am. **90**, 874–884.

Green, D. M. (**1976**). *Introduction to Hearing* (Erlbaum, Hillsdale, NJ).

Guzman, S., and Yost, W. A. (**1996**). ''Perceptual fusion of clicks: The precedence effect,'' Paper presented at the annual meeting of the Association for Research in Otolaryngology, February, 1996.

Haas, H. (**1972**). ''The influence of a single echo on the audibility of speech,'' J. Audio Eng. Soc. **20**, 146–159.

Hellman, W. S., and Hellman, R. P. (**1990**). ''Intensity discrimination as the driving force for loudness. Application to pure tones in quiet,'' J. Acoust. Soc. Am. **87**, 1255–1265.

Irwin, R. J. (**1965**). ''Binaural summation of thermal noises of equal and unequal power in each ear,'' Am. J. Psychol. **78**, 57–65.

Jesteadt, W., Wier, C. C., and Green, D. M. (**1977**). ''Intensity discrimination as a function of frequency and sensation level,'' J. Acoust. Soc. Am. **61**, 169–177.

Leakey, D. M., and Cherry, E. C. (**1957**). ''Influence of noise upon the equivalence of intensity differences and small time delays in two-loudspeaker systems,'' J. Acoust. Soc. Am. **29**, 284–286.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Lim, J. S., Rabinowitz, W. M., Braida, L. D., and Durlach, N. I. (**1977**). ''Intensity perception VIII. Loudness comparisons between different types of stimuli,'' J. Acoust. Soc. Am. **62**, 1256–1267.

Lindemann, W. (**1986**). ''Extension of a binaural cross-correlation model by contralateral inhibition: II. The law of the first wavefront,'' J. Acoust. Soc. Am. **80**, 1623–1630.

Litovsky, R. Y., and Macmillan, N. A. (**1994**). ''Sound localization precision under conditions of the precedence effect: Effects of azimuth and standard stimuli,'' J. Acoust. Soc. Am. **96**, 752–758.

Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A Users Guide* (Cambridge U. P., Cambridge).

McCall, D. D., Freyman, R. L., and Clifton, R. K. (**in press**). ''Sudden changes in spectrum of an echo cause a breakdown of the precedence effect,'' Percept. Psychophys.

Miller, G. A. (**1947**). ''Sensitivity to changes in the intensity of white noise and its relation to masking and loudness,'' J. Acoust. Soc. Am. **19**, 609–619.

Olive, S. E., and Toole, F. E. (**1989**). ''The detection of reflections in typical rooms,'' J. Audio Eng. Soc. **37**, 539–553.

Oxenham, A. J., and Moore, B. C. J. (**1994**). ''Modeling the additivity of nonsimultaneous masking,'' Hearing Res. **80**, 105–118.

Rabinowitz, W. M., Lim, J. S., Braida, L. D., and Durlach, N. I. (**1976**). ''Intensity perception VI. Summary of recent data on deviations from Weber's law for 1000-Hz tone pulses,'' J. Acoust. Soc. Am. **59**, 1506–1509.
Reynolds, G. S., and Stevens, S. S. (**1960**). ''Binaural summation of loudness,'' J. Acoust. Soc. Am. **32**, 1337–1344.
Riesz, R. R. (**1933**). ''The relationship between loudness and the minimum perceptible increment of intensity,'' J. Acoust. Soc. Am. **5**, 211–216.
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.
Seraphim, H. P. (**1961**). ''Uber die Wahrnehmbarkeit mehrerer Ruckwurfe von Sprachschall [On the perceptibility of multiple reflections of speech sounds],'' Acustica **11**, 80–91.
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.
Shaw, E. A. G., and Vaillancourt, M. M. (**1985**). ''Transformation of sound pressure level from the free field to the eardrum in numerical form,'' J. Acoust. Soc. Am. **78**, 1120–1123.
Shinn-Cunningham, B. G. (**1988**). ''Understanding the precedence effect: an examination of the factors affecting its strength,'' unpublished masters thesis, M.I.T.
Shinn-Cunningham, B. G., Zurek, P. M., and Durlach, N. I. (**1993**). ''Adjustment and discrimination measurements of the precedence effect,'' J. Acoust. Soc. Am. **93**, 2923–2932.
Sommerville, T., Gilford, C. L. S., Spring, N. F., and Negus, R. D. M. (**1966**). ''Recent work on the effects of reflectors in concert halls and music studios,'' J. Sound Vib. **3**, 127–134.
Treisman, M., and Irwin, R. J. (**1967**). ''Auditory intensity discriminal scale I. Evidence derived from binaural intensity summation,'' J. Acoust. Soc. Am. **42**, 586–592.
Viemeister, N. F. (**1974**). ''Detection of a noise signal: The effects of duration and masker level,'' J. Acoust. Soc. Am. **76**, S4.
Wallach, H., Newman, E. B., and Rosenzweig, M. R. (**1949**). ''The precedence effect in sound localization,'' Am. J. Psychol. **52**, 314–336.
Yost, W. A., and Soderquist, D. R. (**1984**). ''The precedence effect: Revisited,'' J. Acoust. Soc. Am. **76**, 1377–1383.
Zurek, P. M. (**1979**). ''Measurement of binaural echo suppression,'' J. Acoust. Soc. Am. **66**, 1750–1757.
Zurek, P. M. (**1987**). ''The precedence effect,'' in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer-Verlag, New York).

Ripple density resolution for various rippled-noise patterns

Alexander Ya. Supin, Vladimir V. Popov, Olga N. Milekhina, and Mikhail B. Tarakanov
Institute of Ecology and Evolution, Russian Academy of Sciences. 33 Leninsky Prosp., 117071 Moscow, Russia

(Received 18 February 1997; revised 5 December 1997; accepted 18 December 1997)

Ripple-density resolution was measured in normal humans using rippled noise with a phase-reversal test. The principle of the test was to find the highest ripple density at which an interchange of spectral peak and trough positions (the phase reversal) is detectable. Different rippled noise patterns were used: (i) either frequency-proportional or constant ripple spacing; (ii) various bandwidth; and (iii) either steep or shallow slopes of the spectrum envelope. When tested with frequency-proportional rippled noise, ripple-density resolution as expressed in relative units (the center frequency to ripple spacing ratio) little depended on frequency within a range of 1 to 8 kHz: from 11.4 at 1 kHz to 14.9 at 8 kHz, mean 13.1. These values were virtually independent on noise bandwidth. When tested with constant ripple spacing, the resolution was of similar values taking the relative ripple density at the lower part of the passband. Being measured by noise with steep spectral edges, the resolution was five units higher than it was for shallow-enveloped spectra, thus suggesting some edge effects at the spectrum boundaries. The resolution values obtained were about twice higher than those predicted by peripheral auditory filter tuning. [S0001-4966(98)00604-3]

PACS numbers: 43.66.Fe, 43.66.Jh [WJ]

INTRODUCTION

Resolution of frequency spectra is an important characteristic of the auditory system. It determines the ability to discriminate complex acoustic stimuli including speech. Most of the popular methods of frequency-resolution measurements are based on masking technique with various masker types: tone (frequency-tuning curves, review in Zwicker, 1974), two-tone (Patterson and Henning, 1977; O'Loughlin and Moore, 1981; Glasberg *et al.*, 1984), narrow-band noise (critical bands, review in Zwicker, 1982), notch-noise (Patterson, 1976; Moore and Glasberg, 1981; Patterson *et al.*, 1982; Glasberg *et al.*, 1984; Glasberg and Moore, 1990), or rippled noise (Wilson and Evans, 1971; Houtgast, 1974, 1977; Pick *et al.*, 1977; Pick, 1980; Glasberg *et al.*, 1984). A common feature of these methods is that they use multi-point measures; i.e., the determination of a single resolution value requires several threshold measurements to compute a resolution value from the obtained function. Multi-point measures are time consuming. Perhaps this is the reason why these methods are rarely used in clinical situations where time is short.

A one-point measure is the critical ratio. However, the critical ratio is a poor estimate of frequency resolution because it is more sensitive to the efficiency of signal detection in noise than to frequency tuning. The critical ratio measure becomes more sensitive to frequency resolution with the use of notch-noise as a masker (the ''notch-noise critical ratio'' by Patterson *et al.*, 1982). However, this measure uses a voluntarily chosen notchwidth, and results depend on the notch-width value. Estimation of critical bands by comparison of AM and FM modulation thresholds (Zwicker, 1952) also requires a substantial amount of measurements. Critical band measurement by frequency spacing between tones at which ''roughness'' of the sound disappears (Plomp and Steeneken, 1968; Terhardt, 1968) requires a listener to be well instructed and trained.

Since each of the methods had both advantages and restrictions, elaboration of new methods for spectrum-resolution measurements seemed reasonable. One method purposed in our preceding papers (Popov and Supin, 1984; Supin and Popov, 1987; Supin *et al.*, 1994, 1997) employs rippled noise as probe sound; this signal has a spectrum containing periodically alternating peaks and troughs. Such ''grid'' spectrum is a reliable probe to measure spectral resolution (Wilson and Evans, 1971).

Rippled noise has been widely used to study frequency-selective properties of single units (Wilson and Evans, 1971; Evans and Wilson, 1973; Bilsen *et al.*, 1975; Evans, 1977; Shamma *et al.*, 1995a, b; Shamma and Versnel, 1995), in psychophysical masking studies (Houtgast, 1974, 1977; Pick *et al.*, 1977; Pick, 1980) and in studies of pitch perception (Yost *et al.*, 1977, 1996; Yost and Hill, 1979; Bilsen and Wieman, 1980; Yost, 1982, 1996a, b; Patterson *et al.*, 1996). In these studies, involvement of both temporal processing of rippled noise (pitch discrimination studies by Yost *et al.*, 1996; Patterson *et al.*, 1996) and its frequency-tuning processing (frequency-dependent masking studies by Houtgast, 1974, 1977; Pick *et al.*, 1977; Pick, 1980; ripple-density resolution studies by Wilson and Evans, 1971; Supin *et al.*, 1994) was suggested. But irrespective of underlying neurophysiological mechanisms, rippled noise is a reliable model of complex spectral patterns.

A distinctive feature of our method is the use of a ripple phase-reversal test to find the highest resolvable ripple density. The test principle is as follows. Noise of a certain ripple density is presented to a listener. At a certain instant, it is replaced by a noise of the same intensity, bandwidth, and ripple density but of the opposite position of spectral peaks

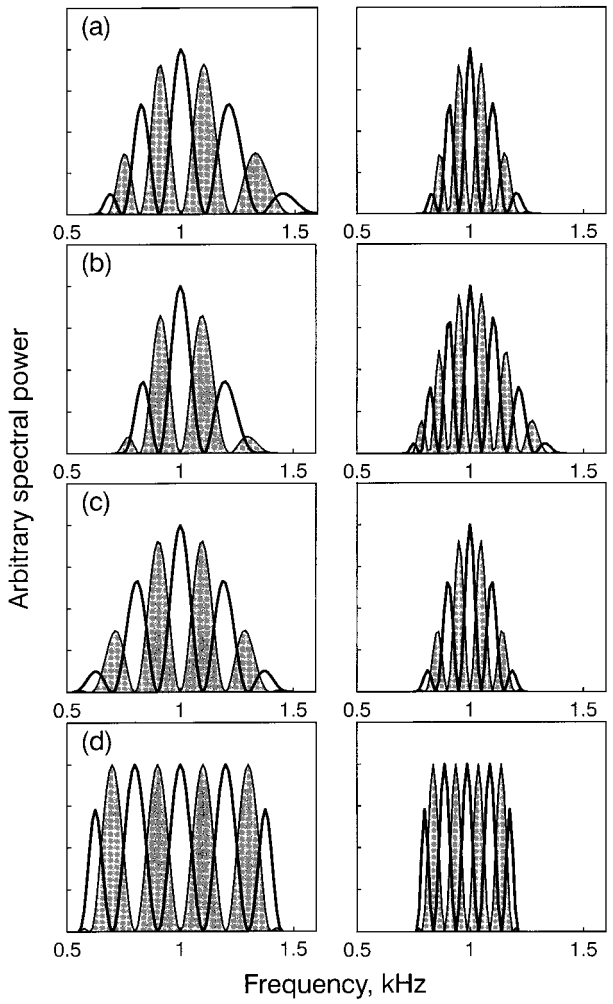


FIG. 1. Filter shapes used for rippled noise generation. Filters with a center frequency of 1 kHz are exemplified. Open and filled areas show two filter forms of opposite peak and trough positions. Left and right columns present filters of different ripple densities to show the bandwidth variation together with the ripple density. (a) Filters with frequency-proportional ripple spacing, shallow boundaries, and bandwidth varying with ripple density. (b) The same with constant octave bandwidth. (c) Filters with constant ripple spacing, shallow boundaries, and bandwidth varying with ripple density. (d) The same with steep boundaries.

and troughs at the frequency scale; i.e., ripple phase reversal occurs. These pairs of spectra are shown as open and filled areas in Fig. 1. The listener can only detect this switch if he discriminates the rippled spectrum structure; suppose it occurs at a ripple density shown in left panels of Fig. 1. If the spectrum structure is not resolvable since the ripples are spaced too densely (suppose it occurs at a density shown in right panels), the switch cannot be detected because the signals before and after the switch are identical in all respects except the peak and trough positions. Thus the highest resolvable ripple density can be found. This principle is very similar to that widely used to measure visual acuity: the phase reversal of a test grid for finding the highest resolvable grid density.

A peculiarity of rippled-noise probes is that they test a resolution of complex spectrum profiles which are characteristic for many natural sounds. In this respect, our approach is close to the spectrum profile analysis proposed by Green (1986, 1987). Other features of the method are: (i) it gives a one-point measure of spectrum resolution; (ii) it does not confound the frequency tuning with the signal detection efficiency; (iii) it requires only a very simple instruction to the listener to detect any change in the noise. However, a number of details had to be defined in order to select proper parameters of rippled-noise probes. In the present study, we addressed ourselves to the problems as follow.

(i) A widely used method of rippled noise generation is to mix a noise with a delayed version of itself. Many rippled-noise studies mentioned above used this principle. This method produces noise with a constant frequency spacing of ripples. However, in a large part of the frequency range, auditory filters are of roughly constant tuning, that is, their bandwidths are proportional to the frequency. For this range, frequency-proportional ripple spacing is more reliable. Frequency-proportional rippled spectra were used in both psychophysical (Green, 1986, 1987) and physiological studies (Shamma *et al.*, 1996a, b) being composed of a number of pure tones. However, those studies were not designed to find a resolvable ripple-density limit, therefore they used rather low ripple densities. In the present study, we compared resolution estimates obtained with noises of both constant and frequency-proportional ripple spacing.

(ii) Dependence of ripple-density resolution on the probe bandwidth cannot be excluded. Therefore, in the present paper, we studied the resolution dependence on noise bandwidth.

(iii) In the preceding studies, we used rippled noise with sharp passband cutoffs. However, sharp spectrum edges may be associated with some specific edge effects. For example, broadband stimuli evoke a pitch sensation associated with the spectrum boundaries (Small and Daniloff, 1967; Fastl, 1980; Kohlrausch and Houtsma, 1992). Some effects at sharp spectrum boundaries may be expected for rippled-noise probes as well. Therefore, in this study, we compared ripple-density resolution for steep and shallow slopes of rippled-noise passbands.

I. MATERIALS AND METHODS

A. Subjects

Four subjects (laboratory staff) were used. They were 25 to 37 years old and had no signs of hearing disease. All of them had practice in psychoacoustic experiments. The experiments were performed in accordance with the guidelines of the Declaration of Helsinki.

B. Stimuli

Rippled-noise stimuli used in this study had either frequency-proportional ripple spacing [Fig. 1(a,b)] or constant ripple spacing [Fig. 1(c,d)]. In the following text, they are referred to as frequency-proportional and constant rippled noise respectively. The frequency-proportional noise had the relative ripple density D constant across the passband, implying

$$D = F_0 / \delta F, \tag{1}$$

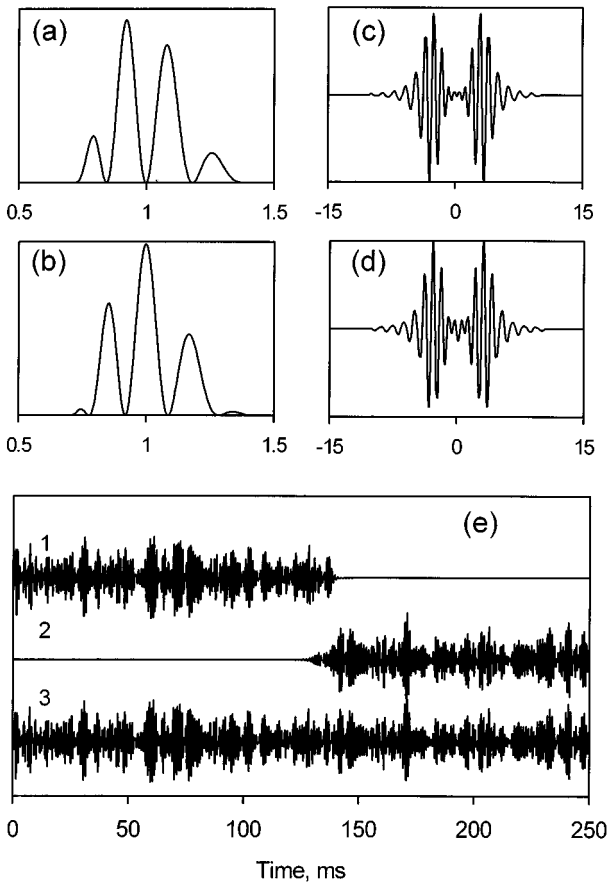


FIG. 2. Generation of rippled noise with phase reversals. (a,b) Examples of two spectra differing by opposite peak trough positions; abscissa—frequency in kHz; ordinate—power in arbitrary units. (c,d) Filter transfer functions obtained by Fourier-transform of these spectra; abscissa—time in ms; ordinate—amplitude in arbitrary units. (e) Samples of the two filter outputs (1,2) and of their sum (3) when wide-band signal was switched from one filter input to the other.

where F_0 kHz is the center frequency, and δF kHz is the ripple spacing. The constant rippled noise had absolute ripple density D_a constant across the passband, implying

$$D_a = 1/\delta F, \tag{2}$$

but its relative ripple density varied across the passband.

Rippled noise was digitally generated at sampling rates of 10, 14, 20, 28, or 40 kHz, keeping the sampling rate at least four times higher than the upper limit of the signal passband but not lower than 10 and not higher than 40 kHz. Generation program simulated steps as follow. A wide-band signal (random digital sequence with the sampling rate, uniformly distributed between ± 1) was filtered by one of two digital filters. Typical frequency responses of the used filters are shown in Fig. 1 and Fig. 2(a,b). These filters determined both noise passband and ripple structure. The two filters had equal passband, level, and ripple density but opposite ripple peak trough positions, as shown. The specified filter forms [Fig. 2(a,b)] were Fourier transformed to obtain filter transfer functions [Fig. 2(c,d)]. The wide-band signal (random digital sequence) was convoluted with one of these transfer functions; the result was a filtered noise [Fig. 2(e)]. Phase reversals were performed by switching the wide-band signal from one filter input to another; filter outputs were summarized. If a noise burst without phase reversals was generated, both of the filters were identical.

Note that phase-reversal switches were made at the filter inputs, not outputs. Otherwise, clicks at phase reversal could arise. With phase reversals at the filter inputs, both transient and sustained parts of the signal were filtered equally, and after the filtering, neither signal discontinuity nor bandwidth difference appeared between the transient and sustained parts of the signal. It is essential also that both of the filters had equal rise–fall times, which is a direct consequence of their equal bandwidths [Fig. 2(c,d)]. Therefore, output signal decay of the first filter and signal rise of the second one precisely compensated one another and the switch did not result in any level change exceeding fluctuations intrinsic in narrow-band noise [Fig. 2(e)]. With this generation technique, experienced listeners could not detect something sounding like clicks or any other short transients at phase reversals; they reported of smooth transition from one sound coloration to another: the higher the ripple density, the more smooth the transition.

Each noise bursts (either with or without phase reversals) lasted 4 s. Phase reversals occurred every 500 ms; i.e., a burst with phase reversals contained four 1/s cycles of switching from one filter to another and back. These parameters were used since it was shown (Supin *et al.*, 1997) that phase-reversal rates higher than 1.5–2/s resulted in decreased ripple-density resolution because of insufficient time to extract phase reversals from fluctuations inherent in noise. In line with the previous data, at a ripple resolution limit, the listeners reported that they need at least a few seconds to listen to the signal and to decide, whether the stimulus did or did not contain some alternations.

Pregenerated noise bursts were played through a 14-bit D/A converter, amplified, attenuated and presented diotically over TDS-7 earphones with the frequency characteristic irregularity of no more than 10 dB within the range of 0.1–6 kHz and 18 dB within the range of 0.1–16 kHz. Sound level was 65–85 dB SPL. Within this range, the listener could choose a level which he considered to be the most comfortable and providing the easiest stimulus discrimination. Most frequently it was a level of 70–80 dB.

C. Procedure

A three-interval two-alternative forced-choice procedure was used. Each trial consisted of three noise bursts, each of 4-s duration with 0.5-s intervals between them. Two trial types alternated randomly: either the first and third bursts within the trial contained phase reversals and the second did not contain them, or the second burst contained the phase reversals and the first and third did not. The listener knew that the first and third bursts were identical and distinct from the second one. The listener's task was to detect any modifications of the noise and to report whether the modifications appeared in the first and third bursts or in the second one. The listener was informed whether the response was true or false. Thus the procedure was two-alternative and differed from the standard one only in the additional third signal which always duplicated the first one. The third signal was

added since all the listeners reported that they were more confident to make a decision in the three-interval procedure rather than in the standard two-interval one: they could make a decision by comparing the first and second bursts and then verify it by comparing the second and third ones.

To find the ripple-density resolution, the ripple density varied by steps as follow: 4, 5, 6, 7, 8, 10, 12, 14, 16, 20, 24, 28, 32 relative units. Except a few cases, each measurement run consisted of a total of 50 trials: 5 randomly alternating ripple densities around the anticipated resolution limit, with each ripple density being presented 10 times. In a few runs, more than 5 ripple densities were tested to trace performance in a wider range. The percentage of true responses was counted for every ripple-density value, and the ripple density corresponding to 75% of true responses was calculated by interpolation between the nearest values. The result was adopted as a ripple-density resolution.

II. RESULTS

A. Narrow-band frequency-proportional rippled noise

Figure 1(a) shows the filters used in this series. Ripple shape (in the power domain) was a cosine function of frequency logarithm. Thus the ripple spacing was frequency proportional and the relative ripple density was constant throughout the passband. The spectrum envelope was also a cosine function of frequency logarithm; it was five times wider than the ripples. Thus the spectrum always contained five ripples. At this spectrum envelope, four ripple peaks around the passband center had levels no less than -4.3 dB *re:* maximum. When ripple density varied, the bandwidth varied correspondingly in such a way that the number of ripples was kept constant: compare left and right panels of Fig. 1(a). The envelope peak was arbitrarily adopted as a center frequency of these spectra.

The bandwidth variation together with the ripple density was used for the following reasons. If the ripple-density resolution is not equal across the noise bandwidth (this possibility can not be excluded in advance), some uncertainty may arise in attributing the resolution obtained to a certain frequency. To minimize this uncertainty, the bandwidth must be as narrow as possible. However, the bandwidth must not be too narrow because ripple phase reversals may change the noise level when the bandwidth and ripple spacing are comparable. It was shown (Supin *et al.*, 1994) that the change in the noise level became insignificant (as small as a fraction of dB) when the spectrum contained 4–5 ripples or more. This number of ripples was chosen as a compromise between requirements to use narrow bandwidth and to have at least several ripples within the passband.

A few typical psychometric curves obtained in conditions of phase-reversal test are shown in Fig. 3. At low ripple densities, the listeners easily detected phase reversals with 100% probability. When the ripple density exceeded a certain limit, the correct response probability fell steeply an then fluctuated around the 50% probability. From these curves, a ripple density corresponding to 75% correct responses could be found with a satisfactory precision. At all stimulus types described below, the psychometric curves were of the same shape.

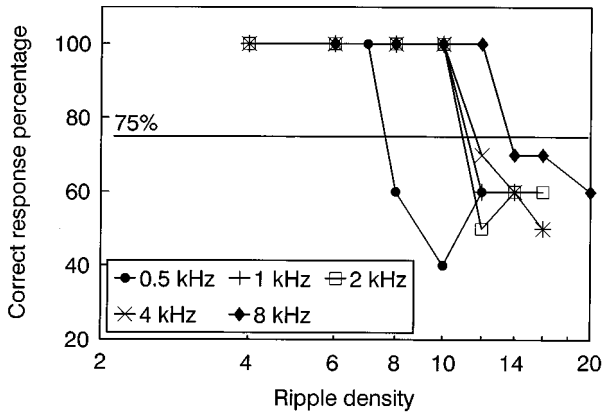


FIG. 3. Typical psychometric curves in conditions of ripple-density resolution measurements. Data of one listener obtained using filters with proportional ripple spacing and shallow boundaries.

The results of measurements with this probe type at various center frequencies are shown in Fig. 4 which presents ripple-density resolution as a function of center frequency. At low frequencies (0.5–1 kHz), relative resolution increased with frequency: from 7.6 ± 0.2 s.d. at 0.5 kHz to 12.0 ± 1.4 s.d. at 1 kHz. Within a higher range of 1–8 kHz, resolution little depended on frequency and only a small trend was observed. Interindividual means changed from 12.0 ± 1.4 s.d. at 1 kHz to 16.4 ± 2.8 s.d. at 8 kHz. Being approximated by a logarithmic regression line, this part of the plot had a mean slope of 1.5 units per octave: from 11.4 at 1 kHz to 15.8 at 8 kHz. Since the trend was rather small, it seemed reasonable to characterize the whole range of 1–8 kHz by an overall mean resolution value by averaging all the data at 1, 2, 4, and 8 kHz. This value was estimated as 13.5 ± 2.5 s.d.

B. Octave-band frequency-proportional rippled noise

Figure 4 shows that resolution expressed in relative ripple density was little dependent on frequency within a rather wide frequency range. This eliminated the need to keep the noise bandwidth very narrow when ripple spacing

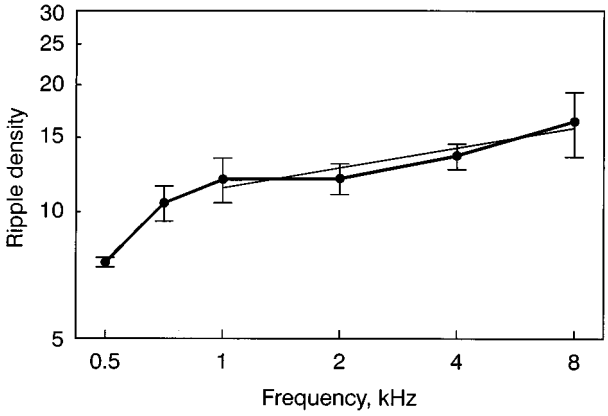


FIG. 4. Ripple-density resolution dependence on center frequency obtained using filters with proportional ripple spacing and shallow boundaries. Solid line—interindividual means ±s.d.; thin straight line—regression line.

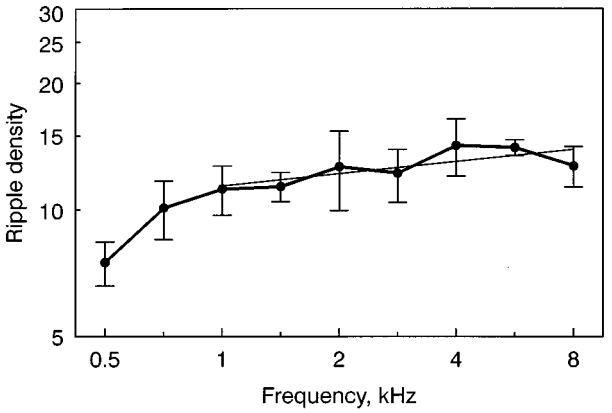


FIG. 5. Ripple-density resolution dependence on center frequency obtained using constant-bandwidth (octave) filters with proportional ripple spacing and shallow boundaries. Designation as in Fig. 3.

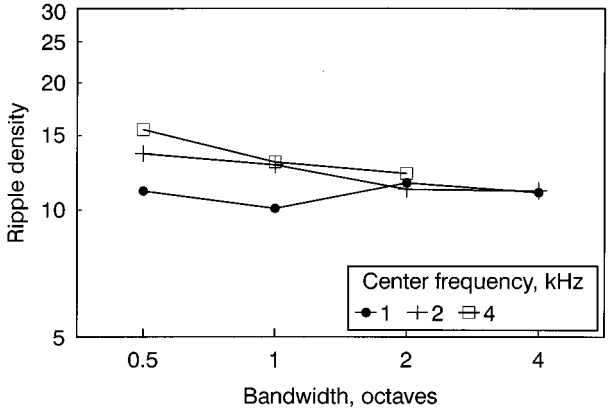


FIG. 6. Ripple-density resolution dependence on noise bandwidth at different center frequencies (1–4 kHz).

was frequency proportional: when both the probe ripple density and resolution are constant across a passband, it becomes less important to attribute a result to a certain frequency within the passband. Therefore in the next series we used frequency-proportional rippled noises with constant (octave) bandwidth.

The filters used in this series are shown in Fig. 1(b). Similarly to the previous series, both the ripple and envelope shape were cosine functions of frequency logarithm. However, the envelope did not vary with ripple density and was always of the octave bandwidth: Note that contrary to Fig. 1(a), the spectra at the left and right panels of Fig. 1(b) are of equal bandwidths.

The results of measurements are shown in Fig. 5. Below 1 kHz, ripple-density resolution markedly rose with frequency. Within a range of 1–8 kHz, resolution was less dependent on frequency. Interindividual means varied from 11.2 ± 1.5 s.d. at 1 kHz to 14.2 ± 1.8 at 4 kHz. Within the range of 1–8 kHz, the approximating regression line had a mean slope of 0.8 units per octave: from 11.4 at 1 kHz to 13.9 at 8 kHz, and the overall mean level was 12.7 ± 2.0 s.d. The difference between the means of the two last series (0.8 ± 0.8 SE) was insignificant.

Because of small difference between the two series, they could be averaged to characterize ripple-density resolution at frequency-proportional probes. Within a range of 1–8 kHz, the average resulted in a regression line with a slope of 1.2 units per octave: from 11.4 at 1 kHz to 14.9 at 8 kHz, and the overall mean level of 13.1 ± 2.3 s.d.

C. Frequency-proportional rippled noise of various bandwidth

In the previous series, the octave bandwidth was chosen. However, a possibility of resolution dependence on noise bandwidth could not be excluded in advance. Therefore, the measurements were repeated at various bandwidths. Filters used in this series were of the same shape as shown in Fig. 1(b), except that the bandwidth varied from 0.5 to 4 octaves.

The results of the measurements are shown in Fig. 6 where ripple-density resolution is presented as a function of bandwidth at a few center frequencies, from 1 to 4 kHz. At a center frequency of 4 kHz, only bandwidths up to 2 octaves were tested. Figure 6 shows that resolution little depended on noise bandwidth, although a small tendency of resolution decrease with the passband increase was noticeable at frequencies of 2 and 4 kHz.

D. Narrow-band constant rippled noise with shallow spectrum edges

This series differed from the preceding ones in the way that the ripple spacing was constant; i.e., both the ripple and envelope shape were cosine functions of frequency. At constant ripple spacing, the relative ripple density varies across the passband. This raises the problem of data attribution to a certain frequency. To minimize the uncertainty, the bandwidth was varied with ripple density in order to keep the bandwidth as narrow as possible but to provide a few ripples within the passband. Figure 1(c) shows the filters used in this series: independently of the ripple density, the filters contained five ripples.

The passband center frequency was arbitrarily taken as F_0 to calculate the relative ripple density according to Eq. (1). However, with the passband containing five ripples, the real relative ripple density varied over a range of ± 2.5: at the lower passband edge, the frequency was $F_0(1 - 2.5\delta F)$, and its ratio to the ripple spacing δF was $F_0/\delta F - 2.5$; respectively, at the upper edge, it was $F_0/\delta F + 2.5$.

The results obtained with this stimulus type are shown in Fig. 7; ripple density is indicated in its relative measure at the passband center. Resolution dependence on center frequency was qualitatively similar to that described above: at low frequencies, resolution increased with the center frequency; at higher frequencies, resolution was less dependent on frequency. Within a range of 1–8 kHz, interindividual means varied from 12.9 ± 2.4 s.d. at 1 kHz to 18.1 ± 1.8 s.d. at 8 kHz. This range was approximated by a regression line with a mean slope of 1.0 units per octave: from 13.2 at 1 kHz to 16.3 at 8 kHz. The overall mean level was 14.9 ± 2.4 s.d. This level is somewhat higher than those obtained with frequency-proportional noises (13.1 ± 2.3 s.d.). The difference was statistically significant: 1.8 ± 0.5 SE. However, taking into account the variation of relative ripple density across

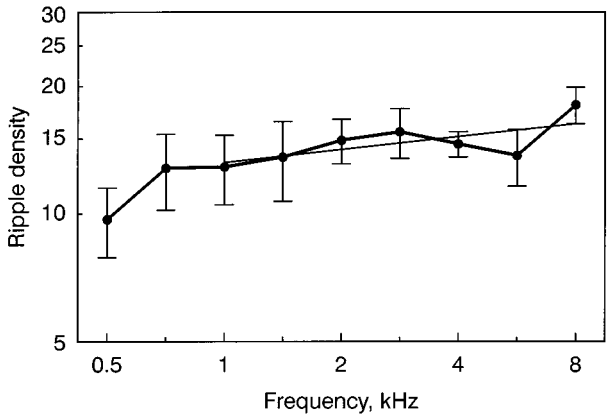


FIG. 7. Ripple-density resolution dependence on center frequency obtained using filters with constant ripple spacing and shallow boundaries. Designation as in Fig. 3.

the passband, the real ripple-density resolution may be within a range of ± 2.5 around the arbitrary level indicated above; i.e., from 12.4 ± 2.4 to 17.4 ± 2.4 s.d.

E. Narrow-band constant rippled noise with steep passband edges

Figure 1(d) shows the filter used in this series. Similarly to the preceding series, the filters had constant ripple spacing, and the bandwidth varied together with ripple density in such a way as to keep the number of ripples constant. However, the spectrum envelope had a flat top and steep edges; its half-level bandwidth was equal to four ripple cycles. This stimulus type was the same as used in our previous study (Supin *et al.*, 1994).

The results obtained with this stimulus type are shown in Fig. 8. The main difference from the preceding series was that in a high-frequency region (2.8–8 kHz), ripple-density resolution reached higher values. Within the region of 2.8–8 kHz, interindividual means were from 17.6 ± 1.8 s.d. (at 5.6 kHz) to 22.0 ± 4.6 s.d. (at 8 kHz). The overall mean level of this range was 19.9 ± 3.4 s.d. Thus the difference between the data obtained with shallow and steep spectrum edges was as large as 5.0 ± 0.6 SE.

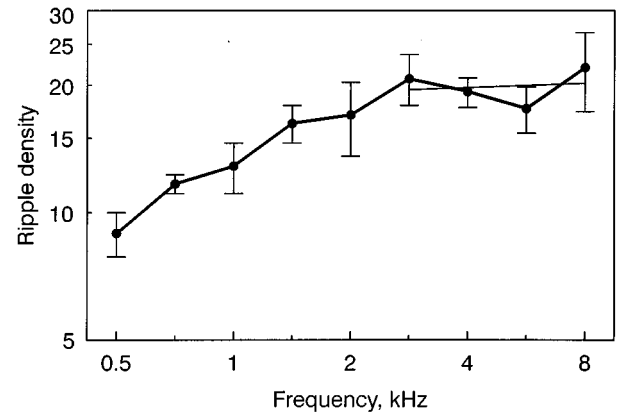


FIG. 8. Ripple-density resolution dependence on center frequency obtained using filters with constant ripple spacing and steep boundaries. Designation as in Fig. 3 (regression line for a range of 2.8–8 kHz).

III. DISCUSSION

A. Comparison of constant and frequency-proportional rippled noises

Small but significant difference was found between ripple-density resolutions obtained with frequency-proportional and constant rippled noise when the latter was characterized by its relative ripple density at the passband center. It is reasonable to suppose that the rippled structure of constant rippled noise was discriminated at a lower part of the passband where the relative ripple density was lower. The real resolution might be down to 2–2.5 units lower than the arbitrary estimates presented above. If suppose that it actually was around 2 units lower, then the data obtained with frequency-proportional and constant rippled noises agree well.

Another possible suggestion is that constant ripple spacing really provides better conditions to discriminate ripple structure than frequency-proportional spacing. Indeed, the auditory system is capable to perceive a characteristic pitch associated with the ripple spacing (Yost *et al.*, 1977, 1996; Yost and Hill, 1979; Bilsen and Wieman, 1980; Yost, 1982, 1996a, b). This pitch is affected a little by phase reversals (approximately, by $\pm 10\%$). In frequency-proportional noise, ripple spacing varies throughout the passband, therefore it may result in a less pitch strength and worse ripple resolution. However, this explanation seems less probable. In frequency-proportional rippled noise, the range of ripple spacing depends on bandwidth. Therefore, one can expect that pitch strength depends on the bandwidth as well. However, ripple-density resolution little depended on noise bandwidth in a very wide range, from 0.5 to 8 octaves. So, ripple-spacing constancy does not seem to play a crucial role in ripple discrimination in our experimental conditions.

Therefore we consider the data obtained with frequency-proportional noises as better ripple-density resolution estimates. Stimuli of frequency-proportional ripple spacing were already used to study discrimination between spectra composed of a number of pure tones (Green, 1986, 1987). However, in those studies, ripple depth discrimination was mostly studied, and ripple densities were rather low. The data presented herein concern mostly the ripple-density resolution.

B. Effect of steep edges of the spectrum envelope

Ripple-density resolution estimates were higher at steep than at shallow spectrum envelope edges. At higher center frequencies, the difference was as large as five units. Two explanations of this difference may be suggested. The first one is that unequal ripple levels within the passband of a cosine-enveloped spectrum result in worse ripple discrimination thus giving lower estimates. This mechanism, however, does not seem probable. In the cosine-enveloped spectrum, four ripples (i.e., the same number as in the steep-edge spectrum) had levels comparable with that of the center peak. It is unlikely that the difference of a few dB prevents resolution of spectral ripples. Furthermore, ripple-density resolution did not improve with widening of cosine-enveloped spectrua from 0.5 to 4 octaves, although at a wide bandwidth, a few

ripples around the center are of almost equal level. This is in contradiction with the suggestion that the ripple level irregularity reduced ripple discrimination.

Another mechanism may be based on edge effects at the spectrum boundaries. It can be seen in Fig. 1(d) that at a sharp spectrum edge, replacing of a trough by a peak results in a shift of the real spectrum boundary. This effect can facilitate the detection of phase reversals. Other processes as a kind of lateral inhibition may also produce edge effects. If a rippled spectrum does not contain steep edges, as in Fig. 1(a)–(c), these edge effects are ''smoothed'' down. This mechanism seems more probable, although we can not explain yet why it expressed itself mostly at higher frequencies (2.8–8 kHz). According to this explanation, ripple-density resolution is around 13 within a more or less uniform spectrum, although at passband edges, finer spectrum structure can be discriminated.

C. Comparison of ripple-density resolution data with some models of rippled-noise processing

Although rippled noise does not have evident temporal structure, its autocorrelation function (ACF) reveals hidden temporal organization. At a ripple spacing of δF, ACF has a prominent peak at a delay $\tau = 1/\delta F$. It was supposed that rippled noise processing is mainly based on the temporal structure (Yost *et al.*, 1996; Patterson *et al.*, 1996). On the other hand, a generally adopted model of the auditory system containing a bank of tuned peripheral filters may be used to explain processing of any spectra including rippled ones. So it seemed reasonable to compare the data obtained herein with predictions of these models.

1. Temporal processing models

A problem of temporal processing models is that they require very high temporal resolution to explain the ripple-density resolution data. For example, at a frequency of 8 kHz and relative ripple density as low as four units, the absolute ripple density is 0.5 kHz^{-1} which corresponds to ACF lag of 0.5 ms. This lag is several times shorter than the ''temporal window'' of the human's hearing (Moore *et al.*, 1988; Plack and Moore, 1990). However, psychometric curves showed that phase reversals were well detectable just at low ripple densities (see Fig. 3); on the contrary, at higher ripple density when ACF lags were comparable with known values of temporal resolution, phase reversals became undetectable.

Furthermore, contrary to wide-band noise [Fig. 9(a)], ACFs of narrow-band noise have similar ''envelopes'' at opposite peak-trough positions and differ only in their fine structure reflecting the center frequency [Fig. 9(b)]. Discrimination of this structure requires extremely high temporal resolution: at 8 kHz, the half-cycle duration is as short as 62.5 μs. This degree of temporal resolution seems problematic. Contribution of the phase-locking mechanisms also seems unlikely at frequencies as high as 8 kHz although ripple-density resolution reached maximal values just at these frequencies.

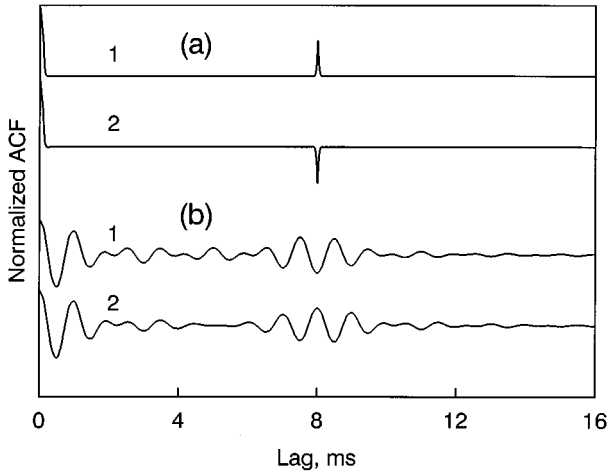


FIG. 9. ACF examples of various rippled noises. (a) Wide-band noise (0–20 kHz) with constant ripple spacing; absolute ripple density is 8 kHz^{-1}. (b) Narrow-band noise with constant ripple spacing and shallow boundaries; center frequency 1 kHz, relative ripple density 8 (absolute density 8 kHz^{-1}). 1 and 2—noises with opposite peak trough positions.

2. Peripheral filtering model

The data presented above can be compared with predictions of a filter-bank model taking a linear approach as a first approximation. Suppose the auditory filter shape is approximated by the rounded exponential (roex) function (Patterson *et al.*, 1982):

$$W(g) = (1 + pg)\exp(-pg) \tag{3}$$

(for $g > 0$), where W is the transmitted power, g is the relative deviation from the center frequency, and p is a parameter determining the filter tuning. The power spectrum of rippled noise at the filter input $P_{\text{in}}(g)$ may be defined as

$$P_{\text{in}}(g) = 1 + \cos 2\pi D(g + \theta), \tag{4}$$

where D is the ripple density and θ determines the phase relation of the spectrum and the filter (at θ equal to 0, $1/D$, $2/D$, etc., a spectrum peak is centered on the filter; at θ equal to $0.5/D$, $1.5/D$, etc., a trough is centered). The power $P_{\text{out}}(g)$ transmitted by the filter is

$$P_{\text{out}}(g) = \int_{-\infty}^{0} P_{\text{in}}(g)W(-g)dg + \int_{0}^{\infty} P_{\text{in}}(g)W(g)dg$$

$$= 4\left[\frac{1}{p} + \frac{\cos 2\pi D\theta}{p(1 + 4\pi^2 D^2/p^2)}\right]. \tag{5}$$

At ripple phase reversals (θ changes by $\pm 0.5/D$), the dB difference δI at the filter output is

$$\delta I = 10 \log \frac{(1 + 4\pi^2 D^2/p^2)^2 + \cos 2\pi D\theta}{(1 + 4\pi^2 D^2/p^2)^2 + \cos 2\pi D(\theta + 0.5)}. \tag{6}$$

At all values of D, $|\delta I|$ is maximal at $\theta = 0$ (ripple peak centered on the filter) or $\theta = \pm 0.5/D$ (ripple trough centered on the filter). $|\delta I|$ decreases with D increase which agrees qualitatively with psychometric curves (see Fig. 3) showing worse ripple resolution at higher D. Suppose that δI becomes undetectable below a certain threshold. This threshold may be set as the contrast threshold for rippled noise which

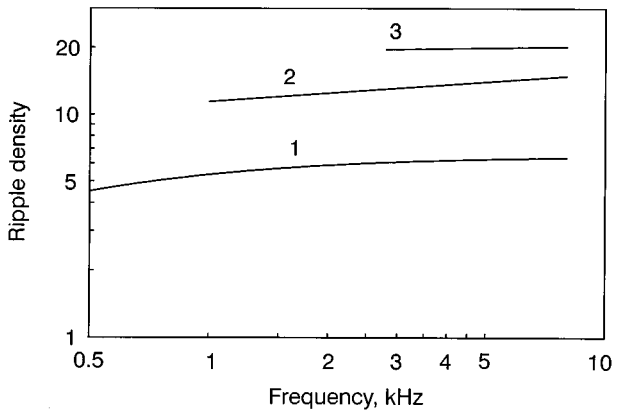


FIG. 10. Ripple-density resolution dependence on center frequency. 1—prediction of the peripheral filter model using Eqs. (7) and (8). 2—experimental data obtained for frequency-proportional rippled spectra within a range of 1–8 kHz. 3—data obtained for steep-edged constant rippled spectra within a range of 2.8–8 kHz (regression lines).

was found to be 1.8 dB (Supin *et al.*, 1994). Close spectral contrast threshold was found by Green (1986, 1987). It follows from Eq. (6) that at the best responding filter ($\theta=0$ or $\theta=\pm 0.5/D$), this δI value appears at $D=p/5.65$. Thus, the model predicts resolution $R=p/5.65$. Eequivalent rectangular bandwidth (ERB) of the roex filter is equal to $4F_0/p$ (Patterson *et al.*, 1982), hence,

$$R=0.71F_0/\mathrm{ERB}. \tag{7}$$

The adjusting factor of 0.71 was found for the roex filter but it little depends on filter form being within a range of 0.65–0.8 for a variety of filter forms.

There are several analytical expressions summarizing numerous data on ERB as a function of center frequency. Among them, a simple function proposed by Glasberg and Moore (1990) can be used:

$$\mathrm{ERB}=24.7(4.37F+1), \tag{8}$$

where ERB is in Hz, and F is the center frequency, kHz. Thus ripple-density resolution predicted by the peripheral auditory filter model can be found as a function of frequency using Eqs. (7) and (8). This function is presented in Fig. 10(1). At the same figure, regression lines approximating the experimental data are shown (2 and 3). Even lower resolution estimates obtained with frequency-proportional probes (2) are twice higher than the values predicted by the peripheral filter model. Estimates obtained with steep-edged spectra (3) still further exceed the prediction.

To explain this disagreement, it may be hypothesized that ripple-density resolution reflects some nonlinear processes sharpening up the internal spectra in the auditory system. Among others, a possible mechanism of such sharpening may be a kind of lateral suppression which is effective to enhance spectral contrast (review in Evans, 1992).

So, the use of either temporal processing or peripheral filter models taken singly has to overcome some difficulties. Further investigations are necessary for better understanding the involved mechanisms.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research, Grant No. 96-04-48036. Authors thank anonymous reviewers for criticism which was helpful to improve the manuscript.

Bilsen, F. A., ten Kate, J. H., Buunen, T. J. F., and Raatgever, J. (**1975**). ''Responses of single units in the cochlear nucleus of the cat to cosine noise,'' J. Acoust. Soc. Am. **58**, 858–866.

Bilsen, F. A., and Wieman, J. L. (**1980**). ''Atonal periodicity sensation for comb filtered noise signals,'' in *Psychophysical and Behavioral Studies in Hearing*, edited by G. van der Brink and F. A. Bilsen (Delft U.P., Delft), pp. 379–382.

Evans, E. F. (**1977**). ''Frequency selectivity at high signal levels of single units in cochlear nerve and cochlear nucleus,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 185–192.

Evans, E. F. (**1992**). ''Auditory processing of complex sounds: an overview,'' Philos. Trans. R. Soc. London, Ser. B **336**, 295–306.

Evans, E. F., and Wilson, J. P. (**1973**). ''Frequency selectivity of the cochlea,'' in *Basic Mechanisms of Hearing*, edited by A. R. Møller (Academic, New York), pp. 519–551.

Fastl, H. (**1980**). ''Pitch strength and masking patterns of low-pass noise,'' in *Psychophysical, Physiological, and Behavioral Studies of Hearing*, edited by G. v. d. Brink and F. Bilsen (Delft U.P., Delft), pp. 334–339.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hearing Res. **47**, 103–138.

Glasberg, B. R., Moore, B. C. J., and Nimmo-Smith, I. (**1984**). ''Comparison of auditory filter shapes derived with three different maskers,'' J. Acoust. Soc. Am. **75**, 536–544.

Green, D. M. (**1986**). '' 'Frequency' and the detection of spectral shape change,'' in *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. D. Patterson (Plenum, New York), pp. 351–358.

Green, D. M. (**1987**). *Profile Analysis: Auditory Intensity Discrimination* (Oxford U.P., New York).

Houtgast, T. (**1974**). ''Masking patterns and lateral inhibition,'' in *Facts and Models in Hearing*, edited by E. Zwicker and E. Terhardt (Springer-Verlag, Berlin), pp. 258–265.

Houtgast, T. (**1977**). ''Auditory-filter characteristics derived from direct-masking and pulsation-threshold data with a rippled-noise masker,'' J. Acoust. Soc. Am. **62**, 409–415.

Kohlrausch, A., and Houtsma, A. J. M. (**1992**). ''Pitch related to spectral edges of broadband signals,'' Philos. Trans. R. Soc. London, Ser. B **336**, 375–382.

Moore, B. C. J., and Glasberg, B. R. (**1981**). ''Auditory filter shapes derived in simultaneous and forward masking,'' J. Acoust. Soc. Am. **71**, 942–945.

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.

O'Loughlin, B. J., and Moore, B. C. J. (**1981**). ''Off-frequency listening: Effects on psychoacoustical tuning curves obtained in simultaneous and forward masking,'' J. Acoust. Soc. Am. **69**, 1119–1125.

Patterson, R. D. (**1976**). ''Auditory filter shapes derived with noise stimuli,'' J. Acoust. Soc. Am. **59**, 640–654.

Patterson, R. D., Handel, S., Yost, W. A., and Datta, A. J. (**1996**). ''The relative strength of the tone and noise components in iterated rippled noise,'' J. Acoust. Soc. Am. **100**, 3286–3294.

Patterson, R. D., and Henning, G. B. (**1977**). ''Stimulus variability and auditory filter shape,'' J. Acoust. Soc. Am. **62**, 649–664.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milory, 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.

Pick, G. F. (**1980**). ''Level dependence of psychophysical frequency resolution and auditory filter shape,'' J. Acoust. Soc. Am. **68**, 1085–1095.

Pick, G. F., Evans, E. F., and Wilson, J. P. (**1977**). ''Frequency resolution in patients with hearing loss of cochlear origin,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, New York), pp. 273–282.

Plack, C. J., and Moore, B. C. J. (**1990**). ''Temporal window shape as a function of frequency and level,'' J. Acoust. Soc. Am. **87**, 2178–2187.

Plomp, R., and Steeneken, H. J. M. (**1968**). ''Interference between two simple tones,'' J. Acoust. Soc. Am. **43**, 883–884.
Popov, V. V., and Supin, A. Ya. (**1984**). ''Quantitative measurement of frequency resolving power of the human hearing,'' Dokl. Akad. Nauk SSSR (Proc. Acad. Sci. USSR) **278**, 1012–1016.
Shamma, S. A., and Versnel, H. (**1995**). ''Ripple analysis in the ferret primary auditory cortex. II. Prediction of unit responses to auditory spectral profiles,'' J. Auditory Neurosci. **1**, 255–271.
Shamma, S. A., Kowalski, N., and Versnel, H. (**1995a**). ''Ripple analysis in the ferret primary auditory cortex. III. Topographic distribution of ripple response parameters,'' J. Auditory Neurosci. **1**, 271–278.
Shamma, S. A., Versnel, H., and Kowalski, N. (**1995b**). ''Ripple analysis in the ferret primary auditory cortex. I. Response characteristics of single units to sinusoidally rippled spectra,'' J. Auditory Neurosci. **1**, 233–255.
Small, Jr., A. M., and Daniloff, R. G. (**1967**). ''Pitch of noise bands,'' J. Acoust. Soc. Am. **41**, 506–512.
Supin, A. Ya., and Popov, V. V. (**1987**). ''Frequency resolving power of the human hearing,'' Fiziologiya Cheloveka (Human Physiology) **13**, 28–34.
Supin, A. Ya., Popov, V. V., Milekhina, O. N., and Tarakanov, M. B. (**1994**). ''Frequency resolving power measured by rippled noise,'' Hearing Res. **78**, 31–40.
Supin, A. Ya., Popov, V. V., Milekhina, O. N., and Tarakanov, M. B. (**1997**). ''Frequency-temporal resolution of hearing measured by rippled noise,'' Hearing Res. **108**, 17–27.
Terhardt, E. (**1968**). ''Über die durch amplitudenmodulierte Sinusstöne hervorgerufene Hörempfindung,'' Acustica **20**, 210–214.
Wilson, J. P., and Evans, E. F. (**1971**). ''Grating acuity of the ear: psychophysical and neurophysiological measures of frequency resolving power,'' in *7th Internat. Congress on Acoustics* (Akademiai Kiado, Budapest), Vol. 3, pp. 397–400.
Yost, W. A. (**1982**). ''The dominance region and ripple-noise pitch: A test of the peripheral weighting model,'' J. Acoust. Soc. Am. **72**, 416–425.
Yost, W. A. (**1996a**). ''Pitch of iterated rippled noise,'' J. Acoust. Soc. Am. **100**, 511–518.
Yost, W. A. (**1996b**). ''Pitch strength of iterated rippled noise,'' J. Acoust. Soc. Am. **100**, 3329–3335.
Yost, W. A., and Hill, R. (**1979**). ''Models of the pitch and pitch strength of ripple noise,'' J. Acoust. Soc. Am. **66**, 400–410.
Yost, W. A., Hill, R., and Perez-Falcon, T. (**1977**). ''Pitch discrimination of ripple noise,'' J. Acoust. Soc. Am. **63**, 1166–1173.
Yost, W. A., Patterson, R. D., and Sheft, S. (**1996**). ''A time domain description for the pitch strength of iterated rippled noise,'' J. Acoust. Soc. Am. **99**, 1066–1078.
Zwicker, E. (**1952**). ''Die Grenzen der Hörbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones,'' Acustica, Akustische Beihefte, AB125-AB133.
Zwicker, E. (**1974**). ''On a psychoacoustical equivalent of tuning curves,'' in *Facts and Models in Hearing*, edited by E. Zwicker and E. Terhardt (Springer-Verlag, Berlin), pp. 132–141.
Zwicker, E. (**1982**). *Psychoacoustic* (Springer-Verlag, Berlin).

Temporal processing of the pitch of complex tones

Louise J. White and Christopher J. Plack
Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, United Kingdom

(Received 8 July 1997; revised 6 January 1998; accepted 16 January 1998)

The effect of tone duration on fundamental frequency ($F0$) discrimination is greater for complexes containing unresolved harmonics than for those containing resolved harmonics [Plack and Carlyon, J. Acoust. Soc. Am. **98**, 1355–1364 (1995)]. Three experiments explored this effect further. The first experiment measured sensitivity (as d') to fundamental frequency ($F0$) differences for two complexes, both with an $F0$ of 250 Hz. The first complex was low-pass filtered at 1875 Hz to create a resolved complex and the second was bandpass filtered between 5500 and 7500 Hz to create an unresolved complex. The harmonics for the resolved complex were selected so that no two harmonics were the same between the two observation intervals. Performance for both complexes was measured for tone durations of 20, 40, 80, and 160 ms. For the unresolved complex, the effect of duration was greater than that for the resolved complex and greater than the predictions of a ''multiple-looks'' model assuming either peripheral (before sampling) or central (after combining samples) sources of variance. The second experiment replicated these results using an $F0$ of 62.5 Hz with the cutoff frequencies of the bandpass filters divided by four, confirming that the effect is related to resolvability and not to spectral region. In the final experiment, $F0$ discrimination for pairs of complexes separated by a temporal gap was measured relative to that for one complex. Performance for the resolved and unresolved complexes was similar: Very little effect of gap duration was observed and the results were consistent with the predictions of the peripheral-variance multiple-looks model. Taken together, the results suggest that the pitch mechanism for resolved harmonics uses a relatively short sampling window of around 20 ms, while the mechanism for unresolved harmonics may use a more complex strategy for optimizing the combination of information over time, perhaps involving a flexible integration time. [S0001-4966(98)04304-5]

PACS numbers: 43.66.Hg, 43.66.Mk, 43.66.Ba, 43.66.Fe [WJ]

INTRODUCTION

Complex tones, such as vowels or musical tones, consist of a series of sinusoidal components (harmonics) whose frequencies are integer multiples of the repetition rate of the complex, also known as the fundamental frequency ($F0$). Complex tones are usually heard as having a single pitch corresponding to $F0$, even though there may be many harmonics present.

The lower-numbered harmonics in a complex tone (harmonic numbers less than about ten) are resolved by the peripheral auditory system; that is, the individual harmonics produce separate peaks in the excitation pattern (Moore and Glasberg, 1983, 1987). Pattern recognition theories (Goldstein, 1973; Srulovicz and Goldstein, 1983; Terhardt, 1974) hypothesize that pitch information is extracted from these lower-numbered resolved harmonics. However, complex tones containing only the higher-numbered harmonics, which are not resolved, also provide a sensation of pitch (Houtsma and Smurzynski, 1990; Moore and Rosen, 1979). The information for this pitch must be temporally based and carried by the complex waveform produced by the interaction of the unresolved harmonics on the basilar membrane. Recent pitch models (Assman and Summerfield, 1990; Meddis and Hewitt, 1991a, b; Moore, 1997) show how $F0$ can easily be extracted from temporal information; for example, by use of the autocorrelation function acting on the synchronized discharge patterns of auditory nerve fibres (phase locking). Furthermore, computational implementations of this approach (Meddis and Hewitt, 1991a, b) produce similar results to human listeners for phenomena such as the missing fundamental (Schouten, 1940, 1970), ambiguous pitch (Schouten *et al.*, 1962), and the pitch derived from inharmonic equally spaced tone complexes (Plomp, 1976), which all involve *resolved* harmonics. Meddis and Hewitt conclude that a timing model based on autocorrelation can account for the pitch produced by both resolved and unresolved harmonics.

Arguing against the single-mechanism view, Shackleton and Carlyon (1994) found perceptual differences in simultaneous comparisons between complexes containing resolved and unresolved harmonics. Using various combinations of bandpass filtering and $F0$, they showed that a comparison of $F0$ between a resolved and an unresolved group is harder than comparisons within a resolvability group (i.e., resolved versus resolved or unresolved versus unresolved). They suggested that this may indicate the existence of two pitch mechanisms, one for resolved harmonics and one for unresolved harmonics. Comparing the output of the two mechanisms results in an additional ''comparison noise'' that reduces detectability.

In addition, Plack and Carlyon (1995) showed that $F0$ discrimination for groups of unresolved harmonics was very poor at short durations, and that the improvement in $F0$ discrimination with increasing tone duration was much greater for unresolved harmonics than for resolved harmonics. They

argued that the different effects of duration for resolved and unresolved harmonics could be evidence for the operation of two pitch mechanisms. They suggested that the mechanism for resolved harmonics measures the frequencies of the individual harmonics and uses these to calculate $F0$. In terms of a temporal mechanism, the repetition rates (i.e., frequencies) of the individual harmonics may be used, rather than the single prominent repetition rate pooled across all harmonics as in the current computational models. In support of their idea, they showed that the effect of duration on $F0$ discrimination for the resolved harmonics was similar to the effect of duration on frequency discrimination for a pure tone (Moore, 1973) in the ''dominant region'' (Plomp, 1967; Ritsma, 1967) of the complex. In contrast, the duration effect for unresolved harmonics was similar to that for a pure tone with a frequency equal to $F0$, suggesting that in this case the overall repetition rate of the complex is the significant parameter.

An important problem to address at this stage concerns what mechanisms underlie the improvements with duration for both resolved and unresolved harmonics. The aim of the present experiments was to go some way toward understanding these mechanisms.

In terms of signal detection theory (Green and Swets, 1966), the detectability of a change in a stimulus parameter is dependent on the size of the change and the variability of the parameter:

$$d' = \frac{\overline{X_s} - \overline{X_n}}{\sigma}, \tag{1}$$

where d' is a measure of discriminability, $\overline{X_s}$ and $\overline{X_n}$ are the magnitudes of the decision variable with (X_s) and without (X_n) the change, and σ^2 is the variance of the decision variable. The decision variable is assumed to vary according to the Gaussian distribution. By measuring psychometric functions, Plack and Carlyon (1995) demonstrated that $F0$ difference (in Hz) is linearly related to d'. In other words, it would be possible to substitute the $F0$s of the comparison and standard intervals in a two-alternative $F0$ discrimination task for $\overline{X_s}$ and $\overline{X_n}$ in Eqs. (1)–(4).

Now, an improvement in discrimination with duration may arise from an accumulation of information over time. In terms of the ''multiple-looks'' model (Viemeister and Wakefield, 1991), discrete short-duration samples of the stimulus are taken and the information from these samples is combined optimally. If the main source of variance is before the sampling process, if each sample varies independently, and if the magnitude of the variance is constant across samples, then the effect on discrimination of combining n samples is given by

$$d'_n = \frac{n(\overline{X_s} - \overline{X_n})}{\sqrt{n\sigma_p^2}} = \frac{\sqrt{n}(\overline{X_s} - \overline{X_n})}{\sigma_p} = \sqrt{n}d'_1, \tag{2}$$

where d'_n is the combined discriminability of n samples, and σ_p^2 is the variance of each sample or ''peripheral'' variance. A greater improvement in duration is predicted if the main source of variance occurs after the samples have been combined. In this case, the total variance is independent of the number of samples so that

$$d'_n = \frac{n(\overline{X_s} - \overline{X_n})}{\sigma_c} = nd'_1, \tag{3}$$

where σ_c^2 is the variance after sampling or ''central'' variance. If both peripheral and central components of variance are present, then the equation becomes

$$d'_n = \frac{n(\overline{X_s} - \overline{X_n})}{\sqrt{n\sigma_p^2 + \sigma_c^2}}. \tag{4}$$

When $\sigma_c^2 \gg n\sigma_p^2$ the central component of variance is dominant and d' increases with n. When $n\sigma_p^2 \gg \sigma_c^2$ the peripheral component of variance is dominant and d' increases with $\sqrt{n}$. So that, in a situation where the central variance is larger than the peripheral variance, the improvement with duration should be large for short durations (low n) but flatten off for large durations (high n) as the relative size of the peripheral variance increases.

In terms of a pitch mechanism which uses only temporal information from the auditory nerve, the peripheral variance may be a result of the stochastic nature of neural discharges: A neuron will not fire at exactly the same phase on each cycle of the incoming waveform (see Rose *et al.*, 1968). The overall accuracy of phase locking can be improved by combining spike information over time. The first two experiments were designed to test whether either basic version (peripheral or central variance) of the multiple-looks model could account for the effects of duration on $F0$ discrimination.

I. EXPERIMENT 1A: EFFECTS OF DURATION ON *F*0 DISCRIMINATION FOR A 250-Hz *F*0

A. Rationale

This experiment attempts to replicate the findings of Plack and Carlyon (1995) showing that the effects of duration on $F0$ discrimination is greater for unresolved than for resolved harmonics. The experiment also tests the hypothesis that the pitch mechanism can be described by a multiple-looks model with either a constant peripheral variance or a central variance. The effect of tone duration on $F0$ discrimination was measured for two complex tones, one consisting only of resolved harmonics and one consisting only of unresolved harmonics. The resolved complex consisted of harmonics chosen so that no two harmonics were the same in the two observation intervals. This prevents listeners basing discrimination on the frequencies of the individual resolved harmonics (Moore and Glasberg, 1990). As mentioned above, the peripheral-variance multiple-looks model predicts that detectability, measured as d', will increase by a factor of $\sqrt{n}$ for every n-fold increase in tone duration. The central-variance multiple-looks model predicts that d' will increase by a factor of n for every n-fold increase in tone duration.

B. Stimuli

In these experiments, the use of the term ''resolvability'' agrees with previous experimental definitions (e.g., Moore

et al., 1984; Houtsma and Smurzynski, 1990; Plomp, 1964; Shackleton and Carlyon, 1994). For example, Plomp (1964) showed that only the first 5–8 harmonics of a complex tone could be ''heard out,'' in the sense that they could be reliably matched to a comparison pure tone. Perhaps the strongest test of resolvability was provided by Shackleton and Carlyon (1994), who showed that a complex containing high-numbered harmonics presented in alternating sine and cosine phase (''ALT phase'') had a pitch an octave higher than the same complex with sine phase harmonics. When ALT-phase harmonics interact they produce a composite waveform with envelope peaks at half the period of $F0$. It is argued that this is heard as a doubling in pitch only when the harmonics interact on the basilar membrane, i.e., when they are unresolved. These results are consistent with a single definition of resolvability. The resolvability of a harmonic is defined by the number of harmonics contained between the 10-dB down-points of an auditory filter centered on that harmonic. If there are less than two then they are classed as resolved and if there are more than 3.25 they are classed as unresolved (Shackleton and Carlyon, 1994). The number of harmonics per filter can be calculated by dividing the filter bandwith (1.8 ERB) by $F0$ (Patterson *et al.*, 1988; Glasberg and Moore, 1990). This definition has been widely used by subsequent studies (e.g., Carlyon, 1997).

The $F0$ of both complexes was 250 Hz. The harmonics were low-pass filtered at 1875 Hz to create the resolved complex (res) and were bandpass filtered between 5500 and 7500 Hz to create the unresolved complex (unres). Prior to filtering, the waveforms were gated on and off with no ramps. However, since the waveforms were subsequently filtered, no gating clicks were audible. Ramps were not used in the main experiment to avoid restricting the usable duration of the tone in any way. The effects of ramping on performance are considered in Appendix B.

For the resolved complex, harmonic numbers 1, 4, 5, 8, 9, 12, 13, etc. were used in configuration A and 2, 3, 6, 7, 10, 11, etc. in configuration B (Moore and Glasberg, 1990). Configurations A and B were randomly assigned to the first and second observation intervals on each trial. Therefore, it would not have been possible for the listeners to perform the task using simple pure-tone frequency discrimination as no two harmonics were the same between the tones being compared. For the unresolved complex consecutive harmonics up to number 50 were used prior to filtering. The level of each harmonic component was 50 dB SPL. The components were added in sine phase and the starting phase of the fundamental was randomized. $F0$ discrimination was measured for complex tones with durations of 20, 40, 80, and 160 ms. A schematic spectrogram of the stimuli is shown in Fig. 1.

The stimuli were generated digitally with 32-bit resolution on a Silicon Graphics workstation. The sampling rate was 48 kHz. Stimuli were filtered using a digital FIR filter with a nominal attenuation of 90 dB per octave. Stimuli were played out through a built-in 16-bit DAC and were presented monaurally to the right ear via Sennheiser HD414 headphones which were connected directly to the headphone output of the computer.

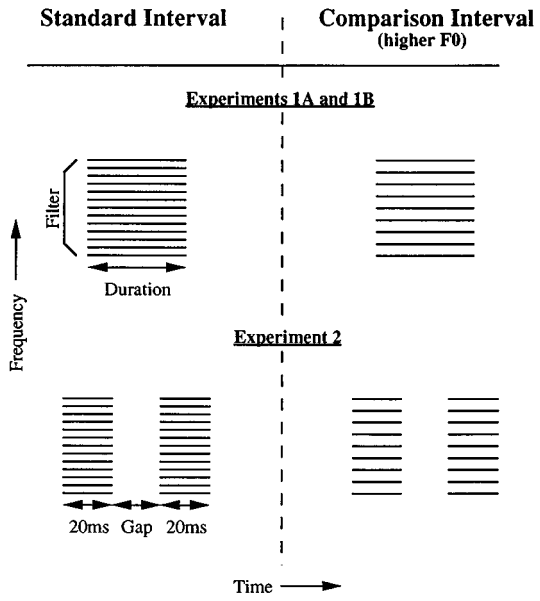


FIG. 1. Schematic spectrograms of the stimuli used in the experiments.

C. Procedure

A two-interval, two-alternative forced choice procedure was used throughout and feedback was provided after every trial. The inter-stimulus interval was 500 ms. Initially, a two-down one-up adaptive tracking procedure (Levitt, 1971) was used to estimate the $F0$ discrimination threshold for the 20-ms duration condition for both the resolved and unresolved complexes. The $F0$ of the complex in the standard interval was fixed at 250 Hz, and the complex in the comparison interval had a higher $F0$, which was varied. The percentage $F0$ difference between the two observation intervals was increased and decreased by a factor of 2 for the first four turnpoints and by a factor of 1.41 thereafter. Sixteen turnpoints were measured and the threshold estimate was taken as the geometric mean of the last twelve. Ten such threshold estimates were made and the final estimate was the geometric mean of these ten. This $F0$ difference, measured independently for each listener and each complex, was then used in a constant stimuli design for all four tone durations. Each block consisted of 100 trials. The conditions were presented randomly until each listener had completed eight blocks of each tone duration (eight repetitions). The overall percent correct was used to calculated d'. The res and the unres conditions were completed separately, the unres conditions were completed first for all listeners.

D. Listeners

Five listeners took part. One was author LW and the others were paid an hourly wage. All listeners had participated in other psychoacoustic experiments and had hearing thresholds within normal clinical limits. They were aged between 27 and 30 years. Listeners were given at least two hours training on the tasks before data collection began. They were tested individually in an IAC single-walled sound-attenuating booth.

TABLE I. The d' scores for the four complex durations in experiment 1A for both the resolved and the unresolved conditions. (The numbers in parentheses are a measure of the variance across 8 repetitions of 100 trials. They were calculated by taking the difference between the mean and the standard error in percent correct and converting this to d'_{error}. The figure in parentheses is the difference between d'_{error} and d'.)

Complex duration (ms)	LW	HT	JB	AD	SW	Mean	s.d.
			Unresolved harmonics				
20	0.80 (0.06)	0.62 (0.12)	0.73 (0.05)	0.48 (0.01)	0.35 (0.08)	**0.59**	0.18
40	1.61 (0.21)	1.49 (0.09)	2.11 (0.09)	1.75 (0.08)	1.15 (0.09)	**1.61**	0.36
80	1.72 (0.13)	2.66 (0.09)	2.50 (0.11)	2.98 (0.27)	1.03 (0.17)	**2.18**	0.79
160	1.86 (0.17)	2.71 (0.08)	2.44 (0.18)	3.43 (0.45)	1.41 (0.07)	**2.37**	0.78
			Resolved harmonics				
20	0.47 (1.00)	0.71 (0.07)	1.22 (0.11)	1.53 (0.17)	1.29 (0.14)	**1.04**	0.44
40	0.96 (0.15)	1.13 (0.09)	2.06 (0.15)	1.88 (0.22)	2.04 (0.12)	**1.61**	0.53
80	1.28 (0.08)	0.90 (0.10)	1.92 (0.13)	2.52 (0.23)	2.20 (0.14)	**1.76**	0.64
160	1.28 (0.16)	0.98 (0.12)	1.81 (0.10)	2.71 (0.19)	2.33 (0.16)	**1.82**	0.72

E. Results and discussion

The threshold $F0$ differences, measured using the adaptive procedure, for the 20-ms resolved complex for LW, HT, JB, AD, and SW were 8.7%, 0.5%, 5.9%, 4.3%, and 2.8%, respectively. The threshold $F0$ differences for the 20-ms unresolved complex were 12.4%, 7.6%, 8.0%, 9.2%, and 6.1%, respectively. The variability between listeners was very large, but this is not unusual for $F0$ discrimination (for example, see Plack and Carlyon, 1995).

The constant-stimuli results are shown in Table I and Fig. 2. The data in Fig. 2 are plotted so that the values of d' for the 40-, 80-, and 160-ms complexes are shown relative to the value of d' for the 20-ms complex. This allows comparison of the increase relative to the 20-ms complex in both conditions, and allows direct comparison with the predictions of the multiple-looks models. The continuous lines show the predictions of multiple-looks models with peripheral (upper lines) and central (lower lines) components of variance.

The data show clear differences between the resolved

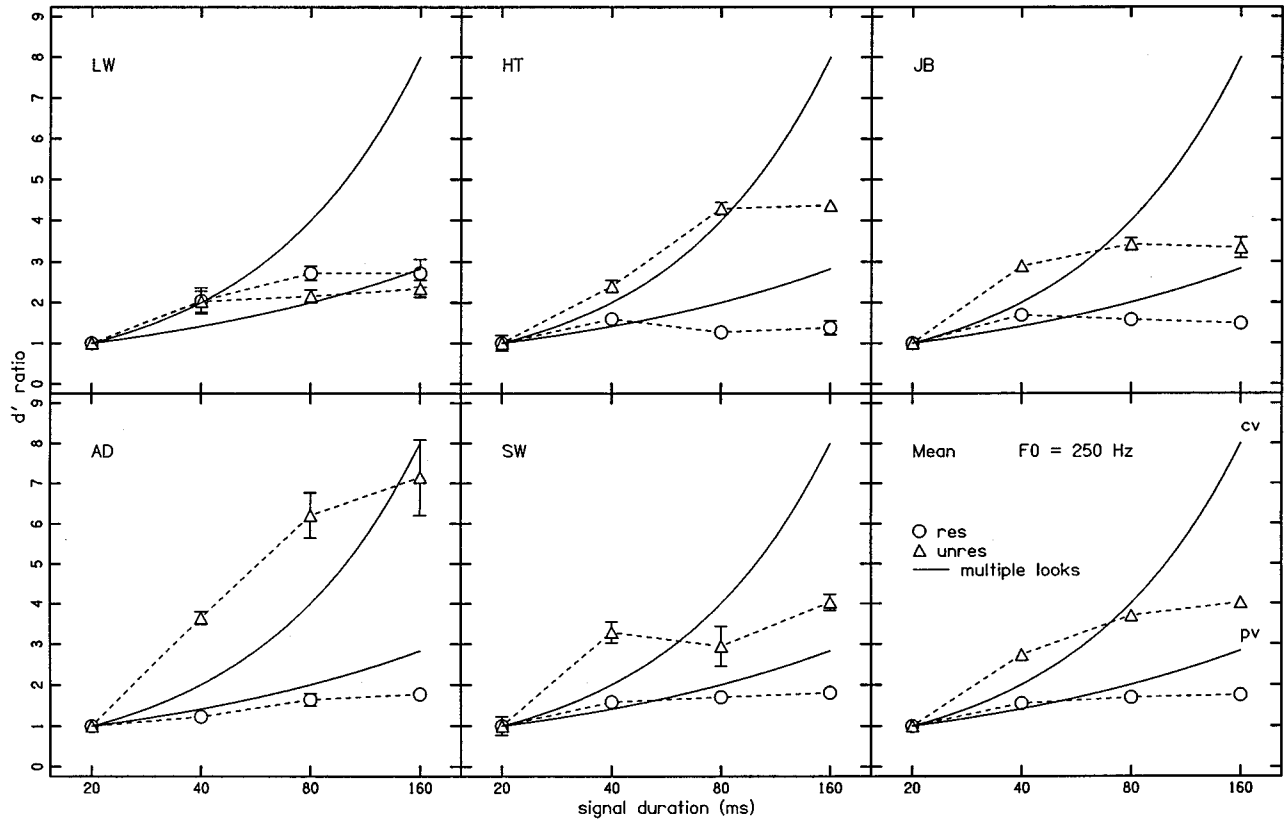


FIG. 2. The results of experiment 1A, showing $F0$ discrimination performance for resolved and unresolved 250-Hz complexes as a function of duration. The data and error bars are plotted relative to the d' value for a 20-ms complex. Error bars for the individual listeners were estimated using the method described in Table I. The predictions of the central-variance and peripheral-variance versions of the multiple looks model are denoted by cv and pv, respectively.

and unresolved complexes for four out of the five listeners. For the unresolved complex, performance increases as the duration of the complex increases up to approximately 80 ms and levels out beyond that. Performance at short durations is better than either model predicts, but at longer durations, performance is worse than the central-variance model (cv) predicts. In contrast, the results for the resolved complex indicate that performance is similar to the predictions of the peripheral-variance multiple-looks model (pv), at least for short durations. A repeated-measures analysis of variance (ANOVA) of all the data (resolvability×duration ×repetition) showed a significant effect of duration overall [$F(3,12)=20.884$, $p<0.01$] and also a significant interaction between resolvability and duration [$F(3,12)=5.113$, $p<0.05$] indicating that the effect of duration is greater for the unresolved harmonics, consistent with the results of Plack and Carlyon (1995).

Four out of the five listeners showed a large improvement in performance for the unresolved complex as duration increased from 20 to 40 ms. The d' scores for the 40-ms unresolved complex were tested against the predictions of both multiple-looks models as follows: The individual d' scores for the 20-ms tone were multiplied by $\sqrt{2}$ and by 2 to create the scores predicted by the two models. Paired sample t-tests were carried out between these predicted scores and the actual d' scores for the 40-ms tone. These showed that the improvement in d' between 20 and 40 ms for the unresolved complex is significantly greater than the predictions of both the peripheral-variance model ($t=6.22$, $df=4$, $p<0.01$) and the central-variance model ($t=3.08$, $df=4$, $p<0.05$). For the resolved complex, there is no significant difference between the measured improvement from 20 to 40 ms and the improvement predicted by the peripheral-variance model ($t=1.28$, $df=4$, $p=0.27$).

It can be seen in Table I that the d' scores for the 20-ms duration were generally lower for the unresolved harmonics than for the resolved harmonics. (A similar pattern was seen in the results from experiment 1B, see below.) This is despite the fact that the $F0$ differences in the constant stimuli presentations were the adaptive thresholds for the 20-ms duration for each listener and each complex type. In other words, all other things being equal, the d' values should have been approximately 0.78 (the 71% point tracked by the adaptive procedure) for all the conditions with a 20-ms duration. Although there is no obvious reason for this discrepancy, it is conceivable that the weak pitch percept produced by the unresolved harmonics (Houtsma and Smurzynski, 1990) made it hard to focus on the cues for discrimination when the $F0$ difference was near threshold, leading to low d' values. The adaptive procedure, by including trials with a large $F0$ difference in the initial part of the adaptive track, may have helped listeners to focus on these cues. It is not clear, however, why performance for the resolved harmonics should be slightly higher than expected. It should be noted that these deviations, while large in terms of d', do not represent a dramatic change in percent correct. For example, a d' increase from 0.59 to 1.04 corresponds to about an 11% increase in correct responses.

Some of the individual d' scores for the unresolved harmonics were high enough to raise the possibility that ceiling effects may have limited performance, and therefore may have artificially limited the improvement with duration. An experiment controlling for these problems is presented in Appendix C.

II. EXPERIMENT 1B: EFFECTS OF DURATION FOR A 62.5-HZ *F*0

A. Rationale and stimuli

The aim of this experiment was to ensure that the effects seen in experiment 1A were dependent on the resolvability of the harmonics and not on the high spectral region of the unresolved harmonics. A complex tone with a 62.5-Hz fundamental was low-pass filtered at 469 Hz to create a resolved complex and bandpass filtered between 1375 and 1875 Hz to create an unresolved complex. In other respects the stimuli were the same as in experiment 1A, including the use of nonoverlapping harmonics.

B. Procedure

The basic procedure was the same as for experiment 1A. A complex with an $F0$ of 62.5 Hz has a period of 16 ms and, unsurprisingly, listeners found an $F0$ discrimination task with a 20-ms complex very difficult to perform. Therefore, adaptive $F0$ discrimination thresholds were measured for a 40-ms complex. The $F0$ discrimination threshold for the 40-ms tone was used to measure d' for complexes with durations of 40, 80, and 160 ms.

C. Listeners

Four listeners with normal hearing took part. Three had taken part in the previous experiment. Ages ranged from 27 to 30.

D. Results and discussion

The $F0$ discrimination thresholds for the resolved 40-ms complex for HT, SW, PL, and AD were 4.3%, 6.6%, 8.2%, and 5.9%, respectively. The thresholds for the unresolved 40-ms complex were 4.5%, 5.5%, 5.7%, and 6.1%, respectively. In this case there was no great difference between the adaptive thresholds for the two complexes. The d' scores are shown in Table II, and plotted relative to those for the 40-ms complex in Fig. 3. Continuous lines show the predictions of the multiple-looks models. As in experiment 1A there is a far greater improvement in discrimination with duration for the unresolved complex than for the resolved complex, for three out of the four listeners. Listener PL did not show a differential effect of resolvability until the tone duration exceeded 80 ms. A repeated-measures ANOVA was performed on the data from all four listeners. This showed a significant improvement in $F0$ discrimination with increasing duration [$F(1,3)=13.561$, $p<0.05$] and a significant interaction between resolvability and duration [$F(1,3)=13.523$, $p<0.05$], again showing that most of the improvement in performance with increasing tone duration is for the unresolved complex.

Following the analysis in experiment 1A, the improvement from 40 to 80 ms for the unresolved complex was

TABLE II. The d' scores for the three complex durations in experiment 1B for both the resolved and the unresolved conditions. (The error values in parentheses were calculated using the same method as in Table I.)

Complex duration (ms)	HT	SW	PL	AD	Mean	s.d.
			Unresolved harmonics			
40	0.62 (0.10)	0.19 (0.06)	0.39 (0.03)	0.66 (0.04)	**0.47**	0.22
80	1.70 (0.14)	1.27 (0.13)	1.28 (0.06)	2.10 (0.10)	**1.59**	0.40
160	2.48 (0.23)	1.65 (0.10)	1.87 (0.08)	2.36 (0.16)	**2.09**	0.39
			Resolved harmonics			
40	1.67 (0.02)	0.91 (0.10)	1.04 (0.05)	0.79 (0.04)	**1.10**	0.39
80	2.96 (0.35)	1.61 (0.18)	3.35 (0.39)	1.37 (0.06)	**2.32**	0.98
160	2.84 (0.22)	2.03 (0.06)	3.47 (0.48)	1.51 (0.08)	**2.46**	0.87

compared with the predictions of both the peripheral-variance and central-variance models. The t-tests showed that performance for the 80-ms tone is significantly better than the predictions of both the peripheral-variance model ($t=9.59$, $df=3$, $p<0.01$) and the central-variance model ($t=6.25$, $df=3$, $p<0.01$).

Some of the d' values for the unresolved harmonics were very low. Listener SW had a value of 0.19 for the 40-ms duration, which corresponds to only 55.4 percent correct. The result raises the possibility that the listener had effectively ''given up'' for this condition and was just responding randomly. However, an analysis using the normal approximation to the binomial distribution (Rosenburg, 1990, page 176) showed that even this low score was significantly greater than chance performance ($p=0.001$).

The results support those from experiment 1A, showing that the improvement in duration for the resolved harmonics is broadly consistent with the predictions of the peripheral-variance model, whilst the improvement for the unresolved complexes does not agree with either multiple-looks model. Furthermore, in contrast to experiment 1A, performance improved substantially for the longer tone durations. Unpaired samples t-tests were carried out on the improvement in d' from 40- to 80-ms tone duration for both the resolved and unresolved complexes for both $F0$'s. Both showed significantly greater improvement for the complex with an $F0$ of 62.5 Hz (resolved harmonics: $t=3.19$, $df=7$, $p<0.05$; unresolved harmonics: $t=3.11$, $df=7$, $p<0.05$). The findings suggest that improvement with duration may depend upon

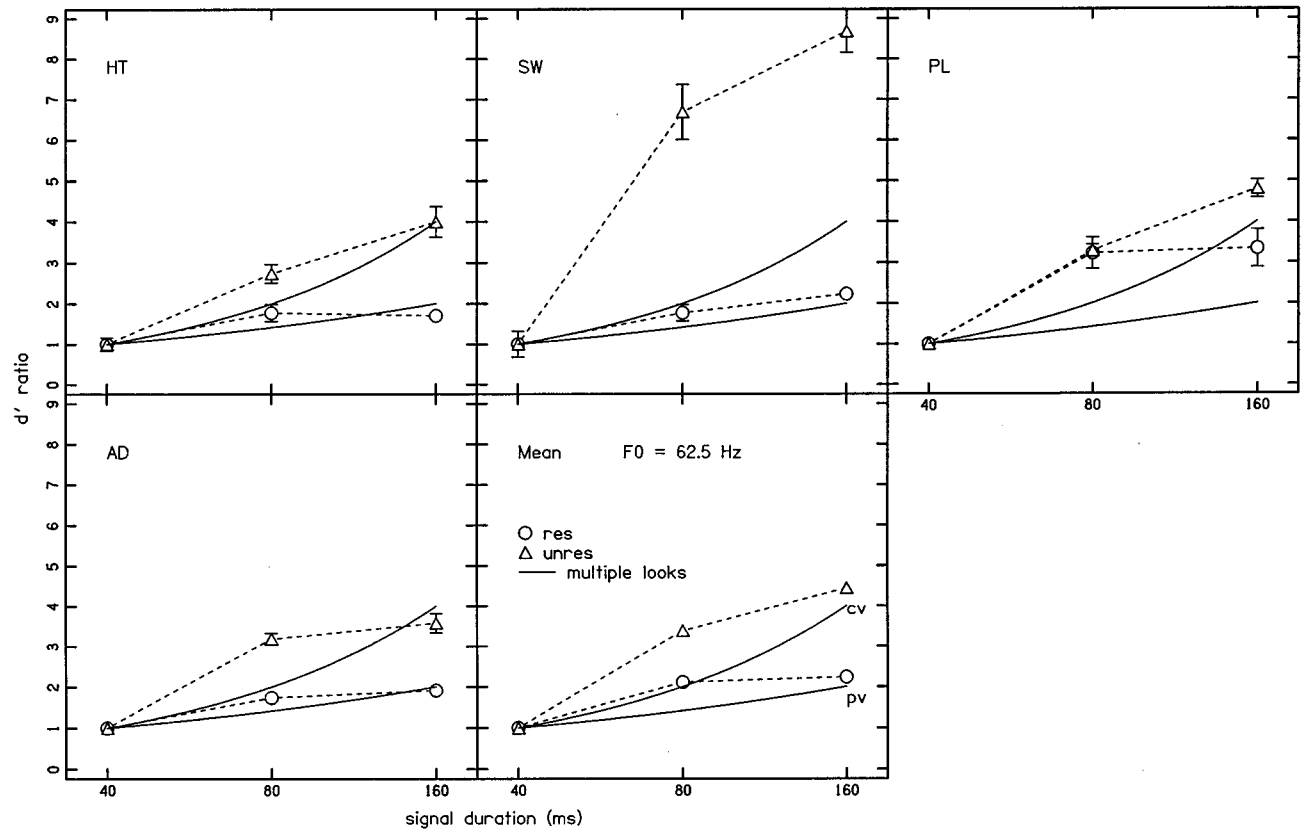


FIG. 3. As Fig. 2, except showing $F0$ discrimination performance for resolved and unresolved 62.5-Hz complexes as a function of duration. The data are plotted relative to the d' value for a 40-ms complex.

the $F0$ of the complex, with a greater duration effect for low $F0$'s.

III. INTERIM DISCUSSION

The results from experiments 1A and 1B are in agreement with those of Plack and Carlyon (1995): The improvement in $F0$ discrimination with increasing tone duration is greater for unresolved than for resolved harmonics. In addition, the improvement with duration for the unresolved harmonics over short durations is greater than that predicted by either the peripheral-variance or the central-variance multiple-looks models. It is possible, however, that the particular assumptions of the model with respect to peripheral and central variance are incorrect while the general concept of discrete sampling is still valid. A modification of the multiple-looks model that may be able to account for the data will be considered in Sec. V A 1. First, however, it is appropriate to consider an alternative explanation for the duration effects, specifically, a mechanism based on a long integration time.

The concept of a long integration time is popular in models of energy detection as a function of duration (e.g., Plomp and Bouman, 1959). With regard to $F0$ discrimination, the approach assumes that $F0$ is computed by *analyzing together* the portion of the stimulus that falls within the integration window, rather than by selectively combining $F0$ information from a number of discrete short-duration samples, as in the multiple-looks approach. From the results of experiment 1A it would appear that, if it exists, the integration period for unresolved harmonics is between 40 and 80 ms for an $F0$ of 250 Hz, as there was little performance improvement for tone durations beyond 80 ms. The results of experiment 1B suggest that the integration period is increased for low $F0$'s.

It is not clear what quantitative improvement in d' with increasing duration is expected for tones within the integration time, since this will depend on the nature of the processing applied to the integrated information. In Sec. V A 2 it will be argued that spectral processing dependent on a continuous integration period may produce an improvement with duration even greater than that predicted by the central-variance multiple-looks model. Whatever the exact nature of this processing, however, experiment 2 was designed to test the general issue of whether $F0$ is determined using a fixed long integration period, as opposed to the combination of discrete short-duration samples in the multiple-looks model.

IV. EXPERIMENT 2: *F0* DISCRIMINATION FOR PAIRS OF TONE BURSTS AS A FUNCTION OF THE GAP BETWEEN THEM

A. Rationale

$F0$ discrimination for two 20-ms tone bursts was measured as a function of the silent interval between them. If the auditory system uses a discrete sampling mechanism similar to either version of the multiple-looks model, then performance should be both independent of gap duration, and similar to performance for one 40-ms burst. On the other hand, if the auditory system uses a long integration window, then performance should get worse as the tone bursts are separated by a duration greater then the integration period. The effect of shorter gaps will depend on the way the information within the integration period is processed. An optimal combination of information from the two bursts may result in no performance deficit compared to one 40-ms burst.

B. Stimuli

Stimulus generation was the same as in experiment 1A. The stimuli consisted of two 20-ms complexes separated by a brief gap. A schematic illustration of the stimuli is shown in the lower panel of Fig. 1. The phase of the $F0$ component was preserved between the two tone bursts so that the second burst started in the phase the first burst would have had if it had been continuous. In order to remove any possible phase cues, the gap duration was randomized by $\pm$ half a period (i.e., 2 ms) in each observation interval independently.

C. Procedure

The frequency difference measured in experiment 1A for the 20-ms complex was used in a method of constant stimuli paradigm to calculate d' for two 20-ms complexes with 5-, 10-, 20-, 40-, and 80-ms separation. In each observation interval listeners heard two tone bursts. They were required to select which interval contained the two bursts with the higher pitch. The five listeners from experiment 1A took part in this experiment.

D. Results and discussion

The d' values for the two 20-ms complexes are shown in Table III and the d' scores relative to that for one 20-ms complex are shown in Fig. 4. Continuous lines show the predictions of the multiple-looks models. The data agree very well with the predictions of the peripheral-variance multiple-looks model (pv): d' is approximately $\sqrt{2}$ times greater for two tones than for one, regardless of the resolvability of the complexes or the gap between them. For the resolved harmonics, these data (and those from experiment 1A) are consistent with the peripheral-variance multiple-looks model, using a sampling duration of less than 20 ms. The results for the unresolved harmonics from experiment 2 alone are also consistent with the peripheral-variance multiple-looks model. However, taken together with the results of experiment 1A, it is clear that neither of the multiple-looks models can account for all the unresolved data. Performance for two 20-ms unresolved bursts with a 5-ms gap (mean relative $d'=1.53$) was much worse than that for one 40-ms burst (mean relative $d'=2.85$). This difference is significant ($t=4.70$, $df=4$, $p<0.01$).

The results for the unresolved harmonics are, at face value, inconsistent with a fixed long integration time, since discrimination at short gaps is much worse than that for the continuous 40-ms tone, and discrimination does not deteriorate at long gaps, which would be expected if the tones were separated by a greater duration than the integration time. Of course, it could be that the integration time is greater than the 120 ms necessary to capture both tone bursts for the 80-ms gap, but this seems unlikely given the results of experiment

TABLE III. The d' scores for the pairs of complex tones for the five gap durations in experiment 2. (The error values in parentheses were calculated using the same method as in Table I.)

Gap duration (ms)	LW	HT	JB	AD	SW	Mean	s.d.
			Unresolved harmonics				
5	1.01 (0.06)	1.14 (0.14)	0.88 (0.11)	0.86 (0.04)	0.53 (0.18)	**0.88**	0.23
10	0.73 (0.15)	1.07 (0.10)	0.47 (0.13)	0.87 (0.04)	0.52 (0.26)	**0.73**	0.25
20	0.95 (0.08)	1.53 (0.23)	0.61 (0.13)	1.06 (0.04)	0.66 (0.27)	**0.96**	0.37
40	0.90 (0.09)	1.15 (0.14)	0.75 (0.07)	0.74 (0.12)	0.47 (0.30)	**0.80**	0.25
80	0.74 (0.05)	1.13 (0.10)	0.63 (0.12)	0.81 (0.04)	0.52 (0.19)	**0.77**	0.23
			Resolved harmonics				
5	0.67 (0.09)	0.94 (0.11)	1.51 (0.11)	1.49 (0.20)	2.02 (0.32)	**1.33**	0.53
10	0.73 (0.05)	0.61 (0.14)	1.77 (0.12)	1.76 (0.24)	1.77 (0.10)	**1.33**	0.60
20	0.71 (0.10)	0.79 (0.08)	1.36 (0.05)	1.66 (0.26)	1.84 (0.13)	**1.27**	0.51
40	0.63 (0.07)	1.00 (0.06)	1.60 (0.13)	1.65 (0.23)	1.75 (0.06)	**1.33**	0.49
80	0.36 (0.10)	0.69 (0.11)	1.42 (0.16)	1.77 (0.21)	1.71 (0.12)	**1.19**	0.63

1A, which are suggestive of an integration time of less than 80 ms. A similar effect is seen in energy detection experiments, where signal threshold decreases with duration at a rate of approximately 8–10 dB per decade for signal durations up to several hundred milliseconds (Carlyon *et al.*, 1990; Florentine *et al.*, 1988; White, 1995). This is more than the 5 dB per decade predicted by a simple multiple looks model using optimum combination of independent observations, although the predictions are closer if the peripheral variance or sample weighting is assumed to vary with time relative to onset (see Viemeister and Wakefield, 1991 and Sec. V A 1). However, pairs of clicks are only integrated when they are separated by less than about 10 ms (Viemeister and Wakefield, 1991), and detectability falls to the predictions of the peripheral-variance multiple looks model for separations greater than this.

V. GENERAL DISCUSSION

A. Multiple looks, integration, and resetting

The results presented here confirm that there is a much greater effect of duration on $F0$ discrimination for unre-

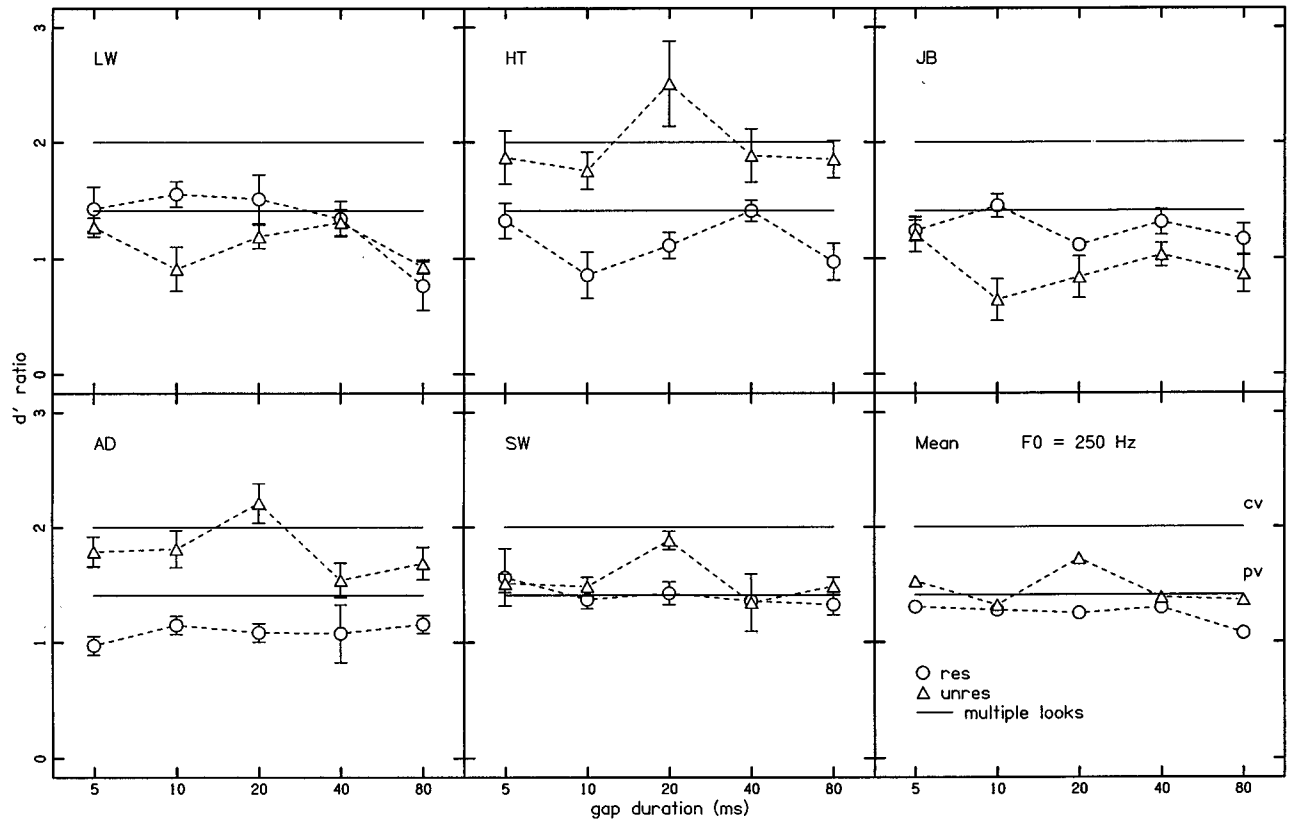


FIG. 4. The results of experiment 2, showing $F0$ discrimination performance for pairs of resolved and unresolved 250-Hz complexes as a function of the gap between the complexes in each pair. The data are plotted relative to the d' value for a single 20-ms complex. The predictions of the central-variance and peripheral-variance versions of the multiple looks model are denoted by cv and pv, respectively.

solved than for resolved harmonics. The data for the resolved harmonics in experiment 2, using gapped stimuli, agree with those for the continuous stimuli in experiments 1A and 1B. Both are broadly consistent with a peripheral-variance multiple-looks mechanism using a short sampling time of less than 20 ms for a 250-Hz $F0$, and of less than 40 ms for a 62.5 Hz $F0$. There is some indication at 250 Hz that performance falls below the model's predictions at long durations. This may reflect a difficulty in processing multiple samples optimally.

The data for unresolved complexes are much more difficult to interpret. For continuous stimuli, neither version of the *standard* multiple-looks model (with a constant peripheral or central source of variance) can account for the data, although it should be emphasized that this model is probably an oversimplification. It will be seen that a modified version of the model can have more success. Three alternative explanations of the results for the unresolved harmonics will now be discussed in detail.

1. Multiple looks with weighted samples

Viemeister and Wakefield (1991) suggested that a large effect of duration could be obtained within the context of the multiple-looks approach by assuming that the samples were weighted such that those near stimulus onset were given less weight than those later on. Such a strategy would be advantageous if there were a greater peripheral variance at stimulus onset. Physiological measurements have indicated that neural *firing rate* is more variable and mean response is greater at onset (Relkin and Pelli, 1987; Teich and Khanna, 1985; Young and Barta, 1986) although it is not clear if neural *synchrony* (i.e., accuracy of phase locking), possibly the relevant variable here, is more variable at onset.

Such a scheme can easily account for the results of experiment 2 if it is assumed that each 20-ms burst elicits a separate onset response. In this case performance should be worse than that for one 40-ms burst, because the high detectability of the latter depends on the final portion of the stimulus. Also, if the main component of variance is peripheral, which it would be if it were determined by the neural onset response, then the $\sqrt{2}$ improvement in d' for two tone bursts compared to one is also predicted, since the model becomes equivalent to the peripheral-variance multiple-looks model described in the introduction.

Despite these successes, the model is unsatisfactory because it cannot explain why the duration dependence for the 250-Hz unresolved complex is different from that for the 62.5-Hz unresolved complex or for the resolved complexes. There would appear to be no good reason why the variance at onset is greater (or extends over a longer portion of the stimulus) for the 62.5-Hz unresolved complex, but is less for both the 62.5- and 250-Hz resolved complexes. So although the model cannot be rejected outright, it does not seem to be applicable generally.

2. Long-term spectral analysis

Another possibility is that the pitch mechanism for unresolved harmonics uses a continuous long integration time, with a duration greater than 40 ms, and that performance depends on some sort of spectral analysis of the integrated information. Since the inner hair cells in the cochlea rectify the input stimulus, and since the refractory period of auditory nerve fibers effectively produces a low-pass filtering of the input, the net effect of the peripheral auditory system is to produce a very rough representation of the envelope of the incoming waveform, particularly at frequencies above 5 kHz where there is no phase locking to the fine structure (see Palmer and Russell, 1986). This being the case, $F0$ may be obtained by performing a spectral analysis of the envelope of the interacting harmonics, which beat together at $F0$. It is a simple mathematical principle that the longer the duration of the sampling window for the analysis, the ''sharper'' the resulting spectrum. This may translate into an improved discrimination threshold (see Srulovicz and Goldstein, 1983). It is also possible that the spectral analysis within the integration window may be interpreted in terms of the autocorrelation function (Meddis and Hewitt, 1991a, b). An integration time is inherent in the definition of autocorrelation; the process requires that a certain duration of signal be autocorrelated with itself. The longer the autocorrelation duration, the more accurate the estimate of periodicity.

To provide a more rigorous evaluation of these ideas, a version of the model developed by Meddis and colleagues (Meddis and Hewitt, 1991a, b) was implemented. The basic model consists of four successive processing stages: (i) simulation of the outer and middle ear transfer functions; (ii) single-channel ''gammatone'' auditory filtering (a center frequency of 6.5 kHz was used); (iii) inner hair cell simulation; and (iv) auditory nerve spike generation using a stochastic generator. The output of the model is a simulation of the action potentials generated in an auditory nerve fiber in response to the stimulus. In the present implementation, the activity was summed over 1000 presentations. A more detailed description of the model is provided in Appendix A.

The inputs to the model were the same digital representations of the stimuli used in experiments 1A and 2 that were presented to listeners during the actual experiments. A 250-Hz $F0$ was used, filtered between 5500 and 7500 Hz. Tone durations of 20 and 40 ms were used in addition to a condition with two 20-ms bursts separated by a 5-ms silent interval. The top panel of Fig. 5 shows the input waveforms and the output of the model. It can be seen from the neural activity patterns that the model predicts strong phase locking to the waveform envelope in all three conditions.

The lower panels of Fig. 5 illustrate the spectral information in the neural activity patterns in two ways: As long-term Fourier spectra, and as autocorrelations computed within an 80-ms rectangular window. The autocorrelation functions have been staggered for clarity, although the y-axis baseline is 0 in each case. Both the spectra and the autocorrelations show the 250-Hz spectral component clearly (in the latter case, it is reflected in a 4-ms periodicity). Considering first the Fourier spectra, it can be seen that there is a much broader peak around 250 Hz for the 20-ms tone than for the 40-ms tone. This suggests that $F0$ discrimination will be superior for the longer tone. The two 20-ms tone bursts separated by a 5-ms gap have a very similar spectrum to that for

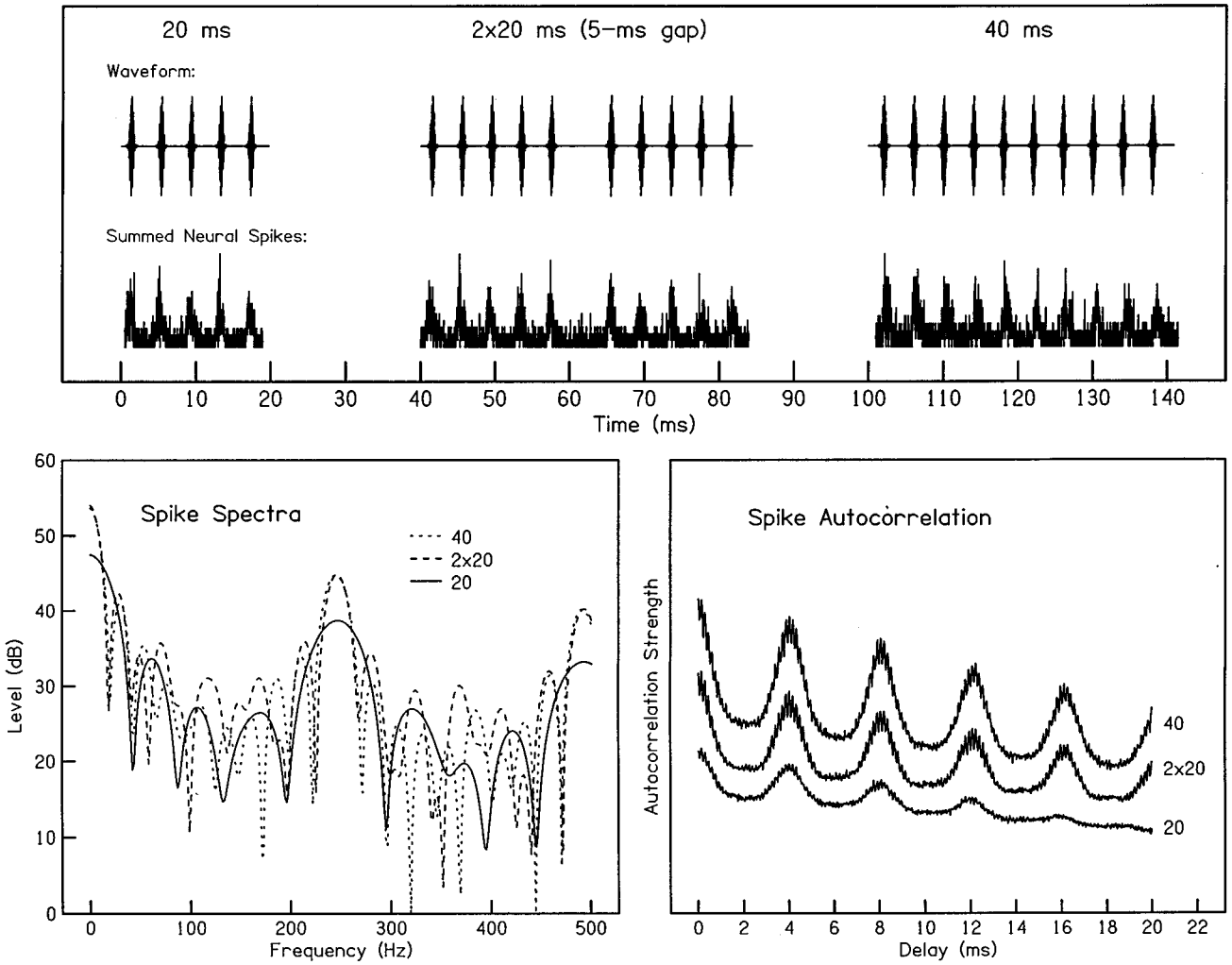


FIG. 5. The top panel shows the waveforms of three stimuli from experiments 1A and 2 with the harmonics of a 250-Hz $F0$ filtered between 5500 and 7500 Hz. Below each waveform is the corresponding output of the auditory-nerve simulation described in the text, consisting of the neural spikes summed over 1000 presentations. The lower panels show the long-term spectra and autocorrelation functions generated from these neural activity patterns.

the continuous 40-ms tone. Although the gap does produce some spectral splatter, the sharpness of the spectrum around the 250-Hz component is relatively unaffected by the gap. On the basis of the spectral information in the spike train, it would be expected that discrimination performance for the two 20-ms bursts would be similar to that for the 40-ms burst. In fact, d' was much greater for the latter.

A similar pattern is seen for the autocorrelation functions. The ratio of the autocorrelation strength at a 4-ms delay to that at a 2-ms delay (i.e., the relative height of the peak at 4 ms) is 1.31 for the single 20-ms tone burst, 1.52 for the two 20-ms bursts, and 1.54 for the 40-ms burst. The lower relative peak size for the 20-ms burst is caused by the integration of a greater proportion of spontaneous neural activity (due to the 80-ms sampling window). It is not clear what the appropriate discrimination statistic is here (for example, how many peaks in the autocorrelation function are taken into consideration when determining $F0$) but it can be said that the 5-ms gap has little effect on the autocorrelation function, which is inconsistent with the results of experiment 2.

Although the evaluation presented here is not quantitative, it seems reasonable to conclude from the relatively poor performance for the gapped stimuli in experiment 2 that the auditory system is not performing a long-term spectral analysis on these stimuli. In addition, the fact that the performance improvement from one 20-ms burst to two 20-ms bursts is almost exactly that predicted by the peripheral-variance multiple-looks model is strongly suggestive of discrete sampling rather than long-duration spectral analysis.

3. Long-term integration with resetting

One way of rescuing the long-term integration/spectral analysis account in the face of the seemingly contradictory results from experiment 2, is to assume that the integration window for unresolved harmonics is not fixed but is able to reset itself when it encounters a discontinuity in the tone. In other words, when the complex tone is continuous the mechanism may use an integration window of around 80 ms, but any discontinuity of, say, 5 ms or greater, causes the system to switch to a short-duration, discrete sampling approach. Such a flexible approach would be useful in the environment, where tones are constantly changing in $F0$ or being interrupted by other sounds. The auditory system's

task should be to derive a single pitch only when it is confident that the $F0$ is stable and coming from a single source. A varying $F0$ should be tracked using short samples in order to follow, for example, the intonation contour of speech. An interrupted sound may be interpreted as a change of source, which should also warrant a separate pitch estimate. It could be that the gaps used in experiment 2 are taken as evidence that the source has changed and the auditory system responds by starting a separate pitch estimate, losing the benefits of long integration, and reverting to a short-sampling-duration multiple-looks strategy to perform the task.

A similar resetting mechanism was proposed by Hafter and Buell (1990) in a study of the adapted binaural system. Interaural time differences (ITDs) were measured for trains of clicks which were either evenly spaced with the same interclick interval or divided into two groups by a longer (5-ms) interclick interval in the middle of the click train. The thresholds for the click train with the gap were lower than for the continuous train indicating that recovery from binaural adaptation had taken place. This resetting effect was also seen when 5-ms bursts of wideband noise or a pure tone were presented in the middle of the train of clicks.

4. Summary

It is clear that the present data are not sufficient to distinguish entirely these three possibilities, which incorporate the two very different approaches of discrete sampling and long integration. At this stage, a long-integration time with resetting appears to provide a reasonable compromise between the two extremes which is also a plausible strategy in response to the dynamic pitch stimuli in the environment.

B. Two pitch mechanisms?

The main differences between the resolvability groups are consistent with the findings of Plack and Carlyon (1995). In addition, it has been shown that the pitch mechanism for the resolved harmonics is broadly consistent with a peripheral-variance multiple-looks model. The implication is that the mechanism for the resolved harmonics uses a combination of short duration samples (less than 20 ms at 250 Hz) to estimate $F0$. For the unresolved harmonics, two broad explanations for the effects of duration have been proposed, both of which seem to imply a different processing strategy compared to that for the resolved harmonics. If the integration time is the same as that for the resolved harmonics then the multiple-looks strategy must be different (Sec. V A 1). If, as seems more likely, the integration time is different, then this implies at least some distinction between the processing of resolved and unresolved harmonics with the same $F0$.

So why does the pitch mechanism for resolved harmonics only use a short integration time? One reason may be that the auditory system does not devote additional resources to a task if the performance gains are not useful. The accuracy with which the $F0$ of a resolved complex can be determined is high even when a short sample is used, so there may not be a perceptual advantage to increasing the integration time. In terms of the hypothesis of Plack and Carlyon (1995), the additional accuracy may come from the increased information rate that can be obtained by using the frequencies of the individual resolved harmonics to help derive $F0$. Alternatively, there may be some ''central limitation'' to performance that imposes a hard limit on the accuracy of pitch estimates, reducing the utility of performance improvements at a more peripheral level.

The hypothesis that the duration effects increase with task difficulty is consistent with the poorer performance, and greater temporal integration, that is seen in the results for the 62.5-Hz unresolved complex (experiment 1B) compared to those for the higher repetition-rate 250-Hz unresolved complex (experiment 1A). Pure-tone frequency discrimination is also less accurate, and more dependent on duration, as frequency is reduced (Moore, 1973). It should be noted, however, that adaptive thresholds for the 40-ms, 62.5-Hz complexes in experiment 1B were roughly independent of resolvability, suggesting that absolute performance may not be the appropriate criterion in every case.

VI. CONCLUSIONS

The main findings of the present experiments were as follows:

(i) The improvement in $F0$ discrimination as a function of duration is greater when the complex is composed only of unresolved harmonics than when it is composed only of resolved harmonics.

(ii) The improvement in performance with increasing duration for an unresolved complex is greater (over short durations) than that predicted by a multiple-looks mechanism with either a stable peripheral (before sampling) or central (after combining samples) source of variance.

(iii) Two 20-ms tone bursts separated by a gap produce performance compared to one 20-ms burst similar to that predicted by the peripheral-variance multiple-looks model, irrespective of gap duration or whether the harmonics are resolved or unresolved.

(iv) The data for the resolved harmonics are broadly consistent with a short sampling window of 20 ms or less.

(v) The data for the unresolved harmonics may reflect a pitch mechanism that uses a long-integration time which is reset in response to discontinuities in the stimulus.

ACKNOWLEDGMENTS

The authors are very grateful to Professor Ray Meddis and colleagues for supplying the LUTEAR software used in the simulation. The authors also thank Dr. Robert Carlyon, Dr. Armin Kohlrausch, and an anonymous reviewer for very helpful criticisms of an earlier draft of this manuscript. The first author is supported by a grant from the Medical Research Council (UK) and the second author by a Royal Society University Research Fellowship.

APPENDIX A

The auditory simulation described in Sec. V A 2 was implemented using the ''LUTEAR'' software developed by Hewitt, Meddis, O'Mard, and Shackleton (Meddis, 1986, 1988; Meddis and Hewitt, 1991a, b; Meddis *et al.*, 1990; O'Mard and Meddis, 1997). Version 1.5.2 of the software was used in the simulation, running on a Silicon Graphics UNIX-based workstation. The basic model has been described in detail elsewhere and only a overview will be provided here. The names of the model stages listed below are the corresponding module names in the software.

(i) ''PreEmphasis_Bandpass.'' The first stage of the model simulates the transfer characteristics of the outer ear and middle ear (Pickles, 1988). The filter is second order bandpass, with 3-dB cutoffs of 450 and 8500 Hz.

(ii) ''BasilarM_GammaT.'' The waveform is then filtered through a fourth-order linear gammatone filter as developed by Patterson and colleagues (Patterson *et al.*, 1988). The filter approximates the spectral characteristics of the auditory filter. Although there is some evidence (Kohlrausch and Sander, 1995) that a gammatone filter does not approximate basilar-membrane phase, this was not thought to be an important consideration here.

(iii) ''IHC_Meddis86.'' This stage simulates the action of the inner hair cell and converts the filtered waveform into a time-varying sequence of ''spike probabilities,'' i.e., the probabilities of a spike being produced post-synaptically in the corresponding auditory nerve fiber (Meddis, 1986, 1988; Meddis *et al.*, 1990). The simulation assumes that this probability is linearly related to the amount of neurotransmitter in the synaptic cleft. The amount of transmitter release depends on the membrane permeability (which is determined by the amplitude of the driving waveform) and the quantity of available transmitter (which is determined by the rate of replenishment).

(iv) ''AuditoryN_SpikeGen.'' The final stage converts the spike probabilities into spike events using a simple random-number generator. It incorporates a refractory period of 1 ms that limits the proximity of each two consecutive spike events. In the simulation, this stage was repeated 1000 times and the outputs summed to provide an ''average'' neural activity pattern, which is, perhaps, a representation of the combined activity across a range of fibers.

With the exception of the gammatone center frequency (6.5 kHz) and the number of summations for the spike patterns (1000) the parameters for each stage of the model are the ''default'' parameters provided with the software. The LUTEAR software can be downloaded on the internet (http://www.essex.ac.uk/psychology/hearinglab/).

APPENDIX B

The waveforms used in the experiment were filtered after gating, so that any spectral splatter produced by the rectangular gating was not audible. However, it is still pertinent to ask whether the abrupt gating used in these experiments had any affect on the pattern of results. Therefore two listeners took part in an experiment with similar conditions to experiment 1A except that the tone durations included 5-ms ramps. An adaptive procedure was used to measure thresholds for a 20-ms complex including 5-ms raised-cosine onset and offset ramps. The threshold values were used in a method of constant stimuli as before to measure d' for 20-, 40-, 80-, and 160-ms complex durations inclusive of 5-ms ramps. As can be seen in Table BI, the data for subject LW are not different from those data collected with rectangular gating and presented in Table I. For subject AD, the improvement in performance from 20 to 40 ms is greater here than it was in experiment 1A. This is not altogether surprising, since if the ramps reduced the effective duration of the complex by a constant amount, the effect would proportionally be much greater for the 20-ms duration. Overall the pattern of data is similar to that for the unramped stimuli, with the improvement in performance with duration being much greater for the unresolved complexes.

TABLE BI. The d' scores for four durations of resolved and unresolved complex tones with 5-ms onset and offset ramps included in the total tone duration. The experiment is described in Appendix B.

Resolved	20 ms	40 ms	80 ms	160 ms
LW	0.63	0.84	1.16	1.32
AD	0.82	1.68	2.89	2.85
Unresolved				
LW	0.68	1.42	1.8	2.18
AD	0.22	1.81	3.53	3.44

APPENDIX C

It is realised that some of the d' scores for the large tone durations, particularly for the unresolved complex, may have been limited by ceiling effects. For example, the d' score for subject AD for the unresolved 160-ms complex was 3.43 (see Table I) which is equivalent to more than 99 percent correct. Therefore a control condition was run, in which the adaptive threshold was measured for a 40-ms signal and d' measured for 40-, 80-, and 160-ms complexes with an $F0$ of 250-Hz bandpass filtered between 5500 and 7500 Hz. Five listeners took part, four of whom had been in experiment 1A. The adaptive thresholds for the 40-ms complex were 2.60%, 3.71%, 3.77%, 2.91%, and 2.51% for subjects HT, LW, SW, PL, and AD. The d' scores for these signals are shown in Table CI. The mean improvement in performance between

TABLE CI. The d' scores for three unresolved complex durations for the experiment described in Appendix C.

Complex duration (ms)	LW	HT	AD	SW	PL	Mean	s.d.
40	0.5	0.9	0.64	1.13	0.32	0.70	0.32
80	1.08	1.17	0.93	1.36	0.39	0.99	0.37
160	1.22	1.46	1.18	1.39	0.53	1.16	0.37

40 and 160 ms (1.66 increase) is similar to that for experiment 1A (1.47 increase) suggesting that ceiling effects did not influence the pattern of results greatly.

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.

Carlyon, R. P. (**1997**). ''The effects of two temporal cues on pitch judgments,'' J. Acoust. Soc. Am. **102**, 1097–1105.

Carlyon, R. P., Buus, S., and Florentine, M. (**1990**). ''Temporal integration of trains of tone pulses by normal and by cochlearly impaired listeners,'' J. Acoust. Soc. Am. **87**, 260–268.

Florentine, M., Fastl, H., and Buus, S. (**1988**). ''Temporal integration in normal hearing, cochlear impairment and impairment simulated by masking,'' J. Acoust. Soc. Am. **84**, 195–203.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hearing Res. **47**, 103–138.

Goldstein, J. L. (**1973**). ''An optimum processor theory for the central formation of the pitch of complex tones,'' J. Acoust. Soc. Am. **54**, 1496–1516.

Green, D. M., and Swets, J. A. (**1996**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Hafter, E. R., and Buell, T. N. (**1990**). ''Restarting the adapted binaural system,'' J. Acoust. Soc. Am. **88**, 806–812.

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.

Kohlrausch, A., and Sander, A. (**1995**). ''Phase effects in masking related to dispersion in the inner ear. II. Masking patterns of short targets,'' J. Acoust. Soc. Am. **97**, 1817–1829.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Meddis, R. (**1986**). ''Simulation of mechanical to neural transduction in the auditory receptor,'' J. Acoust. Soc. Am. **79**, 702–711.

Meddis, R. (**1988**). ''Simulation of auditory-neural transduction: Further studies,'' J. Acoust. Soc. Am. **83**, 1056–1063.

Meddis, R., and Hewitt, M. J. (**1991a**). ''Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification,'' J. Acoust. Soc. Am. **89**, 2866–2882.

Meddis, R., and Hewitt, M. J. (**1991b**). ''Virtual pitch and phase sensitivity of a computer model of the auditory periphery: II Phase sensitivity,'' J. Acoust. Soc. Am. **89**, 2883–2894.

Meddis, R., Hewitt, M. J., and Shackleton, T. M. (**1990**). ''Implementation details of a computational model of the inner hair-cell/auditory-nerve synapse,'' J. Acoust. Soc. Am. **87**, 1813–1818.

Moore, B. C. J. (**1973**). ''Frequency difference limens for short-duration tones,'' J. Acoust. Soc. Am. **54**, 610–619.

Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing* (Academic, New York), 4th ed.

Moore, B. C. J., and Glasberg, B. R. (**1983**). ''Suggested formulae for calculating auditory-filter bandwidths and excitation patterns,'' J. Acoust. Soc. Am. **74**, 750–753.

Moore, B. C. J., and Glasberg, B. R. (**1987**). ''Formulae describing frequency selectivity as a function of frequency and level and their use in calculating excitation patterns,'' Hearing Res. **28**, 209–225.

Moore, B. C. J., and Glasberg, B. R. (**1990**). ''Frequency discrimination of complex tones with overlapping and non-overlapping harmonics,'' J. Acoust. Soc. Am. **87**, 2163–2177.

Moore, B. C. J., and Rosen, S. M. (**1979**). ''Tune recognition with reduced pitch and interval information,'' Q. J. Exp. Psychol. **31**, 229–240.

Moore, B. C. J., Glasberg, B. R., Shailer, M. J. (**1984**). ''Frequency and intensity difference limens for harmonics within complex tones,'' J. Acoust. Soc. Am. **75**, 550–561.

O'Mard, L. P., and Meddis, R. (**1997**) ''Computer exploration of the auditory system with LUTEAR,'' Br. J. Audiol. **31**, 125–126.

Palmers, A. R., and Russell, I. J. (**1986**). ''Phase-locking in the cochlear nerve of the guinea pig and its relation to the receptor potential of inner hair-cells,'' Hearing Res. **24**, 1–15.

Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1988**). *Spiral Vos Final Report, Part A: The Auditory Filterbank* (MRC-APU, Cambridge).

Perdersen, C., and Elberling, C. (**1972**). ''Temporal integration of acoustic energy in normal hearing persons,'' Acta Oto-Laryngol. **74**, 398–405.

Pickles, J. O. (**1988**). *An Introduction to the Physiology of Hearing* (Academic, London), 2nd ed.

Plack, C. J., and Carlyon, R. P. (**1995**). ''Differences in frequency modulation detection and fundamental frequency discrimination between complex tones consisting of resolved and unresolved harmonics,'' J. Acoust. Soc. Am. **98**, 1355–1364.

Plomp, R. (**1964**). ''The ear as a frequency analyzer,'' J. Acoust. Soc. Am. **36**, 1628–1636.

Plomp, R. (**1967**). ''Pitch of complex tones,'' J. Acoust. Soc. Am. **41**, 1526–1533.

Plomp, R. (**1976**). *Aspects of Tone Sensation* (Academic, London).

Plomp, R., and Bouman, R. (**1959**). ''Relation between hearing threshold and duration for pure tones,'' J. Acoust. Soc. Am. **31**, 749–758.

Relkin, E. M., and Elli, D. G. (**1987**). ''Probe tone thresholds in the auditory nerve measured by two-interval forced choice procedures,'' J. Acoust. Soc. Am. **82**, 1679–1691.

Ritsma, R. J. (**1967**). ''Frequencies dominant in the perception of the pitch of complex sounds,'' J. Acoust. Soc. Am. **42**, 191–198.

Rose, J., Hind, J., Anderson, D., and Brugge, J. (**1968**). ''Patterns of activity in single auditory nerve fibers in the squirrel monkey,'' in *Hearing Mechanisms in Vertebrates*, edited by A. de Reuek and J. Knight (Churchill, London).

Rosenberg, K. M. (**1990**). *Statistics for Behavioral Sciences* (Wm. C. Brown, New York).

Schouten, J. F. (**1940**). ''The residue and the mechanism of hearing,'' Proc. K. Ned. Akad. Wet. **43**, 991–999.

Schouten, J. F. (**1970**). *The Residue Revisited* (Sijhoff, Lieden, The Netherlands).

Schouten, J. F., Ritsma, R. J., and Cardozo, B. L. (**1962**). ''Pitch of the residue,'' J. Acoust. Soc. Am. **34**, 1418–1424.

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.

Srulovicz, P., and Goldstein, J. L. (**1983**). ''A central spectrum model: Synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum,'' J. Acoust. Soc. Am. **73**, 1266–1276.

Teich, M. C., and Khanna, S. M. (**1985**). ''Pulse-number distribution for the neural spike train in the cat's auditory nerve,'' J. Acoust. Soc. Am. **77**, 1110–1128.

Terhardt, E. (**1974**). ''Pitch, consonance, and harmony,'' J. Acoust. Soc. Am. **55**, 1061–1069.

Viemeister, N. F., and Wakefield, G. H. (**1991**). ''Temporal integration and multiple looks,'' J. Acoust. Soc. Am. **90**, 858–865.

White, L. J. (**1995**). ''Auditory temporal integration,'' D. Phil., University of Sussex.

Young, E. D., and Barta, P. E. (**1986**). ''Rate responses of auditory-nerve fibers to tones in noise near masked threshold,'' J. Acoust. Soc. Am. **79**, 426–442.

"Central" auditory gap detection: A spatial case

D. P. Phillips, S. E. Hall, I. A. Harrington, and T. L. Taylor
Department of Psychology, Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada

(Received 9 September 1997; accepted for publication 6 January 1998)

Normal listeners were tested for their temporal auditory gap detection thresholds using free-field presentation of white-noise stimuli delivered from the left (L) and right (R) poles of the interaural axis. The noise bursts serving as the leading and trailing markers for the silent period were presented in either the same (LL,RR) or different (LR,RL) auditory locations. The duration of the leading marker was a second independent variable. Gap thresholds for stimuli in which the markers had the same location were low, and usually were independent of the duration of the leading marker. Gap thresholds for the LR and RL conditions were longer. These gap thresholds were sensitive to the duration of the leading marker, and increased as the leading marker duration decreased. This finding is consistent with the hypothesis that a relative timing operation mediates gap detection when the markers activate different perceptual channels. The present data suggest that this timing process can operate on perceptual channels emerging from central nervous system processing. © *1998 Acoustical Society of America.* [S0001-4966(98)02604-6]

PACS numbers: 43.66.Mk, 43.66.Rq [JWH]

INTRODUCTION

Gap detection is one of the standard paradigms for measuring auditory temporal resolution (e.g., Fitzgibbons and Wightman, 1982; Moore, 1985; Moore and Glasberg, 1988). In most studies, the listener is presented with two relatively long (hundreds of ms) noise bursts, one of which contains a brief (a few ms) silent period or ''gap'' at its temporal midpoint, and the task of the listener is to indicate which stimulus contains the gap. Some of the factors that influence the length of the shortest detectable gap (''gap threshold'') have been identified by experiments that have varied different properties of the gap detection stimulus. We thus know that gap thresholds are poorest for very narrow-band and very low-amplitude signals (Eddins *et al.*, 1992; Moore *et al.*, 1993). Explanations of these findings probably lie in the fidelity of the representation of the gap in the neural channels leaving the cochlea, in the number of these channels activated by the stimulus, and in any confusion between envelope fluctuations in the carrier and the inserted gap.

Phillips *et al.* (1997) distinguished two different temporal processes involved in the performance of a gap detection task. They argued that in the traditional gap detection paradigm, the ''temporal'' operation executed is actually a *discontinuity detection* within the perceptual channel activated by the sound. This is because information about the stimulus perturbation constituting the gap can be carried by any single perceptual (or neural) channel representing the stimulus spectrum. In contrast, if the sound marking the leading edge of the gap activates different peripheral neurons from those marking the trailing edge of the gap, then the temporal operation necessarily becomes a *relative timing* of the offset of activity in the perceptual channel representing the leading marker, and the onset of activity in the channel representing the trailing marker. This relative timing operation must be performed centrally, because the auditory periphery contains no lateral connections capable of executing the relative timing operation. Phillips *et al.* (1997) showed that gap thresholds for these ''between-channel'' conditions were grossly larger than those for traditional ''within-channel'' gap detection in the same listeners, irrespective of whether the perceptual channels were defined by stimulus spectrum or by stimulus laterality (ear). Moreover, between-channel, but not within-channel, gap thresholds were highly sensitive to the duration of the leading noise marker: gap thresholds varied inversely with leading element duration for leading marker durations less than about 30 ms. The within-channel and between-channel timing operations thus appear to have different temporal acuities and different sensitivities to stimulus parameters.

The purpose of the present study was to investigate the generality of these phenomena. Specifically, we sought to determine whether the relative timing operation described by Phillips *et al.* (1997) extends to stimulus conditions activating perceptual channels defined by the spatial location of a free-field white-noise source. This point is nontrivial. In the earlier experiments of Phillips *et al.* (1997), activation of different perceptual channels was achieved by selective stimulation of different peripheral neurons. This activation was done by presenting the leading and trailing markers of the gap signal to different frequency sectors of the cochlea within an ear, or to the same frequency sector but in different ears. In the present study, every stimulus element is presented to the whole cochlea on both sides, and what defines the activated channels is the spatial percept arising from computations of interaural disparities in the level, arrival time, and spectra of the sounds at the two eardrums. These computations are performed by higher central auditory nervous system structures.

I. METHODS

A. Participants

Six adult listeners (four female; age range 22–41 years) served as participants in the study. All of the listeners had free-field noise detection thresholds within laboratory norms.

 0001-4966/98/103(4)/2064/5/$10.00

Four of the listeners (the authors) had significant prior experience in auditory psychophysical studies. The remaining listeners had little or no experience in auditory perception studies, and were paid for their participation.

B. Stimuli

Stimulus presentation and data acquisition were controlled by a custom-programmed Macintosh IIfx computer. All stimuli were wideband (20 kHz) noise bursts, generated by a white-noise source (Coulbourn Instruments). The stimuli were linearly ramped with 0.5-ms rise–fall times, including those defining the gaps. The gated stimuli were passed through a passive attenuator, a power amplifier, and were transduced by Minimus speakers (frequency response 0.14–20.0 kHz). The speakers were suspended (at ear level) from a 70-in.-diam, circular track mounted on the ceiling of an Eckel sound-attenuating room (95-in. wide, 83-in. deep, and 80-in. high, internal measurements). The ceiling and walls of this booth had been lined with 1-in.-thick acoustic foam. This foam had 0.5-in.-deep (''egg carton'') ripples to improve sound absorption.

C. Procedure

Each subject was seated below the center of the circular track from which the speakers were hung, and was therefore in the direct field of the sound. Noise detection thresholds were obtained by the method of limits using 300-ms noise bursts. Thresholds were measured separately for the subject facing a speaker directly ahead (0° azimuth) and for speaker locations at $+/-90°$ azimuth (the interaural axis). The noise detection thresholds for the lateral speaker positions were about 3 dB lower than those for speakers in the frontal location, and were close to symmetrical on the two sides. All subsequent testing was done with the speakers located in the lateral positions. Stimulus levels were set at 30-dB sensation level (i.e., 30 dB above detection threshold for the lateral speaker position) for the 300-ms stimulus. This had the consequence that although the spectrum levels of the stimuli were constant, very short leading markers in the gap stimulus were subjectively a few dB quieter than the others. All stimulus elements were, however, audible to all the listeners. Moreover, a previous study in which these loudness differences were compensated for suggested that the compensation was inconsequential to the patterns of data obtained (Phillips *et al.*, 1997).

The general design of the study, and the remaining methodology, was the same as that used by Phillips *et al.* (1997). Each stimulus trial consisted of two sequences of noise bursts, separated by an interval of 300 ms. One of these sequences, the gap ''signal,'' contained a detectable silent period between the leading and trailing element. The other sequence was the ''standard,'' which included a 1.0-ms gap between the leading and trailing markers. This small gap was below the gap detection threshold for each listener. (This undetectable gap in the standard was included so that the within-channel conditions employed stimuli with the same gating transients as the between-channel stimuli.) The duration of the trailing marker in both the signal and the standard was always 300 ms. The duration of the leading marker was an independent variable. In separate blocks of trials, the leading marker durations of the signal and standard were set at 5, 10, 30, 50, 100, or 300 ms.

The second independent variable was the spatial location of the leading and trailing marker in the signal and standard. In separate blocks of trials, the leading and trailing markers of both the signal and standard were presented through the left speaker or through the right speaker; these were the ''within-channel'' (LL, RR) conditions. In separate blocks of trials, ''between-channel'' conditions were evaluated by presenting the leading and trailing markers of the signal and standard from the left and right speakers, respectively, or the reversed order from the right and left speakers. The corresponding conditions are designated as LR and RL, respectively. The general experimental plan was thus a four-by-six, within-subjects design, in which gap thresholds were determined for each of six leading-marker durations, in each of four spatial configurations.

Subjects were seated in the booth facing a visual fixation point at 0° azimuth. Gap thresholds for every stimulus condition were obtained in blocks of two-interval, two-alternative forced-choice trials. On each trial, the order of standard and signal presentation was randomized by the computer, and the task of the subject was to indicate by button press whether the first or second observation interval contained the signal. The intertrial interval was customized for each listener, but was usually in the range from 400 to 800 ms, plus a 500-ms warning light. The listener tracked the gap duration across the block of trials in a two-down/one-up adaptive paradigm (after Levitt, 1971). Briefly, the starting duration of the gap in the signal stimulus was large (around 15–20 ms for the within-channel conditions, and around 80–120 ms for the between-channel conditions). After every two successive correct responses, the size of the gap was decreased by a factor of 1.2, and after every incorrect response, the size of the gap was increased by the same factor. This two-down/one-up procedure was continued until there had been 14 reversals in the direction of change of the gap duration. The gap threshold for that block of trials was taken as the mean gap duration across the last ten trials that produced reversals in the adaptive step. Listeners typically took 40–70 trials to reach this criterion. Over a period of 3 to 4 weeks, every stimulus condition was tested repeatedly until gap thresholds were relatively stable. Subjects typically took six blocks at each stimulus condition to achieve asymptotic performance. Data collected were mean durations of minimal detectable gaps over the last four blocks for each factorial combination of the independent variables.

II. RESULTS

Figure 1 shows the mean data separately for each subject. Gap threshold has been plotted as a function of the leading marker duration. The parameter in each panel is the channel condition, with lines joining rectangles representing the data for the between-channel conditions. Every subject had within-channel gap thresholds in the range from 2 to 3 ms for almost all values of the leading marker duration, and irrespective of speaker location. The only exceptions to this

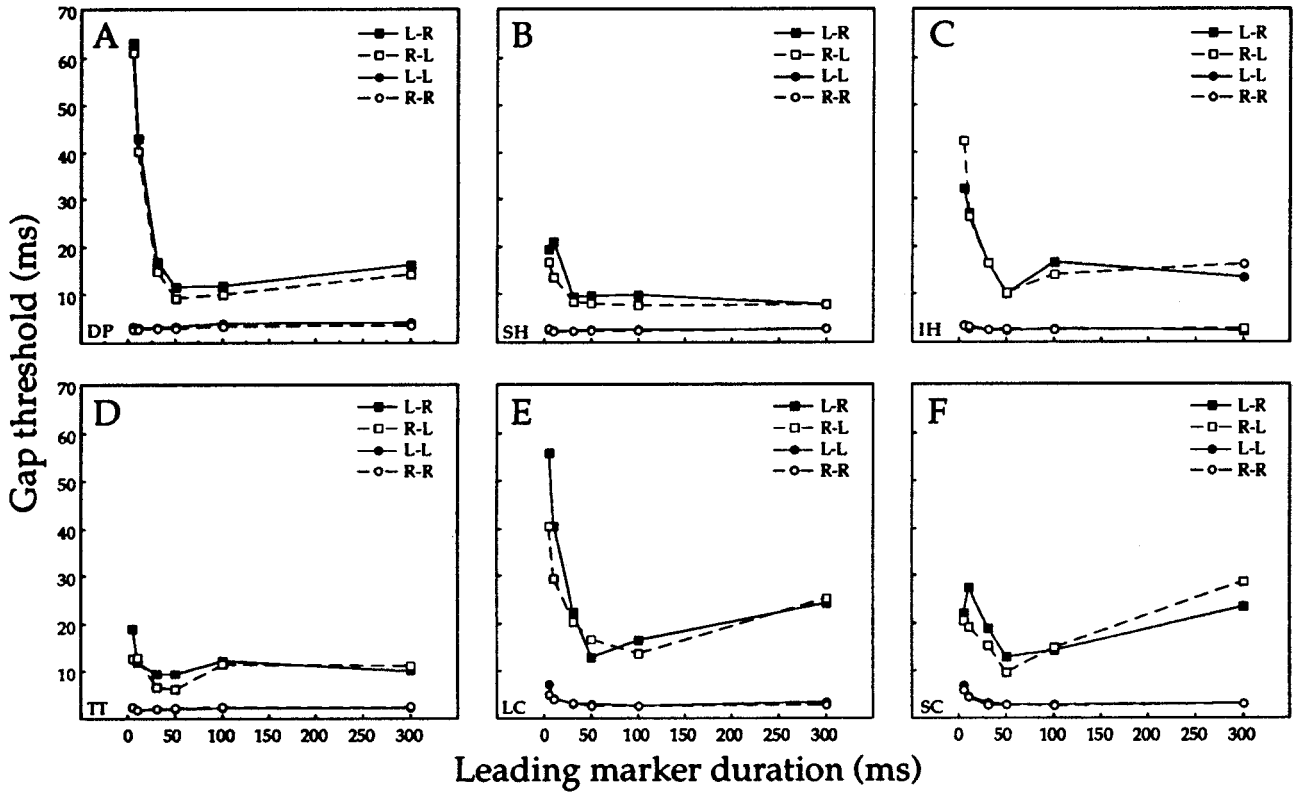


FIG. 1. Panels (a)–(f) show the mean gap thresholds separately for each listener in the study. For each subject, gap threshold has been plotted as a function of the duration of the leading noise marker. The parameter is the sequence of auditory hemifields in which the leading and trailing noises occurred.

generality were the two inexperienced listeners [Fig. 1(e) and (f)], whose LL and RR gap thresholds rose to between 4 and 6 ms for the shortest leading marker durations.

Every listener had longer gap thresholds for the between-channel conditions than for the within-channel conditions. Indeed, the gap thresholds for the two sets of conditions were completely nonoverlapping. Every listener also showed an effect of leading marker duration on the between-channel gap thresholds. There were quite large individual differences in the size of this effect [compare Fig. 1(a) and (d)]. Two subjects [Fig. 1(a) and (e)] had gap thresholds as high as 40–60 ms for the shortest leading markers, while two other subjects, for the same stimulus conditions, had gap thresholds closer to 15 or 20 ms [Fig. 1(b) and (d)]. Two subjects showed relatively large gap thresholds (>20 ms) for the 300-ms leading marker LR and RL conditions [Fig. 1(e) and (f)].

The grand mean data (with standard errors) are shown in Fig. 2. Inspection of these data confirms that performance on the two sets of within-channel conditions (LL,RR) were very similar, as was the performance on the two sets of between-channel conditions (LR,RL), but that the gap thresholds for between-channel conditions were grossly higher than those for within-channel conditions. For the shortest leading-marker durations (5 ms), the between-channel gap thresholds were 32.4 and 35.3 ms for RL and LR conditions, respectively.

The data in Fig. 2 were analyzed in two, two-way analyses of variance, one each for the within- and between-channel data sets. In each analysis, leading marker duration and channel mode (LL and RR for the within-channel case, and LR and RL for the between-channel data) were within-subjects factors. For the within-channel data, the only significant effect was leading marker duration [$F(5,20)=2.93$, $p=0.03$]. This was a small effect (Fig. 2), likely due to two listeners [Fig. 1(b) and (f)]. Neither the main effect of stimu-

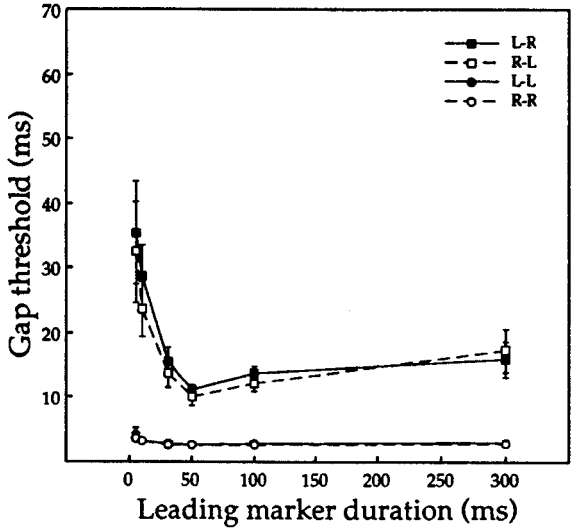


FIG. 2. Grand mean data (with bars indicating standard errors). Gap thresholds were averaged across all listeners, and plotted after the fashion of the data in Fig. 1.

lus laterality (LL vs RR, $p>0.2$), nor the interaction between laterality and leading marker duration ($p>0.1$), approached statistical significance.

For the between-channel data, the only significant factor was again leading marker duration [$F(5,20)=8.27$, $p=0.0001$]. Note in Fig. 2 that the size of this effect was much greater for the between-channel conditions than for the within-channel ones, even in proportional terms. The main effect of order (LR vs RL) approached statistical significance [$F(1,5)=5.17$, $p=0.07$], but the interaction between the two main factors did not ($p>0.2$).

III. DISCUSSION

Phillips *et al.* (1997) distinguished between discontinuity detection ''within'' a perceptual channel, and relative timing of activity ''between'' perceptual channels, as mechanisms mediating gap detection performance. They argued that the former mechanism operates when the stimulus elements marking the onset and offset of the silent gap are contained within the same perceptual channel, and that the latter mechanism is used when the markers are contained within different channels. To ensure that the between-channel timing mechanisms were engaged, Phillips *et al.* (1997) used leading and trailing markers designed to activate different populations of peripheral neurons. They did this by using markers that activated different sectors of the cochlea within an ear, or by using spectrally identical markers presented to different ears via headphones.

The present study used white-noise stimuli in free-field space. The left (L) and right (R) perceptual channels so activated arise from central neural computations executed on the outputs of both cochleas, and thus not from activation of different populations of peripheral neurons. This study found the same general patterns of behavior as those described by Phillips *et al.* (1997). In most listeners, within-channel gap thresholds were short, and were not grossly sensitive to the duration of the leading marker [Fig. 1(a)–(d)]. Between-channel gap thresholds were always long, and were sensitive to the duration of the leading marker [Fig. 1(a)–(f)]. These data suggest that the relative timing operation required for between-channel gap detection can operate on perceptual channels emerging from central processing, as well as on those arising from stimulation of different peripheral neurons.

The absolute values of gap thresholds obtained in the present study were also in accord with previous observations. The grand mean within-channel gap thresholds obtained using wideband noise in this study were in the range from 2.7 to 4.4 ms, which is in agreement with data from other laboratories (Fitzgibbons, 1983; Penner, 1977; Florentine and Buus, 1984; Plomp, 1964; Shailer and Moore, 1983).

For the between-channel conditions, comparisons are more difficult, because few authors have challenged the gap detection mechanisms spatially as we have described. For stimulus settings with long leading and trailing narrow-band markers, Phillips *et al.* (1997) reported gap thresholds between 10 and 20 ms, and the present thresholds for wideband noise mostly fall in the same range. Other authors have used different markers to delimit silent periods, and have reported that the resulting gap detection thresholds are significantly elevated above those for stimulus settings that would be construed as ''within-channel'' in the present study (Fitzgibbons *et al.*, 1974; Formby and Forrest, 1991; Formby *et al.*, 1996; Penner, 1977). Phillips *et al.* (1997) noted that the grand mean gap thresholds for between-channel cases with short (5 ms) leading markers were consistently in the range from 32 to 35 ms (their experiments 2–4). In the present study, the grand mean thresholds for the comparable stimulus conditions were 32.4 and 35.3 ms (RL and LR, respectively).

The similarity of the present thresholds to those of our earlier study is important for a second reason. In the present experiment, stimuli were presented in the free-field, and although the booth in which the experiments were done had its interior surfaces lined with sound-absorbing foam, we had no direct measurements of acoustic reverberation in the room. This raises the possibility that the precedence effect might have played some role in determining the present gap thresholds. However, the data in Fig. 2 are strikingly similar to those obtained in the earlier experiment by Phillips *et al.* (1997, their experiment 4), in which leading and trailing markers were delivered to the left and right ears using headphones. The similarities in both the pattern of data, and the absolute gap thresholds, in the two studies suggests that there is no need to appeal to precedence effect mechanisms to explain the present data.

The present data raise the question of precisely which stimulus cues were used by the current listeners to assign a spatial locus to each stimulus marker, and thus, more generally, of exactly what constitutes a ''perceptual channel.'' The most likely candidate cues are interaural disparities in the spectra, levels, and arrival times of the markers, since these disparities are orderly functions of source eccentricity (see Middlebrooks and Green, 1990; Phillips and Brugge, 1985, for review). However, even if the listeners had, in practice, executed the task monaurally, it would remain the case that differences in the spatial (or other) quality of the percepts generated from the monaural information would constitute the activation of different ''perceptual channels.'' Thus construed as the spatiotemporal patterns of neural activity that generate percepts, different ''perceptual channels'' could be engaged by activation of a single set of peripheral neurons if the relative levels or timing of excitation across that population by two stimuli were different. In this regard, there is clear behavioral evidence for ''channels'' for direction of frequency modulations (Tansley and Regan, 1979; Regan and Tansley, 1979). In that instance, the same peripheral neurons are activated by upward and downward modulations, but the ''channels'' emerge from central computations executed on the peripheral inputs.

Admittedly, this is a very broad conceptualization of perceptual channels, since we usually think of perceptual channels as elements tuned to varying portions of some continuum (e.g., critical bands in the frequency domain). Indeed, between-channel gap detection experiments provide some evidence for tuning of perceptual channels in the frequency domain (Formby *et al.*, 1996; Phillips *et al.*, 1997). There is, however, increasing evidence that this may be too narrow a formulation of the channel concept (Wright and Dai, 1994;

Phillips *et al.*, 1997), *viz.*, that the properties of the channel may depend on the perceptual task required of the listener.

ACKNOWLEDGMENTS

This work was supported by grants from NSERC of Canada to D.P.P. T.L.T. was supported by postgraduate scholarships from NSERC and the Killam Foundation. Thanks are also due to Dr. Craig Formby and Dr. John Grose for their very helpful commentaries on a previous version of this article.

Eddins, D. A., Hall, J. W., and Grose, J. H. (**1992**). ''The detection of gaps as a function of frequency region and absolute noise bandwidth,'' J. Acoust. Soc. Am. **91**, 1069–1077.

Fitzgibbons, P. J. (**1983**). ''Temporal gap detection in noise as a function of frequency, bandwidth and level,'' J. Acoust. Soc. Am. **74**, 67–72.

Fitzgibbons, P. J., Pollatsek, A., and Thomas, I. B. (**1974**). ''Detection of temporal gaps within and between perceptual tonal groups,'' Percept. Psychophys. **16**, 522–528.

Fitzgibbons, P. J., and Wightman, F. L. (**1982**). ''Gap detection in normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **72**, 761–765.

Florentine, M. F., and Buus, S. (**1984**). ''Temporal gap detection in sensorineural and simulated hearing impairments,'' J. Speech Hear. Res. **27**, 449–455.

Formby, C., and Forrest, T. G. (**1991**). ''Detection of silent temporal gaps in sinusoidal markers,'' J. Acoust. Soc. Am. **89**, 830–837.

Formby, C., Sherlock, L. P., and Forrest, T. G. (**1996**). ''An asymmetric roex filter model for describing detection of silent gaps in sinusoidal markers,'' Aud. Neurosci. **3**, 1–20.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Middlebrooks, J. C., and Green, D. M. (**1990**). ''Directional dependence of interaural envelope delays,'' J. Acoust. Soc. Am. **87**, 2149–2162.

Moore, B. C. J. (**1985**). ''Frequency selectivity and temporal resolution in normal and hearing-impaired listeners,'' Br. J. Audiol. **19**, 189–201.

Moore, B. C. J., and Glasberg, B. R. (**1988**). ''Gap detection with sinusoids and noise in normal, impaired, and electrically stimulated ears,'' J. Acoust. Soc. Am. **83**, 1093–1101.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1993**). ''Detection of temporal gaps in sinusoids: Effects of frequency and level,'' J. Acoust. Soc. Am. **93**, 1563–1570.

Penner, M. J. (**1977**). ''Detection of temporal gaps in noise as a measure of the decay of auditory sensation,'' J. Acoust. Soc. Am. **61**, 552–557.

Phillips, D. P., and Brugge, J. F. (**1985**). ''Progress in neurophysiology of sound localization,'' Annu. Rev. Psychol. **36**, 245–274.

Phillips, D. P., Taylor, T. L., Hall, S. E., Carr, M. M., and Mossop, J. E. (**1997**). ''Detection of silent intervals between noises activating different perceptual channels: Some properties of ''central'' auditory gap detection,'' J. Acoust. Soc. Am. **101**, 3694–3705.

Plomp, R. (**1964**). ''Rate of decay of auditory sensation,'' J. Acoust. Soc. Am. **36**, 277–282.

Regan, D., and Tansley, B. W. (**1979**). ''Selective adaptation to frequency-modulated tones: Evidence for an information processing channel selectively sensitive to frequency changes,'' J. Acoust. Soc. Am. **65**, 1249–1257.

Shailer, M. J., and Moore, B. C. J. (**1983**). ''Gap detection as a function of frequency, bandwidth, and level,'' J. Acoust. Soc. Am. **74**, 467–473.

Tansley, B. W., and Regan, D. (**1979**). ''Separate auditory channels for unidirectional frequency modulation and unidirectional amplitude modulation,'' Sensory Proc. **3**, 132–140.

Wright, B. A., and Dai, H. (**1994**). ''Detection of unexpected tones with short and long durations,'' J. Acoust. Soc. Am. **95**, 931–938.

Inter-individual differences in binaural detection of low-frequency or high-frequency tonal signals masked by narrow-band or broadband noise

Leslie R. Bernstein, Constantine Trahiotis, and Erika L. Hyde
Surgical Research Center, Department of Surgery (Otolaryngology), and Center for Neurological Sciences, University of Connecticut Health Center, Farmington, Connecticut 06030

(Received 25 September 1997; revised 1 December 1997; accepted 4 January 1998)

Detection thresholds for either 500-Hz tones or 4-kHz tones were measured for a group of 19 listeners utilizing the interaural configurations NoSo and NoSπ. Both broadband (100–8500 Hz) noises and narrow-band (50-Hz-wide) noises served as maskers. In addition, direct measures of the listeners' sensitivities to changes in interaural temporal differences (ITDs) and interaural intensitive differences (IIDs) were measured using 400-Hz-wide noises centered at 500 Hz or 4 kHz. A rather large range of inter-individual differences in threshold was observed for 4-kHz tonal signals masked by narrow-band noise in the NoSπ configuration. This result is consistent with several sets of data from our previous experiments conducted over more than a decade. A broad range of thresholds was also obtained for 500-Hz tonal signals masked by narrow-band noise in the NoSπ configuration. This outcome, coupled with the fact that the use of a broadband masker did not yield a large distribution of thresholds for the detection of a 500-Hz tone masked by a broad band of noise, suggests that it is the use of a narrow-band masker, per se, that results in a large range of thresholds. Statistical analyses revealed that thresholds in the NoSπ detection tasks were not highly correlated with thresholds measured in the ITD- and the IID-discrimination tasks. Nevertheless, the five listeners who were the most sensitive in the narrow-band NoSπ detection and the five listeners who were the least sensitive in the narrow-band NoSπ detection tasks were those who were the most and least sensitive, respectively, to changes in ITDs and to changes in IIDs. © *1998 Acoustical Society of America.* [S0001-4966(98)05304-1]

PACS numbers: 43.66.Pn, 43.66.Dc [DWG]

INTRODUCTION

In the course of comparing data from several experiments conducted in our laboratory concerning binaural masking-level differences (MLDs), we observed what appeared to be larger inter-individual differences for MLDs measured with high-frequency signals than for MLDs measured with low-frequency signals. All of the data were collected using ''standard'' stimulus conditions and standard psychophysical procedures. Specifically, we noticed that the range of thresholds obtained for 4-kHz tonal signals, masked by Gaussian noise and presented in the NoSπ configuration, was substantially larger than the range of thresholds obtained with the same stimulus configuration when the frequency of the signal was 500 Hz. We became especially aware of such disparities when we evaluated *group-average* thresholds measured across several different crews of three to five listeners who served in experiments spanning a decade or more. Often, the high-frequency NoSπ thresholds appeared to be drawn from two underlying populations of listeners whose average sensitivities differed by about five to ten decibels. Parallel comparisons for stimuli presented in the NoSo configuration revealed a relatively small range of thresholds for both signal frequencies.

Thus the differences in the variability of the MLDs obtained with high-frequency and low-frequency signals appeared to be associated with the processing of the binaural cues available in the NoSπ condition, rather than the processing of the diotically presented, monaurally based cues available in the NoSo condition. One potentially important difference between the high-frequency and low-frequency NoSπ conditions that we examined was the bandwidth of the masking noise. In order to obtain a sizable MLD when the frequency of the signal was 4 kHz, it was necessary to use narrow-band maskers (e.g., Zurek and Durlach, 1987; Bernstein and Trahiotis, 1992). In contrast, when the frequency of the signal was 500 Hz, a broadband masker was typically employed. Given these considerations, the greater variability in thresholds observed with the high-frequency signals could be attributed either (1) to inter-individual differences in listeners' abilities to process fine-structure-based interaural cues at low-frequencies versus their abilities to process envelope-based interaural cues at high frequencies or (2) to inter-individual differences in listeners' abilities to detect tonal signals against the background of narrow-band versus broadband masking noise.

Inter-individual differences similar to those we observed across our own studies can also be observed in data presented by Zurek and Durlach (1987). Their Fig. 3 (p. 462), shows masking-level differences (MLDs) for three listeners obtained with 250-Hz and 4-kHz tonal signals, as a function of the bandwidth of the masker. The data provide no clear answer concerning the degree to which the frequency of the signal and/or the bandwidth of the masker, per se, are responsible for the differences in thresholds across their three listeners. Koehnke *et al.* (1986) also reported inter-individual

differences for data obtained in NoSo and NoSπ tasks for stimuli centered at 500 Hz and 4 kHz (see their Table II, p. 1560). The inter-individual differences they observed with only four listeners were of sufficient magnitude that they discussed their data as being obtained from ''sensitive'' or ''insensitive'' listeners. They too found much larger inter-individual differences in NoSπ conditions than in NoSo conditions.

Because, to our knowledge, there were no data available to provide a definitive set of comparisons that would help one understand the factor(s) responsible for our observations and because we believe it is important to address inter-individual differences in binaural processing in a variety of contexts (e.g., Bernstein and Trahiotis, 1995), we designed a study employing a relatively large number of listeners so that statistical comparisons among several relevant conditions could be made. In addition to measuring masked thresholds, we also measured listeners' sensitivities to changes in interaural temporal differences (ITDs) and interaural intensitive differences (IIDs). This was done to evaluate whether any inter-individual differences observed in the NoSπ thresholds would be correlated with inter-individual differences in sensitivity to ITDs and IIDs, respectively. These are the (confounded) cues that serve as the basis for NoSπ detection which result when a tonal signal is added to masking noise.

I. EXPERIMENT

A. Procedure

1. NoSo and NoSπ detection

The noise maskers were presented for 340 ms (including 20-ms $\cos^2$ rise-decay ramps) at a spectrum level equivalent to 50 dB SPL. Signals were similarly gated, had a total duration of 300 ms, and were temporally centered within the pulsed maskers.

All stimuli were generated digitally with a sampling rate of 20 kHz via a TDT array processor. The noises were constructed using inverse FFT techniques. Tonal signals were generated in the time domain and added to noise waveforms prior to D/A conversion. All stimuli were low-pass filtered at 8.5 kHz (TDT FLT2) and presented via TDH-39 earphones (mounted in MX/41-AR cushions) to listeners seated in single-walled, IAC booths. Nineteen young adults, nine males and ten females, with no evidence or history of hearing loss served as listeners. Six of the listeners were experienced in the sense that they had participated in prior experiments concerning the processing of binaural information. The remaining 13 listeners (one of whom was the third author) were recruited from within the University of Connecticut Health Center and the majority of them were medical students.

The stimuli were presented in a two-alternative temporal forced-choice adaptive task. Each trial consisted of a warning interval (500 ms) and two 340-ms-long observation intervals separated by 450 ms. The observation intervals were marked by a visual display on a computer monitor. Feedback was provided for approximately 400 ms after the listener responded. The tonal signal was presented with equal *a priori* probability in either the first or second interval. Independent masker waveforms were presented during each and every observation interval. The level of the signal was varied adaptively in order to estimate 70.7% correct (Levitt, 1971). The initial step size for the adaptive track was 2 dB and was reduced to 1 dB after two reversals. A run was terminated after 12 reversals and threshold was defined as the average level of the signal across the last 10 reversals. On the rare occasion that the standard deviation of levels across the last ten reversals was greater than 3 dB, the run was discarded and an another estimate was obtained.

TABLE I. The order in which the 8 masking conditions were tested across the 24 experimental sessions.

Sessions	Signal frequency (Hz)	Masker bandwidth (Hz)	Interaural configuration
1–3	500	8400	NoSo
4–6	500	8400	NoSπ
7–9	4000	8400	NoSo
10–12	4000	8400	NoSπ
13–15	500	50	NoSo
16–18	500	50	NoSπ
19–21	4000	50	NoSo
22–24	4000	50	NoSπ

The conditions were presented in the same order for all listeners beginning with detection of 500-Hz tones masked by broadband noise in the NoSo and NoSπ configurations, respectively. These conditions were run first because previous research (Trahiotis *et al.*, 1990) had demonstrated that listeners reach asymptotic ''threshold'' in those conditions almost immediately in that their performance remains stable over many days. This tactic allowed the listeners to become familiar with the psychophysical task in stimulus conditions for which little or no ''learning'' is required or evident.

The ordering of the conditions is shown in Table I. Note that all of the conditions utilizing broadband maskers were run prior to running the conditions utilizing narrow-band maskers. This was done to provide experience with the psychophysical task prior to testing with the narrow-band maskers, where inter-individual differences in performance were of interest. NoSo and NoSπ conditions were interleaved and testing at 500 Hz preceded testing at 4 kHz. This ordering, rather than complete randomization was utilized so that MLDs could be measured while minimizing changes in center frequency from session to session. As also indicated in the table, each of the eight stimulus conditions was run over three consecutive experimental sessions. During each session, nine estimates of threshold were obtained. Therefore, 27 estimates of threshold were obtained before the next condition was visited. In virtually all cases, one session was completed per day. This experimental design allowed us to examine the data for learning effects and to evaluate stability in performance both within and across daily sessions.

2. ITD and IID discrimination

The same group of 19 listeners who participated in the detection experiments participated in the ITD- and IID-discrimination experiments.

a. ITD discrimination. Listeners' sensitivities to *ongoing* ITDs were measured using 400-Hz-wide bands of noise centered at either 500 Hz or 4 kHz. That bandwidth was chosen because it is sufficiently broad so as to permit excellent ITD sensitivity at both 500 Hz and 4 kHz (e.g., Klumpp and Eady, 1956; Smoski and Trahiotis, 1986; Bernstein and Trahiotis, 1994) while being sufficiently narrow so as to confine the stimulus to the two local spectral regions of interest. We did not employ 50-Hz-wide bands of noise as in the detection experiment because it is well known that sensitivity to ongoing ITDs for such narrow-band stimuli centered at 4 kHz is very poor (e.g., Nuetzel and Hafter, 1981; Bernstein and Trahiotis, 1994). Those substantial differences in sensitivity as a function of bandwidth found for ITD discrimination with stimuli centered at 4 kHz are not found for ITD discrimination at 500 Hz. They also are not found for IID discrimination for stimuli centered at 500 Hz or 4 kHz (see Klumpp and Eady, 1956; Yost and Dye, 1988; Smoski and Trahiotis, 1986; Bernstein and Trahiotis, 1995 for data permitting relevant comparisons). Thus it is only for our ITD-discrimination condition at 4 kHz that bandwidth might be expected to affect performance.

Our goal was to investigate inter-individual differences as they show themselves in tasks with parameters of the stimuli that permit, at least in some listeners, excellent performance. That is, we were interested in whether listeners who were ''sensitive''/''insensitive'' in the NoSπ detection task were likewise ''sensitive''/''insensitive'' in other tasks basic to the understanding of binaural processing. Consequently, we chose a bandwidth (400 Hz) that would provide estimates of thresholds that were expected to reveal the ''best'' discriminability for both ITD and IIDs for stimuli centered at 500 Hz and 4 kHz, respectively.

It is also the case that, in a previous study of NoSπ detection at 4 kHz (Bernstein and Trahiotis, 1992), we found that the rank ordering among the thresholds obtained from the three listeners was consistent across bandwidths of 50 Hz, 100 Hz, 200 Hz, 400 Hz, and 800 Hz, even though the mean thresholds across bandwidth spanned a range of about 15 dB. Thus we had no *a priori* reason to believe that the rank ordering of listeners in NoSπ depends on the bandwidth of the masker. This finding, along with the reasoning provided above, led us to choose a bandwidth of 400 Hz for the discrimination tasks.

The noises were presented for 300 ms including 20-ms rise-decay ramps at a spectrum level equivalent to 50 dB SPL. The noises were generated digitally as described above and ITDs were imposed by applying linear phase shifts to the discrete spectral components prior to conversion to the time domain.

The stimuli were presented in a two-cue, two-alternative forced choice task. Each trial consisted of a warning interval (500 ms) and four, 300-ms-long observation intervals separated by 400 ms. Intervals were marked by a visual display on a computer monitor. Feedback was provided for approximately 400 ms after the listener responded. On each trial, four independent tokens of the noise were presented. In the first and fourth intervals, the stimuli were presented diotically. The ongoing ITD (left-ear leading) was presented with equal *a priori* probability in either the second or the third interval. Of course, the remaining interval, like the first and fourth intervals, contained a diotic noise.

The ITD was varied adaptively in order to estimate 70.7% correct. Initially, ITDs were increased or decreased by a factor of 1.585 (i.e., a 2-dB change). After the second reversal, the factor was changed to 1.122 (i.e., a 0.5-dB change). Logarithmic step sizes were employed because the psychometric function relating percent correct to log-ITD is approximately linear (Yost *et al.*, 1971; Saberi, 1995). A run was terminated after 12 reversals and threshold was defined as the geometric mean of the ITDs that defined the last 10 reversals.

b. IID discrimination. Sensitivities to IIDs were measured with stimuli and procedures as described above for ITD discrimination. A given IID was produced by altering the level of the noise across ears symmetrically (up in one ear and down at the other) by IID/2. The step sizes within the adaptive track depended on the magnitude of the IID. For starting values of IID greater than 3.0 dB, the initial step size was 1.0 dB. For starting values of IID less than or equal to 3.0 dB, the initial step-size was 0.5 dB. Following second reversals, the former step size was reduced to 0.5 dB and the latter step size was reduced to 0.2 dB.

c. Order of testing. Threshold ITDs and IIDs were measured directly after measurement of NoSo and NoSπ detection thresholds. For noise centered at 500 Hz, sets of four estimates of threshold were obtained in the following order: ITDs, IIDs, ITDs, and IIDs. The same order was then used for noise centered at 4 kHz. In that fashion, eight estimates of threshold were obtained for each listener and condition and the final ''threshold'' was computed as the mean of the eight estimates.

B. Results and discussion

The presentation and analysis of the data will be directed toward evaluating, for each masking condition, (1) the stability of the 27 thresholds measured across the three daily sessions, (2) the inter-individual variability of the mean of the 27 thresholds, and (3) the degree to which performance in a given masking condition was correlated with performance in the other masking conditions and with performance in the ITD- and IID-discrimination tasks.

1. NoSo and NoSπ detection thresholds

In order to assess, for each condition, the *stability* of the thresholds measured across the three daily sessions, we chose to fit regression lines to the 27 (estimates)$\times$19 (listeners) data points *after removing the listener's own mean* for that condition. In this manner, variability in the data stemming from inter-individual differences in sensitivity was removed. Figure 1 contains the data obtained in all eight masking conditions. Each panel of Fig. 1 represents one masking condition and contains data-points transformed as described above. The solid line in each panel represents the best fit obtained by linear regression. The panels on the left present data obtained in the NoSo conditions; the panels on the right present data obtained in the NoSπ conditions. The top four panels present data obtained with the broadband maskers; the

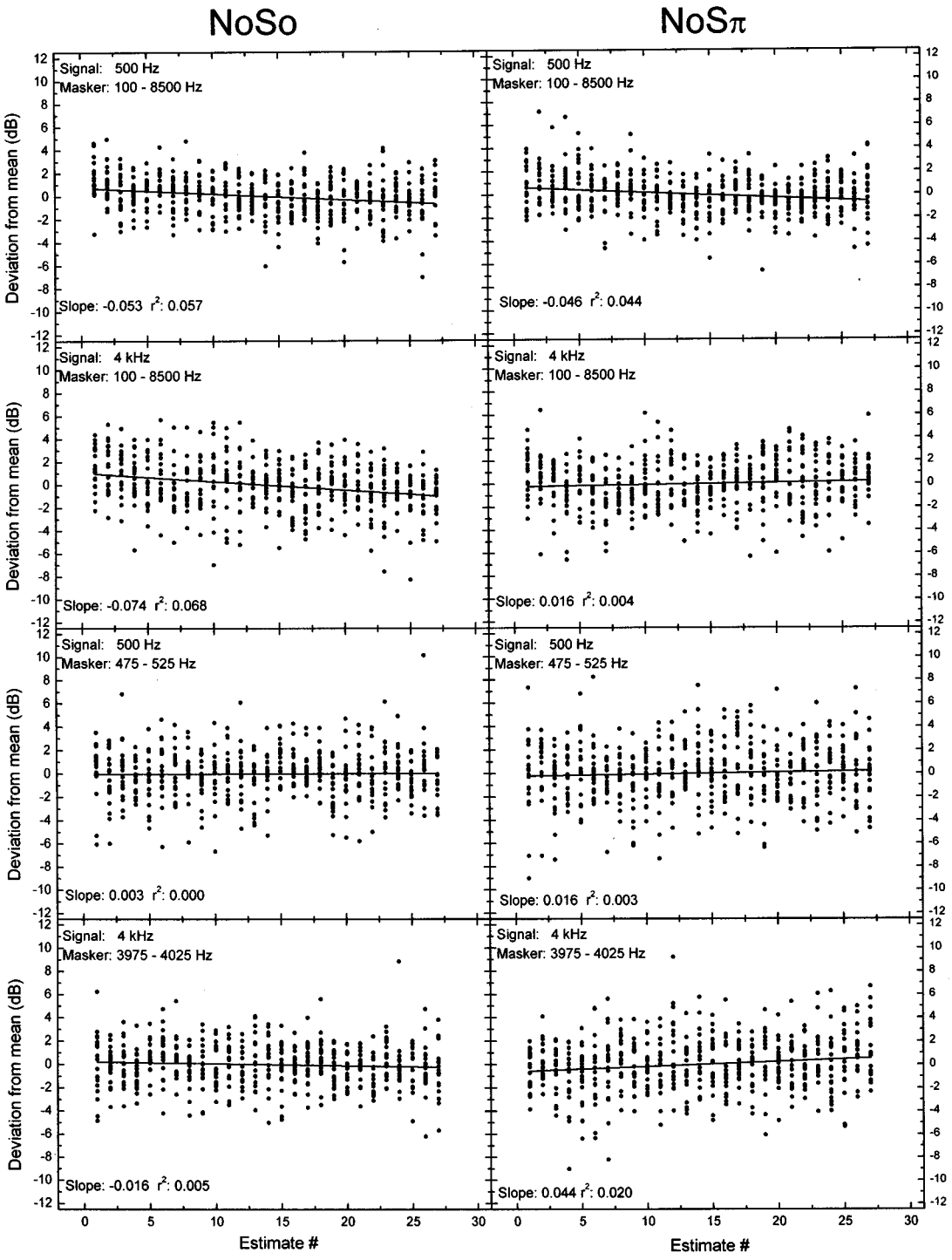


FIG. 1. Each panel contains the 27 (estimates)×19 (listeners) detection thresholds obtained in 1 of the 8 masking conditions. For for each listener, the estimates of threshold were normalized by subtracting the mean threshold for that condition. The normalized values are plotted as a function of the order in which they were obtained. Panels on the left contain the data for the NoSo conditions; panels on the right contain the data for the NoSπ conditions. The top four panels present data obtained with the broadband maskers; the bottom four panels present data obtained with the narrow-band maskers. Data obtained when the center frequency was 500 Hz are plotted in panels above panels containing the data obtained in corresponding conditions when the center frequency was 4 kHz.

bottom four panels present data obtained with the narrow-band maskers. Data obtained when the center frequency was 500 Hz are plotted in panels above panels containing the data obtained in corresponding conditions when the center frequency was 4 kHz.

The slope of the best-fitting line obtained by linear regression and the amount of variance accounted for (r^2) are indicated within each panel. Overall, these measures confirm the visual impression that there was essentially no systematic change in the estimates of threshold obtained across the 27 occasions. All of the fitted lines have a slope close to zero and the amount of variance accounted for was less than 1% for four of the conditions and was between 2% and 7% in the remaining four conditions. This means that the adjusted data

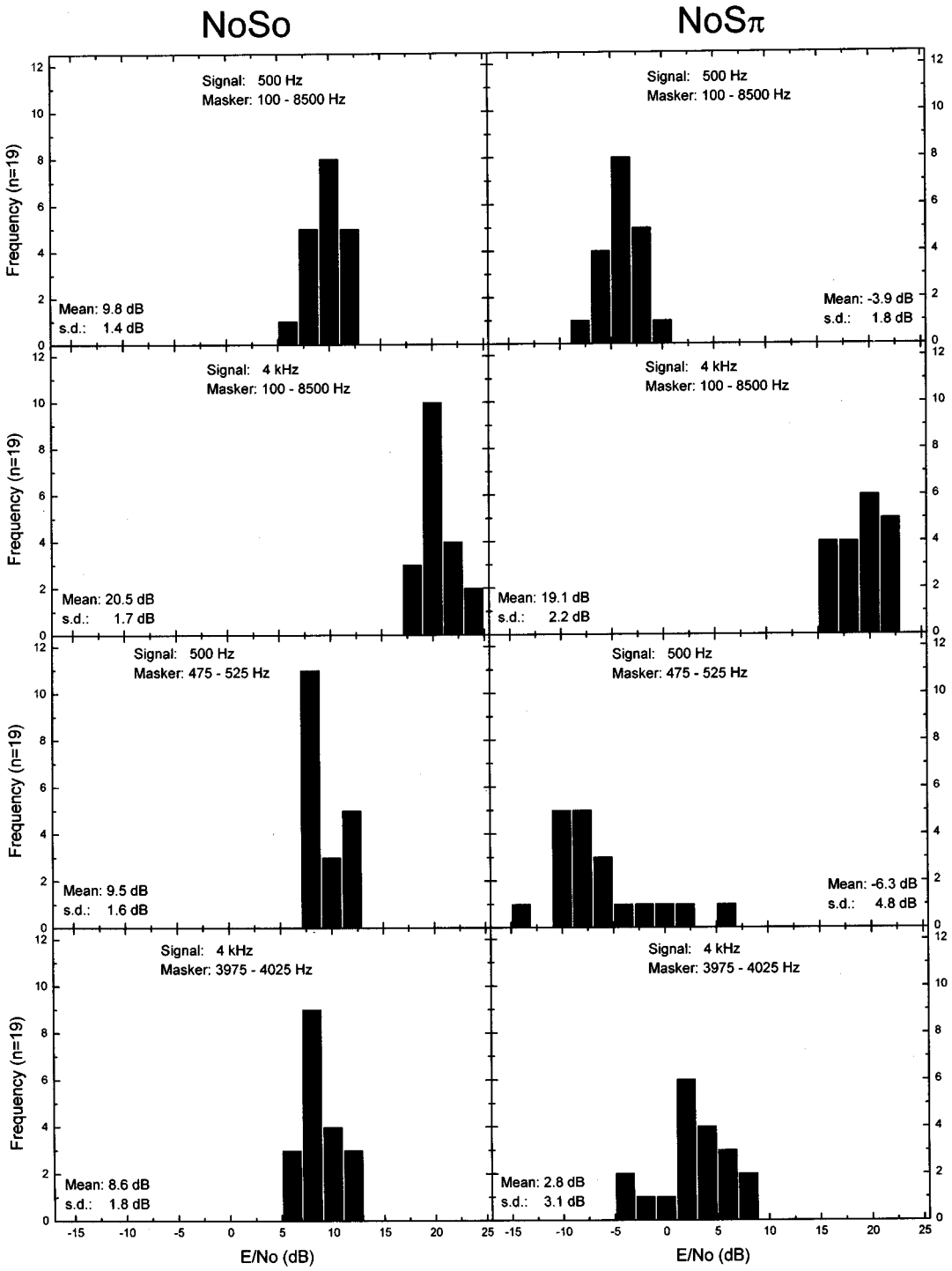


FIG. 2. Frequency distributions of obtained thresholds (E/No in dB) for each of the eight detection conditions. The ordering of the panels in the figure, with respect to the experimental conditions, is identical to that in Fig. 1. The mean and standard deviation of each distribution are indicated within each panel, respectively.

provide virtually no evidence that the estimates of threshold changed over time. This outcome is consistent with the findings of Trahiotis *et al.* (1990) who measured NoSo and NoSπ detection thresholds over several occasions using a broadband masker and a 500-Hz tonal signal. Because there was no evidence that thresholds varied systematically across occasions, the mean of all 27 estimates of threshold for each listener and condition was computed and is considered to be a valid measure of performance.

Figure 2 contains histograms representing the frequency distribution of the mean thresholds obtained from each individual and each condition. The ordering of the panels is identical to that of Fig. 1. The thresholds are presented in terms of E/No in dB, which facilitates direct comparisons between thresholds obtained with the broadband and the narrow-band maskers. For readers who wish to evaluate these thresholds in terms of S/N in dB, the conversion can be made by subtracting 33.9 dB from the E/No thresholds obtained with the

broadband masker or by subtracting 11.6 dB from the E/No thresholds obtained with the narrow-band masker.[1] The overall mean, calculated across the 19 listeners, is shown in each panel as is the standard deviation. Beginning with the data obtained with a broadband masker and a 500-Hz signal (top row), note that both NoSo and NoSπ thresholds are relatively tightly clustered around their respective means of 9.8 dB and -3.9 dB, respectively. The standard deviations of the distributions were 1.4 and 1.8 dB, respectively. The difference in mean threshold of 13.7 dB between these two conditions is the average MLD and is in accord with numerous similar measurements (e.g., Durlach and Colburn, 1978).

The second row contains thresholds obtained when the frequency of the signal was 4 kHz. Once more, the thresholds are tightly clustered about their means with the standard deviations of the distributions being 1.7 and 2.2 dB for the NoSo and NoSπ conditions, respectively. As expected from prior studies (e.g., Durlach and Colburn, 1978), the average MLD in this condition was quite small being only 1.4 dB.

The third row contains the thresholds obtained when the frequency of the signal was 500 Hz and the bandwidth of the masker was 50 Hz. The thresholds for the NoSo configuration are narrowly distributed about the mean of 9.5 dB as evidenced by a standard deviation of 1.6 dB. In contrast, the thresholds obtained with the narrow-band NoSπ configuration are much more widely distributed about their mean of -6.3 dB. The standard deviation of the NoSπ thresholds is 4.8 dB, indicating a much larger range of inter-individual differences as compared to its NoSo counterpart. The average MLD is 15.8 dB, but, as will be seen when Fig. 3 is discussed, the range of MLDs across the 19 listeners was very large.

The same trend concerning variability occurred when a 4-kHz tonal signal was masked by a 50-Hz-wide band of noise (bottom two panels). In this case, the standard deviation of the NoSo and NoSπ thresholds was 1.8 and 3.1 dB, respectively while the average MLD was only 5.8 dB.

A comparison of the ranges of detection thresholds obtained in this study with the ranges of detection thresholds obtained by Koehnke *et al.* (1986) revealed that, in general, the ranges we observed were approximately twice as large. Interestingly, the exception was the data collected in the NoSπ condition at 4 kHz. Koehnke *et al.* obtained a range of about 10 dB which is close to the 11.2-dB range that we obtained with the narrow-band masker. It is not surprising that we observed a larger range of thresholds than did Koehnke *et al.* They tested only four listeners while we tested nineteen. Koehnke *et al.*'s two listeners who were described as ''sensitive'' had ''extensive'' experience in binaural tasks. Their two ''insensitive'' listeners, one of whom had no experience in psychoacoustic tasks, were ''trained'' for 6 h in each task, prior to collection of ''formal'' experimental data. As stated earlier, only 6 of our 19 listeners had any prior experience in psychoacoustic tasks and we provided no ''practice'' prior to collection of formal data. In addition, it should be noted that Koehnke *et al.* utilized 1/3-octave bands of noise as maskers. Although such noises qualify as being ''narrow'' with respect to center frequency, Koehnke *et al.*'s maskers were broader than our ''narrow'' maskers which had a bandwidth of 50 Hz. Perhaps our use of a narrower bandwidth, in and of itself, could have contributed to the greater range of thresholds observed.

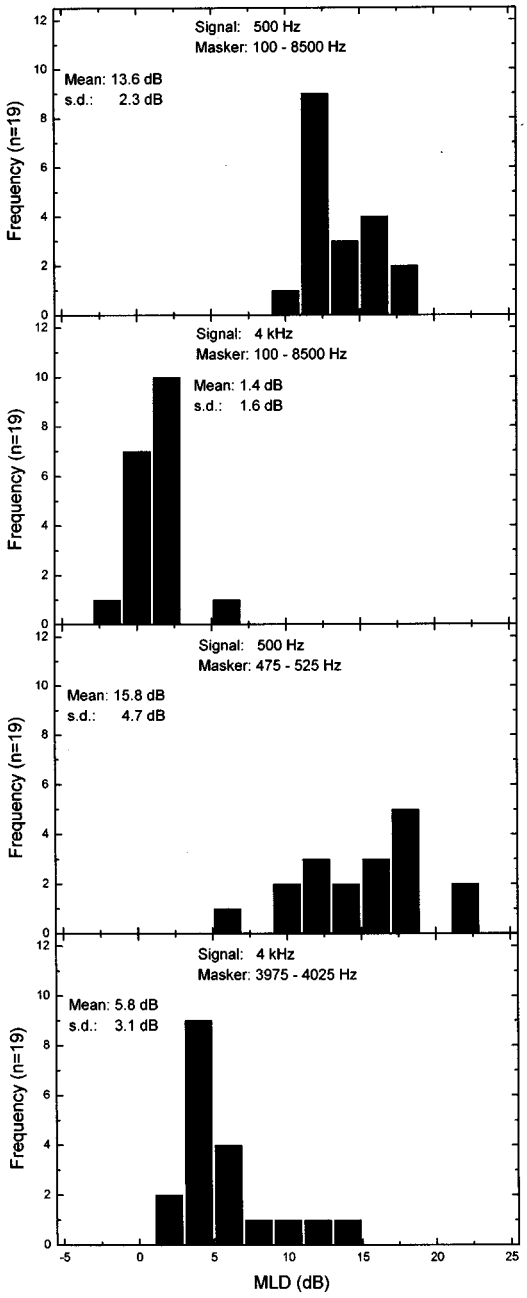


FIG. 3. Frequency distributions of the masking-level differences (MLDs) obtained for each listener in each of the four detection conditions. The ordering of the panels with respect to the center frequency and type of masker is identical to that in Figs. 1 and 2.

Figure 3 displays frequency distributions of the MLDs computed for each of the listeners in each of the four conditions. It is important to stress that these distributions reflect information concerning the *differences between each individual's NoSo and NoSπ thresholds*, information that cannot be gleaned from the distributions in Fig. 2. While the range of MLDs obtained with the broadband maskers (top two panels) is on the order of 8–9 dB, the range of MLDs obtained with the narrow-band maskers (bottom two panels) is much larger being about 18 dB when the frequency of the signal

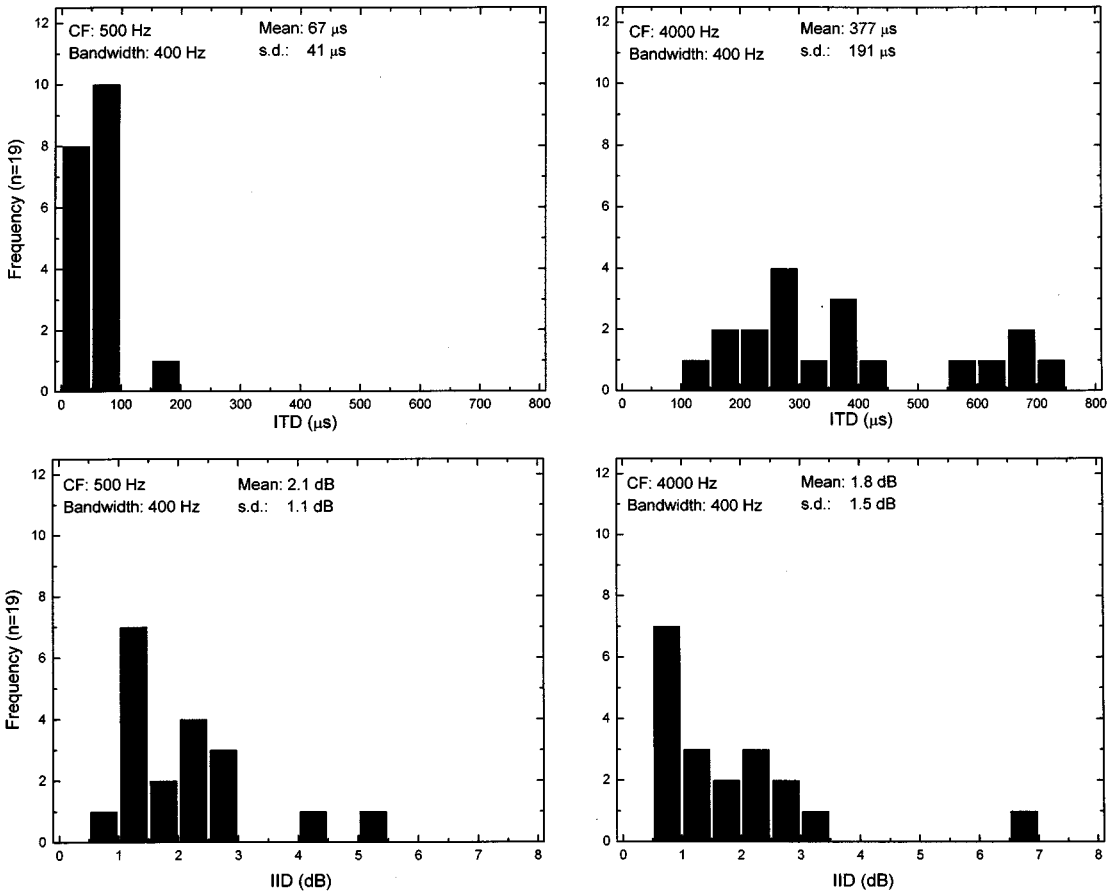


FIG. 4. Frequency distributions of the ITD (top panels) and IID (bottom panels) thresholds obtained for each listener with 400-Hz-wide noises centered at 500 Hz (left panels) and 4 kHz (right panels).

was 500 Hz and about 14 dB when the frequency of the signal was 4 kHz. It is noteworthy that the largest of the MLDs for each stimulus condition that are displayed in Fig. 3 coincide quite well with the magnitude of the MLDs that are typically reported for those stimulus conditions in laboratory investigations employing small numbers of well-practiced listeners (e.g., Durlach and Colburn, 1978; Zurek and Durlach, 1987; Bernstein and Trahiotis, 1992).

Overall, the data presented in Figs. 2 and 3 provide confirmation of observations that we made when comparing data across a variety of studies and groups of listeners. Specifically, $N_0S\pi$ detection thresholds measured with 4-kHz signals and a narrow-band masker yielded large inter-individual differences. As discussed in the Introduction, this could have occurred as a result of (1) inter-individual differences in listeners' abilities to process fine-structure-based interaural cues at low-frequencies versus their abilities to process envelope-based interaural cues at high frequencies or (2) inter-individual differences in listeners' abilities to detect tonal signals against the background of narrow-band versus broadband masking noise. The data obtained with the narrow-band masker and the 500-Hz signal, which also show large inter-individual differences, do not support the first possibility and strongly support the second possibility. That is, it appears that the factor that was responsible for the large inter-individual differences we had observed in the past was the use of narrow band maskers. This conclusion could not have been drawn by considering the literature and prior results. Rather, it required the testing of a reasonably large number of listeners in several pertinent conditions, as was done in this study.

2. ITD and IID discrimination

Figure 4 displays the frequency distributions of the ITD (top panels) and IID (bottom panels) thresholds measured with 400-Hz-wide bands of noise centered at 500 Hz (left panels) and 4 kHz (right panels), respectively. As discussed above, the threshold for each listener was computed by averaging eight estimates. The overall mean and standard deviation calculated across the thresholds for the 19 listeners is indicated at the top of each panel. Note that the threshold ITDs obtained at 500 Hz are tightly clustered and only one listener's threshold exceeded 100 μs. The mean threshold was 67 μs and the standard deviation was 41 μs. In contrast, the threshold ITDs obtained at 4 kHz are much larger, having a mean of 377 μs, and are highly variable, having a standard deviation of 191 μs. Five of the listeners were extremely insensitive and required ITDs of 600–700 μs to reach threshold. Those values of ITD are close to the largest that are produced by external sources of sound placed along the aural axis. The distributions of threshold IIDs at 500 Hz and at 4

TABLE II. Pearson product-moment correlations between the NoSπ detection thresholds obtained in the narrow-band masking conditions and the ITD- and IID-discrimination thresholds obtained at the same two center frequencies, respectively. Asterisks indicate values of correlation significantly greater than zero ($p<0.05$, one-tailed test). The values in parentheses are the part correlations between the respective detection and discrimination thresholds obtained after removing the variability common to the ITD- and IID-discrimination measures at each frequency.

NoSπ	ITD	IID
CF: 500 Hz	0.45* (0.36)	0.28 (0.10)
CF: 4000 Hz	0.45* (0.26)	0.57* (0.43*)

kHz were quite similar having mean values of 2.1 dB and 1.8 dB and standard deviations of 1.1 dB and 1.5 dB, respectively.

As was true for the detection thresholds, the ITD and IID thresholds obtained from the most sensitive listeners are comparable to those reported in a wide variety of studies conducted in many different laboratories (e.g., Klumpp and Eady, 1956; Hafter *et al.*, 1977; Yost and Dye, 1988; Bernstein and Trahiotis, 1994). Perhaps the most surprising aspect of the data in Fig. 4 is that five listeners required more than 550 μs in order to reach threshold at 4 kHz. Once again, the ranges of ITD and IID thresholds that we obtained were approximately twice those observed by Koehnke *et al.* (1986).

3. Correlations among measures of detection and discrimination of ITD and IID

The reader is reminded that the purpose of measuring ITD and IID thresholds was to evaluate whether, and the extent to which, any inter-individual differences observed in the NoSπ thresholds would be correlated with inter-individual differences in sensitivity to ITDs and IIDs. That is, we wished to determine the degree to which listeners were consistent in terms of their relative sensitivity across our NoSπ detection, ITD-discrimination, and IID-discrimination tasks. Both parametric (Pearson product-moment) and nonparametric (Spearman rank-order) correlations were computed. Because both indices led to highly similar values, only the parametric correlations, which have the greater statistical power, are reported.

The Pearson product-moment correlation between the NoSπ narrow-band thresholds measured at 500 Hz and 4 kHz was 0.55, indicating that only 30% of the variance in the thresholds was common. The Pearson product-moment correlation between the NoSπ broadband thresholds measured at 500 Hz and 4 kHz was 0.18, indicating that only 3% of the variance in the thresholds was common. This very low correlation is consistent with the relatively restricted range of thresholds obtained with the broadband maskers shown in Fig. 2.

None of the Pearson product-moment correlations between the broadband NoSπ detection thresholds and the ITD- and IID-discrimination thresholds reached statistical significance. Consequently, the discussion will focus on correlations involving the ITD and IID discrimination thresholds with the narrow-band NoSπ thresholds. Table II contains the matrix of Pearson product-moment correlations between the narrow-band NoSπ detection thresholds and the ITD- and IID-discrimination thresholds obtained at the same two frequencies, respectively. Values marked by an asterisk are those that were significantly greater than zero ($p<0.05$, one-tailed test). The values in parentheses are the part correlations between the respective detection and discrimination thresholds obtained after removing the variability common to the ITD- and IID-discrimination measures at each frequency. Part correlations were used because first-order correlations between ITD- and IID-discrimination thresholds were about 0.4 at both center frequencies. Thus the first-order correlations between NoSπ detection thresholds and each of the two discrimination thresholds could be viewed as somewhat ''contaminated'' or inflated by the common variance between the two discrimination measures. Computation of the part correlations allowed us to determine how well performance on each type of discrimination task (ITD and IID), with the influence of performance on the other task removed, related to performance in the NoSπ task.

For stimuli centered at 500 Hz, the correlation between NoSπ detection thresholds and ITD-discrimination thresholds was 0.45 and was statistically significant. The correlation between NoSπ detection thresholds and IID-discrimination thresholds was 0.28 and was not statistically significant. The part correlations between the same variables were 0.36 and 0.10, respectively, neither of which was significant.

For NoSπ detection at 4 kHz, both the correlations of 0.45 and 0.57 with ITD and IID discrimination, respectively, were significant. The part correlations between the same variables were 0.26 and 0.43, respectively. Only the latter was significant. Figure 5 contains the scatterplots corresponding to the individual, first-order correlations shown in Table II. Visual inspection of the scatterplots does not suggest that strong associations, be they linear or curvilinear, exist between detection and discrimination thresholds.

The rather low first-order correlations coupled with the substantial decreases in the amounts of variance accounted for after removing the variance common to the measures of ITD and IID indicate that, by and large, inter-individual differences were not consistent across our three binaural tasks. In accord with the suggestion of one of the anonymous reviewers, (1) we computed the *multiple* correlation between the performance on the two discrimination tasks and the NoSπ tasks and (2) recomputed the first-order correlations after a logarithmic transformation of the ITD thresholds.

The multiple correlation at 500 Hz was 0.46 indicating that 21% of the variance in performance in the NoSπ task could be accounted for by *combining* the ITD- and IID-discrimination measures. The multiple correlation at 4 kHz was 0.62 indicating that 39% of the variance in performance in the NoSπ task could be accounted for by combining the ITD- and IID-discrimination measures. Both multiple correlations were statistically significant ($p<0.05$, one-tailed test). The first-order correlations obtained after applying a log transformation to the ITD thresholds revealed no greater or weaker relation among the measures. This is in accord

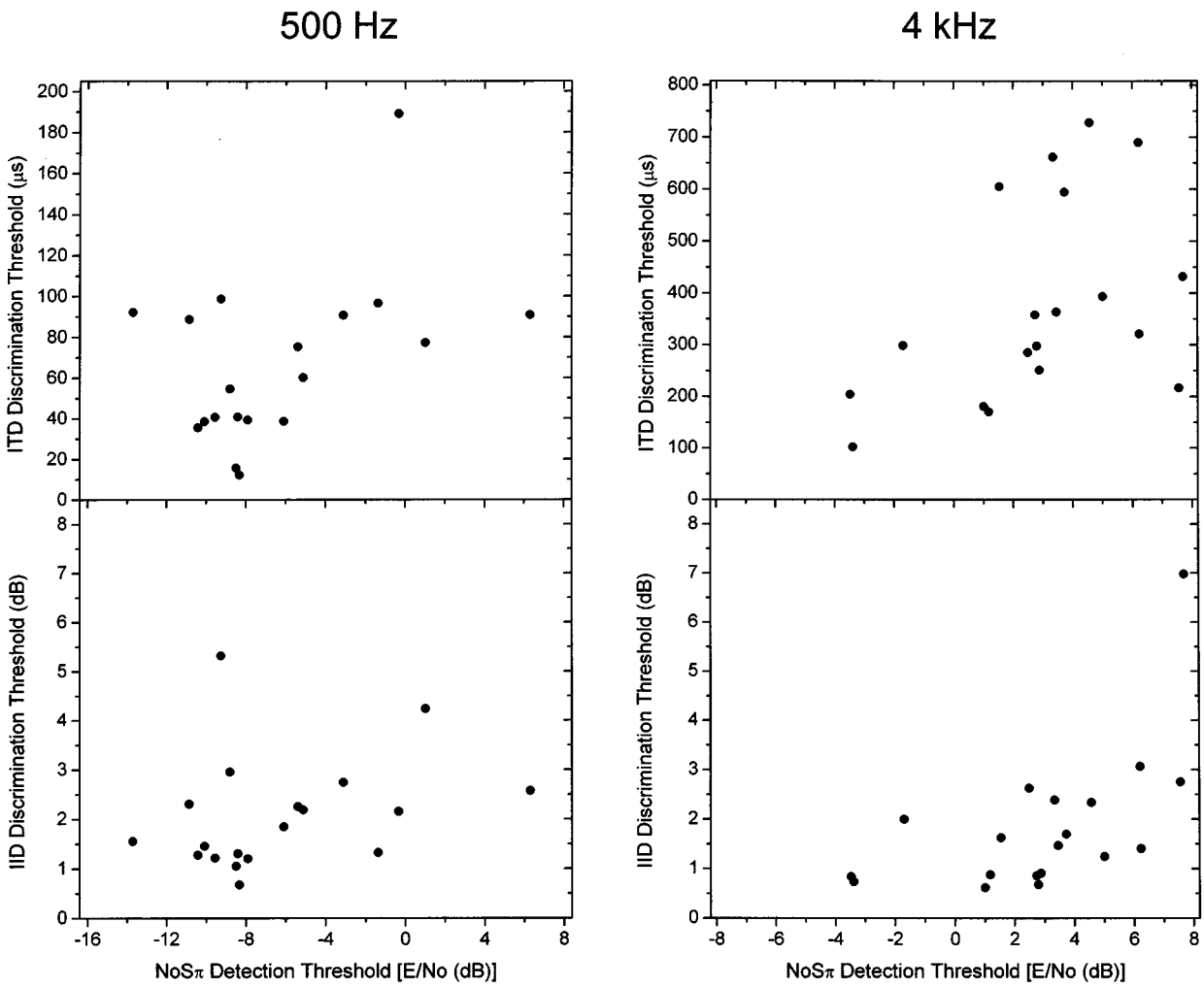


FIG. 5. ITD-discrimination (top panels) and IID-discrimination (bottom panels) thresholds plotted as a function of NoSπ detection thresholds for the nineteen listeners for stimuli centered at 500 Hz (left-hand panels) and stimuli centered at 4 kHz (right-hand panels).

with the lack of curvilinearity observed in the scatterplots of the data in Fig. 5.

Overall, the correlational analyses reveal that inter-individual differences in NoSπ detection are not mirrored in ITD- and IID-discrimination tasks. This suggests that the tasks do not ''tap'' some common, perhaps basic, binaural ability. This outcome may be related to the fact that we measured discrimination of *static* ITDs and IIDs and attempted to relate performance in those tasks to performance in the NoSπ task where ITDs and IIDs change dynamically at a rate proportional to the bandwidth of the masker. Perhaps a future, more complicated experimental design that includes multiple measures of ITD (IID) discrimination (some static and some dynamic) could reveal specific factors that underlie binaural performance.

In addition to our statistical comparisons, we scrutinized the rank ordering of the listeners across the six narrow-band detection and discrimination tasks listed in Table II in order to determine whether the five listeners with the lowest thresholds and the five listeners with the highest thresholds were common across conditions. We found that four of the five listeners who had the lowest 4-kHz narrow-band detection thresholds also were ranked among the five most sensitive listeners in either three, four, or five of the remaining five conditions. On the other hand, among the five listeners who had the highest 4-kHz narrow-band detection thresholds were three listeners who were among those who had the highest thresholds in three of the remaining five tasks. The relatively consistent ranking of thresholds of the ''best'' and ''worst'' performers across the six tasks is responsible for the modest first-order correlations and the smaller part correlations shown in Table II. Consequently, despite the low correlations which indicate low predictability from thresholds in one task to thresholds in another, it appears that listeners who perform at the extremes of the distributions do so consistently. We also checked to see whether the six listeners who had participated in several prior binaural experiments had relatively low thresholds as compared to the 13 listeners who had no prior experience in auditory experiments. Somewhat surprisingly, the thresholds from those 6 listeners essentially spanned the ranges of thresholds obtained in each of the 12 conditions.

Thus the consistencies in listeners' performance across tasks that do exist seem to be a result of there being generally ''sensitive'' or ''insensitive'' binaural listeners, rather than a result of some listeners being differentially adept at discrimi-

nating ITDs or IIDs and relying on their preferred cue in the NoSπ detection task. This is not to say that we did not find some listeners who were relatively more sensitive to ITDs than to IIDs (and vice versa) as did McFadden *et al.* (1973), in their study concerning individual differences in sensitivity to ITDs and IIDs. It is the case, however, that for our group of 19 listeners, knowing the ITD and/or IID thresholds does not afford accurate predictions of the narrow-band NoSπ thresholds at either 500 Hz or 4 kHz.

II. SUMMARY

Using a sample of 19 normal-hearing, young-adult listeners, we found a large range of differences in threshold for 4-kHz tonal signals masked by a narrow band of noise in the NoSπ configuration. This result is consistent with several sets of data from previous experiments conducted over more than a decade (e.g., Durlach and Colburn, 1978). The new data indicate that a broad range of thresholds also occurs when a 500-Hz tone is masked by narrow-band noise with the same stimulus configuration. This outcome, coupled with the fact that the use of a broadband masker does not yield a large distribution of thresholds for the detection of a 500-Hz tone masked by a broad band of noise, suggests that it is the use of a narrow-band masker, per se, that results in a large range of thresholds. This outcome further suggests that the large inter-individual differences in detection thresholds found with narrow-band stimuli at 4 kHz, which prompted this study, cannot be attributed to listeners' differential sensitivity to envelope-based binaural cues at 4 kHz. Statistical analyses revealed that thresholds in the NoSπ detection tasks were not highly correlated with thresholds measured in the ITD- and the IID-discrimination tasks. Nevertheless, the five listeners who were the most sensitive in the narrow-band NoSπ detection and the five listeners who were the least sensitive in the narrow-band NoSπ detection tasks were those who were most and least sensitive, respectively to changes in ITDs and to changes in IIDs.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Steve Colburn, Dr. Wes Grantham, and two anonymous reviewers for their helpful comments and suggestions. Erika L. Hyde's participation, which was facilitated and supported by the ''summer research program'' of the University of Connecticut School of Medicine, was not only scientifically rewarding but was also a distinct pleasure. The senior authors were supported by research Grant Nos. NIH DC-02103 and NIH DC-00234 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

[1]These values were derived by taking into account the nominal bandwidth of the masker (50 or 8400 Hz) and the ''effective'' duration of the signal (290 ms) after gating.

Bernstein, L. R., and Trahiotis, C. (**1992**). ''Discrimination of interaural envelope correlation and its relation to binaural unmasking at high frequencies,'' J. Acoust. Soc. Am. **91**, 306–316.

Bernstein, L. R., and Trahiotis, C. (**1994**). ''Detection of interaural delay in high-frequency SAM tones, two-tone complexes, and bands of noise,'' J. Acoust. Soc. Am. **95**, 3561–3567.

Bernstein, L. R., and Trahiotis, C. (**1995**). ''Binaural interference effects measured with masking-level difference and with ITD- and IID-discrimination paradigms,'' J. Acoust. Soc. Am. **98**, 155–163.

Durlach, N. I., and Colburn, H. S. (**1978**). ''Binaural phenomena'' in *Hearing, Vol. IV, Handbook of Perception*, edited by E. C. Carterette and M. P. Friedman (Academic, New York).

Hafter, E. R., Dye, R. H., Nuetzel, J. M., and Aronow, H. (**1977**). ''Difference thresholds for interaural intensity,'' J. Acoust. Soc. Am. **61**, 829–834.

Klumpp, R. G., and Eady, H. R. (**1956**). ''Some measurements of interaural time difference thresholds,'' J. Acoust. Soc. Am. **28**, 859–860.

Koehnke, J., Colburn, H. S., and Durlach, N. I. (**1986**). ''Performance in several binaural-interaction experiments,'' J. Acoust. Soc. Am. **79**, 1558–1562.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

McFadden, D., Jeffress, L. A., and Russell, W. E. (**1977**). ''Individual differences in sensitivity to interaural differences in time and level,'' Percept. Motor Skills **37**, 755–761.

Nuetzel, J. M., and Hafter, E. R. (**1981**). ''Discrimination of interaural delays in complex waveforms: Spectral effects,'' J. Acoust. Soc. Am. **69**, 1112–1118.

Saberi, K. (**1995**). ''Some considerations on the use of adaptive methods for estimating interaural delay thresholds,'' J. Acoust. Soc. Am. **98**, 1803–1806.

Smoski, W. J., and Trahiotis, C. (**1986**). ''Discrimination of interaural temporal disparities by normal-hearing listeners and listeners with high-frequency, sensorineural hearing loss,'' J. Acoust. Soc. Am. **79**, 1541–1547.

Trahiotis, C., Bernstein, L. R., Buell, T. N., and Spektor, Z. (**1990**). ''On the use of adaptive procedures in binaural experiments,'' J. Acoust. Soc. Am. **87**, 1359–1361.

Yost, W. A., and Dye, R. H. (**1988**). ''Discrimination of interaural differences of level as a function of frequency,'' J. Acoust. Soc. Am. **83**, 1846–1851.

Yost, W. A., Wightman, F. L., and Green, D. M. (**1971**). ''Lateralization of filtered clicks,'' J. Acoust. Soc. Am. **50**, 1526–1530.

Zurek, P. M., and Durlach, N. I. (**1987**). ''Masker-bandwidth dependence in homophasic and antiphasic tone detection,'' J. Acoust. Soc. Am. **81**, 459–464.

Binaural signal detection with phase-shifted and time-delayed noise maskers

Jeroen Breebaart, Steven van de Par, and Armin Kohlrausch
IPO, Center for Research on User-System Interaction, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands

(Received 25 June 1997; revised 2 October 1997; accepted 5 November 1997)

Detection thresholds were measured for interaurally in-phase sinusoids added to a narrow-band dichotic noise masker which was either interaurally phase shifted (NπSo condition) or time delayed (NτSo condition). The signals were spectrally centered in the noise bands and the delay τ equaled half the signal period. Both conditions were tested at 125 and 500 Hz for noise bandwidths of 10, 25, 50, and 100 Hz. In addition, NoSπ and NoSo thresholds were obtained. In contrast to expectations based on the EC theory, no differences in detection thresholds were observed between thresholds in phase-shifted and time-shifted maskers. The results also cannot be explained on the basis of more recent binaural models. The stimuli presented here might therefore serve as a useful validation tool in the development of new binaural theories and models. © *1998 Acoustical Society of America.* [S0001-4966(98)03904-6]

PACS numbers: 43.66.Pn, 43.66.Ba, 43.66.Dc [RHD]

INTRODUCTION

When broadband noise is presented in phase to both ears, and pure tones are presented out of phase to each ear simultaneously (NoSπ condition), the masked threshold is generally lower than for the case when both the noise and the tone are presented in phase (NoSo condition) (Hirsh, 1948; Hafter and Carrier, 1969; Zurek and Durlach, 1987). The difference in detection threshold between the NoSo and the NoSπ condition is referred to as binaural masking level difference (BMLD). The increased sensitivity for out-of-phase signals in a diotic noise is due to the generation of interaural differences by adding the signal to the masker (cf. Zurek, 1991).

If the masker is interaurally phase shifted or time delayed and combined with an in-phase sinusoidal signal (NπSo and NτSo, respectively), the detection threshold is also lower than for the NoSo condition (Webster, 1951; Jeffress *et al.*, 1952). However, the NπSo and NτSo BMLDs are generally smaller than for the NoSπ condition (Hirsh, 1948; Jeffress *et al.*, 1952, 1962; Langford and Jeffress, 1964; Kohlrausch, 1986; Van de Par and Kohlrausch, 1997). The BMLD difference between the NoSπ and the NπSo condition can amount up to about 7 dB for a signal frequency of 167 Hz, and decreases with increasing signal frequency (Jeffress *et al.*, 1962).

One of the theories for interpreting BMLDs is the Equalization and Cancellation (EC) theory, developed by Durlach (1963). The basic idea of this theory is that the auditory system attempts to eliminate the masking components by first transforming the stimuli presented to the two ears in order to equalize the two masking components (E-process). One possible transformation is applying an internal time delay. It is assumed that the E-process is performed imperfectly due to internal errors. Subsequently, the stimulus in one ear is subtracted from the stimulus in the other ear (C-process). For NτSo, where τ equals half the period of the center frequency of the noise and the signal, the cancellation process, after a compensating internal delay, results in an improvement of the signal-to-noise ratio which is equal to the improvement for the NoSπ condition.

For a phase-shifted masker (NπSo), the interaural phase shift has to be compensated by an internal time delay. For a band-limited noise, the consecutive periods of the temporal waveform are not exactly equal. Thus an external phase shift cannot perfectly be compensated by an internal delay. This mismatch of phase shift and time delay will increase with increasing masker bandwidth, due to the decreasing similarity of subsequent periods of the masker. Therefore, the interaural correlation of the Nπ masker after the E-process is smaller than one. This decrease of correlation will be referred to as *decorrelation* of the masker. It is well known that binaural thresholds for an Sπ signal increase with decreasing interaural correlation of a masker (Robinson and Jeffress, 1963). In this respect, and under the assumption that binaural detection for the NπSo condition is degraded by both internal errors and decorrelation (Siegel and Colburn, 1989), NτSo and NoSπ BMLDs should be larger than NπSo BMLDs.

It has been shown previously that for wide-band maskers, the NτSo and NπSo BMLDs are smaller than the NoSπ BMLDs (Jeffress *et al.*, 1962; Langford and Jeffress, 1964), especially at low frequencies. However, a comparison between the NπSo and the NτSo BMLDs for narrow-band maskers as a function of frequency and bandwidth of the noise has never been made. Since the bandwidth and frequency of the masker determine the amount of decorrelation within the framework of the EC theory (and hence the detection threshold), the experiments form a critical test for the EC theory.

To evaluate the hypothesis that the binaural system benefits from the absence of decorrelation in an interaurally time-shifted condition in contrast to an interaurally phase-shifted condition, as suggested by the EC theory, binaural masked thresholds were measured for NτSo and NπSo as a

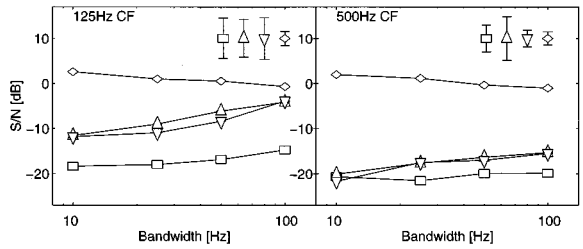


FIG. 1. NoSo (diamonds), NπSo (triangles down), NτSo (triangles up), and NoSπ (blocks) thresholds at 125-Hz (left panel) and 500-Hz (right panel) center frequency as a function of masker bandwidth. Error bars denote the mean standard deviation between subjects for NoSo, NπSo, NτSo, and NoSπ.

function of masker bandwidth. As a reference, thresholds for NoSo and NoSπ were also measured.

I. PROCEDURE AND STIMULI

A 3-interval forced-choice procedure with adaptive signal-level adjustment was used to determine masked thresholds. Three masker intervals of 400-ms duration were separated by pauses of 300 ms. A signal of 300-ms duration was added to the temporal center of one of the masker intervals. Feedback was provided after each response of the subject.

The signal level was adjusted according to a two-down one-up rule (Levitt, 1971). The initial step size for adjusting the level was 8 dB. The stepsize was halved after every second reversal of the level track until it reached 1 dB. The run was then continued for another eight reversals. The median level at these last eight reversals was used as the threshold value. At least four threshold values were obtained for each parameter value and subject.

All stimuli were generated digitally and converted to analog signals with a two-channel, 16-bit D/A converter at a sampling rate of 32 kHz. The masker signals were presented to the subjects over Beyer Dynamic DT990 headphones at a sound pressure level of 65 dB.

The 400-ms masker samples were obtained by randomly selecting a segment from a 2000-ms bandpass-noise buffer. The bandpass-noise buffer was created in the frequency domain by selecting the frequency range from the Fourier transform of a 2000-ms broadband Gaussian noise. After an inverse Fourier transform, the band-limited noise buffer of 2000 ms was obtained.

The 300-ms signals were sinusoids with a frequency equal to the center frequency of the noise masker. In order to avoid spectral splatter, the signal and the maskers were gated with 50-ms raised-cosine ramps. Thresholds are expressed as signal-to-overall-noise-power ratio and are the means of four repetitions per condition and subject. Masked thresholds were measured for NτSo, NπSo, NoSo, and NoSπ conditions where τ equals half the period of the center frequency.

II. RESULTS

Thresholds were obtained from three well-trained subjects with normal hearing. Mean NoSo, NπSo, NτSo, and NoSπ thresholds are shown in Fig. 1 at center frequencies of 125 (left panel) and 500 Hz (right panel). The thresholds for the NoSo condition (diamonds) show a slight decrease with increasing bandwidth for both center frequencies. The slope of 1.3 dB/oct is in line with other data obtained for noise maskers of subcritical bandwidth (de Boer, 1962; Kidd *et al.*, 1989; van de Par and Kohlrausch, 1997). The thresholds for the NπSo (triangles down) and NτSo condition (triangles up) are very similar, while the NoSπ condition (squares) shows lower thresholds. The average difference in detection thresholds between the NπSo and NoSπ conditions is 8 dB at 125 Hz and 4 dB at 500 Hz, where the differences obtained with the 100-Hz masker agree with the results from Kohlrausch (1986) for broadband maskers. The difference between NoSπ and NτSo thresholds for 100-Hz bandwidth at 500 Hz amounts to about 5 dB, a value similar to the 2.9-dB difference measured by Langford and Jeffress (1964) for a broadband masker. Furthermore, there is an increase of the NπSo and NτSo thresholds with increasing masker bandwidth which is stronger at 125 Hz (e.g., 2.1 dB/oct of masker bandwidth) than at 500 Hz (1.2 dB/oct). The mean difference between NπSo and NτSo thresholds amounts to -0.5 dB, implying that NτSo thresholds are slightly higher than NπSo thresholds. According to a one-tailed Student's *t* test on the pooled differences of all subjects and conditions, this difference was significantly different from zero at a 2% significance level.

III. DISCUSSION

According to the EC theory, an Nπ or Nτ masker is equalized by applying an internal time delay. For the Nτ stimulus, the waveforms at the right and left side after the equalization stage are identical, yielding a perfect interaural correlation. For the Nπ stimulus, the phase shift is also compensated by an internal delay, yielding only a partially correlated masker. Therefore, if only this single time delay is used in the detection process, NτSo thresholds would be expected to be lower than NπSo thresholds. This was not found to be true in our experiments.

At 125-Hz center frequency, the increase of thresholds with increasing masker bandwidth is larger than at 500 Hz. This may be related to the fact that the internal delay necessary to compensate for the external phase shift decreases with increasing center frequency, since the optimal internal delay equals half the period of the center frequency of the noise. Note that the temporal fluctuations of the envelope are independent of center frequency, i.e., the interaural correlation decreases only with increasing delay and bandwidth of the masker. Thus for a certain masker bandwidth, the amount of decorrelation for the Nπ masker at a delay of half the period of the center frequency is larger at 125 Hz than at 500 Hz. Furthermore, an increase in the bandwidth of a noise signal results in a stronger damping of the cross-correlation function. Thus at larger internal delays, the decorrelation grows faster with increasing masker bandwidth than at smaller internal delays. Consequently, a smaller effect of masker bandwidth is expected at the higher frequency.

This qualitative analysis of the effects of masker bandwidth and interaural phase relations of the masker made us wonder how large these effects are quantitatively. We there-

fore evaluated the NπSo and NτSo BMLDs predicted by the EC model *as a function of the internal delay*. The rationale for this extension lies in the fact that the NτSo condition shows smaller BMLDs than the NoSπ condition. According to the EC theory, both conditions should have equal BMLDs. Furthermore, the optimal internal delay of 4 ms for the NτSo condition at 125-Hz center frequency is rather large in comparison with plausible delays that occur in daily listening conditions and in comparison with delays found in the neuronal system (Palmer *et al.*, 1990; Caird *et al.*, 1991; McAlpine *et al.*, 1996). In this respect, we hypothesized that the internal delay assessed for detection at the 125-Hz condition might be smaller than the optimal internal delay of 4 ms (cf. also the discussion in van der Heijden *et al.*, 1997).

According to the EC theory, the NoSπ BMLD $f(0,\pi)$, as a function of the internal delay Δ is given by (see Durlach, 1972)

$$f(0,\pi)=\frac{k+\cos(\omega\Delta)}{k-\rho(\Delta)}. \tag{1}$$

Here, k is a factor that represents internal errors of the stimulus representation, ω is the center frequency of the stimulus and $\rho(\Delta)$ represents the autocorrelation function of the noise after peripheral filtering. The NoSπ thresholds were used to determine the internal error term, resulting in $k=1.0202$ (based on an NoSπ BMLD of 20 dB).[1] The NτSo BMLD $f(\tau,0)$ as a function of the internal delay is Δ given by

$$f(\tau,0)=\frac{k-\cos(\omega\Delta)}{k-\rho(\tau-\Delta)}, \tag{2}$$

while the NπSo BMLD $f(\pi,0)$ is given by

$$f(\pi,0)=\frac{k-\cos(\omega\Delta)}{k+\rho(\Delta)}. \tag{3}$$

We computed the autocorrelation functions $\rho(\Delta)$ by filtering the power spectrum of the noise maskers by a fourth-order gammatone filter with an equivalent rectangular bandwidth of 38.2 Hz at 125-Hz center frequency and 78.7 Hz at 500-Hz center frequency (Glasberg and Moore, 1990). The bandwidth of 78.7 Hz at 500-Hz center frequency is in fair agreement with the estimated filter bandwidth as provided by Langford and Jeffress (1964). Subsequently, an inverse Fourier transform and a normalization resulted in the autocorrelation function for the stimuli after peripheral filtering. With these assumptions we calculated the BMLD as a function of the internal delay. The predicted NπSo and NτSo BMLDs at 125-Hz and 500-Hz center frequencies for a masker bandwidth of 10 Hz are shown as a function of the internal delay in the left panel of Fig. 2. The solid and the dotted lines represent the BMLDs for the NπSo and the NτSo condition at 125-Hz center frequency, while the nearly identical dashed and the dash-dotted lines represent the BMLDs at 500-Hz center frequency. The right panel of Fig. 2 shows the

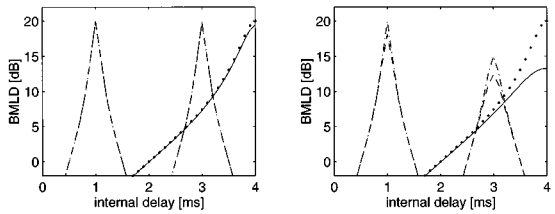


FIG. 2. BMLDs as a function of the internal delay according to the EC theory. NπSo condition at 125-Hz center frequency (solid line); NτSo at 125-Hz center frequency (dotted line); NπSo at 500-Hz center frequency (dashed line); NτSo at 500-Hz center frequency (dash-dotted line). The left panel represents predictions for 10-Hz bandwidth, the right panel those for 100-Hz bandwidth.

BMLDs at 100-Hz bandwidth in the same format. If we assume, as in the EC theory, an optimal delay is used for detection, some striking results in comparison with the experimental data can be summarized as follows:

(1) The EC theory does not predict any differences in BMLD between the three conditions at 10-Hz bandwidth, assuming that optimal internal delays (1 or 3 ms at 500 Hz, 4 ms at 125 Hz) are used.
(2) At 100-Hz bandwidth, the EC theory predicts a difference of 2 dB between the NπSo and the NτSo condition at 500-Hz center frequency, and a difference of 7 dB at 125-Hz center frequency. These differences are not found in our experimental data.
(3) The predicted NτSo BMLDs remain constant with increasing masker bandwidth, while the experimental NτSo BMLDs decrease from 15 dB at 10-Hz bandwidth to 4 dB at 100-Hz bandwidth for the condition at 125-Hz center frequency.

These observations suggest that, within the framework of the EC theory, binaural detection based on a single, optimal delay cannot account for the experimental data presented here. One could argue that a nonoptimal internal delay is assessed for detection. Such an argument is supported by the findings of Jeffress *et al.* (1962), that the NτSo BMLD for a 167-Hz signal reaches its maximum value already for $\tau=0.5$ ms and does not increase for larger noise delays. From Fig. 2, we see that in order to achieve a BMLD of at least 4 dB for the NπSo condition at 125-Hz center frequency and 100-Hz bandwidth, the internal delay must amount to at least 2.5 ms. Furthermore, assuming that delays within a range of 2.5 ms are available for detection, an NτSo BMLD of 20 dB is expected at 500-Hz center frequency and 100-Hz bandwidth, while the experimental data show a BMLD of only 15 dB.

Also modifications to the EC theory, such as proposed by Green (1966), who suggested that besides subtraction also addition is allowed in the cancellation step, cannot account for our data. Clearly, the addition step would totally cancel the masking noise in an NπSo condition, while the subtraction would totally cancel the masking noise in an NoSπ condition and, after an appropriate delay, in an NτSo condition. Consequently, this would result in equal thresholds for all three conditions independent of the masker bandwidth, which does not correspond to our data.[2]

After having concluded that the EC theory cannot account for the results presented here, we tried to explain the results within the framework of cross-correlation models (see Colburn, 1977; Zwicker and Henning, 1985; Lindemann, 1986; Shackleton *et al.*, 1992; Stern and Shear, 1996; Bernstein and Trahiotis, 1996). A quantitative application of these models to the detection conditions of the present study was beyond our intentions. We think, however, that the existing cross-correlation models cannot account for the dependency of the detection thresholds on the masker bandwidth, due to the fact that the simulated neural activity (or crosscorrelation) in these models is not normalized with the stimulus energy. This leads to the problem that the changes in the neural activity that reflect changes in the interaural correlation are much smaller than the variation in the masker activity. To give an example: for a signal-to-noise ratio of -20 dB, the change in the cross correlation due to the addition of an $S\pi$ signal to a narrow-band diotic masker amounts to 0.02 times the masker power. On the other hand, the standard deviation of the masker energy for a 10-Hz-wide noise of a duration of 400 ms amounts to 0.5 times the masker energy (Bendat and Piersol, 1976). Hence without normalization, the change in the interaural cross-correlation function due to the addition of a signal to a narrow-band masker is much smaller than the uncertainty in the correlation function due to stimulus uncertainty. Thus if a change in the (unnormalized) cross-correlation function is used as a cue for detection, only small BMLDs are expected for narrow-band maskers.[3] Furthermore, since the standard deviation of the masker energy decreases with increasing masker bandwidth, the $NoS\pi$ thresholds should decrease with increasing masker bandwidth. This was not found to be true in our experiments; according to Fig. 1, the $NoS\pi$ thresholds remain approximately constant across bandwidth.

In summary, we think that neither the EC theory nor current crosscorrelation models can predict the experimental data presented here. Therefore, the understanding of the binaural processes involved in the experimental conditions of this study still forms a challenge for psychoacousticians.

ACKNOWLEDGMENTS

We want to thank all our subjects for participating in our experiments which meant spending some of their valuable time in the listening booth. We thank G. B. Henning, Andrew Oxenham, and an anonymous reviewer for their valuable comments on earlier drafts of this paper.

[1]Since we are interested in comparing, at each of the two frequencies 125 and 500 Hz, the relation between the three binaural conditions and the influence of masker bandwidth we used, for simplicity, the same value for k at 125 and 500 Hz.

[2]An attempt to explain our data using the vector theory of binaural interaction (Jeffress, 1972) is also bound to fail. As already noted by Jeffress *et al.* (1962), vector diagrams ''do not explain the change of MLD's with frequency nor the rather large difference possible between the MLD's for $NoS\pi$ and $N\pi So$.'' (Jeffress *et al.*, 1962, p. 1125). We can add that, without further modifications, the vector theory would also be unable to predict a change in binaural thresholds with masker bandwidth.

[3]The model by Stern and Shear (1996) incorporates an automatic gain control stage which adjusts and thus eliminates the influence of overall stimulus level. However, as long as this stage does not follow intensity changes on a short-time basis, it does not solve the bandwidth problem described above.

Bendat, J. S., and Piersol, A. G. (**1976**). *Random Data: Analysis and Measurement Procedures* (Wiley-Interscience, New York).

Bernstein, L. R., and Trahiotis, C. (**1996**). ''The normalized correlation: Accounting for binaural detection across center frequency,'' J. Acoust. Soc. Am. **100**, 3774–3787.

de Boer, E. (**1962**). ''Note on the critical bandwidth,'' J. Acoust. Soc. Am. **34**, 985–986.

Caird, D. M., Palmer, A. R., and Rees, A. (**1991**). ''Binaural masking level difference effects in single units of the guinea pig inferior colliculus,'' Hearing Res. **43**, 1–24.

Colburn, H. S. (**1977**). ''Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise,'' J. Acoust. Soc. Am. **61**, 525–533.

Durlach, N. I. (**1963**). ''Equalization and cancellation theory of binaural masking-level differences,'' J. Acoust. Soc. Am. **35**, 1206–1218.

Durlach, N. I. (**1972**). ''Binaural signal detection: Equalization and cancellation theory,'' in *Foundations of Modern Auditory Theory*, Vol. II, edited by J. V. Tobias (Academic, New York).

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' J. Acoust. Soc. Am. **47**, 103–138.

Green, D. M. (**1966**). ''Signal-detection analysis of equalization and cancellation model,'' J. Acoust. Soc. Am. **40**, 833–838.

Hafter, E. R., and Carrier, S. C. (**1969**). ''Masking-level differences obtained with a pulsed tonal masker,'' J. Acoust. Soc. Am. **47**, 1041–1047.

Hirsh, I. J. (**1948**). ''The influence of interaural phase on interaural summation and inhibition,'' J. Acoust. Soc. Am. **20**, 536–544.

Jeffress, L. A. (**1972**). ''Binaural signal detection: Vector theory,'' in *Foundations of Modern Auditory Theory*, Vol. II, edited by J. V. Tobias (Academic, New York).

Jeffress, L. A., Blodgett, H. C., and Deatherage, B. H. (**1952**). ''The masking of tones by white noise as a function of the interaural phases of both components I. 500 Cycles,'' J. Acoust. Soc. Am. **24**, 523–527.

Jeffress, L. A., Blodgett, H. C., and Deatherage, B. H. (**1962**). ''Masking and interaural phase. II. 167 Cycles,'' J. Acoust. Soc. Am. **34**, 1124–1126.

Kidd, G., Mason, C. R., Brantley, M. A., and Owen, G. A. (**1989**). ''Roving-level tone-in-noise detection,'' J. Acoust. Soc. Am. **86**, 1310–1317.

Kohlrausch, A. (**1986**). ''The influence of signal duration, signal frequency and masker duration on binaural masking level differences,'' Hearing Res. **23**, 267–273.

Langford, T. L., and Jeffress, L. A. (**1964**). ''Effect of noise crosscorrelation on binaural signal detection,'' J. Acoust. Soc. Am. **36**, 1455–1458.

Levitt, R. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Lindemann, W. (**1986**). ''Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals,'' J. Acoust. Soc. Am. **80**, 1608–1622.

McAlpine, D., Jiang, D., and Palmer, A. R. (**1996**). ''Binaural masking level differences in the inferior colliculus of the guinea pig,'' J. Acoust. Soc. Am. **100**, 490–503.

Palmer, A. R., Rees, A., and Caird, D. (**1990**). ''Interaural delay sensitivity to tones and broad band signals in the guinea-pig inferior colliculus,'' Hearing Res. **50**, 71–86.

Robinson, D. E., and Jeffress, L. A. (**1963**). ''Effect of varying the interaural noise correlation on the detectability of tonal signals,'' J. Acoust. Soc. Am. **35**, 1947–1952.

Shackleton, T. M., Meddis, R., and Hewitt, M. J. (**1992**). ''Across frequency integration in a model of lateralization,'' J. Acoust. Soc. Am. **91**, 2276–2279.

Siegel, R. A., and Colburn, H. S. (**1989**). ''Binaural processing of noisy stimuli: Internal/external noise ratios for diotic and dichotic stimuli,'' J. Acoust. Soc. Am. **86**, 2122–2128.

Stern, R. M., and Shear, G. D. (**1996**). ''Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of internal delay,'' J. Acoust. Soc. Am. **100**, 2278–2288.

van de Par, S., and Kohlrausch, A. (**1997**). ''A new approach to comparing binaural masking level differences at low and high frequencies,'' J. Acoust. Soc. Am. **101**, 1671–1680.

van der Heijden, M., Trahiotis, C., Kohlrausch, A., and van de Par, S. (**1997**). ''Binaural detection with spectrally nonoverlapping signals and maskers: evidence for masking by aural distortion products,'' J. Acoust. Soc. Am. **102**, 2966–2972.

Webster, F. A. (**1951**). ''The influence of interaural phase on masked thresholds. I. The role of interaural time derivation,'' J. Acoust. Soc. Am. **23**, 452–462.

Zurek, P. M. (**1991**). ''Probability distribution functions of interaural phase and level differences in binaural detection stimuli,'' J. Acoust. Soc. Am. **90**, 1927–1932.

Zurek, P. M., and Durlach, N. I. (**1987**). ''Masker-bandwidth dependence in homophasic and antiphasic tone detection,'' J. Acoust. Soc. Am. **81**, 459–463.

Zwicker, E., and Henning, G. B. (**1985**). ''The four factors leading to binaural masking-level differences,'' Hearing Res. **19**, 29–47.

Effects of sensorineural hearing loss on interaural discrimination and virtual localization

Laura Smith-Olinde
Southwest Missouri State University, Communication Sciences and Disorders, 901 South National Avenue, Springfield, Missouri 65804

Janet Koehnke and Joan Besing
University of South Alabama, Department of Speech Pathology and Audiology, UCOM 2000, Mobile, Alabama 36688

(Received 2 December 1996; revised 12 October 1997; accepted 9 January 1998)

Cross-frequency binaural processing was investigated in listeners with normal hearing (NH) and with bilateral high-frequency sensorineural hearing impairment (IH). In experiment 1 just-noticeable-differences for interaural time and interaural intensity were measured using 1/3-octave narrow-band noises (NBNs) centered at 0.5 and 4 kHz. These stimuli were presented in isolation and in different cross-frequency interaural combinations. IH listeners displayed the best interaural time discrimination when the 0.5-kHz NBN was dichotic and the best intensity discrimination when both bands were dichotic. Both NH listeners (time) and IH listeners (time and intensity) displayed the poorest interaural discrimination when the NBNs were presented simultaneously with interaural differences in only the 4-kHz NBN (0.5 kHz NBN diotic). Localization accuracy was measured in experiment 2 using the 0.5- and 4-kHz NBNs in isolation and with 0.5-kHz target/4-kHz interferer and 4-kHz target/0.5-kHz interferer conditions. Best localization of NH and IH subjects was seen for the 0.5-kHz target, with or without an interferer. Poorest localization of IH subjects was observed for the 4-kHz target and 0.5-kHz interferer. Results suggest that for these IH subjects, localization is most difficult when they are forced to rely on interaural information in a higher-frequency region with conflicting interaural information at low frequencies. © *1998 Acoustical Society of America.* [S0001-4966(98)03604-2]

PACS numbers: 43.66.Pn, 43.66.Qp, 43.66.Rq [RHD]

INTRODUCTION

The ability to efficiently use binaural cues is important in many tasks listeners perform on a daily basis, such as locating sound sources or listening to speech in the presence of background noise. However, it is well established that individuals with sensorineural hearing loss often have difficulty with binaural processing tasks including interaural time discrimination (e.g., Hawkins and Wightman, 1980; Häusler *et al.*, 1983; Smoski and Trahiotis, 1986; Gabriel *et al.*, 1992; Koehnke *et al.*, 1995; Koehnke and Besing, 1996), interaural intensity discrimination (e.g., Häusler *et al.*, 1983; Gabriel *et al.*, 1992; Koehnke *et al.*, 1995), localization (e.g., Jonkgees and Veer, 1957; Tonning, 1975; Colburn *et al.*, 1982; Noble *et al.*, 1994), and binaural detection (e.g., Koehnke *et al.*, 1995; Koehnke and Besing, 1996). These studies have typically employed a single target stimulus, either simple (pure tones) or complex (narrow- or wide-band noise, speech). In natural listening situations, however, other sounds unrelated to the target signal may be present and may interfere with the reception of relevant binaural information in the target signal. The purpose of the present study was to determine the effect of cross-frequency binaural information in conditions where the cross-frequency interaural information is reinforcing and in conditions where the cross-frequency information is conflicting.

There has been a limited number of studies of cross-frequency binaural processing in listeners with normal hearing on basic binaural tasks (McFadden and Pasanen, 1976; Wenzel and Hafter, 1985; Trahiotis and Bernstein, 1990; Heller, 1992; Buell and Hafter, 1991; Buell and Trahiotis, 1993; Dye, 1990; Stellmack and Dye, 1993; Woods and Colburn, 1992). However, these studies have not included individuals with impaired hearing, whose most common complaint is difficulty understanding speech in the presence of interfering noise. Nor have these studies included measurements of the effects of cross-frequency information on performance of everyday listening tasks such as localization.

Buell and Hafter (1991) and Buell and Trahiotis (1993) show integration across the frequency domain for low-frequency tones and SAM high-frequency tones, respectively. In contrast, McFadden and Pasanen (1976) reported no increase in sensitivity to interaural delays with widely spaced narrow-band noise stimuli (above and below 1500 Hz) containing congruent interaural information. They found interaural time just-noticeable-differences (jnds) for the high-frequency signal alone were elevated compared to both the low-frequency signal and both signals played simultaneously. This is consistent with the idea that the binaural system is more sensitive to interaural time information in the low frequencies (Yost *et al.*, 1971; Bernstein and Trahiotis, 1982).

One phenomenon related to the combination of interaural information across frequency is binaural interference. McFadden and Pasanen (1976) reported increases in interaural time JNDs in normal-hearing listeners for a dichotic high-

 0001-4966/98/103(4)/2084/16/$10.00

frequency narrow-band noise when a diotic low-frequency narrow-band noise was presented simultaneously, but no increase in jnds when a dichotic low-frequency narrow-band noise was presented simultaneously with a diotic high-frequency narrow-band noise. Buell and Trahiotis (1993), using SAM high-frequency tones, and Trahiotis and Bernstein (1990), using narrow bands of noise, described similar results; that is, when a lower-frequency signal was the target, no decrement in interaural time discrimination performance was noted with a higher-frequency interferer, but when a higher-frequency signal was the target, the interaural time jnd increased in the presence of a low-frequency interferer. These results suggest that upward spread of masking by an interfering signal may cause peripheral disruption for reception of information in the target signal. Heller (1992), using narrow-band noises, also found greater interference with a low-frequency interferer and a high-frequency target when measuring interaural time jnd's, but when measuring interaural intensity jnd's both high- and low-frequency interferers produced a similar amount of interference. Lack of spectral asymmetry for interaural intensity discrimination suggests that upward spread of masking is not adequate to explain the results of these cross-frequency binaural discrimination studies.

Although many investigators have examined the ability of normal-hearing listeners to localize sound sources (e.g., Wightman and Kistler, 1992), relatively few investigators have studied the localization ability of impaired-hearing listeners (e.g., Jongkees and Veer, 1957; Nordlund, 1964; Colburn *et al.*, 1982; Häusler *et al.*, 1983; Noble *et al.*, 1994). Due to the variety of hearing losses, stimuli, and methods which have been employed in these studies, generalizations regarding the localization abilities of impaired-hearing individuals are difficult to state. Overall, however, these studies indicate that performance of impaired-hearing listeners is less accurate than the performance of normal-hearing listeners.

Most studies of localization in normal- and impaired-hearing listeners have been performed with the target signal in isolation, rather than with a target in the presence of noise. This stimulus paradigm does not allow examination of the effects of combining interaural information across frequency. Therefore, in the present study, a target signal and an interfering signal which differed in spectral and spatial information, were used to measure binaural performance of both impaired-hearing and normal-hearing subjects.

Two experiments were conducted to determine the effects of cross-frequency information on simple and complex binaural processing tasks for the same groups of impaired- and normal-hearing listeners. Experiment 1 included measurement of interaural time and intensity jnds for low- and high-frequency signals with and without cross-frequency information present. In experiment 2, sound source localization of the same low- and high-frequency signals was measured with and without an interfering stimulus with different spectral information present. Because localization is believed to be based on processing of interaural time and intensity information in the signal, it was expected that cross-frequency effects on performance of the basic binaural tasks of experiment 1 might be related to cross-frequency effects on localization ability as measured in experiment 2.

I. EXPERIMENT 1: INTERAURAL DISCRIMINATION

A. Methods

This experiment was designed to measure listeners' interaural just-noticeable-differences (jnd's) in time and intensity for individual noise band targets as well as targets presented with an interfering noise band.

Two groups of subjects participated in the experiment: one group with normal hearing and one with impaired hearing. There were three normal-hearing subjects whose ages ranged from 23 to 36 years. Audiometric testing indicated bilateral audiometric thresholds $\leqslant 5$ dB HL at octave frequencies from 250 through 8000 Hz (ANSI, 1969). There were six impaired-hearing subjects ranging in age from 32 to 64 years. [Audiograms for the impaired-hearing subjects are shown in panel (a) of each Figs. 4 through 9.] All these subjects had bilaterally symmetric pure tone thresholds (within 10 dB between ears), and all had sensorineural hearing loss as evidenced by equivalent air and bone conduction thresholds and normal tympanograms.

The signals used in experiment 1 were individual 1/3-octave narrow-band noises geometrically centered at 500 and 4000 Hz, and combined 1/3-octave narrow-band noises at $500+4000$ Hz. The stimuli were 400-ms bursts with 10-ms $\cos^2$ rise–fall times. Twenty samples of each narrow-band noise were generated at a 40-kHz sampling rate and stored on a CompuAdd 386/33 computer for on-line retrieval. Four experimental conditions were included in the interaural discrimination experiment: (1) 0.5-kHz and 4-kHz narrow-band noises in isolation; (2) both narrow-band noises with congruent interaural time or intensity differences; (3) dichotic 0.5-kHz narrow-band noise with diotic 4-kHz narrow-band noise; and (4) dichotic 4-kHz narrow-band noise with diotic 0.5-kHz narrow-band noise. In the interaural time discrimination task, ''dichotic'' refers to the condition in which a phase delay was imposed on the internal structure of the signal, resulting in signals with unlike phases (or timing information) that are presented to each ear. Interaural time differences ranging from 1 to 1000 μs were created by imposing a phase delay equivalent to the desired time difference. The interaural time difference was present as an ongoing delay with no interaural onset or offset envelope differences. In the interaural intensity discrimination task, ''dichotic'' refers to the condition in which different levels of attenuation were imposed on the signals being sent to each ear. In both interaural time and intensity discrimination ''diotic'' refers to the condition in which the same signal(s) was sent to both ears. In all dual-band signal conditions (2, 3, and 4) the target and interfering sound were gated on and off simultaneously.

Narrow-band noise stimuli were scaled and added using an array processor (Tucker-Davis Technology, AP1) and then played out using a D/A board (TDT DA3) connected to an anti-aliasing filter (TDT FLT3) and programmable attenuators (TDT PA4). All subjects were seated in a sound-treated

TABLE I. Mean narrow-band noise thresholds (dB SPL) for three normal-hearing subjects; individual narrow-band noise thresholds (dB SPL) for impaired-hearing subjects.

Subject	Frequency	Right	Left	Rounded binaural average	Stimulus presentation levels
Normal	500	8.2	6.4	7.0	77.0
hearing	4000	6.4	3.5	5.0	77.0
S1	500	15.6	8.5	10.0	86.0
	4000	58.1	57.9	58.0	86.0
S2	500	10.2	8.4	9.0	77.0
	4000	41.7	35.0	37.0	77.0
S3	500	12.1	10.5	11.0	87.0
	4000	63.0	57.0	59.0	87.0
S4	500	11.8	9.6	10.0	77.0
	4000	33.7	39.7	38.0	77.0
S5	500	33.6	36.4	36.0	97.0
	4000	69.4	56.9	63.0[a]	97.0
S6	500	16.6	10.1	12.0	84.0
	4000	50.2	58.1	56.0	84.0

[a]There was a 12.5-dB difference between ears (right = 69.4, left = 56.9); therefore, 28 dB SL was referenced to the poorer threshold (right ear).

room and listened to stimuli through TDH-49P earphones mounted in MX-41/AR cushions.

Prior to measuring interaural jnd's, narrow-band noise thresholds for each of the two narrow-band noise stimuli were measured for each subject's right and left ears using a two-interval forced-choice (2IFC) adaptive paradigm with feedback. The adaptive procedure consisted of 14 reversals and targeted the 70.7% correct level using the 1-up/2-down Transformed Up-Down method (Levitt, 1971). The stepsize for threshold estimation began at 4 dB and was reduced to 2 dB after the fourth reversal. Each threshold was calculated from the average of the fifth through the fourteenth reversals. The binaural threshold was then calculated by averaging the thresholds obtained for both ears at each frequency and rounding to the nearest dB; those average binaural thresholds served as the reference for calculating sensation levels for subsequent binaural measurements. The stimuli were presented at either 77 dB SPL or 28 dB SL, whichever was higher, for a given subject and may be found in Table I. Both low- and high-frequency stimuli were presented at the same SPL level to both ears with the exception of the interaural level difference.

All interaural time and intensity jnds were obtained using an adaptive 2-cue, 2-interval, 2-alternative forced-choice (2C, 2I, 2AFC) procedure with feedback. For all conditions, intervals one and four were diotic and served as references within and across trials. Interaural differences occurred in either interval two or three with a 0.5 *a priori* probability; therefore, the ''dichotic'' condition occurred on only one of the four intervals, the ''odd'' interval. The paradigm was subject paced in that a response was required before a new trial could begin. The jnd was defined as the 70.7% correct level using the 1-up/2-down Transformed Up-Down method (Levitt, 1971). One sample of noise was chosen randomly for each of the four intervals of every trial. The runs consisted of eight reversals and the jnd was calculated from the average of the fifth through the eighth reversals; each jnd was measured five times.

Implementation of the interaural time delay or interaural intensity increment was controlled by computer algorithm, and the ear which received the odd stimulus (i.e., the time delayed or more intense signal) was fixed within a block of trials (a run) but was randomly varied between blocks of trials. The initial stepsize in interaural time discrimination was changed by a factor of 2; after the fourth reversal the stepsize was changed by a factor of $\sqrt{2}$. Interaural intensity differences ranging from 0.1 to 10 dB were created by presenting half the total interaural difference to each ear, according to Eqs. (1) and (2):

$$\mathrm{SI_R} = I + \Delta I/2, \tag{1}$$

$$\mathrm{SI_L} = I - \Delta I/2, \tag{2}$$

where $\mathrm{SI_R}$ equals the right ear stimulus intensity, $\mathrm{SI_L}$ equals the left ear stimulus intensity, I equals the overall intensity of 77 dB SPL or 28 dB SL, and ΔI equals the total interaural difference. The initial stepsize in interaural intensity discrimination was 2 dB; after the fourth reversal the stepsize became smaller as the interaural intensity difference decreased.[1] The changes in stepsize were implemented in order to make more precise measurements of subjects' jnds. To reduce the possibility that subjects could use the monaural change in signal intensity to perform the task, an overall intensity rove over a 10-dB range was imposed during interaural intensity discrimination testing (Grantham, 1984). The rove was always applied as a decrement in the overall level to insure the signal was never uncomfortably loud for the subjects. Roving levels from 1 to 10 dB in 1-dB increments were chosen at random for each interval of an experimental trial.

All the normal-hearing listeners had experience with masking level difference, virtual localization, and monaural detection with a contralateral cue tasks; therefore, they required only a few runs before their jnd estimates stabilized. The impaired-hearing subjects needed more training than the normal-hearing subjects. Data collection began for each

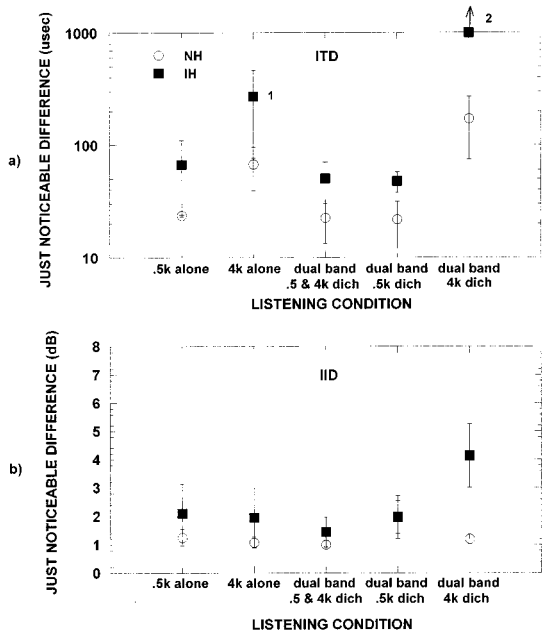


FIG. 1. (a) Average interaural time jnds for the normal- and impaired-hearing subjects. The abscissa represents the signal conditions and the ordinate represents the interaural time jnd in μs. The open circles represent the means of the normal-hearing subjects and the filled squares represent the means of the impaired-hearing subjects. The error bars represent one standard deviation for each signal condition across listeners. 1: Represents mean of four impaired-hearing subjects who could perform interaural time discrimination with the 4-kHz narrow-band noise in isolation. The other two could not perform the task with the maximum interaural difference of 1000 μs. 2: Five of the six impaired-hearing subjects could not perform the task with a 1000-μs interaural time difference (indicated by the arrow). (b) Average interaural intensity jnds for the normal- and impaired-hearing subjects. The abscissa represents the signal conditions while the ordinate represents the interaural intensity jnd in dB. The open circles represent the means of the normal-hearing subjects and the filled squares represent the means of the impaired-hearing subjects. The error bars represent one standard deviation for each signal condition across listeners.

impaired-hearing subject when the standard deviation of five consecutive runs fell to one-half or less of the standard deviation of the previous five runs in a particular condition. For the 0.5- and 4-kHz combined narrow-band noise conditions both the normal- and impaired-hearing subjects subjects received only a few practice runs prior to data collection; neither group had any experience with binaural interference experiments.

B. Results

Group data from the interaural time discrimination experiment are shown in Fig. 1(a). There were four signal configurations in the discrimination experiment: (1) individual noise bands (0.5-kHz alone, 4-kHz alone); (2) combined 0.5-kHz and 4-kHz narrow-band noises with the same interaural information in both bands (dual-band/0.5- and 4-kHz dichotic); (3) both noises played simultaneously with a dichotic 0.5-kHz narrow-band noise and a diotic 4-kHz narrow-band noise (dual-band/0.5-kHz dichotic); and (4) both noises played simultaneously with a dichotic 4-kHz narrow-band noise and a diotic 0.5-kHz narrow-band noise (dual-band/4-kHz dichotic). Figure 1 shows that the normal-hearing group has smaller jnds than the impaired-hearing group for every signal condition and both groups perform most poorly in the dual-band/4-kHz dichotic condition, consistent with the literature (e.g., McFadden and Pasanen, 1976; Trahiotis and Bernstein, 1990). Just noticeable differences of >1000 μs were considered unmeasurable and were not specifically included in the results. Using this criterion, only four of the six impaired-hearing subjects could perform the task in the 4-kHz alone condition, and only one of the six impaired-hearing subjects could perform the task in the dual-band/4-kHz dichotic condition.

A 3-factor ANOVA (group$\times$signal condition $\times$repetition, $2\times3\times5$, between-within-within) performed on the group data reveals a significant difference due to group ($F_{1,105}=23.34$, $p<0.001$); differences due to signal condition and repetition were not significant. For this and all other statistical tests, the *a priori* level of significance was $p<0.05$. The 4-kHz alone and dual-band/4-kHz dichotic conditions are not included in this analysis due to the inability of some of the impaired-hearing subjects to obtain ITD thresholds less than 1000 μs. None of the tests for interaction were significant, indicating that the three factors are independent.

A 2-factor, within-group ANOVA (signal condition $\times$subject, 5×3, blocked on subject) reveals a significant difference for signal condition ($F_{4,8}=7.504$, $p<0.008$) in the normal-hearing subjects. A *post hoc* Duncan's test ($p<0.05$) indicates the normal-hearing group jnd for the dual-band/4-kHz dichotic condition (both narrow-band noises are present but the dichotic information is present in only the 4-kHz signal) is significantly larger than for all other signal conditions, which are statistically similar. The ANOVA run on the data for the impaired-hearing group (signal condition $\times$subject, 3×3, blocked on subject) reveals a significant difference due to subject and no significant differences for the impaired-hearing group for a 0.5-kHz dichotic target with the addition of a second narrow-band noise. A *post hoc* Duncan's test ($p<0.05$) indicates that impaired-hearing subject 6 performed poorer than all other subjects, who performed similarly. Although no statistical test was run on the 4-kHz target signal conditions for the impaired-hearing group, it is clear from the data displayed in Fig. 1(a) that impaired-hearing group performance on interaural time discrimination is degraded when a low-frequency interfering signal is added to a high-frequency target.

Group data for interaural intensity discrimination are shown in Fig. 1(b). The signal conditions are the same as those described for interaural time discrimination. The normal-hearing group has jnds that are smaller than those of the impaired-hearing group in all conditions except dual-band/0.5-kHz dichotic, where the group jnds are equal. This result may be anomalous because one of the three normal-hearing subjects exhibited great difficulty with this signal condition. While the normal-hearing group exhibits similar jnds in all conditions except dual-band/0.5-kHz dichotic, the impaired-hearing group exhibits the smallest jnd in the condition when the same interaural information appears in both narrow-band noises (dual-band/0.5- and 4-kHz dichotic), and the largest jnd when both narrow-band noises are present but the dichotic information is present in only the higher-frequency signal (dual-band/4-kHz dichotic). The dual-band/

4-kHz dichotic result in the impaired-hearing group for interaural intensity discrimination mirrors the dual-band/4-kHz dichotic result for both groups in interaural time discrimination, in which both groups exhibit their largest jnd.

A 3-factor ANOVA (group×signal condition ×repetition, 2×5×5, between-within-within) performed on the interaural intensity discrimination data reveals significant differences due to group ($F_{1,175}=36.38$, $p<0.001$) and signal condition ($F_{4,175}=8.91$, $p<0.001$); differences due to repetition were not significant. There was also a significant interaction between groups across signal conditions ($F_{4,175}=9.55$, $p<0.001$). This, combined with visual inspection of the data in Fig. 1(b), indicates that the normal-hearing group performs better than the impaired-hearing group on some tasks, but both groups perform similarly on other tasks.

A 2-factor, within-group ANOVA (signal condition ×subject, 5×3, blocked on subject) run for the normal-hearing group data reveals no significant differences among signal conditions for either target frequency with the addition of a second narrow-band noise. The differences between the interaural intensity just noticeable differences obtained from the normal-hearing group for the target signals in isolation versus those with an interfering noise band are comparable to those reported by Heller (1992). A similar within-group ANOVA run on the impaired-hearing subjects' jnds reveals a significant difference due to listening condition ($F_{4,20}=17.23$, $p<0.0001$) and subject ($F_{5,20}=7.16$, $p<0.0005$). A *post hoc* Duncan's test ($p<0.05$) indicates the impaired-hearing group jnd in the dual-band/4-kHz dichotic condition is significantly larger than all other signal conditions, which are not significantly different. This suggests that the impaired-hearing subjects have more difficulty discerning the interaural intensity just noticeable difference in a high-frequency target when a low-frequency interferer is present than with a high-frequency target in isolation.

II. EXPERIMENT 2: VIRTUAL LOCALIZATION

A. Method

This experiment was designed to measure the ability of subjects to localize a narrow-band noise source alone or in the presence of a second, interfering narrow-band noise at a different frequency. The same individuals who participated in experiment 1 served as subjects in experiment 2.

Signals for this experiment were processed in two steps. First, 20 samples of 300-ms narrow-band noise geometrically centered at 500 and 4000 Hz were generated with a 20-kHz sampling rate and 10-ms $\cos^2$ rise–fall ramps. Then, these noise samples were convolved with head related transfer functions measured using the KEMAR (Burkhard and Sachs, 1975). Impulse responses were measured at each of KEMAR's eardrums from nine locations spaced 22.5° apart in an arc of 180° in the horizontal plane. During the measurements, a speaker was placed approximately 1 m from KEMAR and moved to an appropriate location. The reverberation time in the measurement room was approximately 0.25 s below 800 Hz and 0.4 s above that frequency. Each of the 40 original noise files (20 at 500 Hz and 20 at 4000 Hz) was convolved with each of the 18 impulse responses and stored on the computer for on-line retrieval.

The nine virtual source locations were spaced 22.5° apart in an arc around the subjects' heads from +90° to −90° azimuth (+90° right, 0° straight ahead, −90° left) in the horizontal plane. The narrow band noises were presented such that with a source at 0° the levels were 77 dB SPL or 28 dB SL. Levels at each ear of the subjects with the source at other locations varied depending on the location of the source relative to the listener's ears (head-shadow effect). Three experimental conditions were included in the virtual localization experiment: (1) localization of individual narrow-band noises; (2) localization of each narrow-band noise as the target with the second narrow-band noise as the interferer, presented randomly from the +90°, −90°, and 0° locations; and (3) localization of each narrow-band noise as the target with the second narrow-band noise as the interferer, presented in a fixed location at +90°, 0°, or −90°, for a total of 10 conditions.

For signal condition (1) the procedure was a single-interval, 9-alternative, forced choice (1l, 9AFC) with feedback. For conditions (2) and (3) the task was a cued, single interval 9-alternative forced choice (1C, 1I, 9AFC), with feedback, with the target presented as a cue at the 0° location, followed 400 ms later by the target plus the interfering sound. Percent correct and root mean square error were calculated for each run and 9×9 confusion matrices were generated for each subject on every run.

Subjects were instructed to press a key from 1–9 on a response keyboard to indicate which one of nine possible locations they perceived as the source of the signal. A drawing of the possible locations relative to the head was mounted in front of the subjects. The paradigm was subject paced in that a response was required before a new trial could begin. All signals were presented via TDH-49P earphones. Subjects received one training run for each of the ten signal configurations prior to beginning data collection. In addition to the training run, all conditions were run twice. Each run included 6 presentations of the target from each location for a total of 54 trials. When the interferer was in a random location, the target was presented from each combination of target and interferer location twice, resulting in 18 trials with the interferer in each location on any one run. Therefore, the individual narrow-band noise and fixed interferer conditions include a total of 108 trials each (54 trials×2 repetitions) while the random interferer conditions include 36 trials each (18 trials×2 repetitions).

B. Results

The data for the localization experiment are plotted in panel *a* of Fig. 2 in root mean square (rms) error, calculated as follows:

$$\text{rms}=\sqrt{\frac{\Sigma_{i=1}^{n}(s_i-r_i)^2}{n}}, \tag{3}$$

where n equals the number of trials per run, s_i equals the stimulus location, and r_i equals the response location. The rms error in degrees indicates the average difference between

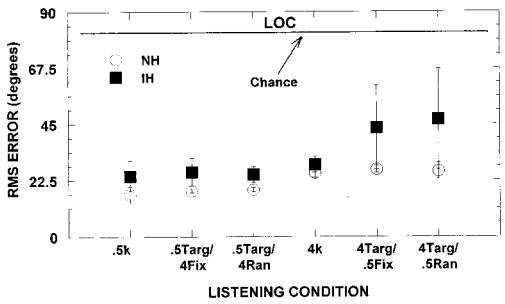


FIG. 2. Mean rms localization errors for normal- and impaired-hearing groups for the six signal conditions collapsed across interferer location. The ordinate represents the rms error in degrees and the abscissa the signal conditions. The three signal conditions with a 0.5-kHz target are to the left side of the panel, and the three with a 4-kHz target are to the right side of the panel. The open circles represent the means for the normal-hearing group and the filled squares the means for the impaired-hearing group.

the perceived source location and the actual source location, without taking direction into account. Within-group ANOVAs comparing the effect of interferer location indicate no significant differences for interferer location, at the $p<0.05$ level, in either group. Therefore, in Fig. 2 the data were collapsed across interferer location within each group. The signal conditions in Fig. 2 are named for the target and interferer; for example, 0.5Targ/4Fix indicates a 0.5-kHz narrow-band noise target with a 4-kHz fixed interferer and 4Targ/0.5Ran indicates a 4-kHz narrow-band noise target with a 0.5-kHz random interferer.

As can be seen in Fig. 2, the normal-hearing group performed similarly for all three conditions at each frequency, suggesting that performance is unaffected by the addition of an interfering signal. The same is true for the impaired-hearing group for the 0.5-kHz target conditions. It is interesting that although five of six impaired-hearing had 500-Hz pure-tone audiometric and 1/3-octave noise detection thresholds within the normal range, they are still less accurate at localizing a 0.5-kHz narrow-band noise than the normal-hearing listeners. When an interfering sound is introduced with the 4-kHz narrow-band noise as target, the impaired-hearing group rms error increases compared to the 4-kHz signal in isolation. The signal with a random interferer yields higher (but not significantly higher) rms errors than with a fixed interferer.

A 2-factor ANOVA was run on group×signal condition (2×6, between-within) collapsed across interferer location (no interferer, fixed interferer, random interferer). The results indicate a significant difference between groups ($F_{1,96}=31.16$, $p<0.05$), a significant difference among signal conditions ($F_{5,96}=9.48$, $p<0.05$), and a nonsignificant interaction of group and signal condition. A *post hoc* Duncan's test ($p<0.05$) indicates: (1) the three signal conditions with a 0.5-kHz target are not significantly different from one another; (2) the 4-kHz alone condition is not significantly different from both 0.5-kHz conditions with an interferer but is significantly different from 0.5 kHz alone; and (3) the two conditions with a 4-kHz target and an interferer are not significantly different from one another but are significantly different from all other conditions.

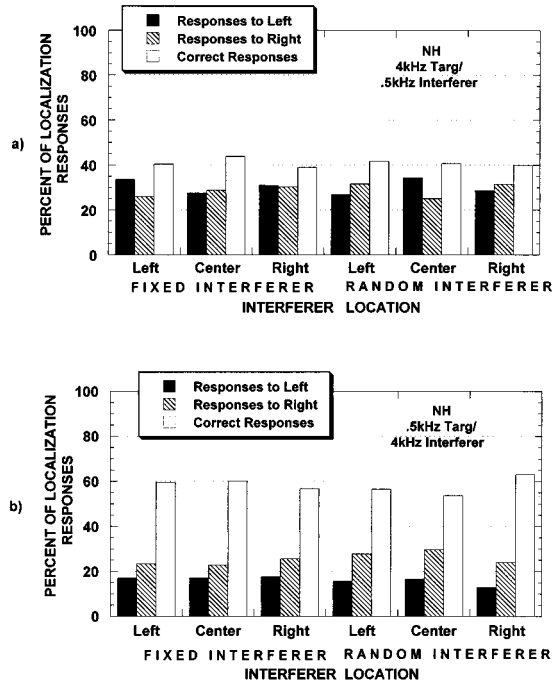


FIG. 3. (a) Normal-hearing group localization performance with a 4-kHz target and 0.5-kHz interferer; leftmost column in each set of three represents the percent of responses to the left of the actual source location, middle column represents the percent of responses to the right of the actual source location correct localization responses, and rightmost column represents percent correct. (b) Normal-hearing group localization performance with a 0.5-kHz target and 4-kHz interferer; leftmost column in each set of three represents the percent of responses to the left of the actual source location, middle column represents the percent of responses to the right of the actual source location correct localization responses, and rightmost column represents percent correct.

Panels (a) and (b) of Fig. 3 contain data which show the percent correct localization responses for the normal-hearing group, as well as the percent of responses which were to the left and to the right of the actual stimulus location, without taking distance into account. The data shown in these two graphs are for both fixed and random interferers. In the 4Targ/0.5Interferer conditions [panel 3(a)] the control group response errors are symmetric about the target location; however, in the 0.5Targ/4Interferer [panel 3(b)] it is apparent that the control group responded more often (incorrectly) to the right than to the left in each condition. These graphs are included primarily for comparison with the performance of the individual subjects with impaired hearing, whose data are located in Figs. 4–9. Note, however, the striking difference in percent correct for the 0.5-kHz target conditions compared to the 4-kHz target conditions.

Overall, the impaired-hearing listeners are less accurate than the normal-hearing listeners at indicating source location. Two-tailed paired t tests run for each group ($p=0.025$) on each interference condition indicate localization accuracy is significantly different with the high-frequency signal than the low-frequency signal for both subject groups.

To explore the relationship between interaural discrimination and localization, rank order correlations using Spearman's rho (r) were calculated for all the subjects with impaired hearing. Spearman's rho was not calculated for the subjects with normal hearing because visual inspection of

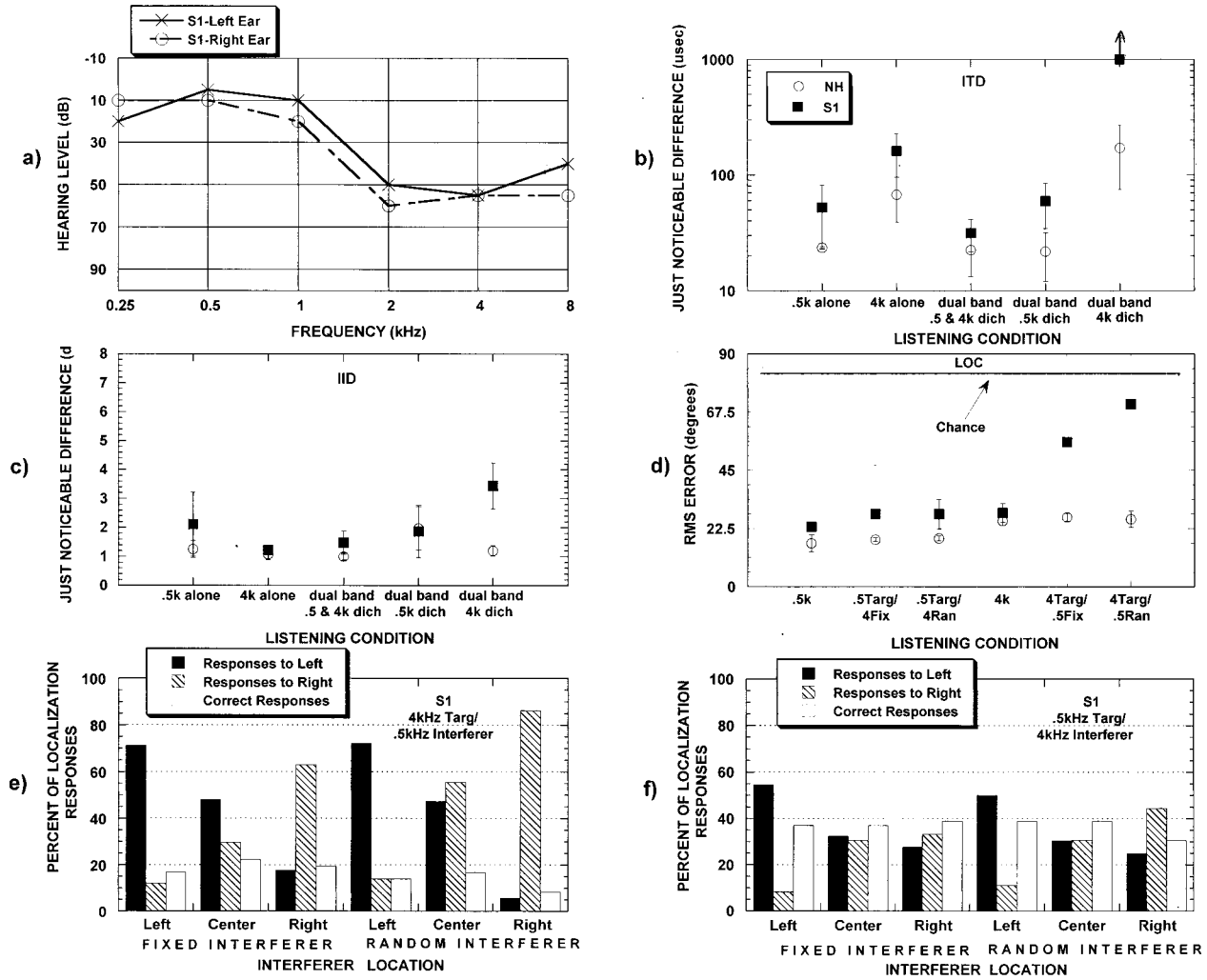


FIG. 4. Results for impaired-hearing S1 in each task, with the normal-hearing group means included for comparison. (a) Audiogram; (b) interaural time discrimination (ITD); (c) interaural intensity discrimination (IID); (d) virtual localization rms error (LOC); (e) 4-kHz target and 0.5-kHz interferer—percent of responses to the left and right of the actual source location and percent correct; (f) 0.5-kHz target and 4-kHz interferer—percent of responses to the left and right of the actual source location and percent correct.

scatterplots of localization accuracy versus interaural discrimination jnds determined that there was no clear relationship between performance on the two tasks. Localization of low-frequency signals is mediated primarily by timing information (in the fine structure, i.e., cycle-by-cycle, of the signal); therefore, the 0.5-kHz time jnd data are compared to the 0.5-kHz and 0.5Targ/4Interferer localization signal conditions. Interaural intensity information is used primarily in localizing high-frequency signals; accordingly, the 4-kHz signal localization data (4-kHz and 4Targ/0.5Interferer) are compared to the 4-kHz interaural intensity jnd data. The results, displayed in Table II, indicate nonsignificant correlations in every case, although interaural intensity performance correlates more highly with localization performance than does interaural time performance.

III. INDIVIDUAL RESULTS

Figures 4 through 9 contain individual data for each subject with impaired hearing and, for comparison, the mean for the normal-hearing group. Each figure (4–9) contains the following information: the upper left panel is the audiogram for that subject; the upper right panel consists of data from the interaural time discrimination task (ITD); the middle left panel contains data for the interaural intensity discrimination task (IID); the middle right panel includes rms error data for the localization task (Loc); the bottom two graphs display localization data plotted to determine the effect of the interferer location on performance. The bottom left panel contains data with a 4-kHz target and 0.5-kHz interferer and the bottom right panel data with a 0.5-kHz target and 4-kHz interferer; the leftmost column in each set of three indicates the percentage of responses to the left of the correct response, the middle column represents the percentage of responses to the right of the correct response, and the rightmost column indicates the percent of correct responses. The three sets of columns on the left side of each graph represent responses with a fixed interferer and the three sets of columns on the right side of each graph represent responses with a random interferer.

Figure 4 illustrates interaural time jnds of less than 100 μs for S1 in the conditions when the dichotic information is

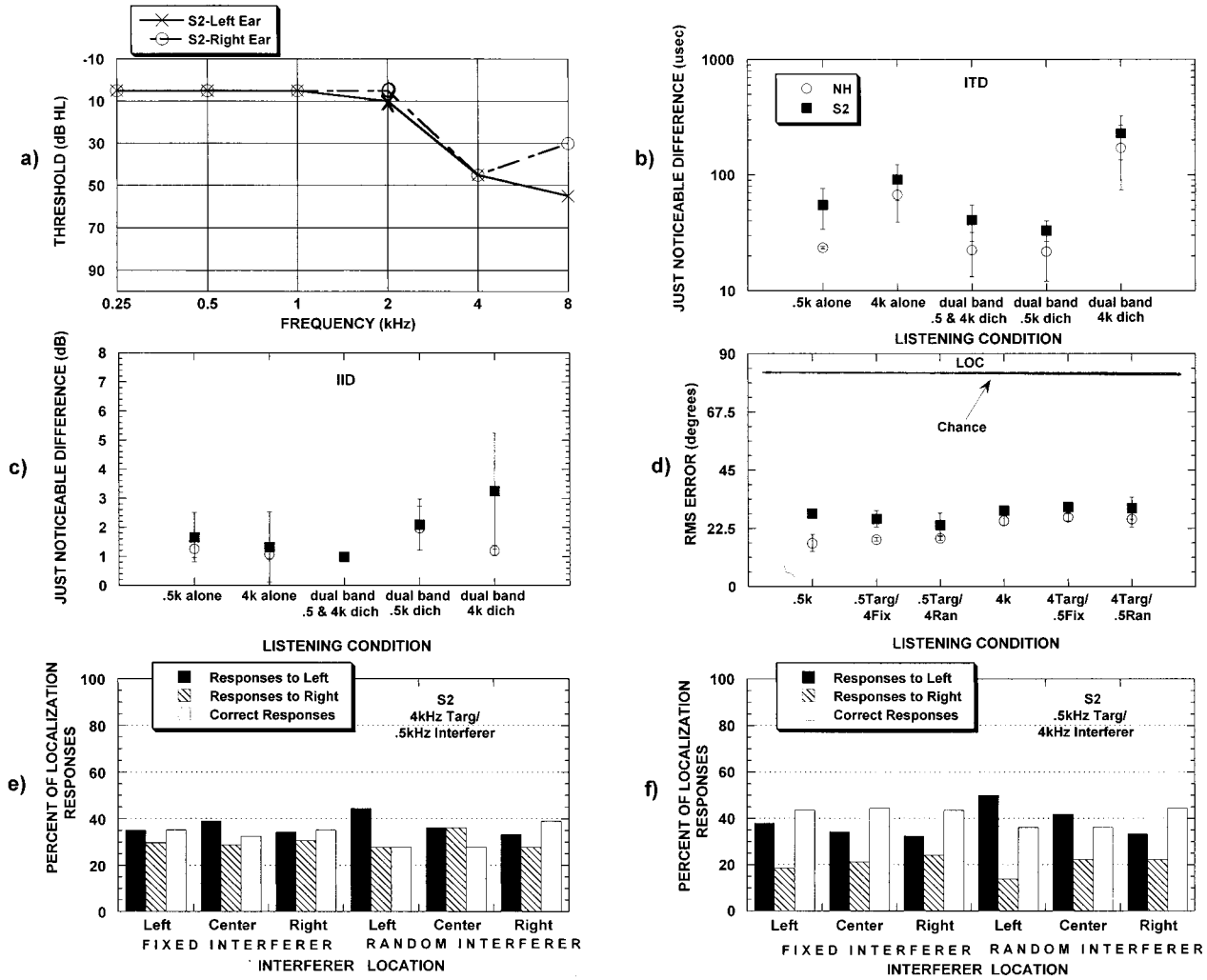


FIG. 5. Results for impaired-hearing S2 in each task, with the normal-hearing group means included for comparison. Panels (a)–(f) same as for Fig. 4.

present in the 0.5-kHz narrow-band noise. Her time jnd for the 4-kHz alone condition is between 100 and 200 μs but she cannot perform interaural time discrimination in the dual-band/4-kHz dichotic condition. Apparently, conflicting interaural time information at 0.5 kHz disrupts S1's interaural time discrimination at 4 kHz. For interaural intensity discrimination S1 has jnds comparable to normal except in the dual-band/4-kHz dichotic condition. Overall, her interaural jnds are only slightly larger than normal except in the dual-band/4-kHz dichotic condition. The rms localization errors for S1 are similar for all tasks except the two with a 4-kHz target with an interferer. S1 localizes the 4-kHz signal more accurately with a fixed than a random interferer, but her performance is much worse in these two conditions than all others. The bar graphs at the bottom of Fig. 4 indicate that the responses of S1 are affected by the virtual location of the interferer in the conditions with 4 kHz as the target. Looking at Fig. 4(e), when the 0.5-kHz interferer is at the left or right ear, her responses are pulled toward the virtual location of the interfering signal, whether the interferer is fixed or random. With a 0.5-kHz target, the effect of the interferer is less pronounced; however, with the 4-kHz interferer at $-90°$ her responses are pulled toward the left of the correct response half of the time.

Unlike each of the other subjects with impaired hearing, S2 (Fig. 5) can perform the interaural time discrimination task in all signal conditions, including dual-band/4-kHz dichotic. His jnds for 4 of the 5 signal conditions are comparable to the normal-hearing listeners' on the interaural intensity discrimination task, although his dual-band/4-kHz dichotic intensity jnd is larger than the normal-hearing listeners'. S2, whose interaural discrimination is only slightly poorer than that of the normal-hearing subjects, has similar virtual localization errors across all signal conditions like the normal-hearing subjects. He is the only subject in either group who demonstrates little difference between the two target frequencies. As can be seen in panel 5(e), S2 split his responses almost evenly to the left and right in every condition with a 0.5-kHz interferer. Panel 5(f) shows he consistently responded more often to the left than to the right with a 4-kHz interferer, whether the interferer was fixed or random.

S3 (Fig. 6) has interaural time jnds <100 μs in all conditions containing 0.5-kHz dichotic information. His interau-

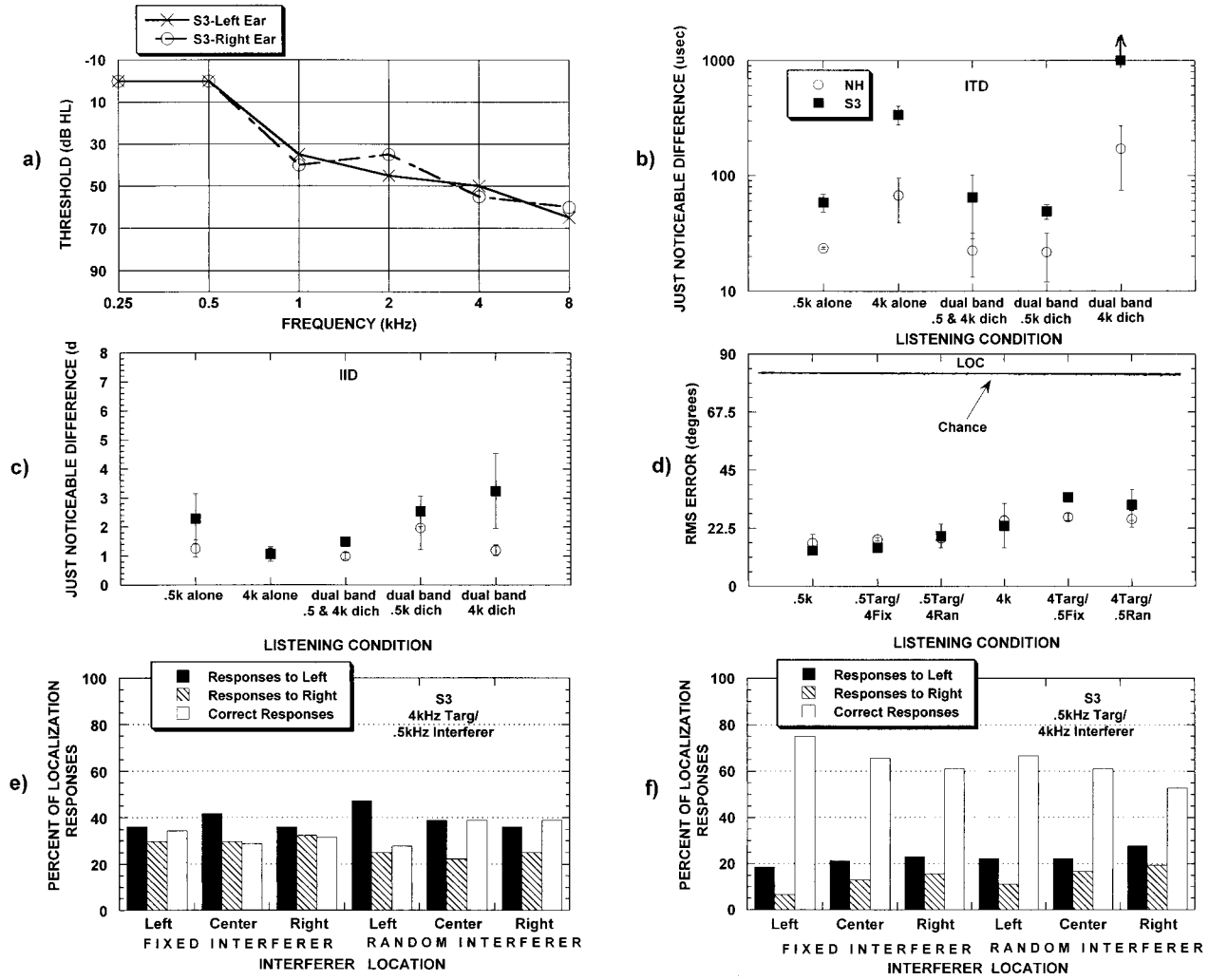


FIG. 6. Results for impaired-hearing S3 in each task, with the normal-hearing group means included for comparison. Panels (a)–(f) same as for Fig. 4.

ral time jnds are much larger than the normal-hearing listeners for the conditions in which the dichotic information is present in only the 4-kHz narrow-band noise, regardless of whether a second noise band is present. Interaural intensity jnds are elevated for the 0.5-kHz alone, dual-band/0.5-kHz dichotic and dual-band/4-kHz dichotic conditions compared to the 4-kHz alone and dual-band/0.5- and 4-kHz dichotic conditions. In the virtual localization task S3 performs comparably to the normal-hearing subjects except with a 4-kHz target and a 0.5-kHz fixed interferer. The data in panels 6(e) and (f) indicate he responds more often to the left than the right of the correct stimulus location when an interfering sound is present; however his performance does not appear to be systematically influenced by the interferer location. S3 is by far the most accurate of the impaired-hearing subjects with a 0.5-kHz target [Fig. 6(f)].

S4 (Fig. 7) has interaural time jnds <100 μs in all conditions containing 0.5-kHz dichotic information. S4 cannot perform interaural time discrimination in either the 4-kHz alone or the dual-band/4-kHz dichotic conditions, even with a 1000-μs interaural difference. Interaural intensity jnds for S4 are between 1.0 and 2.0 dB for all conditions except dual-band/4-kHz dichotic. In the dual-band/4-kHz dichotic condition, where his jnd was >5 dB, he may have been using monaural intensity information rather than interaural intensity information because the roving level was only 10 dB. For virtual localization S4 presents an rms error pattern different from the other experimental subjects; he has the most difficulty with the 0.5-kHz alone and 4Targ/0.5Ran conditions, and the least trouble with the 0.5Targ/4Ran and 4-kHz alone conditions. He exhibits a larger rms error whenever the 4-kHz target is present with an interferer, but little difference is noted between the fixed and random interference conditions. With a 4-kHz target and 0.5-kHz interferer [Fig. 7(e)], he demonstrates a tendency to indicate source location to the right of the correct one, especially when the interferer is located at $+90°$. With a 0.5-kHz target and 4-kHz interferer [Fig. 7(f)] there is little effect of interferer location with a fixed interferer, however, with a random interferer he responds more often to the right than the left of the actual stimulus location.

S5 (Fig. 8) exhibits interaural time jnds <100 μs in all conditions with the dichotic information in the 0.5-kHz narrow-band noise. S5 has an interaural time jnd almost a factor of 10 larger than the normal-hearing listeners in the 4-kHz alone narrow-band noise condition. She cannot perform interaural time discrimination in the dual-band/4-kHz dichotic condition, even with a 1000-μs interaural differ-

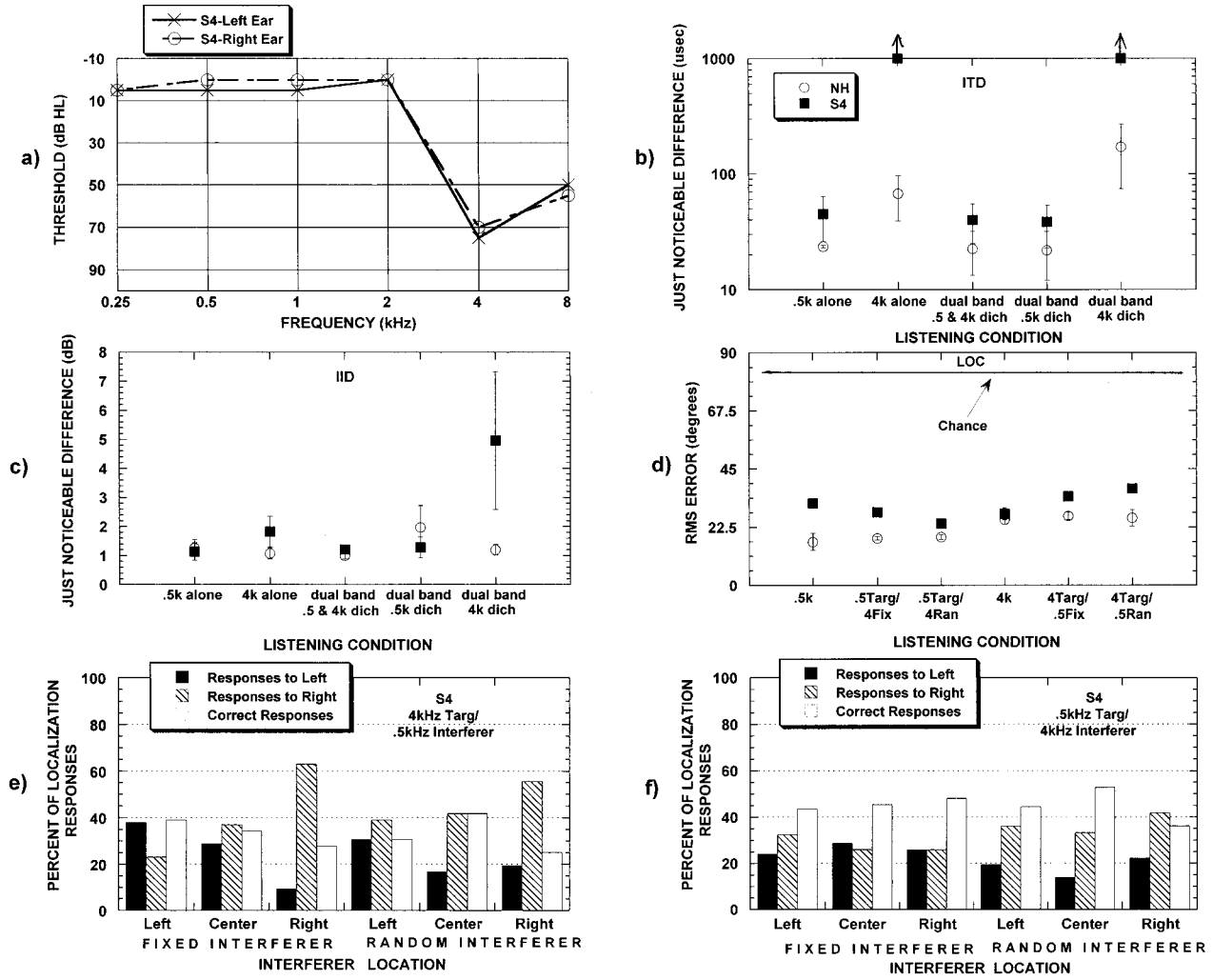


FIG. 7. Results for impaired-hearing S4 in each task, with the normal-hearing group means included for comparison. Panels (a)–(f) same as for Fig. 4.

ence. Interaural intensity jnds for S5 are between 1.0 and 2.5 dB for all conditions except dual-band/4-kHz dichotic, in which her jnd was slightly larger at about 4.0 dB. S5 performs similarly to but slightly poorer than the normal-hearing listeners on the virtual localization task. This subject displays a tendency to respond to the location of the interferer rather than the target in all conditions, whether 0.5 or 4 kHz is the target. Additionally, with a 4-kHz target, she tends to answer more often to the right of the correct stimulus when the interferer is straight ahead [Fig. 8(f)].

S6 (Fig. 9) exhibits interaural time jnds close to 100 μs in all conditions with the dichotic information in the 0.5-kHz narrow-band noise. Like S4, S6 cannot perform interaural time discrimination in the 4-kHz alone or dual-band/4-kHz dichotic conditions, even with a 1000-μs interaural difference. Interaural intensity jnd's for S6 are between 2.5 and 4.0 dB for each condition except dual-band/4-kHz dichotic. In the dual-band/4-kHz dichotic condition his jnd was >5 dB; therefore, S6 may have been using monaural intensity information rather than interaural intensity information. S6 exhibits the greatest difficulty of all the subjects in virtual localization of a 4-kHz target with a 0.5-kHz interferer, as can readily be seen by comparing panel (d) of Figs. 4–9. During data collection he stated some of the sounds appeared to emanate from behind his head. Nonetheless, his rms errors for the 0.5-kHz targets and 4-kHz target in isolation are quite good, although somewhat larger than for the normal-hearing listeners. Variability in his performance is low; however, when an interfering sound is introduced with the 4-kHz target, his performance is degraded and the variability increases. RMS localization error for this subject for the 4-kHz target with an interferer is near chance (about 3.5 locations, or 78.75°, chance$=82°$), but little difference is noted between the fixed and random interference conditions. Figure 9(e) and (f) show that his responses are pulled toward the location of the interfering sound, rather than the target signal, with the interferer at $+90°$ in all four target/interferer combinations. There does not seem to be a consistent pattern with the interferer at $-90°$ or 0°.

If we group the subjects with impaired hearing based on audiometric results and examine performance, we see that similar audiometric configuration is apparently not related to performance on the binaural tasks in the current study. For example, subjects 1, 3, and 6 have hearing within normal limits at 250–500/1000 Hz with sloping losses above these frequencies; none of these subjects can perform interaural

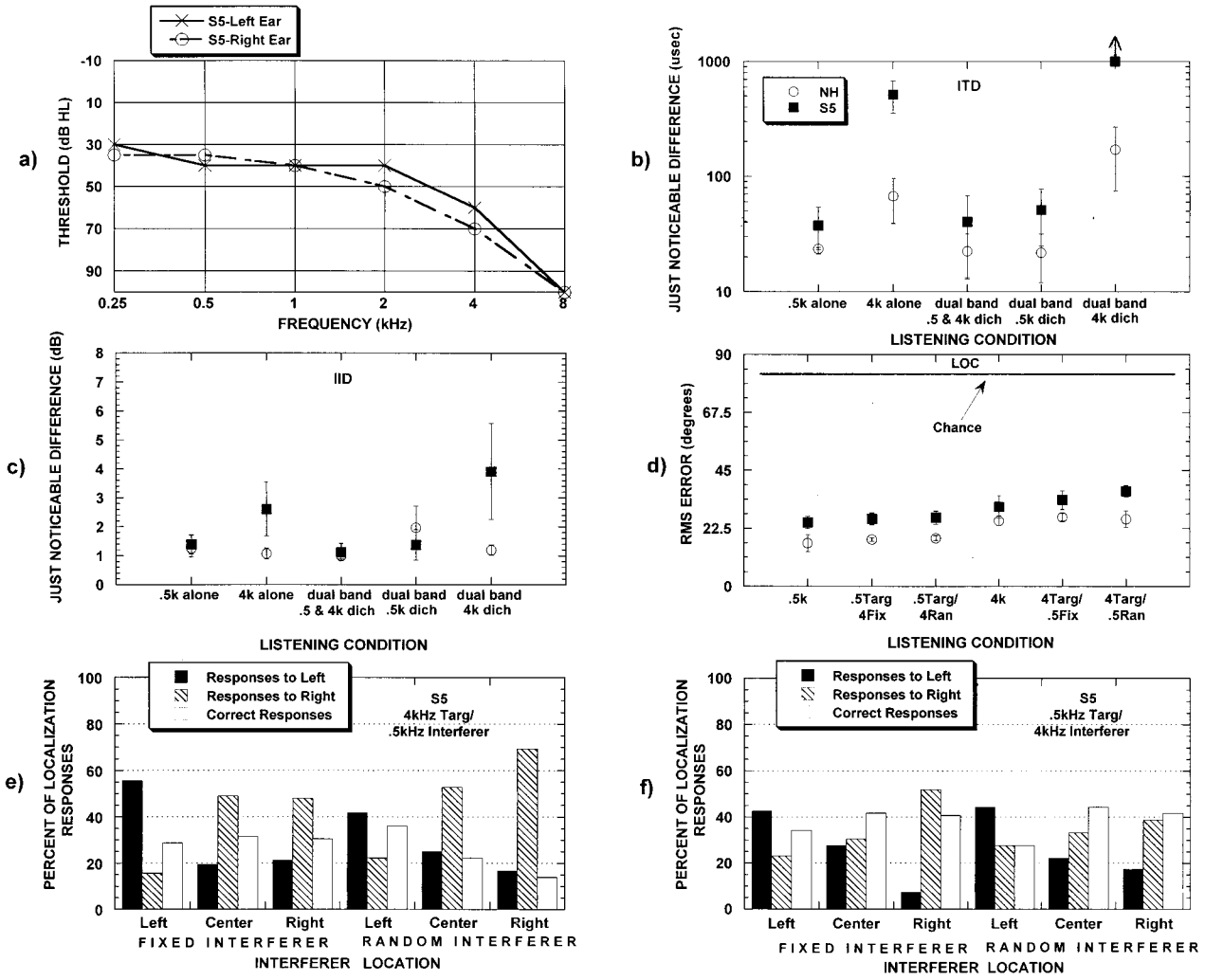


FIG. 8. Results for impaired-hearing S5 in each task, with the normal-hearing group means included for comparison. Panels (a)–(f) same as for Fig. 4.

time discrimination in the dual-band/4-kHz dichotic condition, however, subjects 1 and 3 can perform the task in the 4-kHz alone condition while S6 cannot. For interaural intensity discrimination, subjects 1 and 3 again have similar jnds in each condition while subject 6 has larger jnds in each condition. For virtual localization, subjects 1 and 6 performed similarly (i.e., their responses were pulled toward the interferer in many conditions) and subject 3's performance was similar to the normal-hearing group and unlike that of the other two subjects with similar losses. Likewise, a comparison can be made between subjects 2 and 4, who both have hearing within normal limits from 250 to 2000 Hz and a ''noise notch'' above that frequency; that is, their hearing threshold is poorest at 4 kHz and improves at 8 kHz. For interaural time discrimination S2's jnds were close to those of the control group in each condition, yet S4 could not perform the task with a 4-kHz target signal, with or without an interfering signal. Their performance was similar on interaural intensity discrimination and rms localization error, however S4's incorrect responses were more often to the right while S2's incorrect responses were more often to the left. Subject 5 is not comparable to any of the other subjects with impaired hearing, as she has a mild loss in the low frequencies as well as a sloping high-frequency loss. Despite this fact, it is interesting that her performance on interaural time discrimination is most similar to that of S3; interaural intensity discrimination is most like S4, and localization rms error is comparable to S2.

IV. DISCUSSION

The interaural time jnd's of the normal-hearing subjects for signals in isolation are comparable to those obtained by other researchers (e.g., Klumpp and Eady, 1956; Zwislocki and Feldman, 1956; McFadden and Pasanen, 1976; Hawkins and Wightman, 1980; Smoski and Trahiotis, 1986; Koehnke *et al.*, 1986; Heller, 1992; Koehnke *et al.*, 1995), as are the jnds for the narrow-band noises in isolation obtained for the impaired-hearing subjects (e.g., Hawkins and Wightman, 1980; Smoski and Trahiotis, 1986; and Koehnke *et al.*, 1995). The differences between the interaural time just noticeable differences obtained from the normal-hearing group for the 4-kHz target signal in isolation versus those with an

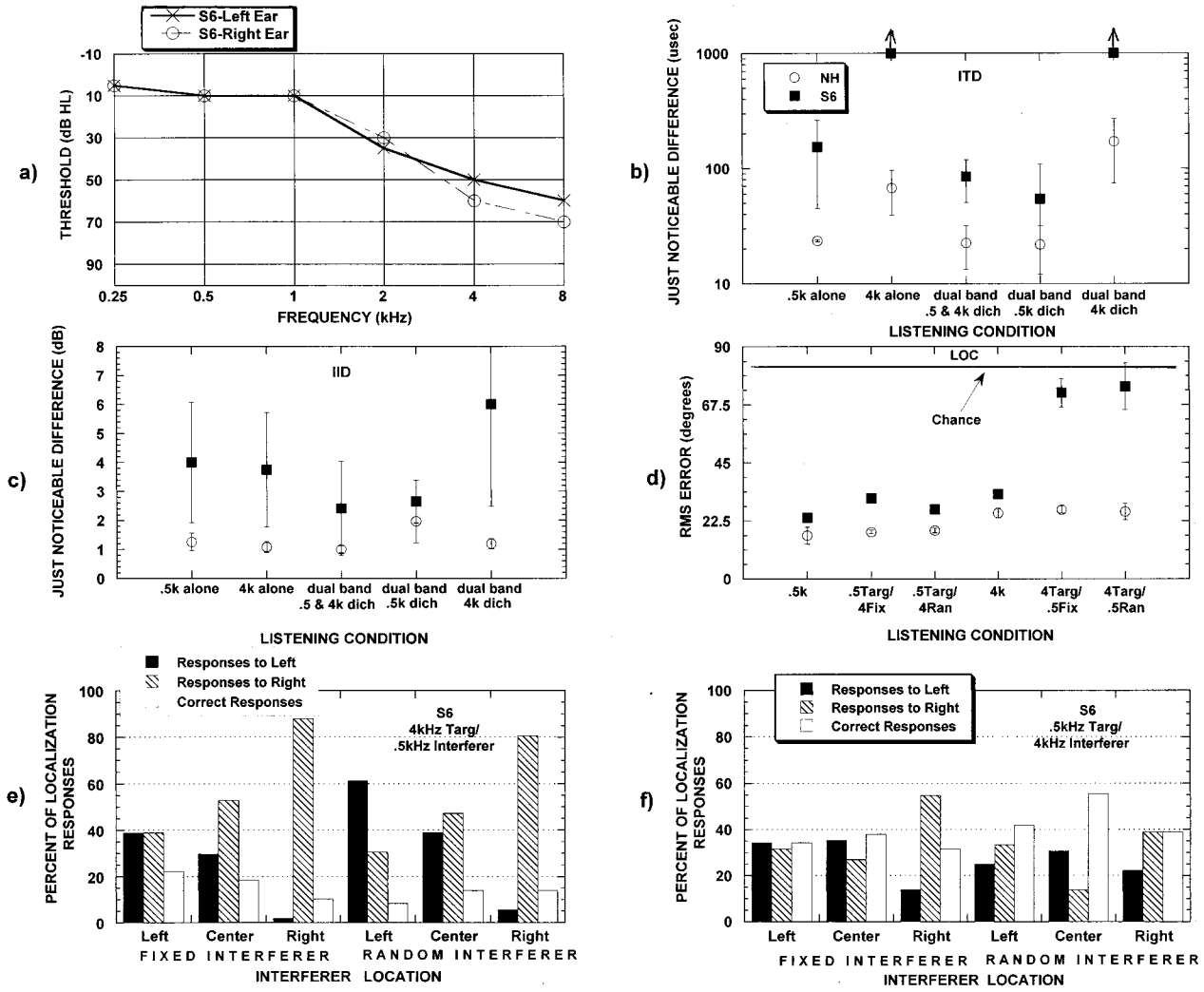


FIG. 9. Results for impaired-hearing S6 in each task, with the normal-hearing group means included for comparison. Panels (a)–(f) same as for Fig. 4.

interfering noise band are comparable to those reported by Heller (1992) and McFadden and Pasanen (1976), who also used widely spaced narrow bands of noise. Other researchers (e.g., Buell and Trahiotis, 1993; Stellmack and Dye, 1993; Woods and Colburn, 1992) who have examined interaural time just noticeable differences report larger just noticeable differences than in the current study. However, in those studies the target and interfering sounds were pure tones and were much closer in frequency than in the current study, indeed, at times the target and interferer resided within the same critical band. Also, duration of the target and interfering tones in Buell and Trahiotis (1993), Stellmack and Dye (1993), and Woods and Colburn (1992) was shorter by about $\frac{1}{2}$ (200 ms) to $\frac{3}{4}$ (100 ms) than the 400-ms signals used in the present study, which, when combined with the frequency spacing differences may account for the differences in results. For both groups, the jnd's for the 0.5-kHz narrow-band noise are smaller than for the 4-kHz narrow-band noise. The interaural time jnd's for the 0.5-kHz alone, dual-band/0.5- and 4-kHz dichotic, and dual-band/0.5-kHz dichotic signal conditions are not significantly different between or within groups. Although statistical comparison is not possible with

TABLE II. Rank order correlation results for the localization signal conditions of the impaired-hearing listeners. Note that none of the Spearman's Rs are statistically different from 0.

Localization signal condition	Discrimination signal condition	Rank order correlation
0.5 kHz	0.5 kHz—Time	−0.3714
4 kHz	4 kHz—Intensity	0.6000
0.5 k Targ/4 k Interferer	1) 0.5 kHz—Time	0.0857
	2) dual band/0.5 kHz dichotic	0.3142
4 k Targ/0.5 k Interferer	1) 4 kHz—Intensity	0.6000
	2) dual band/4 kHz dichotic	0.8142

the 4-kHz alone and dual-band/4-kHz dichotic signals due to the inability of some impaired-hearing listeners to obtain ITD thresholds less than 1000 μs, the average jnd for the impaired-hearing group is four times larger than the normal-hearing group for 4-kHz alone, and five of the six impaired-hearing subjects cannot discriminate 1000-μs interaural time differences with the dual-band/4-kHz dichotic signal.

The inability of some of the impaired-hearing subjects to perform the interaural time discrimination task with the 4-kHz narrow-band noise in isolation is consistent with the data of some of Koehnke *et al.*'s (1995) and Gabriel *et al.*'s (1992) subjects. On the other hand, at least one impaired-hearing subject in the present study (S2 with a mild high-frequency loss) has interaural time jnds comparable to normal-hearing listeners' for the 4-kHz narrow-band noise target. Neither Koehnke *et al.* (1995) nor Gabriel *et al.* (1992) reported impaired-hearing subjects with near-normal interaural discrimination performance across frequency. Intersubject variability is often reported in binaural experiments, especially among impaired-hearing subjects (e.g., Hawkins and Wightman, 1980; Koehnke *et al.*, 1995; Häusler *et al.*, 1983), and was observed in the present study.

The normal-hearing subjects' interaural intensity jnds of about 1.0 dB for narrow-band noises in isolation are equivalent across frequency and are similar to those reported by Heller (1992) and Koehnke *et al.* (1986). The impaired-hearing listeners jnd's of 3–6 dB are comparable to those reported by Gabriel *et al.* (1992) and Koehnke *et al.* (1995). The impaired-hearing subjects' jnd's for the high-frequency narrow-band noise in isolation are comparable to the jnd's for the low-frequency narrow-band noise. This indicates that all the impaired-hearing subjects, regardless of the degree of hearing loss at 4000 Hz, are able to process interaural intensity information. This finding is important because an interaural intensity difference (due to the head-shadow effect) is the primary binaural cue available in high-frequency signals in free-field listening situations. Additionally, good performance on the interaural intensity discrimination task with a 4-kHz signal, in contrast to poor performance on interaural time discrimination with a 4-kHz signal, indicates that the difficulties with time discrimination are not due entirely to the relative sensation level of the signal at 4 kHz.

When a second stimulus band with different spectral content and *congruent* interaural information is added in the interaural time discrimination experiment (i.e., dual-band/0.5- and 4-kHz dichotic versus 4-kHz alone or 0.5-kHz alone), neither the normal-hearing nor the impaired-hearing group demonstrates significantly smaller jnd's, suggesting that interaural time discrimination of the congruent signals is dominated by the 0.5-kHz narrow-band noise. The impaired-hearing group does have a significantly smaller intensity jnd in the dual-band/0.5- and 4-kHz dichotic condition compared to the 4-kHz alone condition, which suggests they can benefit from congruent interaural intensity information in more than one spectral region, despite the fact that the interaural intensity cues at 0.5 kHz are quite small under free-field listening conditions.

It is interesting that the impaired-hearing subjects have similar performances in interaural time and intensity discrimination in the dual-band/4-kHz dichotic condition; i.e., in both cases the jnd for dual-band/4-kHz dichotic increases over 4-kHz alone narrow-band noise. Unlike interaural time discrimination, 4 of 6 impaired-hearing subjects *could* discriminate interaural intensity differences with the dual-band/4-kHz dichotic signal. The two other impaired-hearing subjects may have used monaural intensity cues rather than interaural intensity cues as evidenced by their large jnds relative to the size of the overall roving level. Consistent with reports by other researchers, more of the impaired-hearing listeners in this study were able to perform interaural intensity discrimination than interaural time discrimination for the same signals (Hawkins and Wightman, 1980; Gabriel *et al.*, 1992; Koehnke *et al.*, 1995).

Results with incongruent interaural information across frequency (dual-band/0.5-kHz dichotic, dual-band/4-kHz dichotic) suggest that: (1) normal-hearing listeners can ignore conflicting interaural information in an interaural time discrimination task and (probably) conflicting information in an interaural intensity discrimination task; and (2) impaired-hearing listeners can ignore conflicting interaural time and intensity information with a low-frequency target and a high-frequency interferer (dual-band/0.5-kHz dichotic), but not with a high-frequency target and low-frequency interferer (dual-band/4-kHz dichotic). As seen in Figs. 4–9 [panel (c)] most of the impaired-hearing subjects have greater difficulty detecting interaural intensity differences in a high-frequency signal in the presence of a low-frequency interferer than interaural differences in a high-frequency signal in isolation.

This result, as well as the results for the other dual-band stimuli (dual-band/0.5-kHz dichotic and dual-band/0.5- and 4-kHz dichotic), may have been affected by the relative level of the stimuli in the two bands. The 0.5-kHz narrow-band noise was, on average, 37 dB higher in sensation level than the 4-kHz narrow-band noise (range=21–49 dB) when the impaired-hearing subjects listened to the double-band stimuli, which may have influenced their ability to perform the task. However, comparing the performances of IH S2 and IH S4 (who heard the low- and high-frequency stimuli with less than a 30-dB difference in sensation level) with those of the other four impaired-hearing subjects (who heard the low- and high-frequency stimuli with greater than a 30-dB difference in sensation level) reveals little difference in the ability to perform the tasks. In fact, IH S4, who had the smallest difference in sensation level between the low- and high-frequency stimuli, could not perform the time discrimination task in the dual-band/4-kHz dichotic condition. The possible effects of level are important to consider in understanding how impaired-hearing listeners' auditory systems function in binaural processing tasks. This question is presently being investigated. Preliminary data suggest sensation level is not the only factor at work for the impaired-hearing listeners, and complete results are forthcoming.

Simultaneous gating of the target and interfering signals may also account, at least in part, for the large jnds obtained in the dual-band/4-kHz dichotic condition. Trahiotis and Bernstein (1990) reported larger jnds with normal-hearing subjects for interaural time discrimination of a high-frequency target and low-frequency interfering noise gated

simultaneously than with the interfering noise presented continuously. It is possible that the normal-hearing and impaired-hearing subjects in the present study would also obtain smaller jnd's with a continuous interferer.

There are several other possible explanations for the impaired-hearing listeners' inability to discriminate interaural time differences in the dual-band/4-kHz dichotic condition. First, it is likely among listeners with sensorineural hearing loss of cochlear origin that there is decreased frequency selectivity in the basal region of the cochlea as a result of cochlear damage (Evans and Harrison, 1975; Liberman and Dodds, 1984; Florentine *et al.*, 1980). Studies with two animal models [guinea pig (Evans and Harrison, 1975), cat (Liberman and Dodds, 1984)], and humans (Florentine *et al.*, 1980) support this view. Additionally, listeners with hearing loss of cochlear origin may experience increased upward spread of masking compared to normal-hearing listeners (e.g., Gagne, 1988; Florentine *et al.*, 1980). Upward spread of masking is a peripheral, monaural explanation; i.e., cochlear dysfunction is responsible for the performance measured during the Gagne (1988) and Florentine *et al.* (1980) detection experiments. A cochlear dysfunction component is probably necessary but not sufficient to explain binaural performance in individuals with peripheral hearing loss.

Another explanation for the results of the impaired-hearing subjects' interaural discrimination with the dual-band/4-kHz dichotic signal is to consider that the interference is a central effect. Trahiotis and Bernstein (1990) reported that the jnd in normal-hearing listeners for a diotic, continuous interferer was essentially the same as that for the 4-kHz narrow-band noise in isolation, but the jnd for an interaurally uncorrelated, continuous interferer was almost double that of the 4-kHz narrow-band noise in isolation. Based on these results, the authors argue that a peripheral, monaural explanation is precluded. Heller (1992) also argued against a peripheral, monaural explanation when she demonstrated that interaural time and intensity discrimination tasks were differentially affected in normal-hearing listeners by the frequency of the target and interferer, i.e., target low frequency/interferer high frequency and target high frequency/interferer low frequency.

The normal-hearing listeners in the current study exhibited the same pattern of results as Heller's subjects; however, the impaired-hearing subjects displayed results different from the normal-hearing listeners. Because the jnds for the dual-band/4-kHz dichotic signals in both interaural time and interaural intensity discrimination of the impaired-hearing subjects increased with a high-frequency target and low-frequency interferer compared to the 4-kHz alone signal, upward spread of masking may explain some of their results in the present study, particularly in light of the relative sensation level of the signals in the two frequency bands.

Localization results in the current study for narrow-band noises in isolation agree well with previously published research for both normal- and impaired-hearing listeners (Nordlund, 1964; Tonning, 1975; Noble *et al.*, 1994; Besing and Koehnke, 1995). When an interferer is added in the localization task, some of the impaired-hearing listeners exhibit a tendency to shift the perceived source location toward the position of the interferer, more so with the high-frequency target signal/low-frequency interfering signal condition than the reverse. Heller (1992) reported her normal-hearing subjects tended to shift their identification of source location in the direction of the interfering signal in a lateralization task; in contrast, the normal-hearing subjects in the present study demonstrated only a slight tendency to localize to the right of the correct virtual location for both spectral configurations (target low frequency/interferer high frequency and target high frequency/interferer low frequency).

The results of this virtual localization experiment present some novel information. First, it is interesting to note that the random interferer apparently does not affect rms localization error differently from the fixed interferer for either subject group. It was expected that as signal uncertainty increased (i.e., the interferer is random), subjects would have more difficulty and localization accuracy would decrease. This expectation arises from reports of decrease in performance with increase of signal uncertainty for detection of tones (Wright and Dai, 1994), comodulation masking release (Grose and Hall, 1990), and intensity discrimination (Spiegel *et al.*, 1981). Perhaps because there were only three possible locations for the interferer, most of the listeners rapidly learned, then ignored, the interferer at those locations. This seems true of the normal-hearing listeners, as they show little tendency to shift their responses toward the location of the interfering signal. Although there was a tendency for some impaired-hearing listeners to shift responses toward an interfering signal, overall accuracy was not different in the fixed and random interference conditions.

A second observation regarding the results of the current study is that the addition of an interfering noise band did not significantly affect localization accuracy for most of the subjects. An explanation may be found in studies which have investigated the dominance of the low frequency portion of a broadband signal in localization (Yost *et al.*, 1971; Wightman and Kistler, 1992). Researchers have reported that with a broadband or multiple component signal which emanates from the same source location (i.e., all interaural information is congruent across frequency), listeners rely on interaural information in the low-frequency portion of that signal (Yost *et al.*, 1971). More recently, evidence has been presented which suggests that listeners rely exclusively on interaural time information which is present in the low-frequency portion (Wightman and Kistler, 1992) of a broadband signal, as opposed to interaural intensity and/or monaural spectral information contained in the high-frequency portion, to localize a broadband signal. With a broadband signal restricted to the high frequencies ($>$2.5 kHz), Wightman and Kistler (1992) reported the interaural timing information contained in the envelope of the signal is effectively ignored in favor of the interaural intensity and monaural spectral information. This suggests that the low- and high-frequency signals used in the current study may have been localized using separate mechanisms which did not interfere with one another.

Of the six impaired-hearing subjects in the current study, two (S1 and S6) experienced great difficulty in localizing the 4-kHz narrow-band noise with a 0.5-kHz narrow-band noise

interferer. For subjects 1 and 6 it seems reasonable that this poor localization ability is due to a level effect and their inability to ignore the low-frequency interferer. In terms of the level effect, the interfering signal at 0.5 kHz had a much higher sensation level than the target at 4 kHz, although the 4-kHz narrow-band noise was intense enough to be heard by each subject. The other four subjects appeared able to ignore the low-frequency interferer despite the great difference in sensation level. This issue is also being investigated in our laboratory.

The Spearman's rho analysis of the data of the impaired-hearing group yielded nonsignificant correlations for all conditions. Given the variability among the performances of the impaired-hearing subjects on the interaural discrimination and localization tasks, there does not appear to be a clear, consistent effect of high-frequency sensorineural hearing loss on binaural performance. The possible relationship between interaural discrimination and localization also warrants further investigation.

V. SUMMARY AND CONCLUSIONS

Virtual localization and interaural discrimination of the normal-hearing listeners is good in all conditions. For the impaired-hearing listeners, results of this study do not show a clear or strong relationship between interaural discrimination and localization performance. The variability in performance measured for the impaired-hearing subjects supports the notion that binaural performance is not related to the audiogram. Results of the interaural discrimination experiment suggest the following:

(1) Both normal- and impaired-hearing listeners can ignore a high-frequency interfering signal but not a low-frequency interfering signal when discriminating interaural time differences;
(2) Normal-hearing listeners can ignore a 0.5-kHz narrow-band noise interfering signal in interaural intensity discrimination but cannot ignore a 4-kHz narrow-band noise interferer. Although one of the three normal-hearing subjects had trouble with the dual-band/0.5-kHz dichotic condition, it is likely that most normal-hearing listeners are able to ignore a low-frequency interfering signal in an intensity discrimination paradigm;
(3) In contrast, the impaired-hearing subjects can ignore the 4-kHz narrow-band noise interfering signal but not the 0.5-kHz narrow-band noise interfering signal in interaural intensity discrimination.

Results of the virtual localization experiment suggest the following:

(1) Both normal- and impaired-hearing listeners can ignore high-frequency interference with a low-frequency target signal;
(2) Some impaired-hearing listeners have great difficulty localizing a high-frequency target signal in the presence of a low-frequency interfering signal presented at the same sound pressure level (this may be due to a level effect);
(3) Impaired-hearing listeners have a greater tendency than normal-hearing listeners to shift responses toward an interfering signal, though this tendency does not affect the overall rms localization error.

[1]For interaural level differences between $\geqslant 5$ and <10 dB, the stepsize was 2.0 dB; $\geqslant 2$ and <5 dB, stepsize was 1.0; $\geqslant 1$ and <2, 0.6 dB; and $\geqslant 0.1$ and <1, 0.4 dB.

ANSI (**1969**). ANSI S3.6-1969, ''Specifications for audiometers'' (American National Standards Institute, New York).

Bernstein, L. R., and Trahiotis, C. (**1982**). ''Detection of interaural delay in high-frequency noise,'' J. Acoust. Soc. Am. **71**, 147–152.

Besing, J., and Koehnke, J. (**1995**). ''A test of virtual auditory localization,'' Ear and Hearing **16**, 220–229.

Buell, T. N., and Hafter, E. R. (**1991**). ''Combination of binaural information across frequency bands,'' J. Acoust. Soc. Am. **90**, 1894–1900.

Buell, T. N., and Trahiotis, C. (**1993**). ''Interaural temporal discrimination using two sinusoidally amplitude-modulated, high-frequency tones: Conditions of summation and interference,'' J. Acoust. Soc. Am. **93**, 480–487.

Burkhard, M. D., and Sachs, R. M. (**1975**). ''Anthropometric manikin for acoustic research,'' J. Acoust. Soc. Am. **58**, 214–222.

Colburn, H. S., Barker, M. A., and Milner, P. (**1982**). ''Free-field tests of hearing-impaired listeners: Early results,'' in *Binaural Effects in Normal and Impaired Hearing*, edited by O. J. Pedersen and T. Poulsen, Scand. Audiol. Supplement **15**, 123–133.

Dye, R. H. (**1990**). ''The combination of interaural information across frequencies: Lateralization on the basis of interaural delay,'' J. Acoust. Soc. Am. **88**, 2159–2170.

Evans, E. F., and Harrison, R. V. (**1975**). ''Correlation between cochlear outer hair cell damage and deterioration of cochlear nerve tuning properties in the guinea-pig,'' Proc. Physiol. Soc. **256**, 43–44.

Florentine, M., Buus, S., Scharf, B., and Zwicker, E. (**1980**). ''Frequency selectivity in normally-hearing and hearing-impaired observers,'' J. Speech Hear. Res. **23**, 646–669.

Gabriel, K. J., Koehnke, J., and Colburn, H. S. (**1992**). ''Frequency dependence of binaural performance in listeners with impaired binaural hearing,'' J. Acoust. Soc. Am. **91**, 336–347.

Gagne, J. P. (**1988**). ''Excess masking among listeners with a sensorineural hearing loss,'' J. Acoust. Soc. Am. **83**, 2311–2322.

Grantham, D. W. (**1984**). ''Interaural intensity discrimination: insensitivity at 1000 Hz,'' J. Acoust. Soc. Am. **75**, 1191–1194.

Grose, J. H., and Hall, J. W. (**1990**). ''The effect of signal-frequency uncertainty on comodulation masking release,'' J. Acoust. Soc. Am. **87**, 1272–1277.

Häusler, R., Colburn, S., and Marr, E. (**1983**). ''Sound localization in subjects with impaired hearing: Spatial-discrimination and interaural-discrimination tests,'' Acta Oto-Laryngol. Suppl. **400**, .

Hawkins, D. B., and Wightman, F. L. (**1980**). ''Interaural time discrimination ability of listeners with sensorineural hearing loss,'' Audiology **19**, 495–507.

Heller, L. (**1992**). ''Across-frequency influences in lateralization for interaural time and level differences,'' Unpublished dissertation, University of Pennsylvania.

Jongkees, B. W., and van der Veer, R. A. (**1957**). ''Directional hearing capacity in hearing disorders,'' Acta Oto-Laryngol. **48**, 465–474.

Klumpp, R. G., and Eady, H. R. (**1956**). ''Some measurements of interaural time difference thresholds,'' J. Acoust. Soc. Am. **28**, 859–860.

Koehnke, J., and Besing, J. (**1996**). ''Binaural performance in listeners with impaired hearing: Aided and unaided results,'' in *Proceedings of the Conference on Binaural and Spatial Hearing*, edited by R. Gilkey and T. Anderson (Erlbaum, Mahwah, NJ).

Koehnke, J., Colburn, H. S., and Durlach, N. I. (**1986**). ''Performance in several binaural interaction experiments,'' J. Acoust. Soc. Am. **79**, 1558–1563.

Koehnke, J., Culotta, C. P., Hawley, M., and Colburn, H. S. (**1995**). ''Effects of reference interaural time and intensity differences on binaural performance in listeners with normal and impaired hearing,'' Ear Hear. **16**, 331–353.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Liberman, M. C., and Dodds, L. W. (**1984**). ''Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves,'' Hearing Res. **16**, 55–74.

McFadden, D., and Pasanen, E. G. (**1976**). ''Lateralization at high frequencies based on interaural time differences,'' J. Acoust. Soc. Am. **59**, 634–639.
Noble, W., Byrne, D., and Lepage, B. (**1994**). ''Effects on sound localization of configuration and type of hearing impairment,'' J. Acoust. Soc. Am. **95**, 992–1005.
Nordlund, B. (**1964**). ''Directional audiometry,'' Acta Oto-Laryngol. **57**, 1–18.
Smoski, W. J., and Trahiotis, C. (**1986**). ''Discrimination of interaural temporal disparities by normal-hearing listeners and listeners with high-frequency sensorineural hearing loss,'' J. Acoust. Soc. Am. **79**, 1541–1547.
Spiegel, M. F., Picardi, M. C., and Green, D. M. (**1981**). ''Signal and masker uncertainty in intensity discrimination,'' J. Acoust. Soc. Am. **70**, 1015–1019.
Stellmack, M. A., and Dye, R. H. (**1993**). ''The combination of interaural information across frequencies: The effects of number and spacing of components, onset asynchrony, and harmonicity,'' J. Acoust. Soc. Am. **93**, 2933–2947.
Tonning, F. M. (**1975**). ''Auditory localization and its clinical applications,'' Audiology **14**, 368–380.
Trahiotis, C., and Bernstein, L. R. (**1990**). ''Detectability of interaural delays over select spectral regions: Effects of flanking noise,'' J. Acoust. Soc. Am. **87**, 810–813.
Wenzel, E. M., and Hafter, E. R. (**1985**). ''Lateralization of clicks based on interaural time: Additivity of information across frequency,'' J. Acoust. Soc. Am. **74**, S85.
Wightman, F., and Kistler, D. (**1992**). ''The dominant role of low-frequency interaural time differences in sound localization,'' J. Acoust. Soc. Am. **91**, 1648–1661.
Woods, W. S., and Colburn, H. S. (**1992**). ''Test of a model of auditory object formation using intensity and interaural time difference discrimination,'' J. Acoust. Soc. Am. **91**, 2894–2902.
Wright, B. A., and Dai, H. (**1994**). ''Detection of unexpected tones in gated and continuous maskers,'' J. Acoust. Soc. Am. **95**, 939–948.
Yost, W. A., Wightman, F. L., and Green, D. M. (**1971**). ''Lateralization of filtered clicks,'' J. Acoust. Soc. Am. **50**, 1526–1530.
Zwislocki, J., and Feldman, R. S. (**1956**). ''Just noticeable differences in dichotic phase,'' J. Acoust. Soc. Am. **28**, 860–864.

Diotic and dichotic detection using multiplied-noise maskers

Steven van de Par and Armin Kohlrausch
IPO Center for Research on User-System Interaction, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands

(Received 10 March 1997; revised 2 October 1997; accepted 18 November 1997)

Detection thresholds were measured with a multiplied-noise masker that was in phase in both ears and a sinusoidal signal which was either in phase or out of phase (NoSo and NoSπ conditions). The masker was generated by multiplying a low-pass noise with a sinusoidal carrier. The signal was a sinusoid with the same frequency as the carrier and a constant phase offset, ϑ, with respect to the carrier. By adjusting the phase offset, the stimulus properties were varied in such a way that only interaural time delays ($\vartheta=\pi/2$) or interaural intensity differences ($\vartheta=0$) were present within the NoSπ stimulus. Thresholds were measured at a center frequency of 4 kHz as a function of bandwidth for $\vartheta=\pi/2$ and for $\vartheta=0$. In a second experiment thresholds were measured for a bandwidth of 25 Hz as a function of the center frequency. The results show that narrow-band BMLDs at 4 kHz can amount to 30 dB for the $\vartheta=0$ condition. For this condition, narrow-band BMLDs are also reasonably constant across frequency, in contrast to results obtained with standard Gaussian-noise maskers. For $\vartheta=\pi/2$, BMLDs are restricted to the frequency region below 2 kHz provided that the masker is narrow band, but BMLDs of up to 15 dB are found at 4 kHz if the masker is 50 Hz or wider. The frequency dependence of the binaural thresholds seems to be best explained by assuming that the stimulus waveforms are compressed before binaural interaction. [S0001-4966(98)03804-1]

PACS numbers: 43.66.Pn, 43.66.Ba, 43.66.Dc [RHD]

INTRODUCTION

The detection of a signal in the presence of a masker in a diotic condition can usually be improved by altering the interaural properties of either the masker or the signal. For example, presenting an in-phase masker and an out-of-phase signal (NoSπ) generally leads to lower thresholds than those obtained by presenting both in phase (NoSo). This improvement is termed the binaural masking level difference (BMLD).

As an explanation for the BMLD, the subject's sensitivity to interaural intensity differences (IIDs) and interaural time delays (ITDs) in the NoSπ stimulus has been proposed (Webster, 1951; Jeffress and McFadden, 1971; McFadden *et al.*, 1971). These interaural cues can be of a static nature or can vary over time. Lateralization occurs if IIDs or ITDs are static (Zwislocki and Feldman, 1956; Mills, 1960). In detection experiments using a Gaussian-noise masker in, e.g., a binaural NoSπ condition, the fluctuations in the fine-structure period and the envelope imply that the addition of the out-of-phase signal creates time-varying IIDs and ITDs (cf. Zurek, 1991). With this type of masker it is not possible to assess the individual contributions of IIDs and ITDs to the process of binaural detection.

Several studies have addressed this topic in the past. Hafter and Carrier (1970) and Yost (1972) have measured NoSπ BMLDs using a sinusoidal masker and a sinusoidal signal where the masker and signal were added with well defined phase angles. When the masker and signal are in phase in one ear and out of phase in the other ear, a stimulus is obtained with a static IID. If the signal phase is shifted such that the masker and signal are added with a $\pi/2$ phase difference in one ear and with a $-\pi/2$ phase difference in the other ear, a stimulus is obtained with a static ITD or a static interaural phase difference (IPD). The results from these studies show that in general sensitivity to IIDs is less than to IPDs if the sensitivity is expressed in terms of the signal-to-masker ratio.

McFadden *et al.* (1971) elaborated on this idea by using a noise masker and a noise signal for measuring NoSπ BMLDs. In their experiments the masker noise was shifted in phase over a fixed angle, ϑ, and then used as the signal to be detected. In the same manner as for the studies with sinusoidal maskers and signals, stimuli could be created with only IIDs or IPDs. The results of these experiments showed that some subjects were more sensitive to IIDs and that others were more sensitive to IPDs.

A common factor in all these experiments is that the stimuli could be manipulated such that there was only a static IID or only a static IPD, both leading to a difference in perceived lateralization relative to a reference stimulus. In contrast, in the classical BMLD studies with Gaussian-noise maskers, the IIDs and IPDs are dynamically varying. For these studies, the average value for the IID and the IPD is zero while the variance of the IIDs and IPDs is nonzero which, at medium masker bandwidths, is perceived as a widening of the sound image. Since the binaural system is known to be sluggish (Perrott and Musicant, 1977), the results obtained with static IIDs and ITDs cannot easily be compared to the Gaussian-noise data. Several studies suggest that dynamically varying IIDs and ITDs are processed less efficiently than static IIDs and ITDs in lateralization experiments (Grantham, 1984; Bernstein and Trahiotis, 1996a).

In this study we present results which were measured with stimuli containing either dynamically varying IIDs or

dynamically varying ITDs. For this purpose we used a multiplied-noise masker and a sinusoidal signal. The masker was obtained by multiplying a low-pass noise with a sinusoidal carrier. A sinusoid of the same frequency as the carrier was used as the signal. The properties of this stimulus are such that by manipulating the phase, ϑ, between the carrier and the signal, only IIDs or only ITDs are created in the NoSπ stimulus. Thresholds for the NoSo condition were also measured for the two different phase offsets to serve as a reference for computing BMLDs. Although initially this condition was only considered as a reference to the NoSπ condition, the results for this condition are interesting in their own right and are therefore treated in considerable detail.

With respect to the findings of McFadden *et al.* (1971), it is interesting to compare their finding of differences between the subjects' relative sensitivity to IIDs and ITDs to predictions for our stimulus type on the basis of the interaural cross correlation (ICC). Several researchers have suggested that the ICC is important for describing various aspects of binaural detection for frequencies below about 1.2 kHz (Durlach *et al.*, 1986). The detection of an Sπ signal in the presence of an No masker is thought to result from the decrease in ICC of the NoSπ test stimulus with respect to the No reference stimulus. For the stimuli used in our study, it has been shown that at low frequencies the change in the value of the ICC depends only on the signal-to-noise ratio, but not on the phase ϑ between masker and signal (van de Par and Kohlrausch, 1995). On the basis of this reasoning, the same thresholds at low frequencies are expected for the multiplied-noise conditions that have only IIDs or only ITDs. This prediction conflicts with the assumption that for multiplied-noise maskers with dynamically varying interaural differences, binaural detection can be characterized by two separate parameters, the sensitivity to changes in IIDs and in ITDs, as suggested by the observations of McFadden *et al.* (1971). According to this assumption different thresholds are expected for different values of ϑ where the dependence on ϑ may vary across subject.

The ICC has also been suggested to be relevant at high frequencies. It has been suggested that the envelope correlation is the most important property describing binaural detection above about 1.2 kHz (Bernstein and Trahiotis, 1996b). If the sensitivity to changes in *envelope* cross correlation at high frequencies is the same as the sensitivity to changes in the *waveform* cross correlation at low frequencies, the same thresholds are expected in our experiments at low and high frequencies for the NoSπ conditions that have IIDs only (van de Par and Kohlrausch, 1995). On the other hand, NoSπ stimuli with only ITDs are expected to have low thresholds at low frequencies and thresholds essentially as high as the corresponding NoSo thresholds at high frequencies, due to the loss of fine-structure coding at high frequencies. The experiments of the present study are used to test these theoretical predictions which are based on the assumption that subjects are sensitive to changes in the ICC of the stimulus waveform or the stimulus envelope at low and high frequencies, respectively.

In the first experiment, NoSo and NoSπ thresholds were measured for a fixed center frequency as a function of the bandwidth of the multiplied-noise masker. In the second experiment, the masker bandwidth was fixed and the center frequency of the masker and signal was varied. In both experiments, stimuli were used which either contained only IIDs or only ITDs. For comparison, NoSπ thresholds were also measured with a Gaussian-noise masker as a function of center frequency. In a third experiment additional data were obtained for a limited set of low-frequency conditions that should give us insight whether in general subjects are more sensitive to IIDs or ITDs such as present in a binaural detection task.

I. MULTIPLIED-NOISE STIMULI

The multiplied-noise maskers that are used in our experiments are generated by multiplying a low-pass noise with a sinusoidal carrier. The multiplied noise has periodic zero crossings which result from the multiplication with the sinusoidal carrier. Furthermore, the multiplied-noise spectrum is centered and symmetric around the frequency of the carrier. Above the carrier frequency, we find the spectrum of the initial low-pass noise, below the carrier frequency we find its inverted spectrum. Thus the bandwidth of the multiplied noise is twice the bandwidth of the low-pass noise. If the bandwidth of the low-pass noise is narrow, the two side bands of the multiplied noise can be regarded as two sinusoids which are slowly varying in frequency and in amplitude in a synchronized way. Their amplitudes are always identical while the changes in frequency occur in opposite directions.

The fact that the noise has periodic zero crossings is important for obtaining a stimulus for which the IIDs and the ITDs can be controlled separately. In fact, multiplied noise is a realization of the masking signal Jeffress and McFadden (1968) were looking for. It is possible to add a signal such that it has a controlled phase with respect to the masker. In Fig. 1, the multiplied-noise masker M and a signal S are shown in vector representation. The two panels together form a binaural NoSπ condition. In the left panel, the angle, ϑ, between M and S is zero and, since the signal is interaurally out of phase, in the right panel $\vartheta = \pi$. The addition of the masker and signal in the left and the right panels results in two vectors, L and R, which together represent the binaural stimulus. It can be seen now that the vectors L and R, in nearly all cases, have the same orientation. Only for the case in which the masker vector has a length which is smaller than the signal vector do L and R point in opposite directions. Basically, for this value of ϑ, this stimulus contains only IIDs which vary in time at a rate proportional to the masker bandwidth.

In Fig. 2 a similar picture is shown as in Fig. 1, but now the masker M and the signal S are added with phase differences of $\vartheta = \pi/2$ and $\vartheta = -\pi/2$. Now the vectors L and R have the same length in all cases, but the orientation of both vectors is different. Thus for this value of ϑ, a stimulus is obtained which only contains time-varying ITDs.

In the rest of this paper, the different ways in which the masker and signal can be added to form an NoSπ condition will be identified by the smallest angle, ϑ, between masker

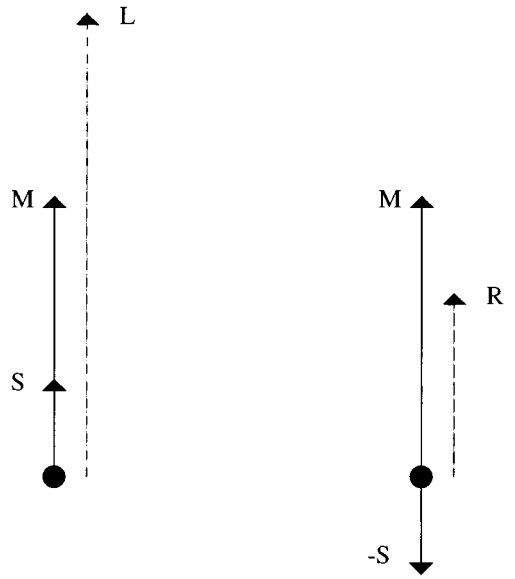


FIG. 1. The vector representation of an NoSπ stimulus. The masker *M* and signal *S* are added in the left ear with zero phase difference, resulting in the vector *L*, and subtracted in the right ear, resulting in the vector *R*. Only an IID is present.

and signal, knowing that the larger value of ϑ is always larger by π.

II. METHOD

A. Procedure

A 3-interval forced-choice procedure with adaptive signal-level adjustment was used to determine masked thresholds. The three masker intervals were separated by pauses of 200 ms. A signal was added to one of these intervals which was randomly selected. The subject's task was to indicate which of the three intervals contained the signal. Feedback was provided to the subject after each trial.

The signal level was adjusted according to a two-down one-up rule (Levitt, 1971). The initial step size for adjusting the level was 8 dB. After each second reversal of the level track, the step size was halved until a step size of 1 dB was reached. The run was then continued for another eight reversals. From the levels of these last eight reversals the median was calculated and used as a threshold value. At least three threshold values were obtained and averaged for each parameter value and subject. Thresholds are expressed as signal-to-overall-noise ratio.

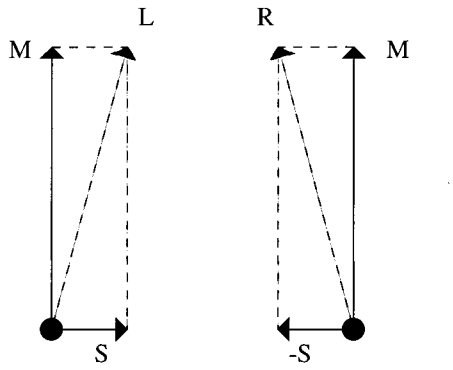


FIG. 2. The vector representation of an NoSπ stimulus. The masker *M* and signal *S* are added in the left ear with $\pi/2$ phase difference, resulting in the vector *L* and added in the right ear with $-\pi/2$ phase difference, resulting in the vector *R*. Only an ITD is present.

B. Stimuli

All stimuli were generated digitally and converted to analog signals with a two-channel, 16-bit D/A convertor at a sampling rate of 32 kHz. The stimuli were presented to the subjects over Beyerdynamic DT 990 headphones at a sound pressure level of 65 dB.

The multiplied-noise stimulus of 400-ms duration was obtained by randomly selecting a segment from a 2000-ms cyclic 2-channel buffer. One channel contained a multiplied-noise which was used as the masker and which was generated by multiplying a cyclic low-pass noise with a sinusoidal carrier. The second channel contained a sinusoid of the same frequency as the carrier with a phase offset ϑ with respect to the carrier and was used as the signal. For the signal only a 300-ms interval which was temporally centered in the masker was used. Both the masker and the signal on- and offsets were gated with 50-ms raised-cosine ramps.

The Gaussian-noise stimulus (for experiment 2 only) of 400-ms duration was obtained by randomly selecting a segment from a 2000-ms cyclic noise buffer. For the signal, a 300-ms sinusoid which was temporally centered in the masker was used. Both the masker and the signal on- and offsets were gated with 50-ms raised-cosine ramps. For each adaptive run, a new noise buffer was generated.

In the first experiment, thresholds were obtained for NoSo and NoSπ for $\vartheta=0$ and for $\vartheta=\pi/2$. In this experiment, masker and signal were centered at 4 kHz. The bandwidth of the masker had the following values: 5, 10, 25, 50, 100, 250, 500, 1000, 2000, and 8000 Hz.

In the second experiment, similar conditions were measured except that now the masker bandwidth was fixed to a value of 25 Hz while the center frequency was set to 125, 250, 500, 1000, 2000, and 4000 Hz. In addition, NoSπ thresholds were measured for a 25-Hz-wide Gaussian-noise masker for the same center frequencies.

In the third experiment, some additional subjects were tested for their sensitivity to NoSπ stimuli with only IIDs or only ITDs. The measurements were conducted at center frequencies of 250 and 500 Hz using maskers with 25-Hz bandwidth.

C. Subjects

Four subjects with no reported history of hearing impairment participated in the first two listening tests. Three subjects had extensive experience in binaural experiments. One subject first got four hours training before data collection started.

Three additional subjects with no reported history of hearing impairment participated in the last experiment.

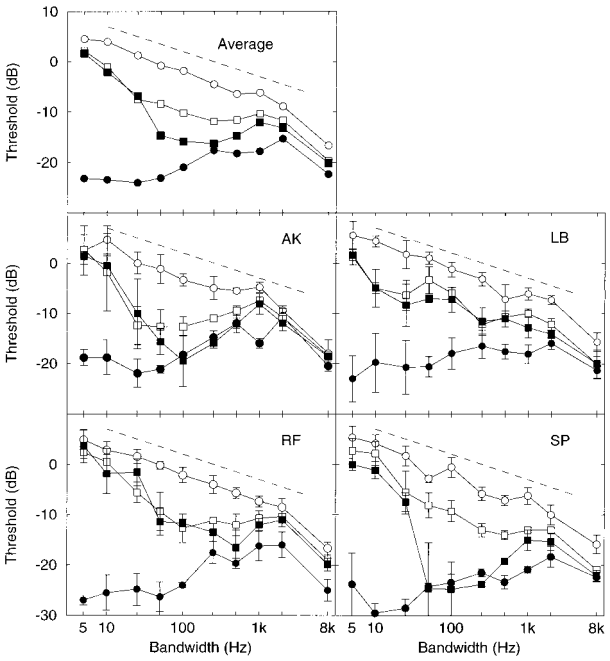


FIG. 3. Thresholds for NoSo (open symbols) and NoSπ (closed symbols) are shown for $\vartheta=0$ (○) and for $\vartheta=\pi/2$ (□) as a function of the bandwidth. Thresholds were measured at a center frequency of 4 kHz and are shown for four subjects in the lower four panels. Average results of the four subjects are shown in the upper panel. The dashed auxiliary lines have a slope of 1.5 dB/oct.

III. EXPERIMENTAL RESULTS

A. The dependence on bandwidth

The results for the experiment at 4-kHz center frequency are shown in Fig. 3. The bottom four panels show individual results for the four subjects; the top panel shows the average data. Thresholds are plotted as a function of bandwidth for NoSo (open symbols) and NoSπ (filled symbols). Circles indicate the conditions with $\vartheta=0$, squares those with $\vartheta=\pi/2$. The dashed auxiliary lines show a slope of 1.5 dB/oct.

The NoSπ thresholds for $\vartheta=0$ (filled circles) are lowest at the narrow bandwidths and increase up to a bandwidth of 2 kHz. Beyond 2 kHz, thresholds decrease with roughly 3 dB/oct. Such a decrease is expected on the basis of peripheral filtering (note that thresholds are expressed relative to the total masker level). However, the effect of peripheral filtering only starts at bandwidths which are much larger than the critical bandwidth observed in monaural experiments (e.g., Glasberg and Moore, 1990). Such a discrepancy has previously been observed with Gaussian-noise maskers (e.g., Sever and Small, 1979; Zurek and Durlach, 1987).

The NoSπ (filled squares) stimulus with $\vartheta=\pi/2$ contains only ITDs. Since several behavioral studies indicate that interaural differences in the stimulus fine structure at frequencies above about 1.5 kHz cannot be processed by the binaural system (e.g., Perrott and Nelson, 1969), no BMLDs are expected for this condition. The data show, however, that at intermediate bandwidths some of the subjects show BMLDs of up to 15 dB (see data of SP).

Comparing the NoSo with the NoSπ condition for $\vartheta=0$ (open and filled circles) we find very large BMLDs at the narrowest bandwidths. For some subjects, BMLDs amount to more than 30 dB. Such values are large compared to what is usually observed with Gaussian-noise maskers at such a high frequency. The large BMLDs result at least partially from the low thresholds for the NoSπ condition. For example, the NoSπ thresholds for bandwidths around 10 Hz found by Zurek and Durlach (1987) with Gaussian-noise maskers at 4 kHz are about -20 dB signal-to-masker ratio while we find about -27 dB for the two most sensitive subjects (RF and SP).

The data show that with increasing bandwidth, BMLDs tend to decrease and at 8-kHz bandwidth we find only small BMLDs. This bandwidth effect is, at least qualitatively, similar to what is observed with Gaussian-noise maskers. However, for 8-kHz bandwidth, most subjects here have a BMLD of about 6–8 dB. Such a BMLD has never been reported for broadband high-frequency conditions with Gaussian-noise maskers. For such a condition BMLDs of only 2–3 dB are observed (Webster, 1951; Zurek and Durlach, 1987; Kohlrausch *et al.*, 1995).

For NoSo we find that thresholds for $\vartheta=0$ tend to decrease over a broad range of bandwidths with a slope of 1.5 dB/oct as indicated by the auxiliary dashed lines. This is consistent with the notion that for this condition the detection process is dominated by an energy cue (Bos and de Boer, 1966; Green and Swets, 1974). For such a detection process, the factor which limits detection is the variability of the energy of successive masker samples. This variability decreases with 1.5 dB/oct of masker bandwidth.

Comparing these results to the other NoSo condition with $\vartheta=\pi/2$ (open squares), we find that thresholds depend on bandwidth in a different way. For bandwidths in the range of 5–50 Hz, thresholds tend to decrease by more than 1.5 dB/oct, while from bandwidths of 250 Hz to 1 kHz thresholds tend to decrease with less than 1.5 dB/oct. In addition, thresholds for the $\vartheta=\pi/2$ condition are lower than for the $\vartheta=0$ condition. For some subjects the difference is more than 10 dB at intermediate bandwidths.

B. The dependence on center frequency

In Fig. 4, the results for the second experiment with a constant masker bandwidth of 25 Hz are shown as a function of masker center frequency. The bottom four panels show results for the four subjects, the top panel shows the average data of the four subjects. NoSo (open symbols) and NoSπ (filled symbols) thresholds are plotted for $\vartheta=0$ (circles) and $\vartheta=\pi/2$ (squares).

For the NoSπ condition with $\vartheta=0$ thresholds are relatively independent of center frequency; only for subject AK is a systematic increase towards higher frequencies seen. In order to compare the masking by a multiplied-noise (with $\vartheta=0$) and a Gaussian-noise, NoSπ thresholds for both conditions are shown in Fig. 5. It can be seen that while the multiplied-noise thresholds are, on average, independent of center frequency, Gaussian-noise thresholds increase towards higher center frequencies. At 4 kHz the difference is on average 9 dB.

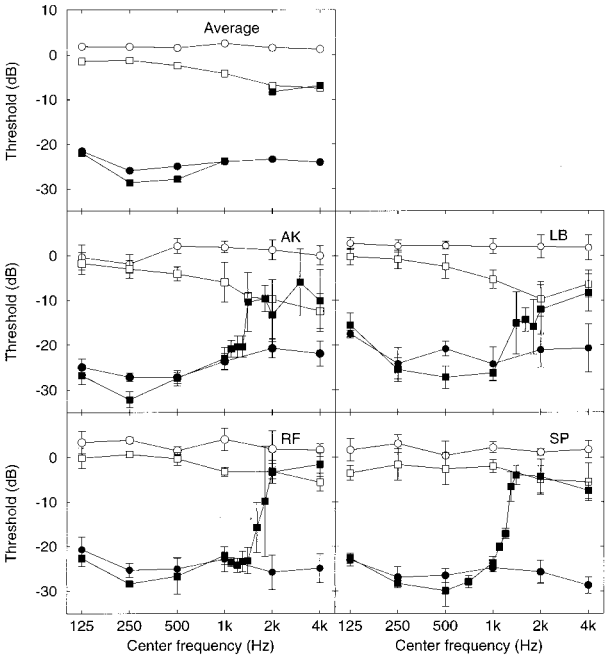


FIG. 4. Thresholds for NoSo (open symbols) and NoSπ (closed symbols) are shown for $\vartheta=0$ (○) and for $\vartheta=\pi/2$ (□) as a function of center frequency. Thresholds were measured at a masker bandwidth of 25 Hz and are shown for four subjects in the lower four panels. Average results for the four subjects are shown in the upper panel.

Returning to Fig. 4, for the NoSπ condition with $\vartheta=\pi/2$ we find low thresholds at low frequencies, and a sharp increase of thresholds around 1.5 kHz. For high frequencies no BMLD is found. These observations are in line with the idea that at high frequencies, the binaural system is not able to process the fine structure of input waveforms. The abrupt transition of thresholds is in line with lateralization experiments where the sensitivity to interaural phase differences in sinusoids is measured (Zwislocki and Feldman, 1956).

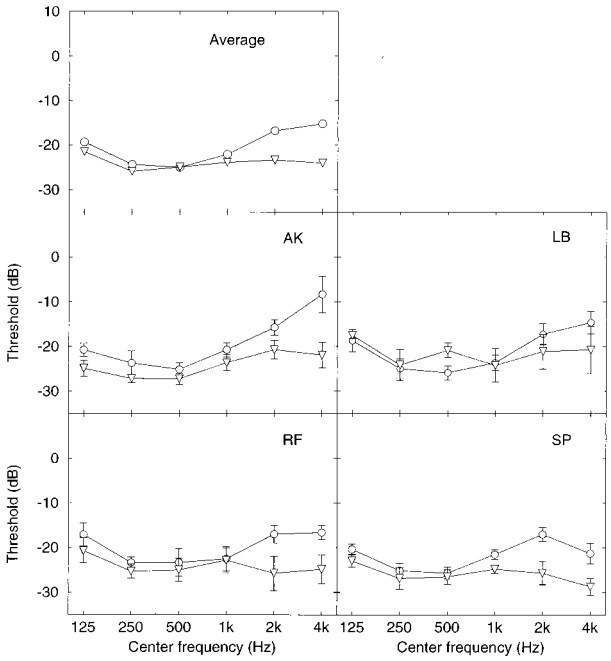


FIG. 5. Threshold for NoSπ for a Gaussian-noise masker (○) and for a multiplied-noise masker (▽) as a function of center frequency. Thresholds were measured for a masker bandwidth of 25 Hz and are shown for four subjects in the lower four panels. Average results for the four subjects are shown in the upper panel.

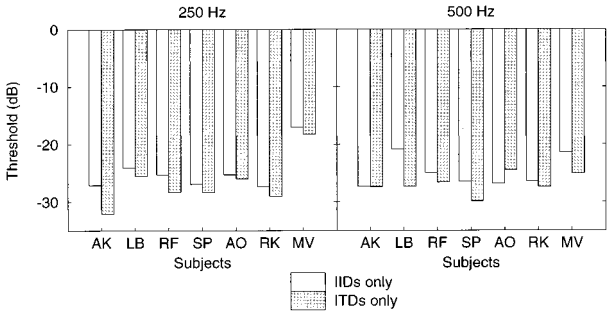


FIG. 6. Individual thresholds for NoSπ for multiplied-noise maskers with $\vartheta=0$ (white bars) and $\vartheta=\pi/2$ (gray bars). The left panel shows 250-Hz data; the right panel shows 500-Hz data.

When the multiplied-noise thresholds for $\vartheta=0$ and $\vartheta=\pi/2$ are compared at frequencies up to 1 kHz, we find that they are rather similar except that thresholds for $\vartheta=\pi/2$ tend to be somewhat lower. When BMLDs are considered there is very little difference between the two phase conditions. This contrasts with the findings of McFadden *et al.* (1971), who found that some subjects are more sensitive to IIDs while others are more sensitive to ITDs. Because the data presented in Fig. 4 are only for a limited number of subjects, the observed relation between the thresholds of the two phase conditions may be mere coincidence.

In order to get a better idea of the processing of IIDs and ITDs, an extra experiment was run with three additional subjects. NoSπ thresholds were measured with a multiplied-noise masker of 25-Hz bandwidth at 250-Hz and 500-Hz center frequency for the two conditions with only IIDs or only ITDs. The results of all subjects are shown in Fig. 6. In the left and right panels the 250-Hz and the 500-Hz thresholds are shown. The white bars show the IID conditions, the gray bars show the ITD conditions. From these data it is clear that thresholds for IID and ITD conditions are fairly similar. However, nearly all ITD thresholds are at least as low as the IID thresholds. There is only one subject, AO for 500 Hz, for whom the ITD threshold is not at least as low as the corresponding IID threshold. Thus even with three additional subjects, we notice a difference between our findings and the results of McFadden *et al.* (1971).

The NoSo thresholds as a function of center frequency in Fig. 4 show that for $\vartheta=\pi/2$ thresholds are substantially lower than for $\vartheta=0$. While the thresholds for $\vartheta=0$ are rather independent of center frequency, thresholds for $\vartheta=\pi/2$ decrease somewhat towards higher frequencies. A similar influence of ϑ on NoSo thresholds was found for different masker bandwidths at 4 kHz (cf. Fig. 3).

IV. DISCUSSION

In this section and the ones that follow the binaural results will be specifically discussed in terms of thresholds and

not in terms of BMLDs. The reason for this specific choice is related to the types of binaural mechanisms that we address. One type of mechanism is based on the subject's sensitivity to interaural time delays and interaural intensity differences, another one is based on the subject's sensitivity to changes in interaural correlation. Both types of models make predictions only on the binaural interaction per se and not on the difference in efficiency of the binaural and monaural processing mode. A model that predicts such differences is the EC theory (Durlach, 1963). Recent papers by Kohlrausch *et al.* (1997) and Hall *et al.* (1998) are of interest in this respect because they measured BMLDs with low-noise noise maskers. Their data show considerably smaller BMLDs in low-noise maskers as compared to Gaussian noise maskers which mainly result from rather low NoSo thresholds. Similarly, considerable dependencies of NoSo thresholds on ϑ were found in the present experiments. The EC theory is not able to account for such results, because it does not incorporate a very elaborate monaural detection mechanism. On the other hand, a model based on sensitivity to changes in IIDs or ITDs or on changes in the ICC only predicts the thresholds for NoSπ and as such makes no implicit inferences on the NoSo thresholds.

The low-frequency data obtained with stimuli with either IIDs or ITDs show a reasonably strong similarity in sensitivity to both binaural cues, at least when expressed in terms of thresholds. This is the case for the four subjects which measured data at various center frequencies as well as for the three additional subjects who measured data at 250 and 500 Hz. This suggests that for the multiplied-noise stimuli with dynamically varying interaural cues, binaural detection is governed by one measure of interaural similarity such as the waveform correlation. It seems unlikely that two separate mechanisms with specific sensitivities to IIDs or ITDs are underlying the detection process such as is suggested by the data from McFadden *et al.* (1971) who used stimuli with static IIDs or ITDs. It seems that our data are inconsistent with those of McFadden *et al.* (1971). The reason for this inconsistency is unclear but it may be related to the fact that we used dynamically varying instead of static IIDs and ITDs. For static IIDs and ITDs, subjects can base detection on a lateralization percept, while for dynamic IIDs and ITDs, subject have to rely on the detection of a broadening sound image, without a clearly lateralized percept being available.

An interesting aspect of the 4-kHz data presented in Figs. 3 and 4 is that for $\vartheta = 0$, BMLDs are in the order of 25 dB for bandwidths of 25 Hz and less and 6 dB for 8-kHz bandwidth. These values are large in comparison with results using Gaussian noise (e.g., Zurek and Durlach, 1987). As already mentioned, these rather large BMLDs are primarily caused by the rather low NoSπ thresholds.

An explanation for these NoSπ thresholds is that, for $\vartheta = 0$, the stimulus contains only IIDs. The addition of the Sπ signal to the No multiplied-noise masker leads to a decrease in the interaural envelope correlation which is larger than the decrease which would have been observed with a Gaussian-noise masker. Assuming that subjects detect the decrease in interaural envelope correlation (Bernstein and Trahiotis, 1992), multiplied-noise thresholds for $\vartheta = 0$ are predicted to be 3-dB lower than Gaussian-noise thresholds (cf. van de Par and Kohlrausch, 1995). Comparing this prediction to the observations which are presented in Fig. 5, we find a much larger difference between comparable Gaussian-noise and multiplied-noise maskers of around 9 dB at the highest center frequency. In Sec. V, this discrepancy will be discussed in more detail by considering the influence of basilar-membrane compression on the coding of the envelopes in the auditory nerve, and the effect this may have on binaural detection.

For the NoSπ multiplied-noise maskers with $\vartheta = 0$, we find thresholds which are relatively independent of center frequency. For narrow-band Gaussian-noise maskers, it is known that BMLDs decrease toward high frequencies (Zurek and Durlach, 1987; Kohlrausch *et al.*, 1995). The difference in thresholds between Gaussian-noise maskers at center frequencies of 500 Hz and 4 kHz varies for the individual subjects from 6 to more than 10 dB. For multiplied noise, the differences are smaller for all subjects and can even be in the opposite direction (cf. subject SP). These observations suggest that the IIDs which are present at low and at high frequencies can be processed with nearly equal efficiency by the binaural system.

For the data with $\vartheta = \pi/2$, the NoSπ stimulus contains only ITDs present within the fine structure. Therefore, at frequencies above about 1.5 kHz, no BMLD was expected for such stimuli. However, Fig. 3 shows that two subjects show substantial BMLDs for bandwidths in the range of 50–500 Hz.

As an explanation for the observed BMLDs we consider the fact that ITDs can be transformed into IIDs when the amplitude spectrum or the phase spectrum of the stimulus is transformed in some way. Sufficiently strong transformations may occur as a result of the peripheral filtering. It has been suggested that the filtering in the basilar membrane is rather steep at the upper flank of the filter (cf. Egan and Hake, 1950). Such an explanation for the conversion of frequency modulation into amplitude modulation has been proposed earlier by Blauert (1981). The addition of an anti-phasic sinusoidal signal to an in-phase multiplied-noise masker leads to interaurally differing modulations in the frequency of the multiplied-noise masker. When frequency modulations are transformed into amplitude modulations, interaural differences in the envelopes are created. For such an explanation to work, it is necessary that the slope of the amplitude or phase spectrum of the filter is steep enough to change the masker properties significantly on a scale of 50–100 Hz, since already for these bandwidths, substantial BMLDs are observed.

In order to further evaluate to what extent such transformations may be expected from basilar-membrane filtering, simulations with a basilar-membrane model (Giguere and Woodland, 1994) are presented in Sec. VI.

The results from the NoSo conditions (cf. Figs. 3 and 4) show, just as for the NoSπ condition, that there is a dependence of thresholds on the phase ϑ. Thresholds for NoSo with $\vartheta = \pi/2$ are as much as 10-dB lower than for $\vartheta = 0$. The reason for this difference is not well understood. There are

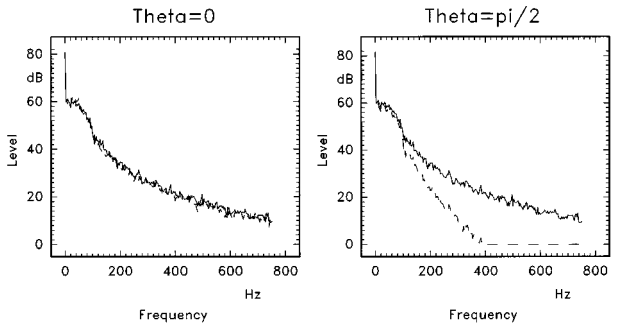


FIG. 7. The envelope spectrum for the $\vartheta = 0$ (left panel) and the $\vartheta = \pi/2$ condition (right panel) are shown. The solid line shows the spectrum of the masker alone; the dashed line shows the spectrum of the masker plus signal.

several properties of the multiplied-noise stimulus which depend on whether the signal is added with $\vartheta = 0$ or $\vartheta = \pi/2$. One such stimulus property is the probability density function of the envelope. This function is changed in a different manner for signals added in $\vartheta = \pi/2$ phase or in $\vartheta = 0$ phase. In Sec. V, it will be shown how this can account for the improved detection for the $\vartheta = \pi/2$ condition by considering the compressive properties of the auditory periphery.

Another stimulus property of interest may be the spectral distribution of the envelope components (Martens, 1982; Green *et al.*, 1992; Dau *et al.*, 1997a, b; Kohlrausch *et al.*, 1997). For the two phase conditions, the power spectrum of the linear envelope is affected in a different way by the presence of the signal. In Fig. 7, the left panel shows envelope spectra for the $\vartheta = 0$ condition and the right panel spectra for $\vartheta = \pi/2$. The solid line shows the spectrum for the masker alone, while the dashed line shows the spectrum for the masker plus signal at a signal-to-masker ratio of -5 dB. For $\vartheta = 0$, the envelope spectrum is hardly changed by the addition of the signal, for $\vartheta = \pi/2$, however, the envelope spectrum is markedly changed. When the signal is present, less high-frequency energy is seen in the envelope spectrum. Thus the envelope spectrum may provide an additional cue to subjects. A model based on such processing has been proposed by Dau *et al.* (1997a, b).

Comparing the NoSo thresholds in Fig. 4 for $\vartheta = 0$ and $\vartheta = \pi/2$, again we observe a large difference. This difference tends to increase towards higher frequencies. If indeed, as was discussed earlier, information about the distribution of the envelope is used for the detection in the $\vartheta = \pi/2$ condition, we may speculate that such a cue can be used better at high frequencies, since at high frequencies, due to the decrease in phase locking (Palmer and Russell, 1986), the envelope is directly available for auditory processing.

V. MODELING EFFECTS OF PERIPHERAL COMPRESSION

In the first part of this section the dependence on center frequency of the narrow-band NoSπ thresholds for Gaussian- and multiplied-noise maskers is treated. As already mentioned, this dependence can be at least partially understood when we consider that above about 1.2 kHz, the coding of the fine structure in the auditory nerve deteriorates (Palmer and Russel, 1986). When it is assumed that at low frequencies, subjects are sensitive to the decrease in the interaural *waveform* correlation due to the addition of the Sπ signal (cf. Durlach *et al.*, 1986), the loss of fine-structure coding toward high frequencies implies that at high frequencies subjects are sensitive to the decrease in the interaural *envelope* correlation (Bernstein and Trahiotis, 1996b). When at low and at high frequencies, the sensitivity to the decrease in correlation is equal, multiplied-noise thresholds for $\vartheta = 0$ are predicted to be independent of center frequency and Gaussian-noise thresholds at high frequencies should be 3-dB higher as compared to low-frequency thresholds (van de Par and Kohlrausch, 1995). However, our data in Fig. 5 show an average increase of 9 dB.

One approach in explaining this could be to assume that the sensitivity to changes in correlation is different at low and high frequencies (Koenke *et al.*, 1986). However, with such an assumption, thresholds for the multiplied-noise masker would be predicted to also increase toward higher frequencies, contrary to the observations. Furthermore, such an assumption is not in line with current thinking which assumes that after the first stages of peripheral transduction, binaural processing is essentially the same at low and at high frequencies (Colburn and Esquissaud, 1976; Bernstein and Trahiotis, 1996b, c; van de Par and Kohlrausch, 1997).

If the probability distribution functions of the envelopes of the two masker types are considered, we find that the Gaussian noise has a Rayleigh distribution while for the multiplied noise, envelope values are distributed according to the positive half of a Gaussian distribution. This implies that for multiplied noise, the probability of finding low envelope values is larger than for Gaussian noise. Interestingly the signal-to-masker ratio is largest when the masker envelope is small. Thus, multiplied noise has relatively more epochs with a high signal-to-noise ratio as compared to Gaussian noise. When we assume that for some reason these low envelope parts of the masker are weighted more in the correlation process than other parts of the masker, this would mean that the multiplied-noise condition could benefit more from such a weighting than the Gaussian-noise condition.

One aspect of the auditory periphery that could account for such a weighting is the compressive response of the basilar membrane (Rhode, 1971; Ruggero, 1992).[1] The effect of such a compressive response would be that interaural differences in the envelope occurring at high-envelope values are reduced in size. As a result, the parts where the envelope has low values are effectively weighted more strongly in the detection process. Therefore the interaural differences in the envelope of a stimulus with a high probability for low-envelope values, like multiplied noise, are affected less by compression than for example Gaussian noise. Several studies showed that at medium levels (40–70 dB SPL), an increase in the stimulus level of 1 dB results in an increase of the basilar-membrane response of only about 0.2 dB (Sellick *et al.*, 1982; Ruggero, 1992). This value has also been derived from studies of auditory-nerve response (Yates, 1990) and behavioral studies in humans using forward masking (Oxenham and Plack, 1997). However, it also seems to be the case that the inner-hair-cell transduction is expansive such that there is a quadratic relation between basilar-

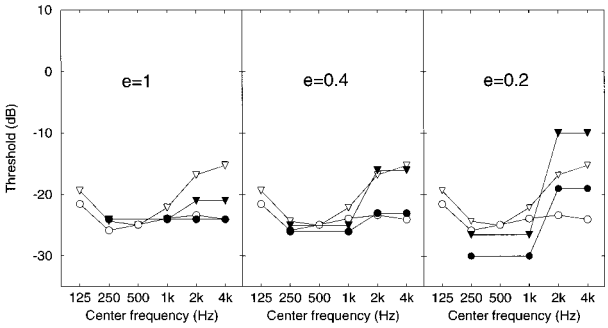


FIG. 8. The open symbols show the average data from Fig. 6; the filled symbols show model predictions with exponents of $e=1$, $e=0.4$, and $e=0.2$.

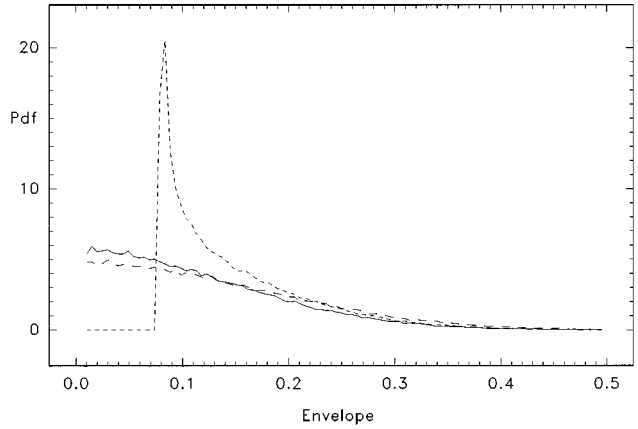


FIG. 9. The envelope probability distribution is shown for a multiplied-noise masker alone (solid line) and for a multiplied-noise masker plus signal with $\vartheta=0$ (long-dashed line), and $\vartheta=\pi/2$ (short-dashed line). The signal-to-masker ratio is -5 dB.

membrane displacement and the firing rate of the auditory nerves (Yates, 1990; Yates *et al.*, 1990). This square-law expansion is also necessary to account for behavioral results from hearing-impaired listeners when it is assumed that the BM is linear (Oxenham and Moore, 1995). The combined effect of the compressive and expansive properties of the auditory pathway can be modeled by raising the stimulus envelope to the power 0.4. This was also the value derived by Oxenham and Moore (1995) in a study on the additivity of masking.

In Fig. 8, the same average data are shown as in Fig. 5. In addition, the thresholds are plotted which are predicted by a simple model which incorporates compression of the stimulus envelope. For this purpose, first the envelope and the fine structure were separated, then the envelope was raised to a certain power (exponent e), after which the compressed envelope and the fine structure were recombined. It is assumed that at low frequencies (below 1.5 kHz) a certain decrease in interaural *waveform* correlation is needed for signal detection, and that at high frequencies (above 1.5 kHz) the same decrease in interaural *envelope* correlation is needed for signal detection. The open symbols show data for Gaussian-noise maskers (○) and multiplied-noise maskers (▽), while the filled symbols show predictions for the same masker types for exponents of $e=1$, $e=0.4$, and $e=0.2$. The just detectable decrease in interaural correlation was set to 0.008.

It is clear that an exponent of 0.4 leads to the best correspondence between data and model predictions. At low frequencies, the compression has very little effect while for high frequencies, Gaussian-noise thresholds increase considerably. Furthermore, the predicted multiplied-noise thresholds are fairly independent of center frequency. The present results suggest that the compressive properties of the basilar-membrane response may be of considerable influence on binaural detection, particularly at high frequencies. According to this model, the frequently observed difference between low- and high-frequency BMLDs for narrowband maskers (Zurek and Durlach, 1987; Kohlrausch *et al.*, 1995), results from two factors, (1) the loss of fine-structure coding at high frequencies, and (2) the compressive response of the basilar membrane to stimulation.[2]

In the following we analyze whether the proposed compression also affects monaural detection. For understanding how compression can affect NoSo thresholds we consider the probability distribution of the stimulus envelope. The envelope of the masker alone has a distribution as is shown in Fig. 9 by the solid line. The distribution corresponds to the positive half of a Gaussian distribution with a mean of zero. Adding the signal to the masker at a signal-to-masker ratio of -5 dB while $\vartheta=0$ leads to the distribution that is shown by the long-dashed line. Both distributions are rather similar, except that for the masker-plus-signal situation larger values of the envelope are more likely than for the masker-alone situation.

In contrast, the short-dashed line shows that the addition of the signal to the masker with $\vartheta=\pi/2$ leads to a marked change in the envelope probability distribution. For the masker-alone situation, envelope values close to zero are likely, while for the masker-plus-signal situation, we expect no envelope values smaller than a certain minimum. Thus for this condition the addition of the signal to the masker leads to a very effective filling of the valleys which were present in the masker. This effect may itself be a cue for effective detection of the signal. The assumption of such a cue does, however, not explain the bandwidth effects which are observed for the condition $\vartheta=\pi/2$.

If the compressive properties of the peripheral auditory pathway are taken into account, we have a mechanism where, effectively, the epochs of the masker where the masker envelope has low values are weighted strongest in the detection process. Because for the $\vartheta=\pi/2$ condition, adding the signal to the masker predominantly leads to changes at the left side of the distribution, signal detection is expected to be better for this condition than for the $\vartheta=0$ condition.

In order to assess whether such a weighting mechanism can account for the NoSo results, a simple detection model was tested. It consisted of a fourth-order gammatone filter centered at 4 kHz followed by a stage where the stimulus envelope was compressed using an exponent of 0.4, and followed by a stage where the stimulus envelope was integrated across the total duration of the stimulus. A stochastic variable with a Gaussian distribution was added to the output of

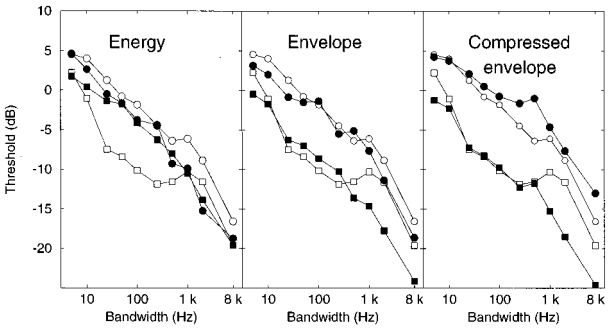


FIG. 10. Observed thresholds (open symbols) and predicted thresholds (filled symbols) for NoSo for $\vartheta = 0$ (○) and $\vartheta = \pi/2$ (□). Predictions are shown for an energy-detection model (left panel), and envelope-integration models without and with compression (middle and right panels, respectively).

the temporal integrator to simulate internal noise with an rms of -23 dB relative to the integrator output. This sum was used as the decision variable in a simulated 3IFC experiment identical to our previously described experiments. The internal-noise level was set to a level, such that the predictions of the above described model for bandwidths up to 500 Hz were best.

For comparison, the same envelope-integration model was tested without any compression and in addition, the squared envelope was used at the input of the integrator. This last case is equivalent to an energy-detection model. For this case, no differences between the $\vartheta = 0$ and $\vartheta = \pi/2$ are expected since masker and signal add incoherently. Notice that the envelope-integration model with compression, without compression and the energy-detection model all correspond to one generic model where the envelope is raised to the powers 0.4, 1, and 2, respectively.

The predictions of the envelope-integration model with compression are shown in the right panel of Fig. 10. In addition, the predictions of an envelope-integration model without compression and the energy-detection model are shown in the middle and left panels, respectively. The open symbols show average experimental data, while the filled symbols show predictions. Values of $\vartheta = 0$ (○) and $\vartheta = \pi/2$ (□) are shown. The energy-detection model predicts no difference between the two phase conditions, just as expected. The envelope-integration models with and without compression show approximately similar results. For bandwidths up to 500 Hz, the predictions are close to the experimental data. The predicted difference between the two phase conditions is slightly smaller than the observations without compression and slightly larger with compression. For the $\vartheta = \pi/2$ condition, at the larger bandwidths, the predictions start to deviate from the observations. This is probably due to the inaccurate modeling of the peripheral filtering. A gammatone filter is a filter with a rather symmetrical amplitude and phase characteristic around its center frequency. Such filtering is expected to preserve the periodic zero crossings of the masker noise much better than a more realistic asymmetrical filter. Thus the phase-dependent interaction between masker and signal is expected to be much more impaired by more realistic asymmetric filtering. Such a filtering effect is expected to especially affect broadband stimuli.

TABLE I. The signal-to-noise ratios needed to obtain a 0.008 decrease in the interaural envelope correlation for an NoSπ condition. The envelope correlation is calculated after basilar-membrane transduction of the stimulus at a point at the basilar membrane where there is sufficient excitation and where the envelope correlation is smallest (i.e., where detection of the signal is best).

	SNR (dB)	
Masker bandwidth (Hz)	$\vartheta = 0$	$\vartheta = \pi/2$
10	−23	[a]
100	−23	−18
250	−23	−21

[a]For this condition, the SNR exceeded the threshold of the NoSo condition.

The simulations with the envelope-integration models show that compression is not essential for obtaining different thresholds for the two phase conditions. Both with and without compression, these models show dependencies on the phase ϑ. This result can be understood once we remember that the three models are different only due to the different powers to which the envelopes are raised before integration. With an energy-detection model, all values of the envelopes are weighted such that no improvement in detection is expected for $\vartheta = \pi/2$, whereas the noncompressed envelope-integration model puts more weight on the low-envelope values in comparison to the energy-detection model.

VI. BMLDS FOR $\vartheta = \pi/2$ AT HIGH FREQUENCIES

In this section we investigate whether the transformations of the stimulus amplitude and phase spectrum that may occur due to basilar-membrane filtering can account for the 4-kHz NoSπ data with $\vartheta = \pi/2$ shown in Fig. 3. In such a scheme, the interaural differences in frequency modulation (FM) of a stimulus that only has ITDs may be transformed into interaural differences in amplitude modulation (AM) which can be processed binaurally at high frequencies. Several of our stimuli were run through a basilar-membrane model (Giguere and Woodland, 1994) and the interaural envelope correlation (IEC) was calculated for various places along the basilar membrane. The place where the smallest value for the IEC was found and where the excitation was at least 20 dB above absolute threshold, was assumed to be the place which is used for binaural detection. It was assumed that a decrease in IEC of 0.008 was needed for signal detection. The signal-to-masker ratio was adjusted such that this value was reached. These signal-to-noise ratios are shown in Table I for stimuli of various bandwidths for $\vartheta = 0$ and $\vartheta = \pi/2$.

For the smallest bandwidth and $\vartheta = \pi/2$, the SNR that was needed for the threshold decrease in the IEC exceeded thresholds for the NoSo condition. Interestingly, at a bandwidth of 100 Hz, thresholds for the two phase conditions differ by only 5 dB, and this difference is even smaller for 250-Hz bandwidth. These simulations suggest that the phase and amplitude transformations which occur at the basilar membrane may be sufficient to explain the results of experiment I.

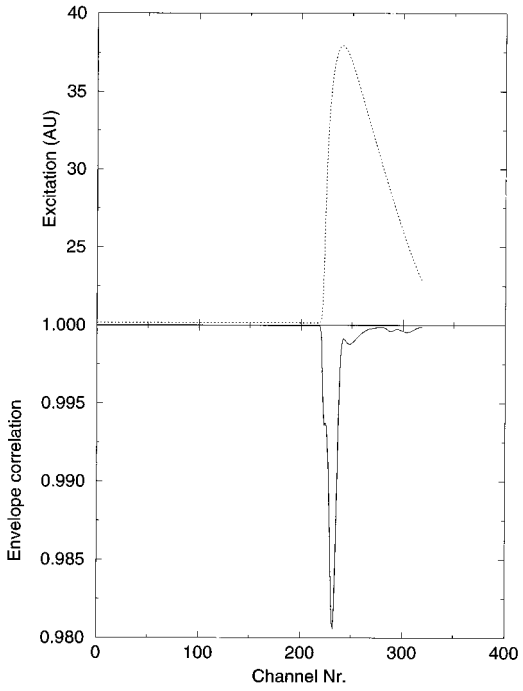


FIG. 11. The interaural envelope correlation predicted by a basilar-membrane model (Giguere and Woodland, 1994) is shown as a function of channel number (solid line). In addition, the excitation pattern is shown (dotted line). The input was a NoSπ stimulus with a 500-Hz-wide masker and $\vartheta = \pi/2$.

In Fig. 11, the interaural envelope correlations found along the basilar membrane are shown (solid line) for $\vartheta = \pi/2$ with a 500-Hz-wide masker. In addition, the excitation pattern is shown (dotted line). The left side of the abscissa shows the part of the basilar membrane which responds to low frequencies (the apex), while the right side shows the high-frequency part of the basilar membrane (the base). According to the present model, the largest decreases in the IEC are found at the segments of the basilar membrane which are tuned to a slightly lower frequency than the center frequency of the masker. Thus the upper flank of the corresponding auditory filter seems to provide the transformations needed for the FM to AM conversion necessary to obtain a decrease in IEC.

VII. CONCLUSIONS

It seems that at low frequencies, the sensitivity to dynamically varying IIDs or ITDs can be singly related to changes in the interaural correlation of the waveforms. Thus there seems to be no need to separately make assumptions about the sensitivity to changes in IIDs and ITDs.

Multiplied-noise maskers with sinusoidal signals lead to very large BMLDs at high frequencies if the phase difference between the carrier of the masker and the sinusoidal signal is zero.

When this condition is measured as a function of the center frequency, we find that thresholds at high frequencies are only slightly higher than at low frequencies. This suggests that IIDs can be equally well processed at low as at high frequencies.

For the NoSπ condition at 4 kHz, for which the envelopes of the waveforms arriving at both ears are identical and which contains ITDs only in the fine structure of the waveform, a significant BMLD is observed. This suggests that, due to the filtering on the basilar membrane, ITDs in the fine structure are transformed into IIDs.

The compressive response of the basilar membrane at the characteristic frequency seems to influence the detection for the NoSπ condition considerably. Model simulations suggest that this compression is responsible for the relatively large increase of NoSπ thresholds for Gaussian noise maskers toward high frequencies which is often observed. An exponent of 0.4 for compression of the stimulus envelope accounts well for the measured results. This value is in line with both the transduction of the basilar membrane and the auditory nerve and with previous psychophysical estimates of peripheral compression.

For the diotic NoSo condition, the phase ϑ between the masker carrier and the signal has a large effect on thresholds. When this phase, ϑ, was $\pi/2$, thresholds were 2–6 dB lower than for $\vartheta = 0$. We propose that this is due to the different manner in which the envelope statistics are changed by the addition of the signal for the two phase conditions. In contrast to an energy-detection model, an envelope-integration model can account for the observed phase effects. Compression of the envelope is not needed in such a model, but it does not impair the predictions either.

ACKNOWLEDGMENTS

We thank Andrew Oxenham and both reviewers for their very useful and encouraging comments on an earlier version of this paper.

[1]Buus *et al.* (1996) have discussed a weighting mechanism to account for results obtained in a CMR experiment and have suggested that models based on envelope comparison may also account for their results if they are modified to include a compressive nonlinearity before the comparison.

[2]Bernstein and Trahiotis (1996c) found that a simple peripheral model followed by an interaural correlator could explain 80% of the variance observed in the psychometric functions which they measured using an NoSπ condition with a Gaussian-noise masker at center frequencies varying from 500 to 2000 Hz. The peripheral model consisted of a half-wave, square-law rectifier followed by a low-pass filter. Comparing the predicted and the observed psychometeric functions shows that for two of the three subjects, the predicted percentages correct for a 2-kHz stimulus at a signal-to-noise ratio around 0 dB are higher than the observed percentages, while for a 500-Hz stimulus, there is not such a clear difference. Since compression according to our analysis is detrimental to binaural detection at high frequencies, the high-frequency psychometric functions predicted by Bernstein and Trahiotis may have been more in line with observations if a compressive stage had been added to their model.

Bernstein, L. R., and Trahiotis, C. (**1992**). ''Discrimination of interaural envelope correlation and its relation to binaural unmasking at high frequencies,'' J. Acoust. Soc. Am. **91**, 306–316.

Bernstein, L. R., and Trahiotis, C. (**1996a**). ''Binaural beats at high frequencies: Listeners' use of envelope-based interaural temporal and intensive disparities,'' J. Acoust. Soc. Am. **99**, 1670–1679.

Bernstein, L. R., and Trahiotis, C. (**1996b**). ''On the use of the normalized correlation as an index of interaural envelope correlation,'' J. Acoust. Soc. Am. **100**, 1754–1763.

Bernstein, L. R., and Trahiotis, C. (**1996c**). ''The normalized correlation: Accounting for binaural detection across center frequency,'' J. Acoust. Soc. Am. **100**, 3774–3784.

Blauert, J. (**1981**). ''Lateralization of jittered tones,'' J. Acoust. Soc. Am. **70**, 694–698.
Bos, C. E., and de Boer, E. (**1966**). ''Masking and discrimination,'' J. Acoust. Soc. Am. **39**, 708–715.
Buus, S., Zhang, L., and Florentine, M. (**1996**). ''Stimulus-driven, time-varying weights for comodulation masking release,'' J. Acoust. Soc. Am. **99**, 2288–2297.
Colburn, H. S., and Esquissaud, P. (**1976**). ''An auditory-nerve model for interaural time discrimination of high-frequency complex stimuli,'' J. Acoust. Soc. Am. **59**, S23.
Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). ''Modeling auditory processing of amplitude modulation. I. Detection and masking with narrowband carriers,'' J. Acoust. Soc. Am. **102**, 2892–2905.
Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). ''Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration,'' J. Acoust. Soc. Am. **102**, 2906–2919.
Durlach, N. I. (**1963**). ''Equalization and cancellation theory of binaural masking-level differences,'' J. Acoust. Soc. Am. **35**, 1206–1218.
Durlach, N. I., Gabriel, K. J., Colburn, H. S., and Trahiotis, C. (**1986**). ''Interaural correlation discrimination: II. Relation to binaural unmasking,'' J. Acoust. Soc. Am. **79**, 1548–1557.
Egan, J. P., and Hake, H. W. (**1950**). ''On the masking pattern of a simple auditory stimulus,'' J. Acoust. Soc. Am. **22**, 622–630.
Giguere, C., and Woodland, P. C. (**1994**). ''A computational model of the auditory periphery for speech and hearing research: I. Ascending path,'' J. Acoust. Soc. Am. **95**, 331–342.
Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hearing Res. **47**, 103–138.
Grantham, D. W. (**1984**). ''Discrimination of dynamic interaural intensity differences,'' J. Acoust. Soc. Am. **76**, 71–76.
Green, D. M., and Swets, J. A. (**1974**). *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966; reprinted by Krieger, New York).
Green, D. M., Berg, B. G., Dai, H., Eddins, D. A., Onsan, Z., and Nguyen, Q. (**1992**). ''Spectral shape discrimination of narrow-band sounds,'' J. Acoust. Soc. Am. **92**, 2586–2597.
Hafter, E. R., and Carrier, S. C. (**1970**). ''Masking-level differences obtained with a pulsed tonal masker,'' J. Acoust. Soc. Am. **47**, 1041–1047.
Hall, J. W., Grose, J. H., and Hartmann, W. M. (**1998**). ''The masking-level difference in low-noise noise,'' J. Acoust. Soc. Am. (in press).
Jeffress, L. A., and McFadden, D. (**1968**). ''MLD's and the phase angle, alpha,'' J. Acoust. Soc. Am. **43**, 164.
Jeffress, L. A., and McFadden, D. (**1971**). ''Differences of interaural phase and level in detection and lateralization,'' J. Acoust. Soc. Am. **49**, 1169–1179.
Koehnke, J., Durlach, N. I., and Colburn, H. S. (**1986**). ''Performance in several binaural-interaction experiments,'' J. Acoust. Soc. Am. **79**, 1558–1562.
Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A. J., and Püschel, D. (**1997**). ''Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations,'' Acust. Acta Acust. **83**, 659–669.
Kohlrausch, A., van de Par, S. L. J. D. E., and Houtsma, A. J. M. (**1995**). ''A new approach to study binaural interaction at high frequencies,'' in *Advances in Hearing Research, Proceedings of the 10th International Symposium on Hearing*, edited by G. A. Manley, G. M. Klump, C. Köppl, H. Fastl, and H. Oeckinghaus (World Scientific, Singapore), pp. 343–353.
Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Martens, J. P. (**1982**). ''A new theory for multitone masking,'' J. Acoust. Soc. Am. **72**, 397–405.
McFadden, D., Jeffress, L. A., and Ermey, H. L. (**1971**). ''Differences of interaural phase and level in detection and lateralization: 250 Hz,'' J. Acoust. Soc. Am. **50**, 1484–1493.
Oxenham, A. J., and Moore, B. C. J. (**1995**). ''Additivity of masking in normally hearing and hearing-impaired subjects,'' J. Acoust. Soc. Am. **98**, 1921–1934.
Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.
Palmer, A. R., and Russell, I. J. (**1986**). ''Phase-locking in the cochlear nerve of the guinea pig and its relation to the receptor potential of the inner hair cells,'' Hearing Res. **24**, 1–15.
van de Par, S., and Kohlrausch, A. (**1995**). ''Analytical expressions for the envelope correlation of certain narrowband stimuli,'' J. Acoust. Soc. Am. **98**, 3157–3169.
van de Par, S., and Kohlrausch, A. (**1997**). ''A new approach to comparing binaural masking level differences at low and high frequencies,'' J. Acoust. Soc. Am. **101**, 1671–1680.
Perrott, D. R., and Musicant, A. D. (**1977**). ''Minimum auditory movement angle: Binaural localization of moving sound sources,'' J. Acoust. Soc. Am. **62**, 1463–1466.
Perrott, D. R., and Nelson, M. A. (**1969**). ''Limits for the detection of binaural beats,'' J. Acoust. Soc. Am. **46**, 1477–1481.
Rhode, W. S. (**1971**). ''Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique,'' J. Acoust. Soc. Am. **49**, 1218–1231.
Ruggero, M. A. (**1992**). ''Responses to sound of the basilar membrane of the mammalian cochlea,'' Curr. Opin. Neurobiol. **2**, 449–456.
Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (**1982**). ''Measurement of the basilar membrane motion in the guinea pig using the Mossbauer technique,'' J. Acoust. Soc. Am. **72**, 131–141.
Sever, Jr., J. C., and Small, Jr., A. M. (**1979**). ''Binaural critical masking bandwidths,'' J. Acoust. Soc. Am. **66**, 1343–1350.
Webster, F. A. (**1951**). ''The influence of interaural phase on masked thresholds I. The role of interaural time-deviation,'' J. Acoust. Soc. Am. **23**, 452–462.
Yates, K. G. (**1990**). ''Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions,'' Hearing Res. **50**, 145–162.
Yates, K. G., Winter, I. M., and Robertson, D. (**1990**). ''Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range,'' Hearing Res. **45**, 203–220.
Yost, W. A. (**1972**). ''Tone-on-tone masking for three binaural listening conditions,'' J. Acoust. Soc. Am. **52**, 1234–1237.
Zurek, P. M. (**1991**). ''Probability distributions of interaural phase and level differences in binaural detection stimuli,'' J. Acoust. Soc. Am. **90**, 1927–1932.
Zurek, P. M., and Durlach, N. I. (**1987**). ''Masker-bandwidth dependence in homophasic and antiphasic tone detection,'' J. Acoust. Soc. Am. **81**, 459–464.
Zwislocki, J., and Feldman, R. S. (**1956**). ''Just noticable differences in dichotic phase,'' J. Acoust. Soc. Am. **28**, 860–864.

Radio frequency hearing: Electrostrictive detection and bone conduction

Wm. R. Bennett, Jr.
Department of Applied Physics, Yale University, New Haven, Connecticut 06520-8284

(Received 24 July 1997; revised 29 October 1997; accepted 20 November 1997)

A hearing aid patented by C. R. Schafer and supposedly based on detection of an amplitude-modulated carrier wave in the auditory cortex was re-examined. It is shown here that the hearing aid actually works by bone conduction of sound. It is concluded that detection of the modulation signal occurs by electrical nonlinearities and electrostriction in the bones of the face and skull. © *1998 Acoustical Society of America.* [S0001-4966(98)02304-2]

PACS numbers: 43.66.Ts, 43.71.Ky [RVS]

INTRODUCTION

During the decade prior to his death in 1987, a Connecticut electrical engineer named Curtis R. Schafer was granted several patents for what he called a ''cortical hearing aid'' (Schafer, 1977, 1980, 1987). The invention claimed to permit ''hearing'' through the use of an AM (amplitude-modulated) carrier frequency in the 15–60 kHz range that was applied to a person's head through electrodes placed on the chin and temple.[1] Schafer used peak-to-peak voltages of 1000 V at 100% modulation, hence, external electric fields of about 6 kV/m. The electrode to the temple was connected in series with a capacitance of about 35 pF. His theory was that amplitude modulation of the electric field in the audio range was detected by nonlinearities in neural junctions in the auditory cortex, thus completely circumventing the normal auditory tract. If the hearing aid really worked as he described, the implications for help to someone with nerve deafness would be little short of miraculous.

Although Schafer's interpretation is in substantial disagreement with the present understanding of the way in which sensations of sound are transmitted by very low-frequency neural pulse chains from the cochlea and interpreted in the auditory cortex (Zwislocki, 1981; Hudspeth, 1989; Hartmann, 1989), his device actually did work. According to Schafer, it was tried by 78 individuals with profound hearing loss, 59 of whom were able to detect speech. In addition, the present author once listened to an entire movement of a Beethoven string quartet over Schafer's original (vacuum tube) apparatus. The fidelity was poor, with considerable distortion, poor bass response, and little sense of spatial direction; nevertheless, the sound was quite ''audible'' and certainly by-passed the outer ear.

Difficulties with Schafer's interpretation of the detection process were analyzed previously by the author (Bennett, 1994). In addition to violation of the accepted theories of sound perception in the brain, there were two other basic problems that made Schafer's theory doubtful: First, from representative data on nonlinear transmembrane current characteristics at the cell level in the brain with the internal fields expected from the Schafer device, rectification of the AM carrier field by brain cells would probably have been very slight. Second, the carrier field inside the brain would most probably have been swamped by Johnson–Nyquist thermal noise at body temperature. As previously noted by the author, it is possible that Schafer's device represents an extension to lower frequencies of a phenomenon reported during World War II by radar technicians who claimed to hear radar pulses. Their reports were not taken seriously until much later (Frey, 1961). Explanations based on direct rf interaction with neurons (Frey, 1962) and thermoacoustic pressure waves (Foster and Finch, 1974) were suggested. The present results suggest that those explanations may have been wrong. The conclusion that ''microwave hearing'' was primarily due to a radiation pressure mechanism was largely based on a comparative analysis of a one-dimensional model in which the motion of a charge imbedded in a homogeneous dielectric medium interacted with a plane wave entering normal to the surface (Lin, 1976, 1978). There, it was found that the radiation pressure mechanism was about 100 times larger at microwave frequencies than one based on electrostriction (elastic deformation of the dielectric material.) But solutions based on direct microwave coupling to the bone were not performed as far as the present author knows.

I. METHOD

In order to study the problem at the low rf range, the apparatus shown in Fig. 1 was used. A commercial solid-state function generator produced a 40-kHz carrier wave that was amplitude modulated by sinewaves at audio frequencies ranging from 200 Hz to 5 kHz. At 100% modulation depth, the open-circuit output voltage from the function generator was about 10-V peak-to-peak. This output voltage was run through a step-up transformer that was self-resonant at about 40 kHz and then fed through a unity-gain vacuum-tube operational amplifier (Tektronix type O) to avoid loading down the solid-state circuitry. The output of the operational amplifier was about 150-V peak-to-peak at 100 percent amplitude modulation and was current limited to about 5 mA. That voltage was applied to the head of the test individual through 1-in. diameter, 1/4-in.-thick, smooth brass electrodes. One electrode was typically placed under the chin as suggested in the Schafer patents (denoted by the electrical ground symbol in Fig. 1), while the other was applied through a thin Teflon spacer to the temple, or other regions of the face and head. The presence of the Teflon spacer was not essential and merely served to reduce the possibility of rf burns on the

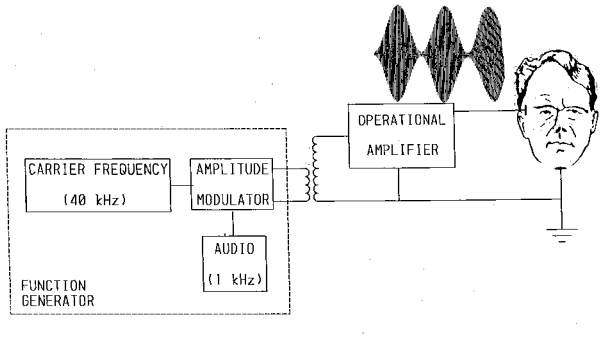


FIG. 1. Schematic diagram of the experimental apparatus.

skin. Tilting the electrode so that it touched the skin on the edge seemed to improve the sensitivity. In addition, it was usually found necessary to move the electrode slowly to sustain the signal. The signal from the electrodes was easily ''heard'' by half a dozen different subjects when they wore ear protectors to dampen out room noise.

The signal could be heard with the electrodes connected between almost any two points across the face and head of the individual. Further experiments showed that the ground electrode could be attached anywhere on the body (e.g., the thumb, or ankle) and the maximum sensitivity was obtained when the second electrode was placed near the end of the jaw bone (under the ear lobe) or on the high bony ridge above the cheek. Optimum sensitivity did not produce significant loading of the signal source by the flow of rf current. Quite the opposite: pressing hard on the electrodes would clip the waveform significantly, with concurrent loss of detected signal and a sensation of pain from current flowing through the skin. A strong sense of left–right directionality was noticed as one electrode was moved across the face or head while the other was held on the chin; the signal seemed to come from the right (or left) when the moving electrode was nearest the right (or left) ear.

A sensitivity calibration was made at optimum conditions (150-V peak-to-peak signal with 100% modulation depth and electrodes placed on the chin and cheek bone) by applying a 1-kHz audio signal from an oscillator through a calibrated headphone while the subject was detecting the signal from the electrodes. By adjusting the oscillator output to achieve the same ''loudness'' level in the brain as the sound level heard normally through the ear, it was found that the equivalent sound level from the Schafer-type device was about 45 dB above the so-called average threshold of hearing ($20\ \mu\text{dyn/cm}^2$) for people with normal hearing (Fletcher and Munson, 1933). The voltage applied by Schafer was about 16 dB higher; the detected signal with his apparatus would presumably be higher by about the same amount, giving an equivalent sound level of about 60 dB (which is equivalent to ''normal conversation at 1 m'').

II. BONE CONDUCTION

The detection phenomenon was studied by attaching a fixed electrode to a body extremity (e.g., the thumb) and by moving the other electrode to various locations. In this way the rf currents are conducted through the body electrolyte to the locale of the second electrode. Because the most detection sensitivity was obtained when the moving electrode was placed on the skin close to bony areas of the face and skull and the least sensitivity when placed on the skin away from bones (e.g., on the puffed-up cheek), it seemed probable that the sensation of sound was produced by bone conduction of audio frequency vibrations to the inner ear (for example, by coupling to the bony shelf supporting the basilar membrane) and that the sound was then detected in the brain in the normal manner by pulses triggered by hair cells in the cochlea that were transmitted through the auditory nerve to the auditory cortex. Further, it seemed probable that the initial auditory vibrations were produced directly in the bone through electrostrictive mixing of the rf electrical currents in the near vicinity of the moving electrode.

The bone conduction interpretation was verified directly by listening to sounds produced in the skull and cheek bones of the test subject with a stethoscope. The audio frequency could be clearly and easily heard through the stethoscope whenever the test subject ''heard'' the detected AM carrier wave in his brain; conversely, such sounds could not be heard through the stethoscope when the electrodes made poor contact and the subject could not ''hear'' the sounds in his head. Sound perception through bone conduction of acoustic vibrations is, of course, well-known. As a double check, it was found that sound could easily be generated in other bones of the body by applying the moving electrode to the skin above them and listening to the sound with the stethoscope. The most sensitivity was obtained in this way across the tibia (shin bone), where the loudness level through the stethoscope was about 10 dB higher using the same apparatus as the maximum found across the bones in the face and skull.

Attempts to hear sounds produced through the stethoscope when the electrodes were placed on isolated, dry bones (e.g., the jaw bone from a sheep) failed, even when the bone had been soaked for many days in a saline solution. This failure may have resulted from inability to obtain a good electrical contact to the inside of the bone. As a further test, a section of quasi-fresh cow femur 18 cm long was sliced lengthwise across one diameter and the experiment repeated. (The bone came from a cow that had been slaughtered a month and a half earlier and stored in a freezer; the bone was tried in the experiment immediately after thawing.) It was necessary to remove some of the bone marrow and place one electrode inside the bone while the other electrode was moved along the outside directly opposite it in order to hear the signal through the stethoscope. The signal was clearly audible, but down by about 10 dB from the sound heard when the same electrodes were applied to the experimentalist's head as shown in Fig. 1. (The stethoscope itself provided adequate isolation from room noise to permit a quick comparison of sound levels in those two cases.) The signal from the cow femur could not be heard through the stethoscope when both electrodes were placed at locations on the outside of the bone.

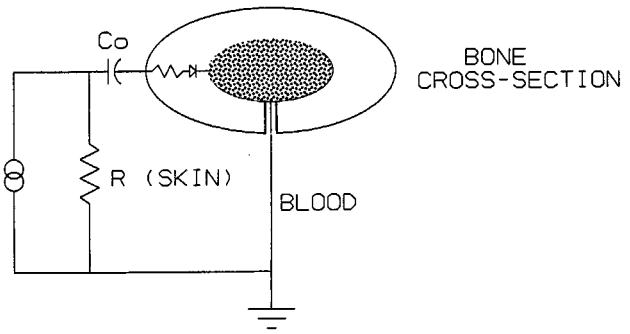


FIG. 2. Capacitive coupling to the bone nonlinearity.

III. THE AM DETECTION PROCESS

First, we should emphasize that AM detection of the modulated carrier wave is an important part of the sound-generation process. Direct application of audio signals at the same peak-to-peak voltage through the operational amplifier to the electrodes in Fig. 1 produced no detectable sound in the brain. This can be understood in terms of the schematic illustration of the paths of electrical conduction to the bone shown in Fig. 2. It is known that the main flow of blood to the bone (hence, the main path of low-frequency electrical conduction from the body electrolyte) is through arteries and veins to the *inner* region of the bone (Williams and Warwick, 1980). Many of these vessels enter at the bone ends, and in some cases the flow is supplemented by major arteries and veins that travel to the inside through oblique holes in the outer nonconducting bone casing. Anatomical studies have shown that the major flow of blood to the bone wall is ''centrifugal'' (from inside out) through a network of small capillaries. Hence, as indicated schematically in Fig. 2, the inner portion of the bone containing collagen fibers running in the lengthwise direction of the bone is essentially at ground potential (i.e., roughly the same electrical potential as the electrolyte elsewhere in the body). However, the outer portion of the bone containing small apatite crystals is an insulator. Because the electrostrictive nonlinearity responsible for the generation of sound is thought to occur by current conduction from the outer regions to the center of the bone, the absence of direct sound generation by low-frequency electric fields in the present experiment is not surprising.

In contrast, rf currents can flow radially from the outer wall through the region of the apatite crystals to the inside of the bone by capacitive coupling. [According to measurements by Shamos and Lavine (1964), the resistivity of the outer sections of human bone is in the order of 10^{11}–10^{12} ohm cm and the dielectric constant for apatite is in the order of 8–9.] Because the carrier frequency is about 40 times the audio frequency and the capacitive reactance varies inversely with the frequency, currents resulting from direct application of the audio signal to the electrodes would be attenuated in respect to the rf carrier by about the same factor (32 dB). The rf carrier clearly provides a much more effective way to implant audio signals into the bone than by direct application of the audio frequency voltage. As is most apparent when the ground electrode is connected to remote regions of the body, the AM detection and sound generation occur within the bone in the immediate vicinity of the non-grounded electrode on the cheek or temple, and the process is most efficient when the bone under that electrode is very close to the skin (i.e., the coupling capacitance is largest).

The specific details of the AM detection process inside the bone are not certain. Piezoelectric effects in bones (e.g., a sudden bending of a bone produces a transitory electric field across the bone in the radial direction) have been investigated by many authors and have been known since the 1950's (Fukada and Yasuda, 1957). Similarly, the converse process of electrostriction in which application of an electric field across a fresh bone produces mechanical movement has also been studied (Aschero *et al.*, 1996) and is to be expected from principles of reversibility. As shown by Aschero *et al.* (1996), there is a tensor relationship between the applied field and the displacement such that sound propagation induced in the bone would not necessarily be in the same direction as the applied field. That these processes involve electrical nonlinearities has also been established. For example, Becker *et al.* (1964) made direct measurements of current–voltage rectification characteristics in human bone samples and found that the forward conduction direction corresponded to current flow from the outer apatite crystals of the bone inward radially to the collagen fibers near the center. Becker and his co-workers, in this and in later articles (e.g., Becker and Brown, 1965; Becker and Bachman, 1966), argued that the bone acts like a PN junction diode. But as Shamos and Lavine (1964, 1966) noted, the semi-conductor junction interpretation probably is a needlessly elaborate mechanism to account for the observed rectification properties. These authors also conclude that the rectification and piezoelectric effects are inherently related.

A. Rectifier characteristics

Both semiconductor and biological rectifiers often have current–voltage characteristics of the form,

$$I = I_0\{\exp[\alpha(V - IR)] - 1\}, \tag{1}$$

where I_0 and α include constants characteristic of the medium, I represents the current through the rectifier containing an internal series resistance, R, and V represents the voltage applied across it. The limiting value of α from various theoretical models is e/kT or about 40 V^{-1} (e is the charge on the electron, k is Boltzmann's constant, and T is the absolute temperature), but is generally found to be quite a bit lower in practice. Attempts by the present author to fit the data of Becker *et al.* (1964) to the functional form of Eq. (1) were not very satisfactory, at least partly due to the lack of data near the origin and the likelihood of contact potential errors; very rough estimates of α from those data were in the range from about 1 to 3 V^{-1}, with values of R in the order of 10 GΩ and I_0 of about 15–40 pA. As has occasionally been found with crystal rectifiers (Torrey and Whitmer, 1948), the slope of the I–V characteristic on a semi-log plot (hence the value of α) decreased sharply at about 0.5 V in the forward direction.

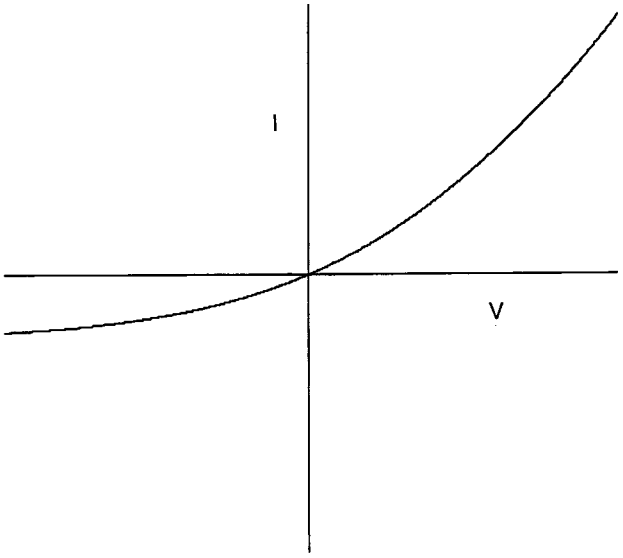


FIG. 3. Current–voltage characteristic of a typical solid-state rectifier in the presence of a large series resistor.

The functional form in Eq. (1) has been well-known in solid-state electronics since the study of crystal rectifiers for radar detection in World War II (Torrey and Whitmer, 1948). Reasons for its occurrence in biological examples have been discussed more recently by Barnes and others (Barnes and Hu, 1977; Barnes, 1986). The form of this characteristic is illustrated in Fig. 3 for conveniently chosen values of the parameters. The characteristic approaches that of an ideal rectifier (low forward resistance and large backward resistance) as α becomes large and R becomes small. Near the origin in Fig. 3, the characteristic has a strong quadratic dependence on V. That is,

$$I = I_0[\alpha V + \alpha^2 V^2/2 + \cdots]/[1 + I_0 \alpha R] \quad \text{for small } V. \quad (2)$$

But at large voltages in the forward direction, the characteristic becomes approximately linear ($dV/dI = R$ at large I). Hence, the characteristic: (1) can rectify at low voltages near the origin; (2) can act as a square-law mixer near the origin; (3) behaves like a linear resistor at large displacements from the origin.

B. AM detection

There are two different ways by which the characteristic in Fig. 3 could detect amplitude modulation.

1. Envelope detection

If significant current flows radially across the bone (say, from the electrode outside the skin to the inside of the bone and then through the blood stream to the distant electrode), the rectifier could act as an envelope detector similar to that found in an AM radio in an equivalent circuit such as that shown in Fig. 4. Here, Co represents the series coupling capacitance to the bone from the electrode. The elements within the dashed box are within the bone and localized to one spot along its length. When operated near the origin of the characteristic in Fig. 3, the circuit permits rectifying the carrier and extracting the audio envelope voltage across the resistance Rt, while the rf carrier itself is conducted back to ground through the capacitance Ct. Voltage across the resistor Rt would then result in an electric field varying at audio frequencies across the bone that in turn would generate sound waves by electrostriction.

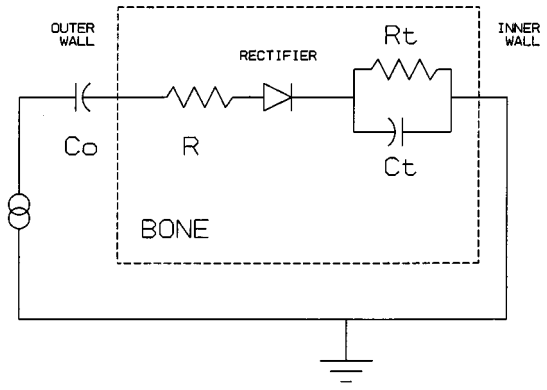


FIG. 4. Equivalent circuit of a hypothetical envelope detection method.

The spectrum for a 100% amplitude-modulated carrier wave of the type produced by the function generator in Fig. 1 consists of a carrier frequency surrounded by sidebands that are reduced by 6 dB in amplitude from the carrier and spaced at the audio frequency from the carrier. The envelope detector extracts one audio signal at the single common difference frequency between the carrier and the upper and lower sidebands.

2. Nonlinear mixing

Here we need only assume that the bone represents a nonlinear resistive element. For example as shown in Eq. (2), the characteristic for the general rectifier in Fig. 3 contains a strong quadratic term near the origin. Consequently, running a current of the form

$$I = I_1 \sin \omega_1 t + I_2 \sin \omega_2 t \quad (3)$$

through that characteristic would generate voltages at both the difference frequency $(\omega_1 - \omega_2)$ and at the sum frequency $(\omega_1 + \omega_2)$. Similarly, running a current through this nonlinearity made from the three spectral components of the 100% amplitude-modulated wave from the circuit in Fig. 1, would result in two difference frequency components—one at the original audio modulation frequency and one at twice the audio frequency. Since they are precisely an octave apart, the resultant wave would seem to have the same pitch as the original audio signal and the mechanism could not easily be distinguished by the apparent sound quality.[2]

To investigate the effect further, the sum of two separate sinusoidal signals of frequencies 40 kHz and 40 kHz minus the audio frequency were substituted for the function generator in Fig. 1. The peak-to-peak voltage between the electrodes was about 100 V in this configuration. Again the audio difference frequency could be heard clearly upon application of the electrodes with the same general results as previously described. The experiment with the stethoscope

on the tibia was repeated and gave the previous results. Clearly, whatever nonlinearity is involved works as an effective single-sideband detector.

One could argue that the present results taken by themselves demonstrate that living bone is indeed a rectifier: Any physically reasonable process would permit expanding the current through the bone in a power series in the applied voltage, for which the first possible nonlinear term is quadratic. (The linear term is just Ohms' Law.) Whether or not even-symmetric terms occur in this expansion depends on the basic properties of the material. In the present work, the difference frequency (and *not*, for example, twice or three times the difference frequency) between two rf sinewaves was produced directly in the bone. That implies that the first nonlinear term in the expansion must be quadratic. But the presence of both a linear and a quadratic term as in Eq. (2) means by definition that the current–voltage characteristic produces rectification. Hence, in the present case the two possible methods of detection outlined above yield fairly similar results. With a quadratic mixer, one does not need the capacitor C_t in Fig. 4 to generate a voltage at the audio difference frequency. However, the existence of the fading effect discussed below suggests that such a capacitance is actually present.

IV. FADING

In both forms of operation, the signal faded out quickly if one of the electrodes was not moved slowly—typically, at about 1 mm/s. (The effect is psychologically similar to producing sound by bowing a violin string.) The effect was independent of the electrode material to the extent that no change occurred when the brass was replaced by gold or silver. The form of the general rectifier characteristic in Fig. 3 suggests a possible explanation of the fading phenomenon that could apply to either mode of detection. Rectification of the signal would result in biasing the current voltage characteristic toward the linear region in Fig. 3. The magnitude of this self-biasing effect would depend on the unknown values of R_t and C_t in Fig. 4. But regardless of whether the rectifier operates as an envelope detector or as a nonlinear mixer, self-biasing to the linear region would kill the detection process.

V. RELATION TO OTHER WORK

In a recent publication, Lenhardt *et al.* (1991) reported some experiments on ultrasonic speech perception. They used an ultrasonic vibrator placed on the skull and a carrier wave in roughly the same frequency range tried in the present electric-field experiments. They suggested that ultrasonic waves were transmitted by bone conduction to a region near the cochlea, where an organ that normally is not associated with sound detection such as the saccule (which normally responds to acceleration and helps provide a sense of balance) might act as a transducer that would send nerve pulses to the brain. They used both suppressed carrier (double-sideband) amplitude modulation and single-frequency ultrasound sources in their work at acceleration levels ranging from 125 to 155 dB above threshold accelerations for detection at audio frequencies.

As Dobie and Wiederhold (1992) pointed out, it would only take a small quadratic nonlinearity somewhere in the transfer path to produce a difference frequency in the normal audio range in the double-sideband experiments that could easily be detected by bone conduction to the cochlea. It is, of course, possible that a mechanical nonlinearity might be present in the bone itself near the ultrasound generator. It is also conceivable that there was enough spectral breadth in the ''single-frequency'' ultrasonic source to permit homodyne detection by such a nonlinearity; that is, if the spectral width of the signal were significantly greater than the lower-frequency limit of normal hearing (say, 20 Hz), different portions of the spectral line could beat together to produce an audible signal in the presence of a nonlinearity. One simple way to check for such a nonlinearity in the bone would be to try the same technique used in the present work: that is, apply the double sideband and other ultrasound signals to a bone like the tibia that is far removed from the normal auditory channels and listen for audio signals generated in the bone itself using a stethoscope.

VI. CONCLUSION

Although it is apparent that Schafer's interpretation based on cortical detection of the AM carrier was wrong, he nevertheless did (unknowingly) discover an effective way to couple sounds into the bones of the skull. The present experiments imply that detection occurs by rf current flow through a nonlinearity in the radial direction of the bone between the outer surface and the inner blood supply. The nature of the nonlinearity also implies that rectification occurs in the bone and that the generation of sound is through electrostriction. Thus a hearing aid based on Schafer's design should help when deafness occurs through malfunction of the middle ear. The method also should be immune to the effects of positive feedback that have plagued more straightforward hearing aids based on audio amplification. Although the voltages required are a trifle awkward to apply, the approach is not invasive. The present results show that single-sideband modulation using one carrier frequency could be used to produce detection of sound in such a hearing aid. Further, the hearing aid could be binaural. For example, the lower sideband could be used to feed an electrode near the left ear with the left channel signal, and the upper sideband could be used to feed an electrode near the right ear. As noted above, a ground connection elsewhere in the body would obviate the awkward necessity to wear an electrode on the chin. But the fading effect still remains a problem.

ACKNOWLEDGMENTS

The author is indebted to Daniel Gould, Peter Kindlmann, Robert Apfel, Todd Brooks, Robert Grober, and Roman Kuc at Yale University for helpful discussions and their willingness to participate in the experiment as subjects. He is

also indebted to James Apuzzo of the Orange Mart butcher store for preparing the sample of cow femur used in the experiment.

[1]Because of the common use of amplitude modulation in radio communication, the term ''rf'' (radio frequency) is used throughout this article to designate the modulated wave, even though the carrier frequencies are below the conventional radio frequency band.

[2]An attempt was made to pick up the sound from the bone with a microphone with the object of Fourier analyzing the sound spectrum. Unfortunately, the entire body radiates like an antenna under the present conditions and the rf field from the body was sufficiently strong that nonlinearities in the microphone available detected the audio signal even when it was not in contact with the skin.

Aschero, G., Gizdulich, P., Mango, F., and Romano, S. M. (**1996**). ''Converse piezoelectric effect detected in fresh cow femur bone,'' J. Biomech. **29**, 1169–1174.

Barnes, F. S. (**1986**). ''Extremely low frequency (ELF) and very low frequency fields: Rectification, frequency sensitivity, noise and related phenomena,'' in *CRC Handbook of Biological Effects of Electromagnetic Fields*, edited by C. Polk and E. Rostow (Chemical Rubber Co., Boca Raton, FL), pp. 121–138.

Barnes, F. S., and Hu, L. S. (**1977**). ''Model for some nonthermal effects of radio and microwave fields on biological membranes,'' IEEE Trans. Microwave Technol. **25**, 742.

Becker, R. O., and Bachman, C. H. (**1966**). ''Bioelectric effects of tissue,'' Clin. Orthopaed. Related Res. **43**, 251–253.

Becker, R. O., Bassett, C. A., and Bachman, C. H. (**1964**). ''Bioelectricity factors controlling bone Structure'' in *Bone Biodynamics*, edited by H. M. Frost (Little, Brown, and Company, Boston), pp. 209–232 (see, especially, Fig. 16).

Becker, R. O., and Brown, F. M. (**1965**). ''Photoelectric effects in human bone,'' Nature (London) **206**, 1325–1328.

Bennett, Wm. R., Jr. (**1994**). *Health and Low-Frequency Electromagnetic Fields* (Yale U.P., New Haven, CT), pp. 130–132.

Dobie, R. A., and Wiederhold, M. L. (**1992**). ''Ultrasonic hearing,'' Science **255**, 1584–1585.

Fletcher, H., and Munson, W. A. (**1933**). ''Loudness, its definition, measurement, and calculation,'' Bell Syst. Tech. J. **12**, 377–431.

Foster, K. R., and Finch, E. F. (**1974**). ''Microwave hearing: Evidence for thermoacoustic auditory stimulation by pulsed microwaves,'' Science **185**, 256–258.

Frey, A. H. (**1961**). ''Auditory response to rf energy,'' Aerosp. Med. **32**, 1140.

Frey, A. H. (**1962**). ''Human auditory system response to modulated electromagnetic energy,'' J. Appl. Physiol. **17**, 689.

Fukada, E., and Yasuda, I. (**1957**). ''On the piezo-electric effect of bone,'' J. Phys. Soc. Jpn. **12**, 1158.

Hartmann, W. M. (**1996**). ''Pitch, periodicity, and auditory organization,'' J. Acoust. Soc. Am. **100**, 3491–3502.

Hudspeth, A. J. (**1989**). ''How the ear works,'' Nature (London) **341**, 397–404.

Lenhardt, M. L., Skellett, R., Wang, P., and Clarke, A. M. (**1991**). ''Human ultrasonic speech perception,'' Science **253**, 82–85.

Lin, J. C. (**1976**). ''Microwave auditory effect—A comparison of some possible transduction mechanisms,'' J. Microwave Power **11**, 77.

Lin, J. C. (**1978**). *Microwave Auditory Effects and Applications* (Charles C. Thomas, Springfield, IL).

Schafer, C. R. (**1977**). ''Hearing aid,'' U.S. Patent 4,052,572.

Schafer, C. R. (**1980**). ''Hearing aid with modulated suppressed carrier,'' U.S. Patent 4,220,830.

Schafer, C. R. (**1987**). ''Cortical hearing aid,'' U.S. Patent 4,711,243.

Shamos, M. H., and Lavine L. S. (**1964**). ''Physical bases for bioelectric effects in mineralized tissues,'' Clin. Orthopaed. Related Res. **35**, 177–188; Table I.

Shamos, M. H., and Lavine, L. S. (**1966**). ''Bioelectric effects of tissue,'' Clinical Orthopaed. Related Res. **43**, 254–255.

Torrey, H. C., and Whitmer, C. A. (**1948**). *Crystal Rectifiers, MIT Radiation Laboratory Series*, edited by S. A. Goudsmit, J. L. Lawson, L. B. Linford, and A. M. Stone (McGraw-Hill, New York), pp. 82–90.

Williams, P. L., and Warwick, R., Eds. (**1980**). ''Blood and vessels of the bone,'' in *Gray's Anatomy* (Saunders, Philadelphia), p. 257.

Zwislocki, J. J. (**1981**). ''Sound analysis in the ear: A history of discoveries,'' Am. Sci. **69**, 184–192.

Auditory and linguistic factors in the perception of voice offset time as a cue for preaspiration

Jörgen Pind
Department of Psychology, University of Iceland, Oddi, Reykjavík, IS-101 Iceland

(Received 21 January 1997; revised 10 December 1997; accepted 7 January 1998)

Previous research [J. Pind, Acta Psychol. **89**, 53–81 (1995)] has shown that preaspiration in Icelandic, an [h]-like sound inserted between a vowel and the following closure, can be cued by Voice Offset Time (VOffT), a speech cue which is the mirror image of Voice Onset Time (VOT). Research has also revealed that VOffT is much more sensitive to the duration of the neighboring vowel than is VOT [J. Pind, Q. J. Exp. Psychol. **49A**, 745–764 (1996)]. This paper explores the hypothesis that it is primarily the perceived quantity of the vowel that is responsible for the effect of the vowel on the perception of preaspiration. This hypothesis is based on the linguistic fact that preaspiration can only follow a phonemically short vowel. This linguistic hypothesis is contrasted with an auditory hypothesis in terms of forward masking. Perceptual experiments show that the perceptual boundaries for preaspiration can be affected either by changing the preceding vowel's duration or its spectrum. If the spectrum of the vowel changes towards that of a long vowel, longer VOffT's are needed for listeners to perceive preaspiration, thus lending support to the linguistic hypothesis. © *1998 Acoustical Society of America.* [S0001-4966(98)03004-5]

PACS numbers: 43.71.Es [WS]

INTRODUCTION

The study of speech perception over the last decades has revealed the immense variability of the speech signal, both in the spectral and temporal domains (Kluender, 1994; Perkell and Klatt, 1986). Yet, listeners show a remarkable ability to understand the message being conveyed by the talker and, in general, are little put off by the variability in the speech signal. Considering the temporal domain, it is clear that numerous factors, extrinsic to the linguistic message, affect the nature of the speech signal at any one moment. The clearest example of this relates to speaking rate. Speaking rate affects the duration of individual speech segments, shortening or lengthening them in accordance with the speaking tempo adopted by the speaker. Some speech cues are of a temporal nature, carrying phonetic distinctions by their durations. This holds, e.g., for Voice Onset Time (VOT), a speech cue which distinguishes voiced from voiceless stop consonants (Lisker and Abramson, 1970). The duration of VOT signals the presence of a syllable-initial voiced or voiceless (aspirated) stop. Another temporal speech cue is that of vowel and consonant durations in a language such as Icelandic. In those cases where speech signal durations carry phonemic distinctions, the question naturally arises as to how the listener can keep track of those durational changes which are of phonetic significance and those which are not.

It has often been pointed out that listening is rate dependent (Miller, 1987) with phoneme boundaries adapting to the speaking rate. Such a change in phoneme boundaries at different speaking rates is well-attested, being found for contrasts such as that between /b–w/ (Miller and Liberman, 1979) and VOT (Summerfield, 1981; Volaitis and Miller, 1992) in English. Such shifts in phoneme boundaries are, of course, also seen in languages other than English, e.g., see Pind (1995c) for the perception of VOT and vowel and consonant durations in Icelandic at different speaking rates. While the findings showing rate-dependent speech perception are not in doubt, the interpretation of these findings has been disputed. The explanation that rate-dependent perception involves an active ''taking-into-account'' mechanism, originally hypothesized by Miller and Liberman (1979), has been adopted by many; others have, however, argued for an explanation in terms of auditory contrast (Diehl and Walsh, 1989) or even proposed that sometimes the concept of rate normalization can be dispensed with as a durational ratio can be seen to provide an invariant cue, e.g., for the perception of duration as cue for quantity in Icelandic, although not for VOT (Pind, 1995c).

Icelandic has two speech cues involved in stop consonant voicing. One of these is VOT and serves to distinguish voiceless unaspirated from voiceless aspirated syllable-initial stops in words such as *dæla* [tai:la], ''to pump,'' and *tæla* [t^hai:la], ''to seduce.'' The other cue is that of Voice Offset Time (VOffT) which is a cue for preaspiration, an h-like sound inserted between a vowel and the following stop closure (Pind, 1995a, 1996a), compare *baka* [pa:ka] ''to bake'' with *bakka* [pahka] ''to turn back.''

I have previously reported studies on the perception of preaspiration in Icelandic (Pind, 1995a, 1996a). These studies mostly investigated the perception of preaspiration using synthetic speech. Preaspiration was synthesized with VOffT, a speech cue which is a mirror image of VOT. Interestingly, the identification functions for VOffT turned out to be quite similar to those for VOT. However, in one respect, the perception of VOT and VOffT did diverge. As far as adjustments to vowel lengthening was concerned, VOffT showed much greater shifts in the location of the phoneme boundaries with increases in vowel duration than did VOT. In one experiment, lengthening the vowel from 132 to 232 ms moved the VOT boundaries in /ba–pa/ continua from 29.7 to 35.5 ms, a shift of a little less than 6 ms. The corresponding

shift for VOffT continua, /ap–ahp/, was close to 22 ms, from 30.7 to 52.6 ms of VOffT (Pind, 1996a).

One interpretation of the differential effect of vowel duration on the perception of VOT and VOffT might be along lines suggested by Bladon (1984). In a paper discussing possible auditory contributions to speech perception, Bladon mentions preaspiration as an example of an auditorily disadvantaged phonetic contrast, noting that it ''would be hard to imagine a speech pattern less favourably designed for the hearer... [and] so, given that preaspiration suffers from an accumulation of auditory handicaps, it would not be a risky prediction that languages would rarely make use of this auditory-phonetic dinosaur (p. 7)'' The auditory factors which Bladon claims put preaspiration at a disadvantage depend amongst other on the following: (a) wide-spread short-term adaptation of neural firings following the vowel, and (b) the fact that preaspiration depends on the detection of offsets. The former factor may also presumably be characterized through asymmetric masking, i.e. the fact that forward auditory masking usually produces a greater effect than backward masking (Moore, 1997). Thus one might hypothesize that the more salient vowel would produce greater masking of aspiration following the vowel (preaspiration) than of aspiration preceding the vowel (e.g., of syllable-initial aspiration of stops). The much greater shifts in the phoneme boundaries for VOffT than VOT could be taken to support Bladon's hypothesis, interpreted psychoacoustically as the effect of forward masking or, in neural terms, as showing short-term adaptation to the preceding vowel.

Another hypothesis, however, also suggests itself. There is, in Icelandic, a fundamental linguistic asymmetry between the speech cues of VOT and VOffT. VOT is syllable initial and can be followed by either a short or a long vowel. Preaspiration, coming as it does *after* the vowel, can only follow a phonemically short vowel. Obviously, when inducing different perceptions of rate by manipulating vowel durations, it is quite likely that changes in vowel duration might affect the perception of quantity rather than rate in those languages which distinguish phonemically long and short segments. (In studies conducted with English listeners it is generally assumed that lengthening of the vowel in syllables like /ba/ and /pa/ will be perceived as a change in utterance rate.)

Clearly, for it to be possible to distinguish between the effects of rate and quantity we need to be able to manipulate them independently by experimental means. Speech segment durations are inherently ambiguous as to their perceptual effect. Lengthening vowels in Icelandic could either be perceived as a change in quantity or in speech rate [under some conditions these effects can in fact be distinguished, see Pind (1986, 1995b)] and is therefore potentially ambiguous. So the question arises as to whether it is possible to change perceived quantity *without* changing speech segment durations. A recent study (Pind, 1996c), using edited natural speech, indicates that this is possible, at least in those cases where the spectral structure of long and short vowels is different.

The present series of experiments is thus concerned with the interaction of auditory and linguistic factors in the perception of preaspiration, testing the hypothesis that there is more to the perception of preaspiration than can be explained in purely auditory terms. In particular, it will be claimed that *perceived quantity*, a language-specific feature, has a major impact on the perception of preaspiration.

The first experiment makes use of sine-wave analogs of VOT and VOffT to investigate the aforementioned asymmetry observed in the effects of vowel duration on the location of the phoneme boundaries for VOT and VOffT. If Bladon's (1986) hypothesis that preaspiration suffers from an ''accumulation of auditory handicaps'' can in fact be used to explain this asymmetry then we would predict that a similar asymmetry should be observed for the nonspeech analogs.

The first experiment is followed by two experiments which establish that the spectral character of the vowel [ɛ] in Icelandic has a marked influence on the perception of its quantity, thus replicating, using synthetic speech, an earlier finding with edited natural speech (Pind, 1996c). This finding is used in the final experiment to manipulate the quantity of vowels apart from their durations, thus enabling us to separate the influence of the two factors of duration and perceived quantity on the perception of preaspiration. Our hypothesis is that changing the quantity of a vowel without changing its duration will also affect the perception of preaspiration, in particular that longer VOffTs are needed to cue preaspiration as perceived quantity tends toward a phonemically long vowel. Such a finding would support the theory that vowel duration primarily influences the phoneme boundaries of preaspiration through its effect on perceived quantity, and thus is a linguistic rather than a psychoacoustic effect.

I. EXPERIMENT 1

Trying to disentangle auditory and linguistic factors in speech perception is no easy task and is complicated by the fact that investigators are not agreed upon the proper control to contrast with speech. Basically, it would seem that auditory stimuli which preserve, to some critical extent at least, the temporal and spectral structure of speech, without being perceived as speech, would form the ideal control. A commonly used nonspeech control is provided by ''sine-wave analogs,'' stimuli where some feature of the formant tracks are replaced by sine-waves. Thus continua of sine-wave tone onset times have been compared with those of speech voice onset time (Hillenbrand, 1984; Parker, 1988; Pisoni, 1977). While it is known that sine-wave analogs can be perceived in a speechlike manner (Remez *et al.*, 1981), thus casting doubt on their use as ''nonspeech'' controls, this seems mainly to occur with longer utterances using complex variations in frequency, which faithfully mirror the formant movements (Darwin and Carlyon, 1995).

In this experiment two-tone stimuli were used in an identification experiment which was modeled on a recent study involving categorization of speech stimuli, namely continua of VOT and VOffT (Pind, 1996a). The two component sine-wave controls simulate just one aspect of the speech continua, namely their voice offset and onset, but did not at all sound speechlike. The results of the speech perception experiment, discussed in the introduction to this paper,

showed a marked asymmetry in the effect of vowel duration on the placement of phoneme boundaries on the VOT and VOffT continua. If Bladon's (1986) hypothesis about the perception of preaspiration is correct, we should expect to find comparable asymmetries in the behavior of the non-speech analogs used in the present experiment.

A. Method

1. Participants

Fourteen undergraduate psychology majors at the University of Iceland participated in the experiment for course credit. All reported normal hearing. Participants in this and other experiments reported here were all native speakers of Icelandic.

2. Stimuli

The stimuli in this experiment were patterned after those originally used by Pisoni (1977). There were six stimulus continua made of two component pure tones, one having a frequency of 500 Hz, the other a frequency of 1500 Hz, this latter tone being 12 dB lower in intensity than the lower frequency tone. Both tones had 5 ms long cosinusoidal attacks and decays.

The six continua were made by varying two parameters: (a) The two tones were offset (by shortening the lower frequency tone) either at the start of the tones or their end. In the former case we have a continuum of tone onset times (TOTs), in the latter a continuum of tone offset times (TOffTs). The tone onsets and offsets were varied in 5-ms steps ranging from 0 to 50 ms in each of the six continua. (b) The second variable was the overall duration of the higher frequency tone. This was fixed at respectively 132, 180, and 232 ms, using the same durations as in the experiments reported in Pind (1996a). The stimuli were generated using hardware and software from Tucker-Davis Technologies and presented to subjects on line over Sennheiser HD 513 circumaural headphones in a sound-treated room. The peak presentation level was approximately 75 dB SPL.

3. Procedure

The listening test took place in one session. Half the participants listened to the tone onset continua followed by the tone offset continua, for the other half the order was reversed. Each half of the experiment consisted of two parts, a training phase and a testing phase. The training phase consisted of 4 blocks of 24 stimuli each. In the first two training blocks only the endpoints of the stimulus continua were used, i.e., stimuli having either 0- or 50-ms TOT or TOffT. In training blocks 3 and 4 stimuli having 5- and 45-ms of TOT or TOffT were also used. Participants were told to press button ''A'' on the computer keyboard if the tone sounded ''continuous'' and ''uniform'' throughout its duration and to press ''B'' if the tone sounded somehow ''discontinuous'' or ''broken'' to them. During the training phase participants received visual feedback as to the correctness of their responses, response ''A'' having been defined as the correct response for TOTs and TOffTs of 0 and 5 ms, and response ''B'' being defined as correct for the stimuli having 45 and 50 ms of TOT or TOffT.

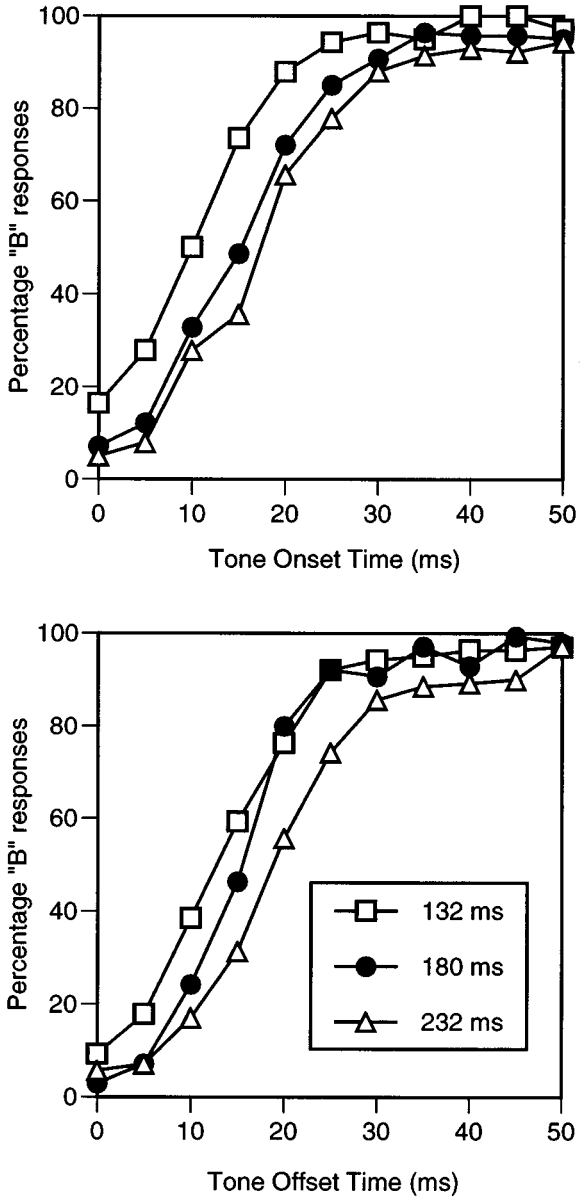


FIG. 1. Pooled identification curves for all 14 participants in experiment 1. The upper figure shows the results for the Tone Onset Time continua, and the lower figure the results for the Tone Offset Time continua.

Following the training phase came 10 blocks, each consisting of a single randomization of the 33 different stimuli. The experimental session lasted approximately one hour.

B. Results and discussion

Pooled identification curves for all 14 participants are shown in Fig. 1. In general, the shapes of the identification curves for TOT (upper panel of Fig. 1) and TOffT (lower panel) are quite comparable. In both cases overall duration of the stimuli is shown to exert an influence on the location of the response boundaries.

Perceptual boundaries were calculated for 13 of the 14 individual participants using the probit method (Finney,

1971) (one participant was left out since his responses did not cross the perceptual boundary in the 232-ms TOffT condition). On average the boundaries in the TOT continua were found to lie at 9.3, 14.0, and 16.3 ms, respectively in the 132-, 180-, and 232-ms long stimuli. The comparable values for the TOffT stimuli were 12.1, 13.8, and 17.6 ms. The individual boundaries were subjected to a two-way repeated measures ANOVA, stimulus series TOT vs TOffT×stimulus durations (132, 180, and 232 ms). Results showed no effect of stimulus series [$F(1,12)<1$] but a highly significant effect of stimulus duration, $F(2,24)=24.39$, $p<0.001$. The interaction of series and duration, $F(2,24)=1.46$, was not significant, $p>0.10$.

Pairwise comparisons (all using the Bonferroni correction) of the average perceptual boundaries within each stimulus series, TOT and TOffT, showed that for the TOT series only the 132- and 232-ms stimuli had significantly different perceptual boundaries [$F(1,12)=20.37$, $p<0.01$]. For the TOffT series the 180- and 232-ms stimuli showed significantly different perceptual boundaries [$F(1,12)=12.96$, $p<0.05$] as did the 132- and 232-ms stimuli [$F(1,12)=16.54$, $p<0.05$].

The purpose of the present experiment was to serve as a nonspeech analog, modeled on the experiment of Pisoni (1977), of a previous study (Pind, 1996a) comparing the perception of VOT and VOffT in Icelandic. In that study the effect of three vowel durations, 132, 180, and 232 ms, shifted VOT boundaries in /ba–pa/ continua from 29.7 ms to 35.5 ms of VOT. Shifts for VOffT boundaries in /ap–ahp/ continua were from 30.7 ms to 52.6 ms of VOffT. The total VOT shift thus amounted to 5.8 ms whereas the total VOffT shift amounted to 21.9 ms.

Two explanations of this discrepancy in the behaviour of VOT and VOffT with increasing vowel length have been discussed in the introduction to this paper. On the one hand it is possible that VOffT is more prone to auditory masking than is VOT (Bladon, 1986) coming as it does after the vowel rather than before it. If this were so one would predict that a similar effect would show up in the perception of the TOffT continua. The result of the present experiment, showing no significant difference between the responses of listeners to the two continua of TOT and TOffT with increases in stimulus length, argues against this hypothesis.

It should of course be acknowledged that the precise role which nonspeech analogs can play in speech research is open to debate. Thus the values for the perceptual boundaries are appreciably lower in these sine-wave analogs than in the previous speech study. Even so, the fact that the shifts in the perceptual boundaries shown in the present study (amounting to 7 ms) is roughly the same as for the perception of VOT in the speech study (6 ms) does lend credibility to the position of Diehl and his associates that some kind of auditory contrast underlies at least some of the effects of speaking rate seen in speech perception (Diehl and Walsh, 1989). This explanation will not suffice, however, for the shifts seen in the phoneme boundaries for VOffT, and still leaves us with the question as to why VOffT shows much greater sensitivity to vowel duration changes than does VOT. This issue will be further explored in the following experiments.

II. EXPERIMENT 2

The following two experiments bear on the following issue: Is it possible to change the perceived quantity of a vowel in Icelandic without changing its duration? Results of a recent study (Pind, 1996c), using edited natural speech, show that this can be accomplished for some vowels by changing their spectral character. The purpose of the following experiment was to gather production data which were then used as models for the definition of synthetic speech tokens used as stimuli in experiments 3 and 4. First, some brief remarks about the phonetics of Icelandic will make it easier for the reader to follow the logic of the following experiments.

Lexical stress in Icelandic always falls on the first syllable of noncompound words. This first, stressed, syllable shows distinctions of *quantity*. In general, if the vowel of the stressed syllable is followed by a single (or no) consonant, it is phonemically *long*, otherwise it is phonemically *short* (Benediktsson, 1963), cf. words such as *fara* [faːra], ''to go,'' *farga* [farka], ''to destroy.'' Note that quantity is dependent on the number of consonants following the vowel in the the same syllable. A stressed vowel followed by a succession of one of p, t, k, or s followed by one of v, j, or r is long, as in the word *lepja* [lɛːpja], ''to lick.'' Here the consonant cluster is seen to split across the syllable boundary with only ''p'' belonging to the first syllable. In those cases where a nongeminate consonant follows a vowel in the first syllable, the quantity of the consonant is complementary to that of the vowel, a long vowel being followed by a short consonant and vice versa. Thus contrast *sek* [sɛːk], ''guilty (fem.acc.sg.)'' with *segg* [sɛkː], ''man (acc.sg.).''

Icelandic has eight vowel monophthongs [i, ɪ, ɛ, a, ɔ, œ, ʏ, u]. Studies of vowel formants in the monophthongs (Garnes, 1976; Pétursson, 1974) have shown that five of these [i, ɪ, a, ʏ, u] are spectrally quite similar whether long or short whereas the three central vowels [ɛ, ɔ, œ] are spectrally dissimilar in their long and short varieties. Thus Pétursson (1974) reported an $F1$ of 550 Hz and an $F2$ of 2400 Hz for the phonemically long [ɛː]; the corresponding values for a short [ɛ] were 750 Hz and 2050 Hz. These differences are well beyond the jnd's for formant frequencies reported in the literature (Flanagan, 1972; Kewley-Port and Watson, 1994).

While the perception of quantity in Icelandic in general is mediated through a relational speech cue of vowel to rhyme ratio (Pind, 1986, 1995c) this only holds for vowels which are spectrally similar whether short or long. Thus Pind (1996c) has shown that in the case of the vowel [ɛ] the spectral dissimilarity of the long and short phonemes overrides the relational cue to quantity. Studies of Swedish (Hadding-Koch and Abramson, 1964) and Thai (Abramson and Ren, 1990) have yielded comparable results.

In this experiment, production data (segment durations and formant values) were gathered for a number of word tokens containing the vowel [ɛ], both its short and long variants. It was expected that a pattern similar to that found in Pind (1996c) would also emerge in the present series of words, namely a marked difference in the location of formant 1 and 2 depending on whether the vowel was short or long. If this were so, then it should in fact be possible to

affect the perceived quantity of a vowel by changing its spectrum rather than its duration and thus also affect the perception of preaspiration if this is affected by perceived vowel quantity rather than vowel duration (see experiments 3 and 4). The purpose of this small-scale production study is thus primarily to analyze a few word tokens to serve as models for synthetic speech stimuli used in experiments 3 and 4.

A. Method

1. Stimuli

The stimulus words chosen for this production experiment were the following: *sek* [sɛːk] ''guilty (fem.sg.acc.),'' *sekur* [sɛːkʏr] ''guilty (masc.sg.acc.),'' *segg* [sɛkː] ''man (acc.sg.),'' *seggur* [sɛkːʏr] ''man (nom.sg.),'' *sekk* [sɛhk] ''to sink (1st pers.sg.),'' and *sekkur* [sɛhkʏr] ''to sink (3rd pers.sg.).'' Notice that the word stems *sek* [sɛːk] and *segg* [sɛkː] contrast in terms of quantity whereas the word *sekk* [sɛhk] contains preaspiration.

2. Procedure

The author read 20 tokens of each of the stimulus words onto digital audiotape at a normal utterance rate. The words were read from a list which contained the six words intermixed in randomized order. Each line of the list contained five words and was read as a single utterance. For the analysis the first and last words from each utterance were discarded. The middle three words were low-pass filtered at 4.5 kHz and recorded at 10 kHz on a PC computer through a DA-converter from Tucker-Davis Technologies. All measurements were carried out using the CSRE software (Jamieson *et al.*, 1992).

The measurements proceeded as follows. First the duration of these segments were measured: (a) Duration of the first (stressed) vowel; (b) duration of the preaspirated segment (in those words having preaspiration); and (c) duration of the closure of the first syllable. The boundaries of the segments were drawn using traditional acoustic criteria (Shoup and Pfeifer, 1976) and mostly made from waveform displays (supplemented occasionally by spectrograms). Thus the beginning of the vowel was fixed at the beginning of voicing following the word initial [s]. The end of the vowel and the beginning of the following closure was drawn at the end of the last clear voicelike pulse belonging to the vowel preceding the closure. The end of the closure was easily measured from the burst accompanying the release of the closure. Estimating the boundary of vowel and the preaspirated segment proved most problematic. The vowel was considered to end when aspiration noise could be discerned in the waveform for the vowel. The spectrograms, which were displayed on the computer screen alongside the waveform displays, were helpful in locating the beginning of preaspiration. After the durations of vowels had been obtained from these measurements the midpoint of each vowel was calculated and the cursor placed as near to this midpoint as possible. Vowel formant frequencies were then measured at that point using an LPC-analysis of order 16 (on 256 samples). The formant extraction was done by the CSRE program.

TABLE I. Average durations (rounded to the nearest 5 ms) for the individual segments of the stressed rhymes measured in experiment 2. The final column gives the duration of the rhyme which is the sum of the previous three columns.

Word type	No. of syllables	Segment			
		Vowel	Preaspiration	Closure	Rhyme
[sɛːk]	1	225		175	400
	2	180		160	340
[sɛkː]	1	110		290	400
	2	85		260	345
[sɛhk]	1	90	120	225	435
	2	70	125	180	375

B. Results and discussion

Table I shows the average durations of vowels, closures, and preaspiration in the words measured along with the total rhyme duration (this being the sum of the durations of vowel and closure and, when appropriate, preaspiration). As expected the stressed rhyme is shorter in the two-syllable words (average duration 353 ms) than in the one-syllable words (412 ms). Note that the average rhyme durations are roughly comparable in the three word types, 370 ms in V:C words and 372.5 ms in VC: words (V=vowel, C=consonant). The rhyme is slightly longer in VhC type words, i.e., words containing preaspiration, 405 ms. These broadly comparable durations suggest that the rhymes may all be viewed as consisting of three segments. Such an analysis would view the phonemically long vowel as a double vowel and similarly for the long consonant (which is in fact how the Icelandic orthography presents long consonants as can be seen in the spelling of the word *seggur* [sɛkːʏr]). Consistent with this view is also the fact that preaspiration is of full segment duration, thus differing in this respect quite clearly from postaspiration as cued by VOT (Thráinsson, 1978).

In the words sek(ur) [sɛːkʏr] and segg(ur) [sɛkːʏr] the ''complementarity'' of vowel and consonant duration emerges clearly from these data. A long vowel is followed by a short consonant and vice versa. This can also be seen in Fig. 2 where the durations of vowels and closures in these words have been plotted for each word token measured in this experiment. Two clusters emerge in this figure, one corresponding to the V:C type words, the other corresponding to the VC: type words.

The average formant frequency values, measured in the middle of the stressed vowels, are shown in Table II. Looking at $F1$ and $F2$ it can be seen that the short/long distinction emerges clearly in $F2$ with an $F2$ of 1620–1630 Hz in the short vowel but at 1790 Hz in the long vowel. The results for $F1$ are somewhat mixed, the frequency being 470 Hz in the long vowel but 515 and 545 in the short vowels. The values for the short vowels are obviously more centralized than for the long vowel as expected. These values were used to set the boundaries on the $F1$ and $F2$ frequencies in the synthetic stimuli used in experiments 3 and 4, with $F1$ varying from 475 to 550 Hz and $F2$ from 1600 to 1800 Hz.

This experiment has confirmed the finding reported ear-

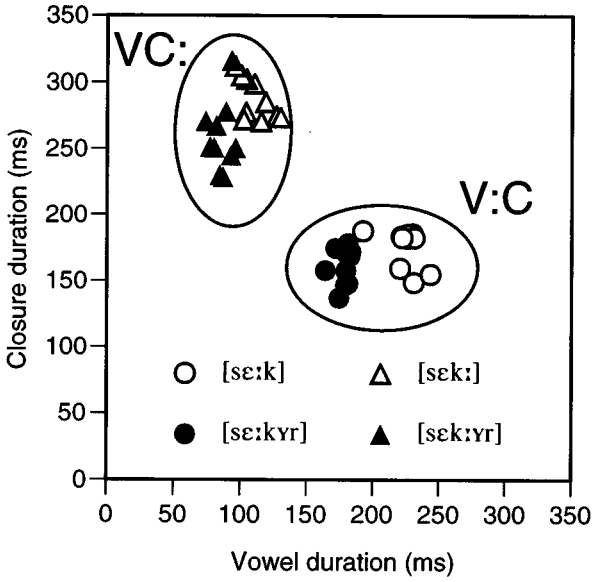


FIG. 2. The distribution of vowel and consonant (closure) durations in V:C and VC: type words. The complementary nature of the quantity opposition in Icelandic stands out in this figure. Long vowels are followed by short consonants (triangles) and short vowels are followed by long consonants (circles). Ellipses have been drawn around the data for the two types of syllables.

lier of a considerable difference in formant frequencies for long and short [ɛ] in Icelandic. This quality difference does not, however, seem to affect the durational differences of words containing long and short vowels and consonants. In view of the finding reported in Pind (1996c) that vowel quality is of paramount importance in the perception of vowel quantity, at least for the vowel [ɛ], it would not perhaps have come as a surprise were the durational patterns of words containing long and short [ɛ] to show somewhat less extreme values than in the case of those vowels which can only depend on duration to signify quantity. This does not, however, appear to be the case. Thus the vowel to rhyme ratio for the data gathered here turns out to be 0.54 for words containing a long vowel and 0.26 for the words containing a short vowel. This is comparable to values found in studies involving other vowels (Pind, 1986, 1995c).

III. EXPERIMENT 3

Previous research on the perception of quantity in Icelandic (Pind, 1996c) using edited natural speech has shown that the spectral composition of the vowel can have a marked effect on the perception of quantity in the vowel. The purpose of the present experiment was to replicate the earlier experiment but this time using synthetic speech, giving better control of the makeup of the stimuli for the experiment. The stimuli used were modeled on the words measured in experiment 2.

TABLE II. Average formant frequencies (rounded to the nearest 5 Hz) for the stressed vowel [ɛ] in the words measured in experiment 2. Each number shows the average of 20 measurements taken in the middle of the vowel. These formant frequencies were measured using LPC analysis.

Word type	*F*1	*F*2	*F*3
[sɛ:k]	470	1790	2495
[sɛk:]	545	1620	2390
[sɛhk]	515	1630	2495

A. Method

1. Participants

Thirteen undergraduate psychology majors at the University of Iceland participated in the experiment for course credit. All reported normal hearing.

2. Stimuli

The stimuli for this experiment were modeled on the words [sɛ:k] and [sɛk:] investigated in the production study reported here as experiment 2. The stimuli were synthesized using the Sensimetrics SenSyn™ synthesizer, a version of the Klatt cascade/parallel formant synthesizer (Klatt, 1980; Klatt and Klatt, 1990). The synthesizer was run in the cascade configuration. Three stimulus continua were made by varying the steady-state frequencies of *F*1 and *F*2. The following *F*1/*F*2 pairings were used: 475 and 1800 Hz: 510 and 1700 Hz; and 550 and 1600 Hz. The first pair had frequencies lying close to those measured for the long [ɛ:], the last pair had frequencies lying close to those measured in the short [ɛ]. Within each of these continua the duration of vowels and closures were varied so that they spanned the range from a durationally clear V:C type stimulus to a clear VC: type stimulus. A total of 14 stimuli made up each continuum so the total number of stimuli for this experiment was 42 (3 formant pairs×14 V/C durations).

The stimuli started out with a 100-ms fricative segment [s]. This was synthesized by fricative noise in the frequency region above 2000 Hz with the major peak lying around 3700 Hz. This fricative segment had a rising amplitude contour for the first 20 ms and a falling contour for the last 20 ms before the onset of the vowel at 100 ms.

The fricative segment was followed by the syllable rhyme, consisting of a vowel and following closure. The end of the rhyme was marked by a noise burst, centered around 1800 Hz, and thus appropriate for the velar place of articulation. The total rhyme duration was 400 ms. The final burst was approximately 30 ms long.

Within the rhyme the vowel continua ranged in duration from 108 to 264 ms. Since the duration of the closure was such that the total rhyme duration was always constant, i.e., 400 ms, the closures can be seen to range from 292 to 136 ms. The individual steps of the vowel/closure durations were 12 ms long. Thus vowel and closure durations were as follows for the first stimuli of each continua: 108+292 ms, 120+280 ms, 132+268 ms, and so on until the final durational values of 264+136 s. The vowel to rhyme ratios, V/(V+C) (Pind, 1986, 1995c), range from 0.27 to 0.66. Stimuli near the former values have durational values appropriate for a VC: type word, [sɛk:], stimuli near the other end have values characteristic for V:C words, [sɛ:k].

The stimuli were synthesized with five vowel formants. The frequencies of *F*3–*F*5 were fixed at 2600, 3250, and 3700 Hz. The steady-state values of *F*1 and *F*2 have been

given before. At the [s] to [ɛ] boundary the frequency of $F1$ rose from a starting frequency of 250 Hz to the appropriate steady-state value over 40 ms. At the [ɛ] to closure boundary 48-ms transitions characterized $F1$, $F2$, and $F3$. $F1$ fell to 250 Hz, $F2$ rose to 1950 Hz, and $F3$ fell to 2300 Hz. These transitions were all linear.

The fundamental frequency during the voiced portion of each stimulus was fixed at 125 Hz and the synthesizer was set to use an update interval of 4 ms. The stimuli were synthesized at 10 kHz and low-pass filtered at 4.5 kHz before being routed to the earphones.

3. Procedure

The experiment was carried out under computer control, with each participant being tested individually. The experiment was split into seven blocks. The first block contained 10 stimuli, 5 tokens of a stimulus which was a clear example of the word [sɛːk] followed by 5 tokens of an unambiguous example of the word [sɛkː]. Participants were informed of the nature of these stimuli and their order. This was done to briefly familiarize them with the character of the synthetic speech. Following this came a practice block (without feedback) containing a randomization of all the 42 stimuli used in the experiment. This was then followed by 5 blocks, each containing two tokens of the 42 stimuli in randomized order, for a total of 10 presentations of each stimulus. Participants listened to the stimuli in a sound-treated room. The stimuli were played at a comfortable listening level over Sennheiser HD-530-II circumaural headphones.

The participants indicated their responses either by clicking with a mouse on one of the words *sek* [sɛːk] or *segg* [sɛkː] which were written in large type on the screen of the computer or by pressing appropriately designated keys on the keyboard. The experiment was self-paced.

B. Results and discussion

Figure 3 shows the pooled identification curves for all 13 subjects. It is evident that the spectrum of the vowel [ɛ] had a marked effect on the location of the identification curves. Average phoneme boundaries, based on the total responses of all subjects and calculated using the method of probits (Finney, 1971), shows the phoneme boundary to lie at 153.3 ms in the first stimulus series ($F1=475$ Hz; $F2=1800$ Hz), at 178 ms in the second stimulus series ($F1=510$ Hz, $F2=1700$ Hz), and at 223 ms in the third stimulus series ($F1=550$ Hz, $F2=1600$ Hz). This is obviously a marked shift in the location of the phoneme boundaries. These boundaries were, however, analyzed for the total results of all subjects rather than for each subject individually. The reason for this is that the identification curves did not cross the phoneme boundaries for all the participants in the third stimulus series. For four of the participants the curves did not cross the phoneme boundary, for two of the subjects the crossover was marginal while seven subjects showed reliable crossover in the third stimulus series. This reflects the different extent to which individual participants make use of spectral versus temporal speech cues. Those participants who do not show a crossover in their perceptual

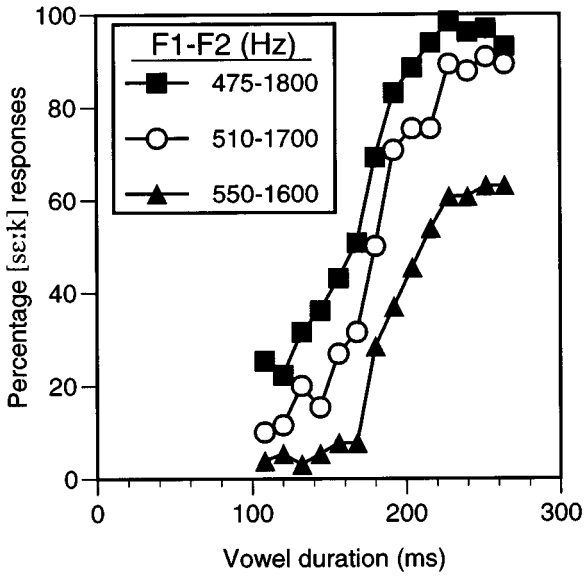


FIG. 3. Pooled identification curves for all 13 participants in experiment 3. The legends denote the values of the steady-state values of the first two formants in Hz.

classifications with increasing vowel duration are obviously giving much greater weight to the spectral cue than the durational cue. Similar individual differences were earlier reported in Pind (1996c).

Since not all listeners showed responses which crossed the phoneme boundaries the following statistical analyses are based on the number of [sɛːk] responses rather than on calculated phoneme boundaries. In series 1 the number of [sɛːk] responses averaged 92.9 across all subjects, 74.4 in series 2, and 44.5 in series 3. A one-way repeated measures ANOVA shows these numbers to be significantly different, $F(2,24)=15.41$, $p<0.0001$. Pairwise comparisons (all using the Bonferroni correction) reveal each comparison to show a significant difference ($p<0.001$).

The present experiment replicates, using synthetic speech, the finding reported in Pind (1996c) that the spectral character of a vowel can significantly influence the perception of quantity in the syllable to which the vowel belongs. Earlier studies have shown quite clearly that the major cue to quantity in Icelandic is relational, defined by the ratio of vowel to syllable rhyme duration (Pind, 1986, 1995c). Interestingly, this cue is potentially invariant over transformations of rate and experimental studies have indeed shown that listeners make use of such an invariant ratio in perception of speech at different rates (Pind, 1995c). Two qualifications, however, need to be made in regard to this relational invariant. First, as the present experiment has shown, the spectral makeup of the vowel is important in those cases where there is a clear spectral difference between the long and short vowel phonemes. This is so even though all the vowels were released and thus it might have been expected that listeners could make use of the relational invariant in the stimuli. Obviously, the fact that they did not, suggests that the spectral differences will override the temporal cue in those cases were spectral differences are sufficiently great between long and short vowels. This has been demonstrated here for the

vowel [ɛ] and presumably also holds for the vowels [ɔ] and [œ] in Icelandic. Second, it has recently been shown that in those cases where listeners are not able to perceive the whole syllable rhyme, in particular in syllables with unreleased closures, and are thus unable to calculate the relational cue, they become more susceptible to external influences, e.g., of the rate in a precursor sentence (Pind, 1996b). This is as expected and underscores the important role which a relational invariant can play in the perception of temporal speech cues.

IV. EXPERIMENT 4

The Introduction to this paper cited the finding that the perceptual boundaries for VOffT are influenced by vowel duration to a greater extent than is VOT. Two potential explanations for this finding were offered.

(1) The longer phoneme boundaries for VOffT with longer vowels could presumably be explained in psychoacoustic terms, showing the operation of forward masking or neural short-term adaptation in speech perception (Bladon, 1986).
(2) A ''linguistic'' explanation is also possible (Pind, 1995a, 1996a), namely that the lengthened boundaries for VOffT with increases in vowel duration might be due to the fact that preaspiration always follows a phonemically short vowel. As vowel duration is increased the likelihood of perceiving a long vowel increases, which would oppose the perception of the syllable as containing preaspiration.
(3) To this may be added a third hypothesis, previously entertained by the present author (Pind, 1995a), which posits that the perception of preaspiration is strictly a matter of the ratio between the durations of VOffT and the preceding vowel yielding an invariant VOffT/vowel ratio at the phoneme boundary for preaspiration (this hypothesis differs from the second in excluding any contribution of the preceding vowel's spectrum).

Assuming that experiment 1 can be taken to argue against the first explanation we still need to experimentally decide between the latter two possibilities. To do so requires a method whereby we could influence the perceived quantity of the vowels *without* having to change their duration. The fact that the spectrum of the vowel influences the perceived quantity makes it possible to accomplish precisely this experimental manipulation, namely to manipulate the perceived quantity of a vowel without changing its duration. This experiment tests the hypothesis that the phoneme boundaries for VOffT can be moved by changing the quantity of the vowel, either by changing its duration (while keeping rhyme duration constant) or by changing its quality to reflect either a short or a long vowel. Changing the quality of the vowel towards that of the phonemically long vowel is hypothesized to have the same effect as increasing the duration of the vowel while keeping quality constant.

A. Method

1. Participants

The participants were 13 undergraduate psychology students at the University of Iceland who participated in the experiment for course credit. All reported normal hearing. Seven of these participants had earlier taken part in experiment 3.

2. Stimuli

The stimuli for this experiment were based on those used in experiment 3. Two stimuli from each of the three continua used in the previous experiment were used to make up six continua varying in VOffT. For each of the three $F1/F2$ series used in experiment 3 two pairs of vowel and closure durations were chosen, namely 108 ms+292 ms and 240 ms+160 ms (the former values are appropriate for a VC: type word, the latter for a V:C type word).

Each of these six stimuli formed the basis of a separate VOffT continuum. VOffT was synthesized with a combination of aspiration and frication, the latter in the frequency region of $F2$ and $F3$, while simultaneously cutting off the amplitude of voicing and increasing the bandwidth of $F1$ from 60 to 200 Hz. This was done in nine 8-ms steps with VOffT ranging from 0 to 72 ms in each stimulus continuum. In all other respects the stimuli were synthesized in the manner described previously for experiment 3. Note that the VOffT is synthesized from the *end* of the vowel. A VOffT of 8 ms thus has the final 8 ms of the vowel replaced by aspiration, a VOffT of 16 ms similarly replaces the last 16 ms of the vowel by aspiration, and so on.

A total of 10 stimuli made up each continuum so the total number of stimuli for this experiment was 60 (3 $F1/F2$ pairs×2 V/C durations×10 VOffT values).

3. Procedure

The experiment was carried out under computer control, using the same setup as in experiment 3. The experiment was split into seven blocks. The first block contained 21 stimuli, 7 tokens each of clear examples of the words *sek* [sɛk], *segg* [sɛkː], and *sekk* [sɛhk]. Following this the three words were played three times in a row. Participants were informed of the nature of these stimuli and their order. This was done to briefly familiarize them with the character of the synthetic speech. Following this came a practice block (without feedback) containing a randomization of all 60 stimuli used in the experiment. This was then followed by five blocks, each containing two tokens of the 60 stimuli in randomized order, for a total of ten presentation of each stimulus. Participants listened to the stimuli in a sound-treated room. The stimuli were played at a comfortable listening level over Sennheiser HD-530-II circumaural headphones.

The participants indicated their responses either by clicking with a mouse on one of the words *sek*, *segg*, or *sekk*, which were written in large type on the screen of the computer, or by pressing appropriately designated keys on the keyboard. The experiment was self-paced.

B. Results and discussion

In this experiment participants had three response categories to choose from. Figure 4 shows pooled identification curves for all three response categories broken down by vowel duration and $F1/F2$ formant frequencies. The upper

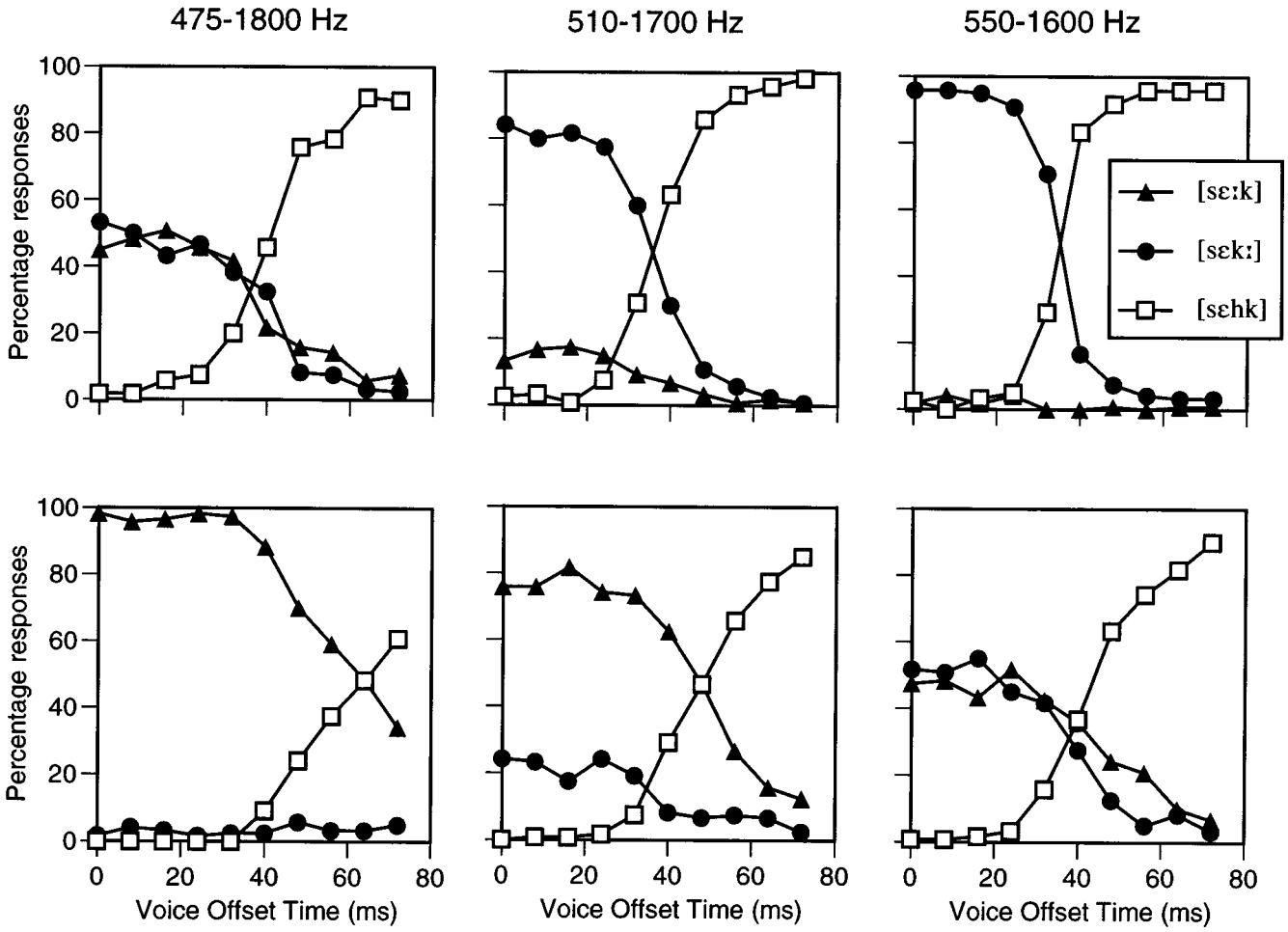


FIG. 4. Pooled identification curves for all 13 participants in experiment 4. The upper three figures show the responses for the continua with the shorter vowel (108 ms), and the lower three figures the responses for the continua with the longer vowel (240 ms). From left to right the figures denote the responses with changing steady-state values of $F1$ and $F2$ as shown at the top.

three panels show the results for the three continua with the 108-ms long vowels, and the lower three panels the results for the three continua with the 240-ms long vowels. Calculating phoneme boundaries for the total results as presented in Fig. 4, using the method of probits, yields the following values, expressed in ms of VOffT. In the stimuli with a 108-ms vowel: 41.8 ms for the 475/1800 series, 34.8 ms for the 510/1700 series, and 31.5 ms for the 550/1600 series. The comparable values for the continua with a 240-ms long vowel are 67.4, 49.7, and 45.3 ms. These values show that with longer vowels listeners need more VOffT to perceive preaspiration (Pind, 1995a, 1996a). These boundary values also reveal the effect of vowel quality on the location of the perceptual boundaries for preaspiration. As hypothesized, longer VOffTs are needed to cue preaspiration following a vowel which has the formant values appropriate for a long vowel (475/1800) than after a vowel which has the formant patterns appropriate for a short vowel (550/1600).

Because it is not possible to calculate phoneme boundaries for the responses of all participants in the conditions with the longer vowels, since the responses of individual participants did not in all cases cross the phoneme boundaries, results were analyzed in terms of the number of preaspirated responses. For the shorter (108 ms) vowels, these averaged 39, 47.7, and 51.9 (out of a total of 100) for the three $F1/F2$ series (in the order 475/1800 Hz, 510/1700 Hz, and 550/1600 Hz). For the longer vowel continua (240 ms) the average number of preaspirated responses were respectively 16, 31, and 36.6.

A two-way repeated measures ANOVA (vowel duration versus formant values) showed both main effects to be significant, duration, $F(1,12)=50.44$, $p<0.001$; formants, $F(2,24)=17.29$, $p<0.001$. The interaction was not significant, $F(2,24)=1.55$, $p>0.10$. The hypothesis for this experiment was that in both duration conditions we could expect to see the number of preaspirated responses increasing as the formant values approached those for a short [ɛ]. This would imply that the number of preaspiration responses (PR) should be ordered as followed, PR-550/1600 >PR-510/1700>PR-475/1600. Checking for this, using two pairwise comparisons for each vowel duration and employing the Bonferroni correction, reveals the following: For the stimuli with a 108-ms long vowel: PR-550/1600>PR-510/1700, $F(1,12)=2.77$, n.s.; PR-510/1700>PR-475/1600, $F(1,12)=8.89$, $p<0.05$. For the stimuli with a 240-ms long vowel: PR-550/1600>PR-510/1700, $F(1,12)=12.3$, $p<0.05$; PR-510/ 1700>PR-475/1600, $F(1,12)=16.64$, $p<0.01$. These results can therefore on the whole be taken to

support the hypothesis advanced, that the perception of preaspiration depends on the perceived quantity of the vowel. Since preaspiration in Icelandic can only follow a short vowel it is to be expected that lengthening of the vowel will call for longer VOffTs to cue preaspiration. This should hold whether the lengthening is accomplished through durational or spectral means as this experiment has verified.

V. GENERAL DISCUSSION

The impetus for the research reported in this paper was the earlier finding (Pind, 1995a, 1996a) that VOffT shows much greater sensitivity to vowel durations than does VOT. Previous research, mostly in English, had shown that the perception of VOT was susceptible to the influence of the duration of the following vowel. With increases in vowel duration following the word-initial VOT the phoneme boundaries separating voiced from voiceless (or nonaspirated from aspirated) stops moved to longer values of VOT. This has commonly been interpreted as showing the effect of rate normalization (Miller, 1981; Summerfield, 1981) on the assumption that a longer vowel, in English, reflects a slower speaking rate.

It had previously been shown that preaspiration, an [h]-like sound between a vowel and a following closure, can be cued by VOffT, a speech cue which is the mirror image of VOT (Pind, 1995a, 1996a). Interestingly, VOffT shows much greater effect of vowel duration than does VOT. The question is why this should be so. An intriguing explanation was suggested by Bladon's (1986) hypothesis that preaspiration is an ''auditorily disadvantaged'' speech cue considering its position in the syllable since following the much louder vowel it would easily be susceptible to neural short-term adaptation. This could explain the different effects of vowel lengthening on the perception of VOT and VOffT.

As mentioned earlier another hypothesis also suggests itself, namely that this different behavior of VOT and VOffT rather stems from the different location of these two speech cues in the syllable. Since VOT is word-initial it can be followed by either a phonemically long or short vowel. VOffT, being located in the syllable rhyme, is however more constrained in that it can only follow a phonemically short vowel. Lengthening the vowel in front of VOffT, while keeping the rhyme duration constant, would eventually change the vowel's quantity from short to long which would, irrespective of the perceived rate, lessen the likelihood of the listener perceiving preaspiration.

To distinguish between Bladon's auditory hypothesis and our own ''linguistic'' hypothesis a number of experiments were conducted. Experiment 1 showed that nonspeech analogs of VOT and VOffT, made of two sine-wave tones having different onset and offset asynchronies (TOT and TOffT), did not reveal any difference in perception. These nonspeech analogs were modeled after the speech stimuli of Pind (1996a), which had clearly revealed the different response of VOT and VOffT to vowel lengthening. Although the total length of the sine-waves did influence the perceptual boundaries in the expected direction, both for the TOT and TOffT continua, the effect was comparable in both cases and did not reveal the differences earlier found for the speech continua. This result argues clearly against the auditory hypothesis.

The other experiments of the paper dealt with the two ways in which it is possible to change the perception of quantity in Icelandic. The first, and obvious, way is to change the durational relationship of the vowel and the following consonant which is the primary cue to vowel quantity in Icelandic. Lengthening the vowel, while keeping the rhyme duration constant, will thus predictably influence the perception of VOffT and move the perceptual boundary for preaspiration to longer values of VOffT (Pind, 1995a, 1996a). Is this an auditory contrast effect or specifically an effect of the perceived quantity? Experiment 4 addressed this issue by using the other cue for quantity, vowel quality, to change the perceived quantity. It turns out that this has the *same effect* on the perception of preaspiration as increasing the duration of the vowel. Thus it does not matter whether the vowel quantity is changed by durational or spectral means, in both cases the effect on the perception of preaspiration is comparable. This finding follows quite naturally from the ''linguistic'' hypothesis that the major reason for the big shifts in the phoneme boundaries for VOffT as compared to VOT with changes in vowel duration can most naturally be explained in ''speech-specific'' terms, as reflecting the influence of the perceived quantity on the perception of preaspiration. This result also argues against our own earlier hypothesis (Pind, 1995a) that the perception of preaspiration is mediated through a relational invariant of VOffT to vowel duration. That hypothesis fails to take into account the effect of vowel spectrum on perceived vowel quantity now established.

The experiments of the present paper were specifically aimed at testing the hypothesis advanced by Bladon (1986) regarding the perception of preaspiration and to inquire whether a fundamental difference in ''audibility'' of VOT and VOffT could be used to explain the more notable influence of vowel duration on the perception of VOffT as compared to VOT. We have argued that the results of the present series of experiments do not accord with Bladon's hypothesis.

It is possible, however, that a different auditory hypothesis might be of some relevance to the present results. It has been demonstrated that the frequency of $F1$ at voicing onset has an effect on the location of the phoneme boundaries for VOT, a finding which has also been demonstrated in animal studies (Kluender, 1991; Kluender and Lotto, 1994; Summerfield and Haggard, 1977). The lower the $F1$ at voicing onset the longer the VOT needed to cue a voiceless (aspirated) stop. It is of course possible that this effect is also contributing to the effect seen here in experiment 4 since the long [ɛ] is associated with a lower $F1$ than the short [ɛ]. But note that from Fig. 4 it can be seen (since three response categories were used in the experiment) that the lowering of $F1$ never has the effect of just affecting the perception of VOffT as a cue for voicing: It simultaneously affects the perception of preaspiration and quantity (as reflected in the changing proportion of [sɛk/sɛkː] responses). If the effect of $F1$ were limited to affecting VOffT one would not nec-

essarily expect any changes in the perceived quantity. Obviously, though, this is an issue deserving of further study, and experiments comparing the effect of $F1$ frequency on the perception of VOT and VOffT in Icelandic are currently being undertaken.

ACKNOWLEDGMENTS

This research was supported by the Icelandic Science Foundation and by the Research Foundation of the University of Iceland. I am grateful to Thórunn H. Halldórsdóttir for help in running the experiments and for carrying out the measurements in experiment 2. I am grateful for helpful comments from Dr. Winifred Strange and an anonymous reviewer which have improved the paper.

Abramson, A. S., and Ren, N. (**1990**). ''Distinctive vowel length: duration vs. spectrum in Thai,'' J. Phon. **18**, 79–92.

Benediktsson, H. (**1963**). ''The nonuniqueness of phonemic solutions: Quantity and stress in Icelandic,'' Phonetica **10**, 133–153.

Bladon, A. (**1986**). ''Phonetics for hearers,'' in *Language for Hearers*, edited by G. McGregor (Pergamon, Oxford), pp. 1–24.

Darwin, C. J., and Carlyon, R. P. (**1995**). ''Auditory grouping,'' in *Hearing*, edited by B. C. J. Moore (Academic, San Diego), pp. 387–424.

Diehl, R. L., and Walsh, M. A. (**1989**). ''An auditory basis for the stimulus-length effect in the perception of stops and glides,'' J. Acoust. Soc. Am. **85**, 2154–2164.

Finney, D. J. (**1971**). *Probit Analysis* (Cambridge U.P., Cambridge).

Flanagan, J. L. (**1972**). *Speech Analysis Synthesis and Perception* (Springer-Verlag, Berlin).

Garnes, S. (**1976**). *Quantity in Icelandic: Production and Perception* (Helmut Buske Verlag, Hamburg).

Hadding-Koch, K., and Abramson, A. (**1964**). ''Duration versus spectrum in Swedish vowels: Some perceptual experiments,'' Studia Linguistica **18**, 94–107.

Hillenbrand, J. (**1984**). ''Perception of sine-wave analogs of voice onset time stimuli,'' J. Acoust. Soc. Am. **75**, 231–240.

Jamieson, D. G., Ramji, K., Kheirallah, I., and Neary, T. M. (**1992**). ''CSRE: A speech research environment,'' in *Proceedings of the Second International Conference on Spoken Language Processing*, edited by J. J. Ohala, T. M. Neary, B. L. Derwing, M. M. Hodge, and G. E. Wiebe (University of Alberta, Edmonton), pp. 1127–1130.

Kewley-Port, D., and Watson, C. S. (**1994**). ''Formant-frequency discrimination for isolated English vowels,'' J. Acoust. Soc. Am. **95**, 485–496.

Klatt, D. H. (**1980**). ''Software for a cascade/parallel formant synthesizer,'' J. Acoust. Soc. Am. **67**, 971–995.

Klatt, D. H., and Klatt, L. C. (**1990**). ''Analysis, synthesis, and perception of voice quality variations among female and male talkers,'' J. Acoust. Soc. Am. **87**, 820–857.

Kluender, K. R. (**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. (**1994**). ''Speech perception as a tractable problem in cognitive science,'' in *Handbook of Psycholinguistics*, edited by M. A. Gernsbacher (Academic, San Diego), pp. 173–217.

Kluender, K. R., and Lotto, A. J. (**1994**). ''Effects of first formant onset frequency on [-voice] judgements results from auditory processes not specific to humans,'' J. Acoust. Soc. Am. **95**, 1044–1052.

Lisker, L., and Abramson, A. S. (**1970**). ''The voicing dimension: Some experiments in comparative phonetics,'' in *Proceedings of the Sixth International Congress of Phonetic Sciences* (Academia, Prague), pp. 563–567.

Miller, J. L. (**1981**). ''Effects of speaking rates on segmental distinctions,'' in *Perspectives on the Study of Speech*, edited by P. D. Eimas and J. L. Miller (Erlbaum, Hillsdale, NJ), pp. 39–74.

Miller, J. L. (**1987**). ''Rate-dependent processing in speech perception,'' in *Progress in the Psychology of Language*, edited by Andrew W. Ellis (Erlbaum, London), Vol. III, pp. 119–157.

Miller, J. L., and Liberman, A. M. (**1979**). ''Some effects of later-occurring information on the perception of stop consonant and semivowel,'' Percept. Psychophys. **25**, 457–465.

Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing* (Academic, London), 4th ed.

Parker, E. M. (**1988**). ''Auditory constraints on the perception of voice-onset time: The influence of lower tone frequency on judgements of tone-onset simultaneity,'' J. Acoust. Soc. Am. **83**, 1597–1607.

Perkell, J. S., and Klatt, D. H. (**1986**). *Invariance and Variability in Speech Processes* (Erlbaum, Hillsdale, NJ).

Pétursson, M. (**1974**). ''Peut-on interpréter les données de la radiocinématographie en fonction du tube acoustique à section uniforme?'' Phonetica **29**, 22–79.

Pind, J. (**1986**). ''The perception of quantity in Icelandic,'' Phonetica **43**, 116–139.

Pind, J. (**1995a**). ''Constancy and normalization in the perception of Voice Offset Time as a cue for preaspiration,'' Acta Psychol. **89**, 53–81.

Pind, J. (**1995b**). ''Rate-dependent processing of VOT: auditory contrast or rate normalization,'' in *Proceedings of the XIIIth International Congress of Phonetic Sciences*, edited by K. Elenius and K. Branderud (Stockholm), Vol. 2, pp. 538–541.

Pind, J. (**1995c**). ''Speaking rate, VOT and quantity: The search for higher-order invariants for two Icelandic speech cues,'' Percept. Psychophys. **57**, 291–304.

Pind, J. (**1996a**). ''Rate-dependent perception of aspiration and preaspiration in Icelandic,'' Q. J. Exp. Psychol. **49A**, 745–764.

Pind, J. (**1996b**). ''Rate-dependent perception of quantity in released and unreleased syllables in Icelandic,'' Speech Commun. **19**, 295–306.

Pind, J. (**1996c**). ''Spectral factors in the perception of vowel quantity in Icelandic,'' Scand. J. Psychol. **37**, 121–131.

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.

Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (**1981**). ''Speech perception without traditional speech cues,'' Science **212**, 947–950.

Shoup, J. E., and Pfeifer, L. L. (**1976**). ''Acoustic characteristics of speech sounds,'' in *Contemporary Issues in Experimental Phonetics*, edited by N. J. Lass (Academic, New York), pp. 171–224.

Summerfield, Q. (**1981**). ''On 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.

Thráinsson, H. (**1978**). ''On the phonology of Icelandic preaspiration,'' Nord. J. Ling. **1**, 3–54.

Volaitis, L. E., and Miller, J. L. (**1992**). ''Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories,'' J. Acoust. Soc. Am. **92**, 723–735.

Mechanical impedance of a piano soundboard

N. Giordano
Department of Physics, 1396 Physics Building, Purdue University, West Lafayette, Indiana 47907-1396

(Received 4 August 1997; revised 29 December 1997; accepted 5 January 1998)

An experimental study of the mechanical impedance Z of a piano soundboard in the frequency range $\sim 50-10^4$ Hz is reported. The results differ significantly from the behavior reported by Wogram above a few kHz, but are consistent with the measurements of Conklin. The data presented here are also in good agreement with the predictions of our recent numerical calculations. Those calculations found that the soundboard ribs have an important effect on the frequency dependence of Z above a few kHz, and our measurements confirm that prediction. © *1998 Acoustical Society of America.* [S0001-4966(98)02704-0]

PACS numbers: 43.75.Mn [WJS]

INTRODUCTION AND BACKGROUND

The physics of the piano[1] continues to attract attention, as there remain many aspects of the instrument which are incompletely or poorly understood. A longstanding goal has been the development of a theoretical model, most probably a numerical/computational one, which can be used to calculate the sound produced by a piano, and to help provide an understanding of how different aspects of the instrument contribute to its performance. There now exist theoretical models which can account fairly well for the measured vibrational properties of the strings, and for the manner in which the hammer excites these vibrations.[2–6] However, the situation regarding the vibrational properties of the soundboard is much less satisfactory. Detailed measurements of the soundboard behavior at frequencies below a few hundred Hz are available,[7–10] and here the experiments seem to be well understood theoretically.[10] However, the results in the literature for the behavior in the range above a few hundred Hz (which is, of course, extremely important for the musical performance) are contradictory. The primary goal of the present work was to resolve these contradictions, and obtain reliable results for the mechanical impedance Z of the soundboard over a wide frequency range ($50-10^4$ Hz). Since we hope to ultimately construct a computational model of the piano, we also wanted to compare the results for Z with our recent theoretical calculations.[11] According to those calculations, the soundboard ribs have an important, and quite distinctive, effect on the frequency dependence of the impedance, and we will see that our measurements verify this prediction.

As we have just noted, previous studies of soundboard vibrations can be grouped into two categories. Work at low frequencies, typically at and below a few hundred Hz, has generally involved experimental determinations or theoretical calculations, or both, of the frequencies and eigenvectors of the lowest few normal modes. Here the measurements[7–10] are in good agreement with general arguments[1] and with detailed calculations,[10] so the situation at low frequencies is quite satisfactory. The behavior at higher frequencies, where the response is characterized by many overlapping modes, is usually discussed in terms of the mechanical impedance, which is the ratio of an applied force to the velocity which it produces. As far as we know, the first results for Z were those of Wogram[12,13] who conducted extensive measurements over a wide frequency range ($50-10^4$ Hz). He studied the behavior (i.e., applied the force) at many locations on the soundboard, and also investigated the effects of the plate and the ribs. He found that while the measuring location (i.e., the point where the force was applied and the velocity measured) had a significant effect at low frequencies, where the lowest few normal modes could be readily separated, the behavior above about 500 Hz was surprisingly (at least to the author) similar at all locations, as Z decreased by nearly a factor of 10 in going from a 1 to 10 kHz. Moreover, the ribs were found to have very little effect on Z, especially in the region above 1 kHz. The only other experimental results for Z are those of Conklin.[1,14] He reported that Z is approximately constant from low frequencies up to his highest frequency, which was a few kHz, in sharp contrast to the rapid decrease in Z found by Wogram, as just described.

On the theoretical side, we know of only one calculation of the soundboard impedance at high frequencies. It was carried out by the author,[11] and yielded the following features. First, the results were consistent with the measurements of Conklin, as Z was approximately independent of frequency from 100 Hz to a few kHz. Second, at higher frequencies Z was found to decrease significantly, by typically a factor of 3–5 at 10 kHz, but the decrease was not as large as found experimentally by Wogram. Third, it was possible to show that this decrease of Z with increasing frequency was caused by the ribs. A (hypothetical) soundboard without ribs should, according to the author's calculations, exhibit an impedance which is approximately independent of frequency. The ribs cause Z to decrease significantly at high frequencies, a result which can be readily understood intuitively (as will be discussed further below), but which is in sharp contrast to the experiments of Wogram.

From this brief review of previous soundboard studies, it is clear that there are a number of important discrepancies among the experimental and theoretical results in the literature. The primary goal of the present experiments was to resolve these discrepancies, so that an accurate, and experimentally verified, soundboard model can be developed. As will be seen in the following sections, we believe that this

goal has been realized. Our new results for Z are in good agreement with those of Conklin (although ours extend to somewhat higher frequencies), and with the calculations of Ref. 11.

I. EXPERIMENTAL SETUP

In this paper we will be concerned with the mechanical impedance which is defined by

$$Z=\frac{F}{v}, \tag{1}$$

where F is the applied force and v is the velocity of the soundboard.[15] It is simplest to imagine that F is a sinusoidal function of time at a frequency f; the impedance is thus characterized by a magnitude and a phase, which are both functions of frequency. Throughout this paper we will assume that v is the velocity at the point where F is applied, so that Z corresponds to the driving point impedance.[15] This impedance can, of course, depend on the location of the driving point, and in our experiments we will compare the behavior of Z at several different locations. The dependence on position will turn out to provide very useful information.

In order to measure Z we applied a force to the soundboard of a fully assembled upright piano[16] using a mechanical shaker,[17] which was essentially just a small speaker. The force was transmitted through a piezoelectric force sensor[18] which was attached (as described below) to the soundboard, and the resulting motion was measured with a piezoelectric accelerometer[19] which was mounted (also to be described in more detail below) very close to the force sensor, typically within ~ 1 cm. To obtain results as a function of frequency, the oscillator which powered the shaker was swept continuously, and the rectified outputs of the force sensor and accelerometer were recorded by a computer, along with the frequency. The relative phase of the force and accelerometer signals was measured with a simple phase comparison circuit[20] and also recorded. Note that the force was always applied perpendicular to the plane of the soundboard, and the acceleration was measured in the same direction.

A crucial aspect of the experiment is the method by which the force sensor and the accelerometer are attached to the soundboard. Wogram's measurements employed a device, called an impedance head, in which the force sensor and the accelerometer are incorporated into one unit, which is then connected to the soundboard. Any such mechanical connection is essentially a ''spring,'' and the impedance head and this mounting ''spring'' will have a resonant frequency. It is crucial that this resonant frequency be well above the range of interest, as otherwise it can dominate the measured response. It has been suggested to the author[21] that such a problem may have affected the results of Wogram at high frequencies, and we will discuss this possibility further after we have presented our results. We mention these points here since they were an important factor in our experimental design, and led us to employ separate force and acceleration sensors, rather than an impedance head. This made it possible to compare the results with different mounting schemes for both sensors, and to show that our results were independent of the mounting scheme.

After some trial and error, we found that the following mounting schemes were satisfactory up to nearly 10 kHz. (1) The most convenient approach was to use a threaded metal block (Al and stainless steel were both used with success) which was glued to either the bridge or the soundboard with a thin layer of superglue (varnish was also suitable, but took longer to dry). Different blocks were tried (with thicknesses of either 3.2 or 6.3 cm), and the sensor was screwed directly into a threaded hole in the block. Note that separate blocks were used for the force and acceleration sensors. (2) A second approach was to hold the sensor in place with a thin layer of wax. These two mounting methods gave very similar results up to typically 9 kHz. At higher frequencies they sometimes differed; a warning sign of problems (i.e., erroneous results) was generally a rapid increase in the measured impedance. We believe that this occurred due to a mechanical ''decoupling'' of the accelerometer from the soundboard. The measured acceleration was then smaller than the true value, yielding an erroneously large value of Z. We have belabored these details since we believe that they are crucial for obtaining a correct result, and that they may be helpful for others conducting similar experiments.

II. RESULTS

Measurements of the impedance were performed at several different locations. Figure 1(a) shows these locations, along with the positions of the ribs and the bridges. Note that this is a view from the back of the piano; the bridges were, of course, on the opposite side. These experiments were performed with a fully assembled piano; the plate and strings were all in place. This greatly limited access to the treble bridge, so for measurements at points appropriate for most notes it was much more convenient to attach sensors to the ''back'' of the soundboard, on the side opposite the bridges. We therefore first checked to see that the impedance measured at corresponding points on opposite sides is approximately the same (as one might expect). A convenient location for this test was at the bridge end of string G3 (key 35), since this bridge point was easily accessible. We compared the impedance measured with sensors on the bridge with the results when the same sensors were attached to the back of the soundboard, and the two results were very similar. The behavior found with sensors mounted on the back of the soundboard is shown in Fig. 2 (all of the rest of the results shown in this paper were also measured on the side of the soundboard opposite the bridges). Here we plot $|Z|$; results for the phase will be presented below. Since $Z=F/v$, a soundboard resonance gives rise to a minimum in Z. The fundamental mode for this soundboard is at ~ 109 Hz while the second and third modes are at ~ 189 and 250 Hz. These values are typical.[7,9,10,12,1,11] The overall, or ''average,'' value of $|Z|$ at low frequencies (below 1 kHz) also agrees well with the levels in the range 1–5 kg/s reported by many previous workers. Another noteworthy feature is the decrease in $|Z|$ which starts around 2–4 kHz. We will see similar behavior at other measuring points, and it turns out that this

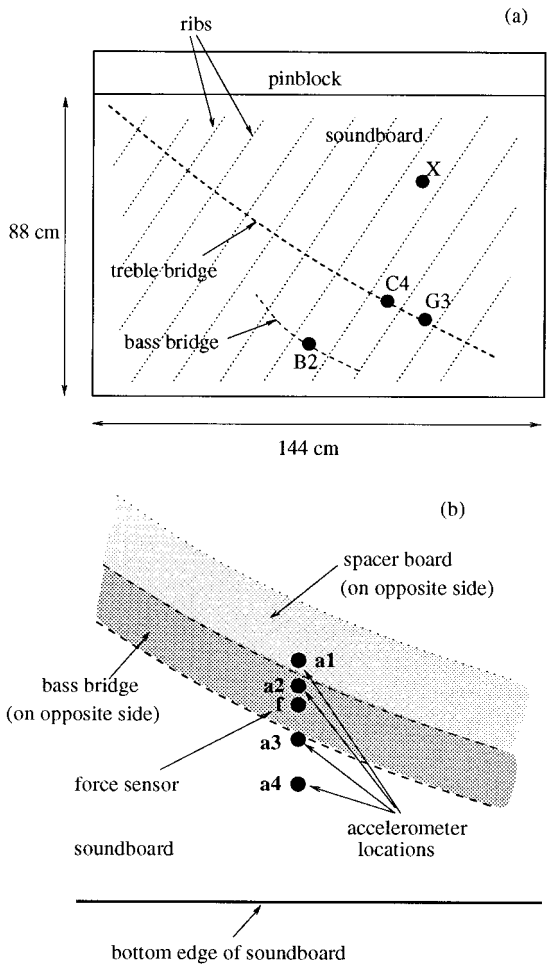


FIG. 1. (a) Diagram of the soundboard of our upright piano, approximately to scale, as seen from the back of the piano. The filled circles show the locations where the impedance results were measured, and the dotted lines show the positions of the ribs. The positions of the bridges, which are on the opposite side of the board, are also shown. (b) Expanded view of the region near location **B2** in (a), again as seen from the back of the piano. The force sensor was located at **f**, and the accelerometer was mounted at several different locations, including **a1**, **a2**, **a3**, and **a4**. The location of the bass bridge, which was on the opposite side of the piano, is indicated. The bass bridge was elevated off the surface of the soundboard by a spacer board. The location of this spacer is also shown; note that the spacer board did not extend to the lower edge of the bass bridge, but the two overlapped only slightly. As a result, the main body of the bass bridge was suspended away from the soundboard. Hence in the lightly shaded region the soundboard was in contact with the spacer, while in the heavy shaded region the thickness was simply that of the soundboard.

was also observed in our recent numerical calculation of Z. It is therefore useful to discuss its proposed origin; this will also aid in understanding other results which will be presented below.

For a board of uniform thickness, exact analytic calculations suggest,[22] and our numerical results confirm,[11] that the background level of $|Z|$ is a constant, independent of frequency. However, a real soundboard does not have a uniform thickness, since ribs and bridges are attached, and the thickness of the board itself also varies somewhat with position. Our calculations[11] showed that ribs (and presumably also the bridges[23]) have a profound effect on Z, as can be understood from the following argument. Stiffening the soundboard, which can be accomplished by increasing its

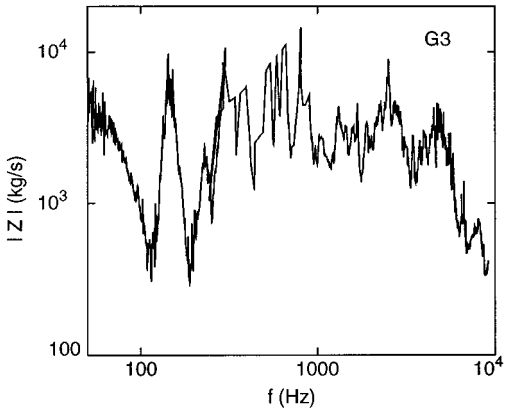


FIG. 2. Impedance $|Z|$ measured at location **G3** [see Fig. 1(a)]. This is the location where note G3 drives the soundboard. For this measurement, both of the sensors were screw mounted.

Young's modulus, making it thicker, or adding ribs, will increase its impedance, since a stiffer board will exhibit a smaller velocity for a given applied force. However, the ribs make the board appear stiffer mainly for wavelengths comparable to or greater than the rib spacing. Modes with shorter wavelengths can ''fit between'' the ribs, and therefore experience a softer board than do the long wavelength vibrations. For this reason, we expect $|Z|$ to decrease when the frequency is high enough that the wavelength is comparable to the rib spacing. This general argument explains the behavior found numerically for boards with and without ribs,[11] and is consistent with the results seen in Fig. 2. We will see below that this argument accounts well for several other features of our experimental results.

The behavior in Fig. 2 can also be compared with that found by Wogram at his measuring point 11. He too found the fundamental mode to be close to 100 Hz, with an average low-frequency impedance level near 1000 kg/s. However, he also found that starting at about 1 kHz, $|Z|$ decreased monotonically to a value near 150 kg/s at 10^4 Hz. This is in contrast to our result that $|Z|$ remains approximately constant up to about 3 kHz. Note that Conklin also found $|Z|$ to be approximately constant in this range, in agreement with our results (although he did not appear to report the absolute magnitude of the impedance). These differences suggest to us that Wogram's results are in error above 1 kHz. The behavior he found could have been caused by an effective decoupling of his impedance head from the soundboard at high frequencies. We have already mentioned how decoupling affected our own measurements, and shown that this was only a problem for us at frequencies close to 10 kHz. In our case decoupling caused the accelerometer motion to be smaller than that of the board, so the apparent v was smaller and the derived Z larger than the true value. In Wogram's experiment it seems likely that the entire impedance head became decoupled.[21] In that case the measured v would be larger and the derived Z smaller than if the impedance head were still well coupled to the board. This decoupling would also mean that Wogram's measured impedance would be insensitive to the properties of the soundboard, and this would explain why he also did not observe any significant fluctuations in $|Z|$ above about 1 kHz. In contrast, we observed such fluctua-

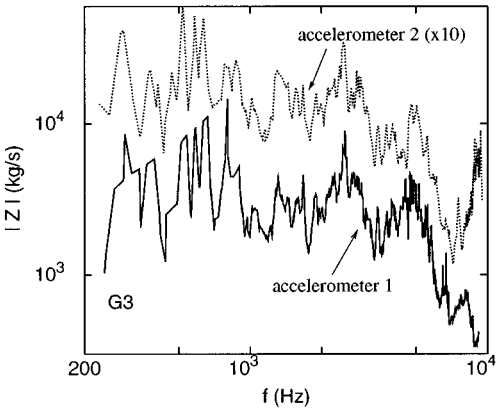


FIG. 3. Impedance $|Z|$ measured at location **G3** [see Fig. 1(a)], using two different accelerometers and mounting schemes. Accelerometer No. 1 was attached with the screw arrangement while accelerometer No. 2 was attached with wax; the force sensor was screw mounted. The result for accelerometer No. 2 has been displaced upwards by a factor of 10 for clarity.

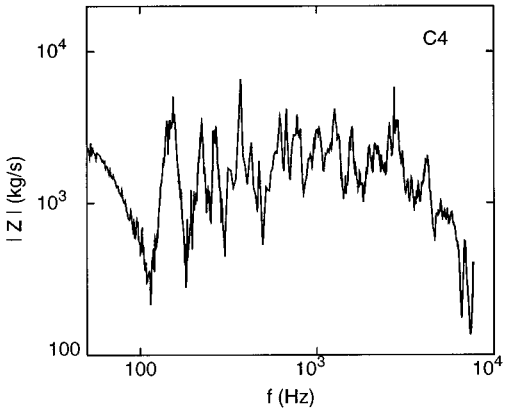


FIG. 4. Impedance $|Z|$ measured at location **C4** [see Fig. 1(a)]. This is the location where note C4 drives the soundboard. For this measurement the force sensor was screw mounted, while the accelerometer was mounted with wax.

tions over our entire measurement range; such fluctuations were also observed in the measurements of Suzuki and Conklin.

We have emphasized the importance of checking that the sensor to soundboard coupling was not influencing the results. One way to do this is to compare results obtained with different accelerometers, and with different accelerometer mounting schemes. This is done in Fig. 3, where we show measurements carried out at the same location but with different accelerometers. Accelerometer No. 1 was screw mounted to a metal block glued to the soundboard, while accelerometer No. 2 (which was also much less massive than No. 1) was attached with wax. The two results agree extremely well up to about 8 kHz. At higher frequencies the apparent impedance measured with the wax mounted accelerometer begins to increase, while this is not seen with the screw mounted sensor. We attribute this to a decoupling of the wax mounted sensor from the board; as we have just noted, this would lead to an erroneously large apparent value of $|Z|$. We show this effect here, as it is important in understanding the measurements and their range of validity. In the results which follow we will show only data for which we have obtained the same result with both screw and wax mounted sensors. In these cases we have confidence that the accelerometers were always well coupled to the board. We have also considered the coupling between the force sensor and the board. Several measurements (not shown here) with the wax coupling scheme demonstrated that the screw mounting approach was reliable up to our highest frequencies.

Figure 4 shows results at location **C4**; this is opposite the point where string C4 (note 40) contacts the bridge [see Fig. 1(a)]. The behavior is, as expected, quite similar to that found at **G3** (the two locations are separated by only 15 cm). The lowest few modes are at the same frequencies (as expected), and the overall magnitude of $|Z|$ is about the same. Here the high-frequency decrease in $|Z|$ begins at about 2 kHz, which is a little sooner than seen at **G3**. This relatively small difference is probably caused by the fact that location **C4** is spaced slightly farther from the ribs than **G3** [see Fig. 1(a)]. This will, according to the argument given above, mean that at location **G3** it is necessary to go to slightly higher frequencies to strongly excite modes which take advantage of the effectively softer region between the ribs.

Figure 5 shows results for location **B2**, which corresponds to a note whose string terminates at the bass bridge. Here the force was applied to the soundboard at the bridge location (but on the opposite side of the board), and the acceleration was measured at several nearby positions. Figure 1(b) shows in detail these positions relative to the bridge, and to the spacer board which laid between the bridge and the soundboard. On our soundboard, the top edge of the bass bridge was mounted on the spacer board. However, over most of its width the bass bridge did not contact the soundboard, but was suspended in cantilever fashion away from it. Hence in the lightly shaded region of Fig. 1(b) the spacer board was in direct contact with the soundboard, while in the heavily shaded region the effective thickness was that of only the soundboard. The results labeled a1, a2, and a3 in Fig. 5 were obtained with accelerometers at the corresponding locations in Fig. 1(b). The behavior with the accelerom-

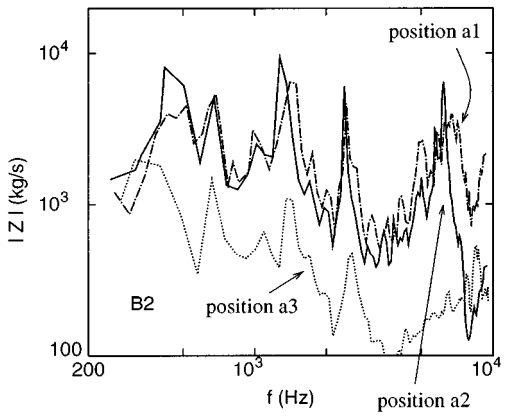


FIG. 5. Impedance $|Z|$ measured near location **B2** [see Fig. 1(a)]. The three results were obtained with accelerometers mounted at positions **a1**, **a2**, and **a3**, as shown in Fig. 1(b). In all three cases the force sensor was screw mounted at **f**. The accelerometer was screw mounted for the measurement at location **a2**, and mounted with wax at **a1** and **a3**. The result at location **a4** was similar to that found at **a3**.

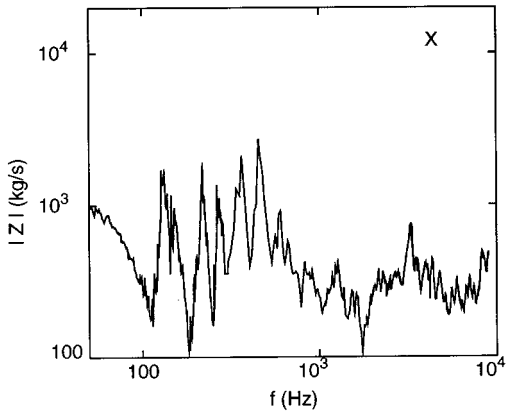


FIG. 6. Impedance $|Z|$ measured at location **X** [see Fig. 1(a)]. This location is far away from the bridges. For this measurement the force sensor was screw mounted, while the accelerometer was mounted with wax.

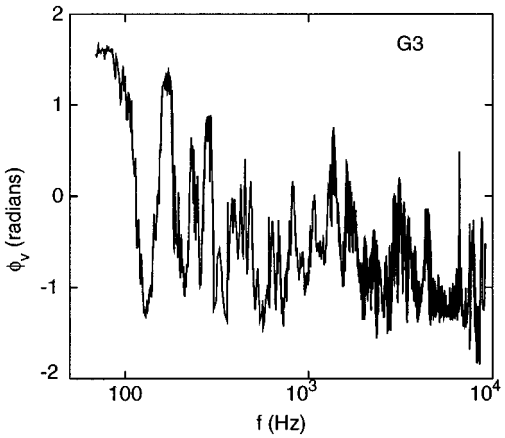


FIG. 7. Relative phase of the velocity and the force, ϕ_v, as a function of frequency at location **G3**.

eter at **a3** is qualitatively similar to that seen earlier for locations **G3** and **C4**; the impedance is of order 1000 kg/s at low frequencies (here below about 1 kHz) and decreases at higher frequencies. The same general behavior is observed at accelerometer location **a2**, with the important difference that the decrease in $|Z|$ does not begin until a much higher frequency, approximately 7 kHz. For location **a1** there seems to be no decrease in the impedance up to our highest frequency. This can all be understood qualitatively with the ''frequency dependent stiffness'' argument given above. As shown in Fig. 1(b) and noted above, the board is thickest at location **a1**, since here the spacer is in direct contact with the soundboard. This greatly increases the bending stiffness (this stiffness varies as the cube of the thickness). As a result, the high-frequency modes which fit between the ribs will tend to have nodes at this location and hence not contribute to Z, and this keeps the impedance large over our entire frequency range. Location **a2** is just off the edge of this thicker region, so the stiffness is not quite as large. This allows $|Z|$ to decrease, at least somewhat, at our highest frequencies, although this decrease occurs at a higher frequency than observed at **G3** and **C4**. Measuring point **a1** is farthest from the thicker portion, and so the stiffness is smallest there. Correspondingly, we observe that the impedance is smaller over much of the frequency range shown, and the decrease in $|Z|$ occurs at a much lower frequency.

In addition to supporting the notion that the ribs greatly influence the frequency dependence of Z, the results in Fig. 5 also demonstrate conclusively that the decrease in $|Z|$ we have observed at high frequencies is not due to problems with the instrumentation. This is an important observation, given the differences between our results and those of Wogram (and also the comments of Conklin[21]).

The importance of the stiffness added by the ribs and bridges can also be seen from the results in Fig. 6, which shows the impedance measured at location **X** in Fig. 1(a). This location is far from the bridges and approximately equidistant from the nearest ribs. According to our arguments, the board should be effectively softer, and hence $|Z|$ lower, than found at the locations considered previously. This is precisely the behavior observed in Fig. 6. The usual low-frequency modes are clearly seen, but the overall level of the impedance at low frequencies is a factor of 3 or more smaller than at locations **G3**, **C4**, and **B2**. This level decreases somewhat about 600 Hz, which is much earlier (in terms of frequency) than found at the other locations, and this is again in good accord with our model.

The impedance (1) is, of course, a complex quantity. So far we have discussed only the magnitude of Z, but it is worthwhile to consider also its phase. Figure 7 shows the phase of the velocity for measurements at location **G3**. For this plot we have assumed that the force and velocity are given, respectively, by $f = f_0 \sin(\omega t)$ and

$$v = v_0 \sin(\omega t + \phi_v), \tag{2}$$

so that ϕ_v is the phase angle by which the velocity differs from the force. The corresponding phase angle for Z is then $-\phi_v$. Here we plot ϕ_v since its behavior can in certain limits be readily understood in analogy with a harmonic oscillator.[15] For example, at the lowest frequencies, below the fundamental (f_0), we observe that $\phi_v \sim \pi/2$, as expected for an oscillator in the ''stiffness'' dominated regime.[15] As one passes through resonance at the fundamental, ϕ_v decreases and passes through zero on its way to a value near $-\pi/2$ at frequencies somewhat above f_0. This pattern is repeated for the lowest few modes, but as the frequency reaches a few hundred Hz the mode overlap obscures this behavior. A noteworthy point here is that at high frequencies the ''background'' value of ϕ_v is approximately constant with a value near ~ -1 rad. This is quite different from the behavior reported by Wogram,[12] who found that ϕ_v was very close to $\pi/2$ above about 1 kHz. The behavior of ϕ_v is important since it determines, in combination with $|Z|$, the effectiveness with which energy is coupled from the strings to the soundboard (the energy transmitted should vary as $\sim \cos \phi_v$). Wogram's value of $\phi_v \sim \pi/2$ implies that this coupling is very small, i.e., that very little of the string energy is converted to sound above ~ 1 kHz, which does not seem (to the author) like a reasonable result. The measurements in Fig. 7 imply that the energy transfer between the string and soundboard remains substantial up to our highest frequencies. We also note that the results of our calculation[11] are in good agreement with the measurements in Fig. 7.[24]

III. SUMMARY AND OUTLOOK

We have presented new results for the mechanical impedance of the soundboard of an upright piano. Our experiments resolve several discrepancies and puzzles that had existed in the literature; in particular, the extensive data of Wogram appear to be seriously in error above about 1 kHz.[25] Our results fully support the results of Conklin, although we find that there are significant and instructive variations of Z at frequencies above the range he studied. These variations can be readily understood in terms of the model calculation reported in Ref. 11. A key to this understanding is to realize that the ribs make the soundboard stiffness and hence the impedance depend rather strongly on frequency, as vibrational modes with wavelengths smaller than the rib spacing experience an effectively softer board than seen by the longer wavelength modes. Similar effects make Z dependent on the measuring location.

The model considered in Ref. 11 was quite simplified in order to test the method and help reveal the essential physics. Now that the present experiments confirm the basic correctness of that approach, it should be possible to refine the model and make a more detailed, quantitative comparison with the experiments. We plan to extend the calculation to treat the specific soundboard geometry of our piano, including the actual ribs and bridges. We hope that the resulting model will then be useful for computational studies of string-soundboard motion.

ACKNOWLEDGMENTS

We are grateful to P. F. Muzikar for useful comments, especially concerning the manuscript, to H. A. Conklin, Jr., and G. Weinreich for very instructive discussions regarding the experimental situation, and to H. W. Jackson for providing a copy of Ref. 10. This work was supported in part by the NSF through Grant No. PHY-9722031.

[1] For a very complete review see H. A. Conklin, Jr., ''Design and tone in the mechanoacoustic piano. Part I. Piano hammers and tonal effects,'' J. Acoust. Soc. Am. **99**, 3286 (1996); ''Design and tone in the mechanoacoustic piano. Part II. Piano structure,'' *ibid.* **100**, 695 (1996); ''Design and tone in the mechanoacoustic piano. Part III. Piano strings and scale design,'' *ibid.* **100**, 1286 (1996).

[2] A. Chaigne, ''On the use of finite differences for musical synthesis. Application to plucked string instruments,'' J. Acoust. **5**, 181 (1992).

[3] A. Chaigne and A. Askenfelt, ''Numerical simulations of piano strings. I. Physical model for a struck string using finite difference methods,'' J. Acoust. Soc. Am. **95**, 1112 (1994).

[4] A. Chaigne and A. Askenfelt, ''Numerical simulations of piano strings. II. Comparisons with measurements and systematic exploration of some hammer-string parameters,'' J. Acoust. Soc. Am. **95**, 1631 (1994).

[5] X. Boutillon, ''Model for piano hammers: Experimental determination and digital simulation,'' J. Acoust. Soc. Am. **83**, 746 (1988).

[6] D. E. Hall, ''Piano string excitation. VI: Nonlinear modeling,'' J. Acoust. Soc. Am. **92**, 95 (1992). This one of a series of papers on piano hammers.

[7] I. Nakamura, ''The vibrational character of the piano soundboard,'' Proc. 11th ICA, Paris (1983), Vol. 4, p. 385.

[8] H. Suzuki, ''Vibration and sound radiation of a piano soundboard,'' J. Acoust. Soc. Am. **80**, 1573 (1986).

[9] J. Kindel and I-C. Wang, ''Modal analysis and finite element analysis of a piano soundboard,'' in Proceedings of the 5th International Modal Analysis Conference (Union College, Schenectady, New York, 1987), p. 1545.

[10] J. Kindel, ''Modal analysis and finite element analysis of a piano soundboard,'' Masters thesis, University of Cincinnati, 1989.

[11] N. Giordano, ''Simple model of a piano soundboard,'' J. Acoust. Soc. Am. **102**, 1159 (1997).

[12] K. Wogram, ''Acoustical research on pianos. Part I: Vibrational characteristics of the soundboard,'' Das Musikinstrument **24**, 694 (1980); **24**, 776 (1980); **24**, 872 (1980).

[13] K. Wogram, in *The Acoustics of the Piano*, edited by A. Askenfelt, Royal Swedish Academy of Music Publication No. 64 (Stockholm, 1990), p. 83.

[14] We should also note that Suzuki (Ref. 8) has reported some results concerning the mechanical response up to a few kHz. While his measurements do not yield values for the mechanical impedance, they imply that this impedance is approximately constant up to his highest frequencies. Suzuki's results are thus consistent with the findings of Conklin (Ref. 1).

[15] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments* (Springer-Verlag, New York, 1991).

[16] The piano was an upright made by the Charles Fredrich Stein Company, and was approximately 60 years old. It appeared to have relatively new dampers, but the hammers and strings were probably original. For all of our measurements the dampers were always in contact with the strings.

[17] The shaker was a model V102 from Ling Dynamical Systems.

[18] The force sensor was a model 209C01 from PCB Piezotronics.

[19] Two different accelerometers, both obtained from PCB Piezotronics, were used. One (model 352B68) had a mass of 2.0 g and had a base which was designed to be screwed into the object to be studied, while the other (model 352A10) had a mass of 0.7 g and was attached with a small amount of wax.

[20] *The Art of Electronics*, edited by P. Horowitz and W. Hill (Cambridge U.P., Cambridge, England, 1989), 2nd ed.

[21] Details of this concern, and a very instructive discussion of the difficulties that can be encountered in such measurements, were communicated to the author by H. A. Conklin, Jr. Similar ideas were also conveyed to us by G. Weinreich.

[22] L. Cremer and M. Heckl, *Structure-Borne Sound* (Springer-Verlag, New York, 1973).

[23] The model calculation in Ref. 11 considered a soundboard with ribs, but for simplicity ignored the effect of the bridges. However, the presence of bridges does not change any of our qualitative arguments.

[24] Two comments need to be made concerning the phase results in Ref. 11. First, in that paper we plotted the phase of Z, which is $-\phi_v$ in the notation of the present paper. Second, the results for the phase reported in Ref. 11 have a minor error. In the notation of the present paper, negative values of ϕ_v were in Ref. 11 erroneously given a positive sign. Correcting this error yields results in good agreement with the measurements in Fig. 7, especially for the average value of ϕ_v at high frequencies (above $\sim$1 kHz).

[25] Our measurements also indicate that Wogram's results for the p/v, the sound intensity normalized by the soundboard velocity, underestimate the sound production at high frequencies. We will present our results for the sound production elsewhere.

Acoustical and physical dynamics of the diatonic harmonica

Henry T. Bahnson and James F. Antaki
Department of Surgery, University of Pittsburgh, School of Medicine, 3550 Terrace Street, Pittsburgh, Pennsylvania 15261

Quinter C. Beery
Department of Otolaryngology, University of Pittsburgh, School of Medicine, 3550 Terrace Street, Pittsburgh, Pennsylvania 15261

(Received 18 February 1997; revised 24 December 1997; accepted 5 January 1998)

The harmonica is arguably the most widely played instrument in the world, yet there is a surprising paucity of published studies of its acoustics or physical dynamics. The typical diatonic harmonica and the physical forces involved in its natural function are described, and simple observations of the harp's functions are reported. The speaking of the reeds, naturally, when producing a bend, and when speaking as an overblow or overdraw is discussed and investigated by simple stopping of the reeds, by videostroboscopic analysis, and by recording vibration of the reeds with displacement gauges. The reeds of the ten hole harmonica can be made to vibrate at varying frequencies depending on the size and shape of the player's vocal tract. Three different modes of speaking from each hole and its pair of reeds are revealed and studied: first, naturally in a closing mode, either blown or drawn; second, as a bend, either blown or drawn, with pitch in the interval between the two notes in the hole; and third, as an overblow or overdraw in an opening mode with a pitch outside the interval between the two natural notes of the hole. This dynamic interaction allows the player to speak with the instrument perhaps as with no other. [S0001-4966(98)02404-7]

PACS numbers: 43.75.Pq [WJS]

INTRODUCTION

The name, *Harmonica*, has been used to designate many instruments over the years, some being reed instruments related to the instrument under discussion, and others seemingly unrelated, such as the glass harmonica attributed to Benjamin Franklin, which consisted of a set of glasses of varying size, tuned by varying the amount of water within. The modern harmonica—a.k.a. harp, mouth organ, French harp, Mundharmonica (in Germany), pocket piano, and Mississippi saxophone—is a descendant of instruments first reported and patented in the 1820's.[1] The principle embodied in the harmonica can be traced to the Orient where variations of the Chinese sheng existed for several millennia. All of these instruments use a vibrating free reed.

Structure and function of the harmonica

The active element of the harmonica is a metal reed with one end fastened to the surface of a thin metal plate; the other end is free to vibrate. All but the fastened end of the reed overlies a slot in the reed plate that is just large enough to allow the reed to vibrate freely in the slot [Fig. 1(a), (b)]. The reed is activated by the flow of air across it. It vibrates at a frequency which is near its natural (i.e., plucked) frequency. This frequency is determined by the mass and stiffness of the reed in cooperation with its associated acoustical system. The throttling action of the reed on the flow of air causes a periodic fluctuation of the otherwise steady air pressure provided through the mouth, thus creating sound.

The body of the harmonica is a comb, made of wood, plastic, or metal, with prongs separating a row of ten or more chambers. One plate of reeds, the *blow reeds*, which are activated by positive pressure in the comb (i.e., blowing), is fastened to the top side of the comb with the reeds inside. A second plate of reeds, the *draw reeds*, which are activated by negative pressure, is fastened to the bottom side of the comb with the reeds outside. The reed slots on the blow- and draw-reed plates are arranged to coincide with each other on opposite sides of the chambers of the comb. Thin metal covers over the reed plates provide small resonating chambers which further modify the generated sound. The reeds can be arranged in various sequences of tones that encompass several octaves. Reeds of the popular ten hole, diatonic harmonica, tuned in the Richter model in a major key, are arranged as shown in Fig. 2. With this arrangement, one can produce the major chord of the key of the harp by blowing on any three adjacent holes. Drawing on the first four holes produces the dominant chord of the key. In order to accomplish this harmony, certain tones are omitted in the scale of the lowest and highest octaves, as can be seen in Fig. 2. The 10-hole diatonic harmonica has a range encompassing 3 full octaves, 22 diatonic tones, but since there are only 20 reeds in the 10 holes and 1 tone is repeated, there are 3 ''missing'' natural tones in the 3 octaves of the diatonic scale.

The arrangement of reeds in the typical diatonic harmonica is such that each of the three octaves is played differently. For the first six holes, comprising most of two octaves, the draw note is higher in pitch than the corresponding blow note. For these six holes, it is the draw note that can be *bent* down or flatted through air pressure and manipulation of the vocal tract by the player. For the remaining four holes, the blow reed is higher in pitch; here it is the blow note that can be bent. In each case, it is the higher note of the hole that

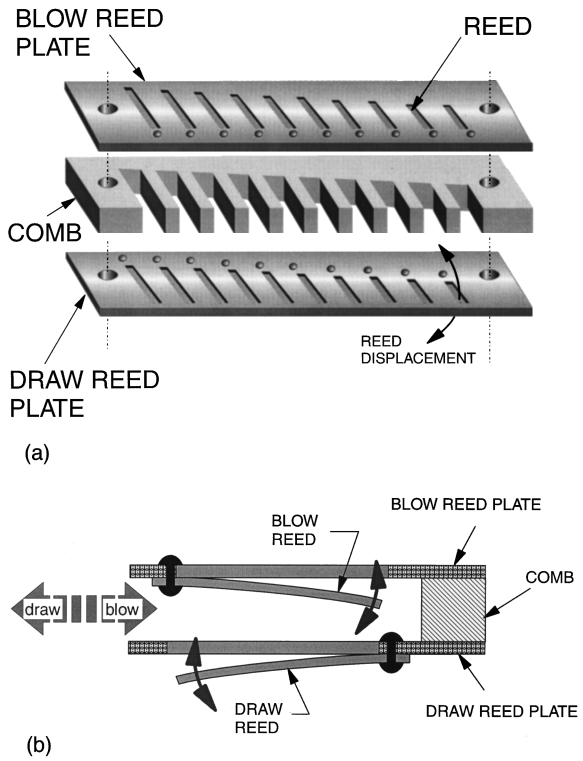


FIG. 1. (a) Exploded view of the ten hole diatonic harmonica showing the upper, blow reed plate, the lower, draw reed plate and the separating comb. Upper and lower covers of the reed plates are not shown. Reeds are mounted on the inside surface of the blow reed plate and the outside of the draw reed plate. Thus blowing closes the blow reed into the blow reed plate and opens the draw reed out of the draw reed plate. The reverse occurs on drawing. It is the closing reed that speaks on natural playing. (b) Schematic, cross sectional view of reeds and plates of a hole of a diatonic harmonica. Reeds vibrate in slots cut out of the reed plate. Blowing into the hole closes the blow reed and opens the draw reed. Drawing closes the draw reed and opens the blow reed.

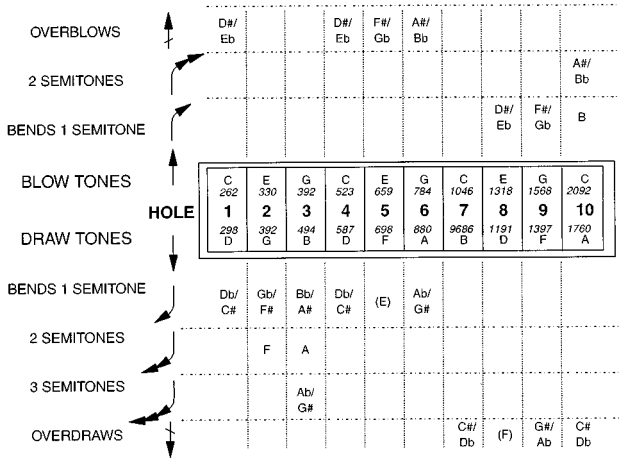


FIG. 2. Notes and approximate frequencies of a ten hole, diatonic harmonica, key of C. The upward arrow indicates blow; the downward arrow, draw. Blow and draw bends are indicated by bent shafts, up and down, respectively; number of semitones bent is indicated by number of arrow heads. Overblow and overdraw are indicated by slashed arrows.

appears to be bendable, whether blown or drawn. The often stated rule is that one can bend the pitch of the higher note in a hole down to a semitone above the lower note of the respective hole. It should be pointed out that bending produces virtually continuous pitch control within the interval between the blow and draw notes of the hole, in contrast to overblowing, to be discussed later.

In order to play the basic, natural tones of the key of the harp, no special configuration of the vocal tract is typically required. The player simply blows or draws through the appropriate hole. Vibration of the reeds occurs as a result of the interaction of aerodynamic forces acting on the reed and the mechanical properties of the reed itself. It is well known that a plucked reed will oscillate at a natural frequency determined by its mass distribution and stiffness. In order to cause self-sustaining oscillations, air must flow past the reed in a manner which reinforces this vibration. St. Hilaire *et al.*[2] have shown that such pressure instabilities could result when air flow is sufficient to cause boundary layer separation at the edge of the reed surface. When this occurs, the aerodynamic drag force on the reed increases as the gap decreases, and the reed moves against the air flow. Conversely, the drag force decreases when the gap increases, and the reed moves with the flow. This will thus inject energy into the reed motion, giving it a ''kick.'' This phenomenon can also be described in terms of acoustical impedance, as outlined by Johnston.[3] The resulting frequency of oscillation will differ from the plucked frequency because of the coupling with the acoustical inertance and compliance of the air flow and airway of the player.

A reed vibrating in this fashion with the reed attached to the high pressure side of the reed plate and being forced primarily *into* the reed plate is said to function as a *closing reed*. Simultaneously the companion reed, overlying the same cavity but attached to the outer, low pressure side of its reed plate, is blown open. It vibrates only minimally when the natural note is played and functions as an *opening reed*. Consequently, the frequency of vibration of the opening reed is determined primarily by the closing reed. When the player blows through a hole, the blow reed is a closing reed, and the draw reed is an opening reed. When the player draws on the hole, these actions are reversed. Alternatively, one can define the blow reed as the reed which speaks primarily when the player blows through a given hole. The reverse occurs on drawing. The primary speaking reed is the closing reed in each case of simple blow and draw. (It is worth pointing out that despite the names, *opening* and *closing*, both reeds do in fact open and close their respective reed slots in the course of their vibration. Accordingly, this nomenclature convention unfortunately introduces some confusion. However, inasmuch as this designation is well established, we will adopt this terminology to designate the mode of operation of a particular reed.)

The discovery that missing notes of the diatonic scale and some notes of the chromatic scale can be fashioned by bending is attributed to the African-Americans of the southern United States in the last half of the nineteenth century. Modern instructions on this technique have been described as: ''...draw in sharply, drop the jaw a bit, along with the

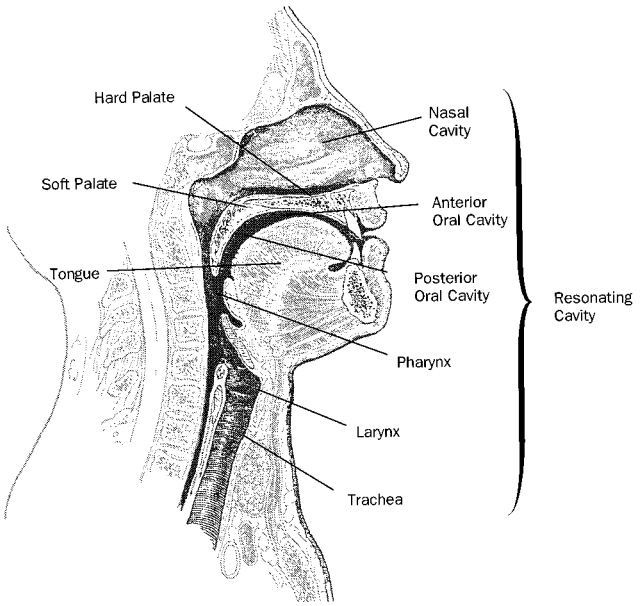


FIG. 3. Cross section of the vocal tract.

tongue, causing the air to dart downward to the floor of the mouth;''[4] ''...suck the air through the harp towards the tip of [the] tongue;''[5] ''[inhale] with more force and [cause] the air stream to move to the bottom of the mouth and throat;''[6] or, most descriptively, ''[take] in air from your shoe.''[7] Glover[8] instructs that if one visualizes drawing a column of air, like a straw, over the tongue and throat, then bending consists of ''bending that column of air down to the front part of [the] lower jaw.'' Palmer[9] relates that Sonny Terry's technique was: ''*If you constrict the flow at any point along its passage from reeds to lungs, you change the resultant pitch of the reeds*. Most harpmen *choke* the harp to achieve this effect. They narrow the opening in their throats to constrict the air flow....there is a simultaneous increase in pressure from the lungs.'' Most practical is the advice of Baker,[10] ''...the tongue and throat movements when bending notes is (analogous to) different vowel sounds,... like forming AAH to OOH,... or EEE to OOO for the high notes.'' We interpret these instructions as attempts to describe alterations of the volume and configuration of the vocal tract—here used to describe the lips, cheeks, mouth, pharynx, and trachea (Fig. 3). Unfortunately for the novice, the oropharyngeal configurations required in bending are not identical to any of those of normal speech.

There is an unfortunate dearth of reports in the scientific literature about the acoustics of the harmonica, or the physiology involved in playing it. Perhaps this is due to its humble nature, for there are many reports, and indeed volumes, on more ''noble'' instruments. In the nearly unique article concerning physics of the harmonica by Johnston,[3] all of the references cited are to observations on other instruments, mostly woodwind and reed instruments. He cites no references to previous studies of the harmonica. Subsequent to Johnston's article, in the second edition of *Harp Handbook*, 1991,[10] Steve Baker described his experience based on simple observations and logical deductions, but his description lacks the benefit of vigorous scientific procedure. Our studies elaborate on Baker's observations and couple these with Johnston's theoretical examination of the physical dynamics of harmonica reeds.

I. FUNCTION OF HARMONICA REEDS

A. Preliminary observations

Key to most of the expression on the harmonica is the ability of the musician to play a single note. Pursing the lips to the size of a single hole is the commonly used technique. Tongue blocking is another technique in which the lips cover several holes but the tongue blocks one or more of these holes. The player forces air through a hole on one or both sides of the tongue. Another method, curling the tip of the tongue into a tube that surrounds the hole, limits use of the tongue for other purposes, such as tongue blocking for octaves. However, curling the tongue permits playing with the covers of the harmonica removed. Consequently, the curled tongue technique was used in several of our studies. By such playing with the covers off and alternately stopping one or the other reed with a finger, one learns that stopping the draw reed while blowing has no effect on pitch and only a slight increase in loudness of the blown tone. This is most noticeable with more vigorous playing. The findings are reversed on drawing and stopping the blow reed. This suggests that stopping the reed that is not primarily speaking decreases the leak of air from the chamber, and that the natural blow and draw tones come predominantly from the blow and draw reeds, respectively. The other reed of the hole contributes minimally if at all.

As mentioned above, the technique of bending notes is perceived as a lowering of pitch as the player appropriately alters the vocal tract. Draw bends are obtained in the lower holes 1 through 6 where the higher pitched tone is drawn, and blow bends are obtained in holes 7 through 10 where the higher pitched tone is blown. Alterations of the vocal tract involved in bending tones principally consists of arching and/or thickening the tongue at various places along its length (anterior or posterior). This has the effect of altering the volume and shape of the resonant cavity.

So-called *overblow* and *overdraw* tones are a relatively recent addition to playing. Such tones can be elicited from all holes of the diatonically tuned harmonica (cf. Fig. 2), but they are used for holes 1, 4, 5, and 6 blow and 7, 9, and 10 draw to produce notes that are otherwise missing on the diatonic harmonica (a flatted third—holes 1 and 4 blow; a sharped fourth and a flatted seventh—holes 5 and 6 blow; a flatted second—holes, 7 and 10 draw; and a flatted sixth—hole 9 draw). Stringent control of breath and positioning of the player's vocal tract are required in order to play overblow and overdraw tones. The novice attempting to play overblow tones can be very rapidly frustrated, for the consequence of an incorrect posture of the vocal tract is either silence, or an inharmonious sound. The first recording of such a tone is credited to Blues Birdhead (James Simons) in 1929,[11] but full use and perfection of the tones awaited the skill and persistence of Howard Levy who plays them with as much facility as he plays the bends. Consequently he is able to use the naturally diatonic harmonica as a fully chromatic instru-

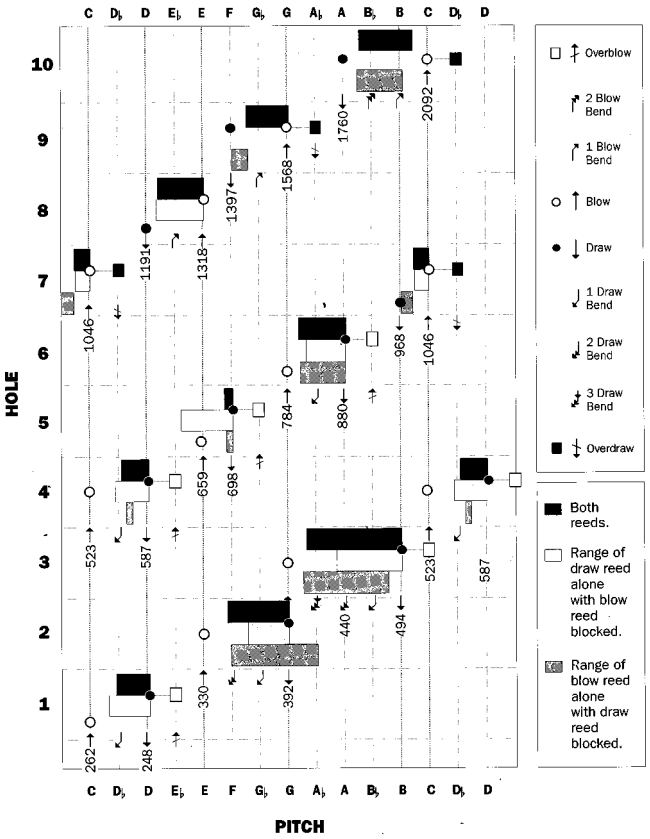


FIG. 4. Tones that can be modulated from blow and draw reeds together and alone on a Golden Melody Harmonica (Hohner) key of C, observed by blocking reeds with a finger. (+) and (−) indicate cents above or below the notes indicated along the abscissa. ○ is blow note, ● is draw note. Symbols otherwise are as in Fig. 2. Bars indicate range of frequency obtained by bending: black, straight with both reeds free; light gray, blow and draw bends with draw reed alone with blow reed blocked; dark gray (lowermost of the 3 bars), draw reed alone with blow reed blocked. Bars on holes 1–6 indicate draw bends, and on 7–10 blow bends. A more skillful player might increase the ranges with more precise configuration of the vocal tract. Note that both reeds are usually active with bends. Pure overblow and overdraw tones are obtained from a single reed (see overblows and overdraws for holes 1, 4, 5, 6, 7, and 9). Note that in holes 1 to 6 the draw note is higher pitched than the blow note, and bends are drawn. In notes 7 to 10 the reverse is true, i.e. the blow note is higher pitched than the draw note, and bends are blown rather than drawn.

ment. The resulting musical effects that this enables can be quite extraordinary.

The preliminary qualitative observations of reed motion with the covers off prompted a more rigorous investigation of reed and function. This was accomplished by three techniques: analysis of pitch production, videostroboscopic observation, and precision, noncontacting measurement of the displacement of the reeds.

B. Pitch production and reed primacy

Figure 4 illustrates the range of pitch that can be obtained from each hole of the diatonic harmonica as it is usually played, i.e., draw bending holes 1 to 6 and blow bending holes 7 to 10. The pitch was determined with a digital tuner (Korg, DT-2, Tokyo) and is expressed as cents above or below the targeted pitch. It is convenient to refer to harmonica reeds as functioning *primarily* and *secondarily*, because although actions of the blow and draw reeds are related, one reed usually dominates and is the *speaking* reed. Some simple initial studies were conducted to identify reed primacy by alternate blocking of one or the other of the two reeds of the hole while playing various notes—straight, bent, and overblown. Figure 4 also illustrates the range of pitch that can be obtained from each reed when the opposing reed of the hole is blocked. Within certain limits both reeds of a given hole can be induced to vibrate over a range of frequencies between the natural frequencies of the two reeds of a given hole.

When the blow and draw notes in the hole are two semitones apart, for example hole 4, the bent tone can be obtained from either reed with the proper configuration of the vocal tract. This suggests that the bent tone normally comes from either or both reeds, and that the reeds can be made to share primacy in producing the bent tones. When the tones in the hole are four semitones apart as in hole 3, the higher pitched reed (whether blow or draw) is primary for the first portion of the bend and lowering of pitch one semitone. When the full, three-semitone bend is obtained, however, the sound is produced entirely from the lower pitched reed which becomes primary. More specifically, the first semitone, draw bend of hole 3 (B flat on a C harp) is produced primarily from the draw reed. The second semitone, A, can be produced from either reed alone and is most easily produced by both together. The third semitone bend, A flat, comes from the lower pitched blow reed. Similar reed function is found in holes two and ten where the notes are three semitones apart. (In hole 10 the bend is a blow bend.)

The rule that one can bend the tone of a reed down to a semitone above the lower note in the hole is a convenient one for the player since that is what appears to happen. To be more accurate, however, one should recognize that primacy shifts to the lower pitched reed which is modulated up. Thus bends more properly must also be considered as involving an upward as well as downward modulation of pitch. Furthermore, when the higher pitched reed is blocked, it is possible to bend the lower pitched reed down. When the higher pitched reed is unblocked during one of these bends, the bent note is rapidly squelched.

Greater skill in bending allows wider range of control of all modes of reed action. A more skillful player might extend the range of bending beyond those shown in Fig. 4, as well might one do with a differently configured or tuned harp. Nonetheless, it appears that whether bending a tone up or down with both reeds free, the pitch can be bent most easily in the direction of the other note in the hole and is much more difficult (if not impossible) when the lower pitched reed is blocked.

In summary of these observations of reed function, it appears that both draw and blow bends involve both reeds to produce pitches in the interval between the natural pitches of the two reeds. The higher pitched reed bends down and the lower pitched reed bends up. The bent tone seems to come primarily from the reed with the natural pitch closest to the tone being played. There is more overlapping function of the lower pitched reeds, as in hole 3, than in the higher ones, as in hole 9. These conclusions require more quantitative validation. Accordingly, the speculation and subjectivity associ-

TABLE I. Stroboscopic examination of blow and draw reeds with various modes of playing. (+++) indicates vigorous action; (++) is moderate action, and (+) is minimal action. D Bend, DD Bend, and TD Bend indicate a draw bend, double draw bend, and triple draw bend respectively.

Hole	Action	Appx. note	Blow reed	Draw reed
1	Blow	C	+++	0
	Draw	D	++	+++
	D Bend	C#	++	++
	Overblow	D#	+	++
2	Blow	E	+++	0
	Draw	G	0	+++
	D Bend	F#	++	+++
	DD Bend	F	++	++
3	Blow	G	++	0
	Draw	B	0	+++
	D Bend	A#	+	+++
	DD Bend	A	++	++
	TD Bend	G#	++	+ to ++
	Overblow	C	0	++
4	Blow	C	++	+
	Draw	D	+	+++
	D Bend	C#	++	++
	Overblow	D#	0	+++
5	Blow	E	++	+
	Draw	F	+	++
	Overblow	F#	+	++
6	Blow	G	+++	0
	Draw	A	0	+++
	D Bend	G#	+++	0
	Overblow	A#	0	+++

ated with these preliminary observations that anyone can make can be elucidated by measurement of reed dynamics.

C. Stroboscopic evaluation of reed primacy

The above observations, obtained by blocking reeds, were confirmed by direct observation and by video recording the reed motion through a magnifying fiberoptic videoendoscope having a 70° angled tip (Nagashima) under stroboscopic illumination (Bruel and Kjaer, Rhinolarynx stroboscope, 4914). When the frequency of the strobe was offset slightly from the frequency of the tone being played, reed activity could be observed as a slow motion portrayal of its movement.[12] Analysis of the video tape recordings of such studies, summarized in Table I, shows that with simple blow and draw, the designated blow or draw reed vibrates freely while the secondary reed barely moves or vibrates minimally, the displacement being estimated at a twentieth or less that of the primary reed. At the beginning of a draw bend, the draw reed was observed to vibrate primarily, but as the tone dropped a whole tone or more, the blow reed began to vibrate and to become primary, even though the air flow continued to be drawn. In this mode, the blow reed functioned as an opening reed. Comparable observations were made with the blow bends. These shifts in reed primacy were most clearly seen in holes where the pitches of the blow and draw tones are more widely separated, as in holes 2 and 3 or 10. Somewhat similar observations were reported independently by Thaden,[13] Baker,[10] and other observant players.[14]

Overblows and overdraws are examples of things not being exactly what they seem on the harmonica. It would be logical to think, and was commonly believed, that the overblow notes come from the blow reed on which the player is blowing and that the pitch jumps up two or three semitones as one continues to blow. In actual fact, our simple, initial studies of reed blocking with a finger showed that the overblow tone actually comes from the draw reed of the hole which abruptly begins to vibrate as an opening reed, with a pitch that is only a semitone higher than the draw reed's natural frequency. Conversely, an overdraw tone was observed to come from the blow reed of the hole. In addition, stroboscopic analysis of the reeds during overblows showed that the blow reed becomes almost still, in spite of intensified vibration of the draw reed. Stopping the blow reed while playing an overblow tone also makes the tone purer, louder, and easier to hit. This observation led to the development of a modification of the typical diatonic harp that makes it easily fully chromatic.[14,15] The pitch of both overblow and overdraw tones is outside the interval between natural pitches of the blow and draw notes of the hole, in contrast to bends which are almost completely within these intervals. Hence, there appear to be three modes of useful function of each reed: closing as in the natural modes of simple blow or draw, opening as in bending with the pitch within the interval of the two tones, and an overblow or overdraw as an opening reed with a pitch above that interval.

D. Reed primacy by dynamic vibration measurements

Dynamic measurements of reed vibration were performed to better quantify the observations obtained from the above stroboscopic evaluation. These experiments were conducted with a Hohner Golden Melody harmonica with covers removed, key of C, upon which were mounted two precision, noncontacting, proximity sensors (KD-2400, Kaman Instruments Corp., Colorado Springs), one over the blow reed and one over the draw reed. A specially constructed fixture allowed these sensors to be positioned to measure simultaneously the motion (displacement) of both the blow and draw reeds of any hole [see Fig. 5(a)]. The frequency response of the system was flat from DC to 10 kHz (± 3 dB). The output of the reed displacement transducers was recorded digitally at 10 000 Hz by a computer workstation (Apollo 3500, Apollo Computer Inc., Chelmsford, MA) which implemented high-speed data acquisition software (Significat, Hudson, MA). These signals were further processed in real time by a custom built circuit which measured the frequency and peak-to-peak amplitude of both signals of reed displacement. The latter data were recorded on a PC/386 with commercial acquisition software (LabTech Notebook, Laboratory Technologies, Inc., Wilmington, MA).

In addition to observations with simple oral playing, a resonating volume chamber was placed in the air stream adjacent to the harp [see Fig. 5(b)]. This device acted as a rudimentary equivalent of the vocal passage, similar to a system described by Johnston.[3] With the simple plunger the volume of the resonating chamber was easily altered, and the

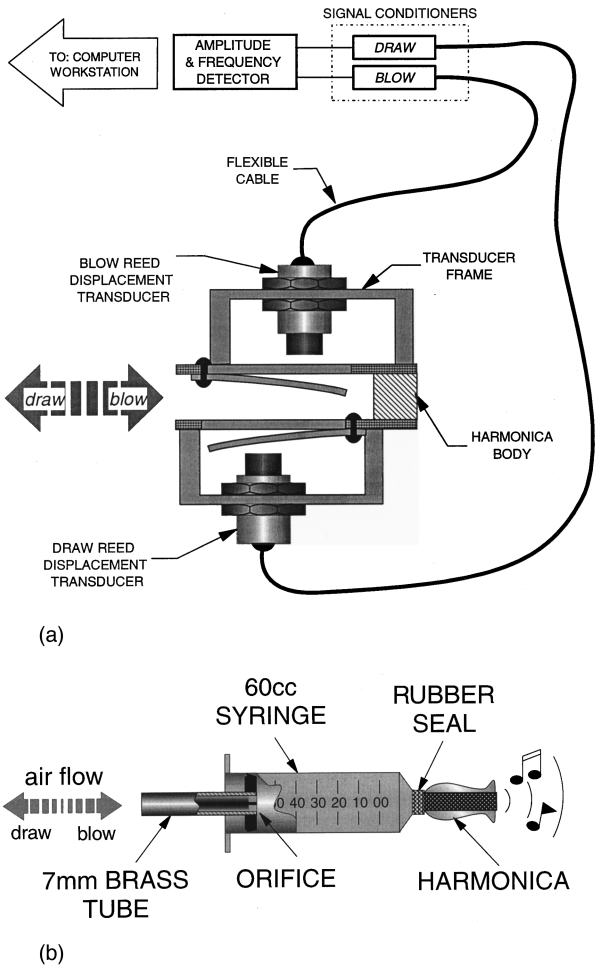


FIG. 5. (a) Schematic of experimental apparatus for measuring and recording reed displacement. (b) Sketch of variable resonating volume chamber (60-cc syringe) used to crudely simulate vocal tract, modified from device described by Johnston (Ref. 10).

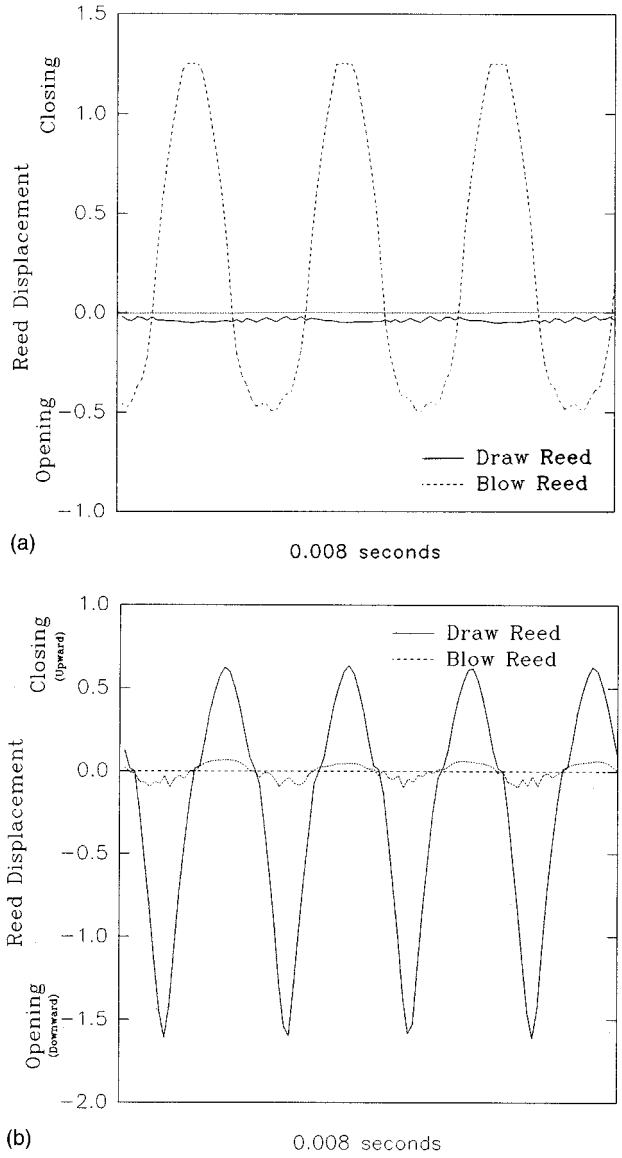


FIG. 6. (a), (b) Reed displacement with time for straight blow (a) and draw (b) notes on hole 3. Reeds move in parallel. Displacement values are relative only.

reeds were made to vibrate in any of the three different modes of speaking (blow and draw; blow and draw bend; and overblow and overdraw) using regulated air pressure available in the laboratory.

Raw displacement data were displayed in compressed form in order to reveal the relative displacement envelope of the two reeds during various maneuvers (for example, see Figs. 7 and 12). The uncompressed display of the same data allowed detailed analysis of the shape of the trajectory of the individual reed displacement. The relationship between relative amplitudes and frequencies of the two reeds could be observed during bends and overtones from the real time display of the signal processor circuit.

Measurement and display of reed vibration provided several qualitative observations as well as quantitative results. Although all 10 holes were studied, we present below detailed results for hole 3, which were taken to represent the general behavior observed for the remaining holes. It should be noted, however, that the two reeds of hole 3 provide a wider interval of tones than any other hole (six semitones from G, a straight note, to C, an overblow). Although unique, the hole appears to encompass all modes of action.

When a natural note was obtained by either blowing or drawing, only the primarily speaking blow or draw reed was significantly active. There was only a minimal amount of the secondary reed vibration [Fig. 6(a),(b)]. This confirmed the observations made with the strobe light listed in Table I. In addition to the peak-to-peak amplitude of vibration, it was also informative to consider the relative displacement of the nominal position of the vibrating reed.

When playing a low blow note [such as 3-hole blow, Fig. 6(a)], the nominal, or ''average,'' position of the blow reed moved away from the comb. In other words, the blow reed appeared to travel farther upward, into the reed slot, out of the comb, than downward. The draw reed, although nearly idle during blowing, appeared to displace slightly upward, into its respective reed slot and into the comb. The low draw notes [such as 3-hole draw, Fig. 6(b)] were characterized by a primarily closing draw reed, but the blow reed appeared to vibrate, albeit slightly, either about the nominal position or slightly away from the comb. The high blow and draw notes

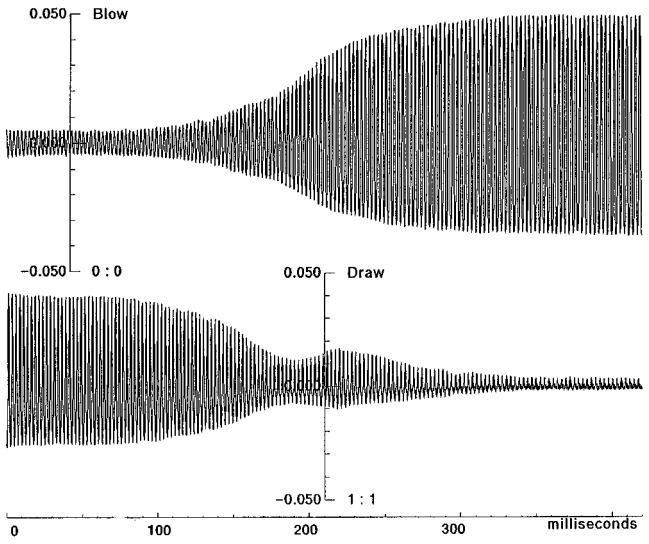


FIG. 7. Relative displacement of blow and draw reeds entering a bend on hole 2, from G to approximately F#. Note, the draw reed speaks initially and as a closing reed, but the blow reed takes over as an opening reed, as the player continues to draw but with a different embouchure.

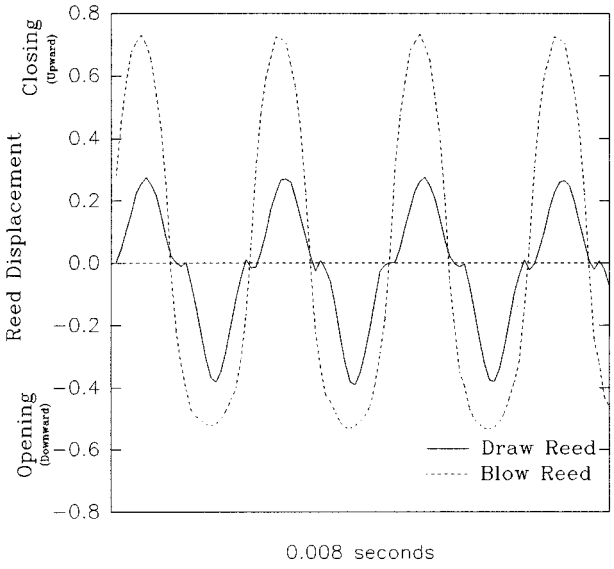


FIG. 8. Relative displacement of blow and draw reeds during a two-semitone bend of hole 3. Note significant vibration of both reeds, and parallel motion of the reeds.

behaved similarly, with roles reversed. These observations contradicted the intuitive prediction that the nominal position of both reeds of a hole should move outward on blowing with positive pressure in the hole and inward on drawing with negative pressure in the hole.

Intuition would likewise dictate that the relative phase between the two reeds would result in their contrary motion: outward with blowing and positive pressure in the hole and inward with drawing and negative pressure in the hole. However, this was observed to be the case only for the high blow notes. When any of the ten holes was drawn, the reeds were observed to oscillate in parallel with one another. Since the draw reed did not vibrate appreciably for the low blow notes, it was difficult to distinguish the relative phase. Parallel motion of reeds can be interpreted as both reeds acting simultaneously to close, or open, their respective reed slots. The level of pressure also appeared to bear a slight influence on the phase relation for some of the holes. For example, when high pressure was applied on the 3-draw or 3-draw bend, the motion of the closing reed was observed to lag slightly further behind the opening reed.

The displacement patterns of the blow and draw reeds displayed disparate degrees of departure from simple harmonic motion. In most cases, the blow reed appeared to follow a sinusoidal trajectory, whereas the draw reed displayed a notable amount of superimposed third harmonic (second overtone; see Figs. 6(b) and 8, for example). The degree of departure of both displacement patterns from a sinusoidal shape was more prominent for higher air pressure and flow (not shown in figures) than for lower pressure and flow.

Dynamic measurements of reed displacement were recorded while *entering* and *exiting* a 3-draw-bend (3DB). A compressed view of the reed displacements upon entry to the bend is demonstrated in Fig. 7. As the bent note was established, the amplitude of vibration of the (higher pitched) draw reed gradually diminished in favor of increased vibration of the blow reed. A detailed view of reed displacement at the point of a two-semitone bend (low A on a C harp) is shown in Fig. 8. Comparison with the pure blow and draw waveforms demonstrates that the bent note results in an almost composite of the two. The wave shapes and relative position of both the blow and draw reed displacements were observed to resemble those occurring for their respective straight notes. (This is worthy of recognition since the blow reed is no longer operating as a closing reed, but operates as an opening reed.)

Figure 9 depicts data for hole 2 played by mouth as a double draw bend is entered. The amplitude of vibration is plotted as a function of frequency. With proper configuration

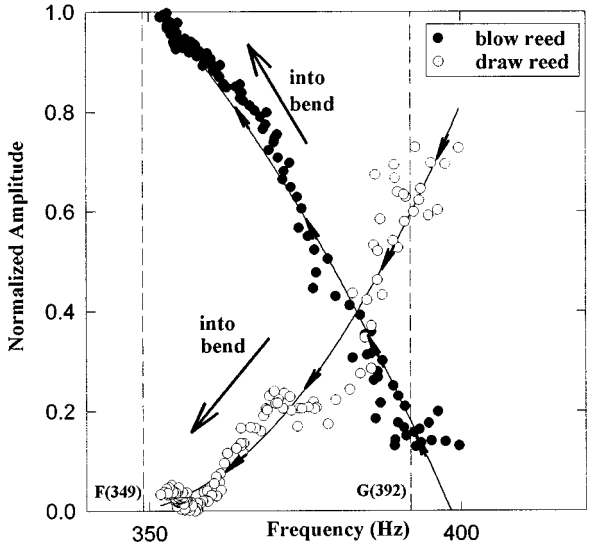


FIG. 9. Amplitude of reed displacement versus frequency while entering a double draw bend from a straight draw on hole 2 of a 10 hole diatonic harmonica. As frequency decreases, the amplitude of vibration of the draw reed decreases as the amplitude of the blow reed vibration increases.

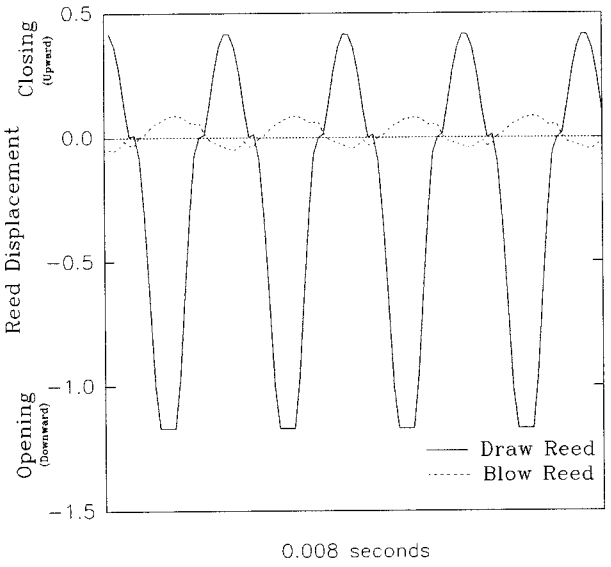


FIG. 10. Relative displacement of reeds with overblow on hole 3. Almost all action is on draw reed. (Clipping of waveform at bottom is due to saturation of the amplifier.)

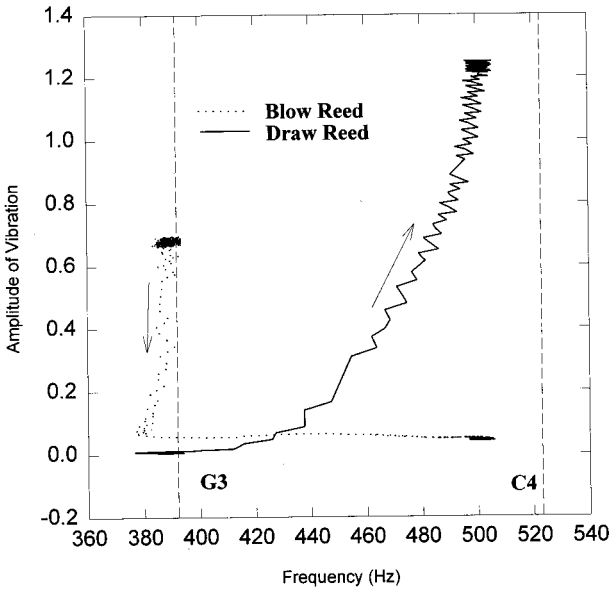


FIG. 11. Relative displacement of reeds during onset of overblow on hole three (not used in normal playing). Note: blow reed, with pitch of G, speaks initially but abruptly becomes silent as the draw reed, with natural pitch of B, speaks with (overblow) pitch of C.

of the vocal tract, a wide range of pitch was possible, confirming observations shown in Fig. 4. As the bend was entered and frequency dropped, the amplitude of vibration of the draw reed diminished while the amplitude of the blow reed increased. Thus the act of bending appeared to divert or transfer the reed activity from the draw (closing) reed to the blow (opening) reed.

The nominal positions of the reeds during the low draw bends (holes 1 to 6) were somewhat counter-intuitive. The nominal position of both reeds was displaced *away* from the comb [see Fig. 6(a), for example]. In other words, the maximal outward excursion of the reeds was surpassed by their maximal inward excursion—counter the principal air flow and against the primary pressure gradient. The high note blow bends (holes 7 to 10) did not evidence such anomalous behavior. However, for both draw bends and blow bends, counterintuitive parallel motion of the reeds was demonstrated. Most intriguingly, this did not occur in holes 9 and 10. As with the straight notes, the phase relation between the reed motions was observed to depend on the magnitude of air flow. Particularly, for the low draw bends, it was possible to achieve phase from 0 degrees (parallel motion) to 90 degrees phase lag between the reeds, depending on the intensity of the tone.

It was possible to obtain overblow tones with the volume-chamber apparatus, in most instances with a small volume. The overtones achieved in this fashion were observed to occur abruptly, as they do with normal playing of overblow notes. The transition from blow to overblow would ''pop'' into place as one changed the volume rather than slide smoothly, as is characteristic of the bent tones. The apparatus allowed more pure and complete overblows to be achieved than could be obtained with oral playing. Figure 10 shows the detailed waveforms of reed displacement for a sustained hole-3-overblow. The minimal activity of the blow reed noted here was not always easily obtained and sustained when the harp was played by mouth. For all holes studied, the nominal position of the reeds was observed to move outward with respect to the comb, concurrent with the air flow. This was especially true for the vigorously vibrating draw reed.

The relative phase of the reeds for the overblows studied (hole 3 to 6) was observed to depend upon the location of the displacement transducer. When placed near the roots of the two reeds of hole 5 and 6, the phase relation evidenced parallel motion; whereas, when placed at the tips, the motion was in opposition. The opposite was true for hole 4. Hole 3 was inconclusive due to the negligibly small amplitude of vibration of the blow reed.

The distinguishing characteristics of the overblow as compared to the bend are further illustrated by comparing the amplitude versus frequency plot of an overblow (Fig. 11) with that of a bend (Fig. 9). Unlike the bend, which displays a smooth and gradual transition of reed primacy as the frequency changes, the overblow demonstrates a much sharper drop in the amplitude of oscillation of the closing, blow reed and simultaneous rise in amplitude of oscillation of the opening, draw reed. The interval between natural frequencies of the two reeds is far more devoid of activity in the overblow as compared to the bend.

The resulting overlap in frequency response of the two reeds during bending allows the player not only to bend more easily, but to slide between bent and straight notes. Consequently, this provides more opportunity to introduce expression into the notes being played. An example of this is depicted in Fig. 12 which shows a tracing of the amplitude of reed vibration compressed in time and recorded during a 6-draw to 6-draw-bend, played by Howard Levy.

In some cases, it was possible to obtain, and maintain, a dissonant overblow. In this case, the vibration of the two reeds remained synchronized but assumed different fundamental frequencies. Figure 13(a) and (b) shows, respectively,

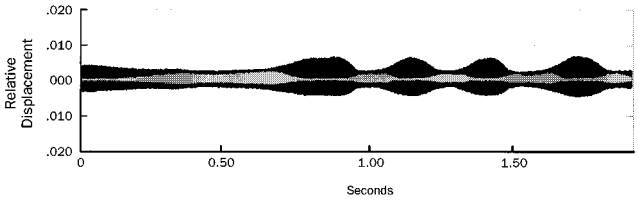


FIG. 12. Relative displacement amplitude of blow and draw reeds during several partial entries and exits of hole 6 draw bend. Note reciprocal shifting activity of the two reeds. Note the blow reed is the dominant actor in the draw bend, as indicated in Table I.

a consonant and dissonant overblow obtained in hole 3. In the case of the dissonant 3-overblow, the draw reed, with a natural pitch of B (494 Hz), was induced to vibrate at 518 Hz, but the blow reed with natural pitch of G (392 Hz) vibrated at 358 Hz (approximately F) with an apparent 25% modulation in its period from cycle to cycle. In a consonant overblow, the draw reed was observed to vibrate at 523 Hz (C_5) and the blow reed vibrated at the same frequency but with diminished amplitude. In normal oral playing such dissonance is common during overblows, and is produced even more frequently by less skillful players.

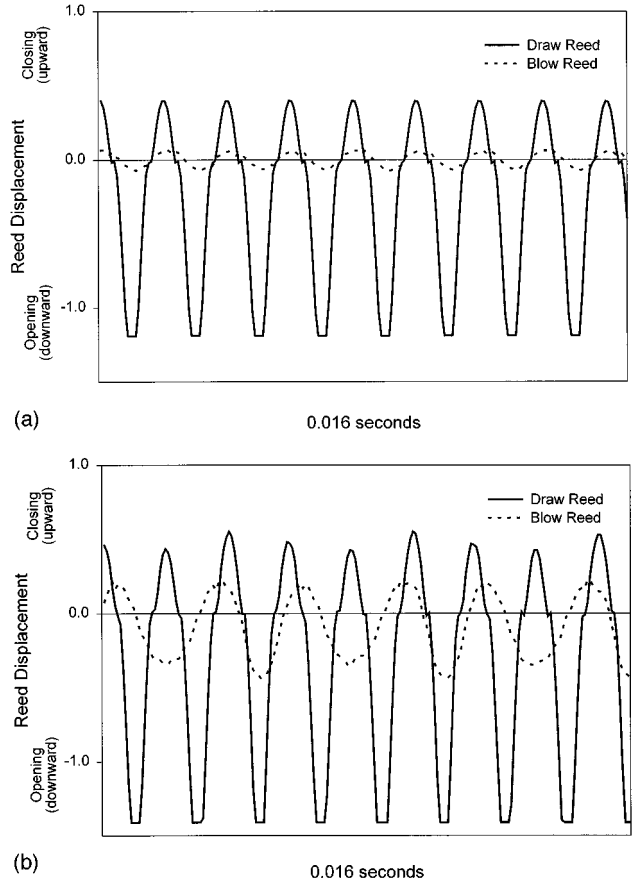


FIG. 13. (a), (b) Relative vibration of reeds during a consonant (a) and dissonant (b) overblow on hole 3. During the dissonant overblow, the draw reed, with a natural pitch of B (494 Hz), vibrated at 518 Hz, but the blow reed with natural pitch of G (392 Hz) vibrated at 358 Hz (approximately F) with an apparent 25% modulation in its period from cycle to cycle. In a consonant overblow, the draw reed vibrates at 523 Hz (C_5) and the blow reed vibrates at the same frequency but with diminished amplitude.

II. DISCUSSION

For over a century harmonica players have manipulated the harmonica to produce a wide variety of tones, glides, and slurs. The resulting effects reflect the dynamic interplay of physics and acoustics of the harmonica coupled to the body of the musician.

To understand the dynamics of music production with the harmonica it is essential to understand the behavior of the instrument. In these studies we have focused on the dynamics of the reeds themselves and of their function as a player achieves desired notes through the combination of blowing, drawing, bending, overblowing, and overdrawing. The information obtained has shown how an array of sounds can be accurately produced with a single channel and a pair of reeds. For example, in the third hole of the popular diatonic harmonica, two natural notes (blow and draw), one-, two-, and three-semitone bends, and an overblow can be produced with the two reeds by alteration of the vocal tract of the player. The new group of tones introduced in the last quarter century, namely overblows and overdraws, have been shown to be effected primarily by only one reed which functions as an opening reed—the normal function is closing—at a pitch outside the interval of the two notes in the hole. Better understanding of these new tones provided the motivation for the authors to undertake the studies described here.

The common rule that bent notes are limited to within a semitone of the opposite note of the hole was substantiated in these studies. This is partially explained by the theoretical model of Johnston which describes the range of pitches which are producible from a single reed. His analysis demonstrated that closing reeds can be flatted a semitone or so below their natural resonant pitch. Johnston's theoretical model unfortunately did not account for this possible coupling between the two reeds. The reed-blocking experiments reported above demonstrated that the presence of a lower pitched, opening reed in the same hole appeared to diminish the potential range by imposing a lower limit on the bend. Similarly, the presence of a *higher* pitched opening reed was found to virtually obliterate the blow bend of the closing reed in holes 1–6 and the draw bends in holes 7–10.

Due to the simplifying assumptions associated with Johnston's model, it is also not capable of accounting for many of the nuances of the construction of the harmonica itself. Therefore, additional modeling will be necessary in order to understand how the geometry, physical properties, and air flow combine to produce the observed phenomena involved in producing tones on the harmonica. In spite of this instrument's apparent simplicity, there is a multitude of modifications that have been introduced, or proposed, to alter the timbre and improve the ease of achieving certain notes. The aerodynamics associated with increasing the reed plate thickness, the shape of the air hole in the comb, or the orientation of the reed pairs, for example, would make an interesting topic of study if a more elaborate mathematical model were developed.

Our quantitative measurements of reed dynamics help substantiate some of the suspected relationships between reed primacy and the fundamental tone that is generated. Since the trajectory of the reed determines the flow of air, the

fundamental frequency of the reed displacement determines the frequency of the tone that is generated. However, the reed motion was observed to contain overtones. The cause of this distortion could be explained by the nonlinearity of either the displacement measurement system or the reed oscillator itself. In the case of the draw reed, a significant inflection in the displacement tracing occurred in close proximity to the zero-crossing point of the reed with respect to the reed plate. The calibration of the displacement transducers did not reveal any discontinuity at the zero-crossing point which could explain this inflection. Accordingly, this ''hesitation'' in the reed motion is most likely caused by either the aerodynamic disturbance or instability of the jet of air which results in the constricted reed slot.[2]

The paradoxical behavior observed with respect to the parallel motion of some of the reed pairs may appear counterintuitive and initially seems to contradict the fact that the blowing pressure or drawing pressure causes one reed to open when the other one closes, and vice versa. However, it can be explained through the small-signal theoretical model of Johnston[3] which predicts the relative phase of the reed motion with respect to the acoustic pressure in the hole. As would be expected from a nonlinear dynamic system, the phase relation at higher amplitudes (greater pressure or air flow) was observed to differ slightly from the low amplitude response. In particular the closing reed was observed to lag slightly further behind the opening reed for the draw and draw-bends when high pressure was applied. The slight variation of phase observed in response to increased air pressure can easily be explained by the well-known influence of damping on resonant frequency and phase angle; see Ref. 16, for example.

Care was taken to position the displacement sensors consistently in a similar relative location along the length of each of the reeds. The finite spatial resolution of the sensor resulted in an averaging effect of the displacement measurement. For most of the reeds studied, the shape of the displacement trajectory did not appear to depend upon the location of the transducer. However, the overblow in hole 3 and 6 did demonstrate such a dependence. This observation can be taken to imply that one, or both of the reeds during these overblows assumes an inflected shape. This would further imply that an overtone would be present. Basic beam theory would predict the frequency of this overtone to be 6.26 times the fundamental.[16] The absence of this observation for the other maneuvers could be attributed to the absence of this overtone or to the limited spatial resolution of the sensor, as mentioned above.

The counterintuitive observations relating to the nominal position of the reeds may be understood by considering the effect of air flow velocity upon the local gradient of pressure. According to the Bernoulli principle, increased velocity along a streamline of the flow, for example created by the constriction between the reed and reed slot, causes the pressure to decrease; therefore, it is possible to create a partial vacuum with positive flow. These results also shed some light as to the influence of reed offset upon ease of achieving desired notes. The common practice of adjusting the resting position of the reed, or reed offset, by gently bending the reed into or out of the plate may, in effect, relocate the reed to an equilibrium position more amenable to bending or overblowing/overdrawing—but at the expense of increasing the difficulty of playing the natural tone.

The graphs of amplitude versus frequency of reed vibration that were generated during bent, overblown, and overdrawn notes were quite informative for demonstrating the distribution of reed primacy. As the linear acoustical model of Johnston[3] would theoretically predict, the process of bending involved a smooth transition of primacy, whereas the overblows and overdraws demonstrated a more discontinuous, sudden jump in primacy.

The harmonica is nearly unique among musical instruments in that the vibrating oscillator, the reed, which produces the sound, is alternately upstream and downstream from the controlling resonating volumes. We have attempted to qualify and quantify some of the phenomena of reed function as the reeds react to this acoustical coupling. Johnston,[3] who provided the first scientific study of this phenomenon, demonstrated that changes of pitch are effected by altering the acoustic impedance of the vocal tract, which in turn drives both reeds of the channel of the harmonica.

Similarly, the present investigation has centered primarily on the harmonica itself. However, in order to understand the dynamic function of harmonica playing, the human vocal tract must also be studied. Clinch and associates[17] have shown with the clarinet, saxophone, and recorder that for good quality notes resonant frequencies of the vocal tract must match the frequency of the desired note. The harmonica requires pitch control of another order, for the acoustic impedance of the vocal tract controls not only quality of tones but it also raises or lowers the pitch in order to achieve certain notes. A wide variety of air flow rates, vocal cavity volumes and vocal configurations are compatible with the production of natural tones from the harmonica. To produce a more pure natural tone requires not only adequate shape of the vocal tract, but good musical sense and a sensitive auditory-vocal feedback mechanism. It appears that the volume and shape of the vocal tract have only a slight effect on the natural tones, but the mode and frequency of the fashioned notes—bent, overblown, or overdrawn—are caused by changes in the vocal tract.

In a preliminary study of one aspect of the vocal tract, namely the volume of the oral cavity, a simple experiment was conducted in which a player, lying supine, played specific fashioned notes and then held the configuration of the vocal tract while the oral cavity was filled with water and the required volume was recorded. Reproducible results were obtained with practice. The results of this preliminary experiment are shown in Fig. 14. It is probable that actual playing volumes were larger than measured, since when water was instilled, constriction of the glossopharynx and larynx occurred in order to suppress the swallow reflex and prevent aspiration. The volume of the anterior oral cavity was found to be inversely related to pitch as modified by bending. This relationship was not found with the straight notes which could be obtained with a wide range of airway volumes. Changes in configuration of the oral tract, however, determine the timbre of notes, which is a subject of further study.

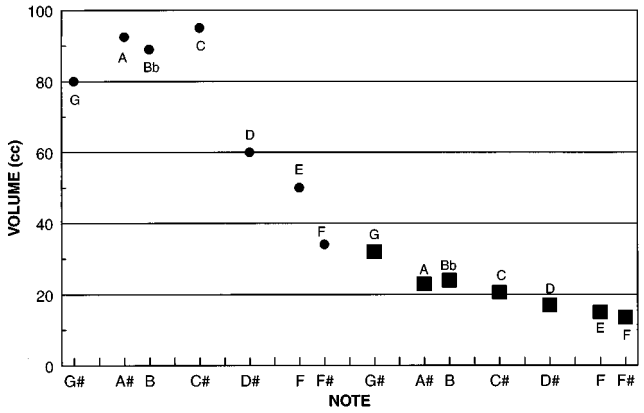


FIG. 14. Volume of oral tract used to play the scale with 1 and 4 draw bends. Harps keyed in the scale of F were used on which hole 1 draw bend (circles), or for the higher scale, hole 4 draw bend (squares) were played. The resulting scale is in the key of F#. The player held the configuration of the vocal tract as the volume of water required to fill the oral cavity was introduced and measured. Pitch of notes was inversely related to volume.

Upon delving into the physics and physiology of harmonica playing, the extent of the unknown phenomena concerning this instrument became increasingly evident. Each experiment, more times than not, revealed new ground yet unexplored. The authors now realize that this report does not include many of the additional features of sound affecting the timbre, modulation, or voicing. Most notably absent in these studies was the use of cupping of the hands. Most harpmen will use this technique to introduce vibrato and other forms of expression, as well as to alter the timbre of the sound. The answer to these and other questions regarding the physiology of harmonica playing will be further elucidated through systematic analysis of coupled interaction of the harmonica with the vocal tract—as revealed by videoendoscopic, ultrasonic, and fluoroscopic imaging.

III. CONCLUSION

Artistry of harmonica playing is related to the interplay between the player and the instrument. In this project, the physical behavior of the reeds of the harmonica was investigated. Mechanical and acoustic aspects of reed dynamics were elucidated during blowing and drawing, overblowing and overdrawing, and blow and draw bending. Through knowledge of reed function, the artistry and process of ''speaking'' with the harmonica can be better understood, and possibly mastery of this instrument will be enhanced.

In the final analysis, the music created by the harmonica consists of more than the acoustic and physical function of the reeds. There is a synergy which causes the whole to be greater than the sum of the component parts and an artistry which cannot be quantified—dynamics which give personality to the instrument, reflecting individuality of the musician. As with speech which varies from person to person, there are certain tonal elements of harmonica playing which are similarly individualized. This dynamic interaction allows the player to speak with his instrument perhaps as with no other. Just as no two voices are exactly alike, each player imparts his own timbre, and one cannot expect to emulate exactly the musical tonality of another. This helps keep the harmonica interesting, and indeed has helped to sustain its enduring prominence throughout the world

ACKNOWLEDGMENTS

The authors wish to acknowledge the willing and expert assistance of Ricardo Schultz and Lee Denault who performed much of the computer processing of the experimental data. The authors are also highly grateful to Howard Levy for participating in several unusual experiments involved in these studies of harmonica physics, phonetics, and physiology.

[1] K. Field, *Harmonicas, Harps, and Heavy Breathers* (Fireside, New York, 1993), Chap. 1, pp. 19–30.
[2] A. O. St. Hilaire, T. A. Wilson, and G. S. Beavers, ''Aerodynamic excitation of the harmonium reed,'' J. Fluid Mech. **49**, 803–816 (1971).
[3] R. B. Johnston, ''Pitch control in harmonica playing,'' Acoust. Aust. **15**, 69–75 (1987).
[4] P. Duncan, *Blues Harp for Diatonic and Chromatic Harmonica* (Mel Bay Publications, Pacific, MO, 1982), p. 17.
[5] T. Glover, *Blues Harp* (Oak Publications, New York, 1965), p. 32.
[6] P. Duncan and C. Musselwhite, *Power Blues Harp* (Mel Bay Publications, Pacific, MO, 1990), p. 13.
[7] S. McCloskey, *Harping It Easy* [cassette tape] (Hohner, Richmond, 1983).
[8] T. Glover, *Blues Harp Songbook* (Oak Publications, New York, 1973), p. 10.
[9] F. Palmer, *The Harp Styles of Sonny Terry* (Oak Publications, New York, 1975), p. 47.
[10] S. Baker, *The Harp Handbook* (Hermann Demmler, Ludwigsburg, 1991), p. 20ff.
[11] W. T. Yerxa, ''Bending and overblowing—All one donut,'' Harmonica Information Press **4**, 35–42 (1994).
[12] M. Hiramo and D. Bless, *Videostroboscopic Examination of the Larynx* (Singular Publishing Group, San Diego, 1993).
[13] J. Thaden, ''Doctor Diatonic,'' Harmonica Horizons **5** (1990).
[14] W. T. Yerxa, ''The Overblow Harp,'' Harmonica Information Press **4**, 4–7 (1994).
[15] H. T. Bahnson, U.S. Patent No. 5,739,446 (1993).
[16] S. Timoshenko, D. H. Young and W. Weaver, Jr., *Vibration Problems in Engineering* (Wiley, New York, 1974), p. 74ff.
[17] P. G. Clinch, G. J. Troup, and L. Harris, ''The importance of vocal tract resonance in clarinet and saxophone performance: A preliminary account,'' Acoustica **30**, 280–284 (1982).

Broadband measurements of phase velocity in Albunex® suspensions

Joel Mobley, Jon N. Marsh, and Christopher S. Hall
Laboratory for Ultrasonics, Department of Physics, Washington University, St. Louis, Missouri 63130-4899

Michael S. Hughes and Gary H. Brandenburger
Mallinckrodt, Inc., 675 McDonnell Boulevard, Hazelwood, Missouri 63042

James G. Miller
Laboratory for Ultrasonics, Department of Physics, Washington University, St. Louis, Missouri 63130-4899

(Received 7 May 1997; accepted for publication 30 November 1997)

The phase velocities in suspensions of Albunex®, an ultrasound contrast agent, are determined using the technique of broadband phase spectroscopy. In these suspensions, phase velocities were measured as a function of Albunex concentration in Isoton II (buffered saline). The dispersion over the measurement bandwidth (1–15 MHz) grew with increasing Albunex concentration, exhibiting a 221-m/s change (17% increase) in phase velocity between 1.0 and 3.8 MHz at the highest concentration reported (32 μL Albunex/8 mL Isoton). The dispersion behavior of the Albunex suspensions is described using a model of encapsulated gas bubbles. The influence of the dispersion in Albunex on broadband pulses is discussed, as well as the potential impact this dispersion may have on the formation of ultrasonic images of the heart. © *1998 Acoustical Society of America.* [S0001-4966(98)04303-3]

PACS numbers: 43.80.Cs, 43.80.Vj, 43.20.Hq, 43.35.Bf [FD]

INTRODUCTION

Broadband ultrasonic techniques have been used to extract a wide variety of ultrasonic material parameters including phase velocity, attenuation, and backscatter and reflection coefficients. These techniques measure the desired parameters over a broad spectrum by analyzing the Fourier transforms of captured time-domain signals. The broadband phase spectroscopy technique, developed by Sachse and Pao,[1] has been successfully applied to phase velocity measurements in solids.[1–5] We have applied the phase spectroscopy approach in the context of a through-transmission substitution technique. We used two focused immersion transducers in a water tank to obtain broadband measurements of the phase velocities in suspensions of the contrast agent Albunex.

In suspensions, knowledge of phase velocities and their dispersion (i.e., frequency dependence of the phase velocity) can be required in a number of circumstances, especially those that require modeling of propagating ultrasonic fields. Contrast agents are suspensions used in medical imaging to modify the ultrasonic scattering properties in a specific anatomic region. In echocardiography, these agents are used to increase the backscatter from blood-filled cavities and vessels in order to aid in the delineation of blood–tissue boundaries in the chambers and to detect perfusion of the coronary vasculature. The phase velocities and dispersion characteristics of these agents impact the propagation of the ultrasonic imaging field. Dispersion could cause spreading of an imaging pulse, both lateral to the axis of propagation, from refraction between contrast perfused and unmodified regions, and along the axis of propagation, due to propagation through a contrast-filled volume. When the attenuation properties are also considered, this spreading could become a splitting of the pulse. Local contrast-modified changes in the phase velocity can also alter boundary reflections due to changes in the impedance differences at the interface. We report the phase velocities for Albunex® diluted in a saline solution (Isoton II®, Coulter Corp.) over a range of volume concentrations from 1 μL of Albunex in 8 mL of Isoton to 32 μL of Albunex in 8 mL of Isoton. Albunex is a contrast agent approved in the United States for use in clinical echocardiography and consists of gas bubbles encapsulated in shells of denatured human albumin, a blood protein. The majority of the microbubble volume in Albunex comes from shells with radii from 2–5 μm. At concentrations approaching those which may be relevant *in vivo*, we find that the dispersion in Albunex suspensions is significant and reaches a level sufficient, in principle, to affect feature registration in images formed using ultrasonic pulses with energy in the 1–4 MHz range.

In Sec. I of this paper, we describe the general approach to determining phase velocity as well as specifics of the data acquisition, data reduction, and samples. Section II presents the phase velocity results. Section III contains a discussion of the general features of the results, possible implications of the Albunex results on broadband pulse propagation and imaging, the impact of diffraction and boundary effects on the results, and nonlinearity.

I. METHODS

A. Phase velocity relation from broadband phase spectroscopy

Broadband ultrasonic spectroscopic methods use the Fourier transforms of captured time-localized signals to derive material parameters over a range of frequencies. In this

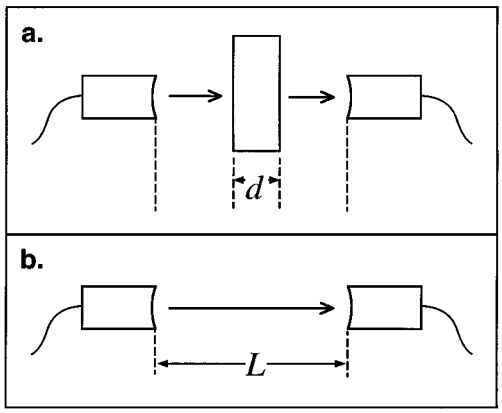


FIG. 1. (a) The relationship of the transducers and sample for the acquisition of the through-sample waveform. The sample has thickness d. (b) The setup for acquiring the reference trace. The transducers are separated by distance L for both acquisitions.

work, the phase velocities over a broad spectrum were determined by comparing the phase functions of the Fourier transforms of through-sample and reference signals. Following the approach of Sachse and Pao[1] the phase velocity relation can be derived by applying the spectroscopic technique in the context of a through-transmission substitution measurement. The measurement procedure is diagrammed in Fig. 1. The sample is placed in the region between two transducers in an immersion tank. An ultrasonic pulse is sent through the sample and the resulting received waveform is captured. The sample is then removed and a pulse is propagated through the water only. The derivation assumes that the host, sample, and receiver all respond linearly and that the sample faces are normal to the direction of propagation.

The transmit pulse is taken to be a time-localized excitation and is expressed as a superposition of unidirectional plane waves with harmonic time dependence, unbounded in the $\hat{y}$ and $\hat{z}$ directions, propagating along the positive $\hat{x}$ axis. At a distance x from the source plane, the signal can be written

$$f(x,t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}F(\omega)\exp(-\alpha(\omega)x)\times\exp\left[i\omega\left(t-\frac{x}{c(\omega)}\right)\right]d\omega, \tag{1}$$

where $c(\omega)$ is the speed of propagation for planes of constant phase and $\alpha(\omega)$ is the attenuation coefficient. The following Fourier transform pair can be read directly from Eq. (1):

$$f(x,t)\Leftrightarrow F(\omega)\exp(-\alpha(\omega)x)\exp\left(-i\frac{\omega x}{c(\omega)}\right). \tag{2}$$

The transform of the water path reference signal captured by the receiver can be shown to be

$$F_w(\omega)=|F(\omega)|\exp(i\phi_0)\exp(-\alpha_w(\omega)L)\times\exp\left(-i\omega\frac{L}{c_w(\omega)}\right), \tag{3}$$

where $F(\omega)=|F(\omega)|\exp(i\phi_0)$ and L is the transducer separation. Similarly, the Fourier transform of the captured through-sample signal can be shown to be

$$F_s(\omega)=T(\omega,d)H(\omega,d)F(\omega)\exp(-\alpha_w(\omega)[L-d])\times\exp\left(-i\omega\frac{L-d}{c_w(\omega)}\right), \tag{4}$$

where $T(\omega,d)$ is the transmission factor due to the sample/host interfaces, $H(\omega,d)$ is the transfer function of the sample, and d is the sample thickness. Here, $T(\omega,d)$ and $H(\omega,d)$ are expressed explicitly as follows:

$$H(\omega,d)=\exp(-\alpha_s(\omega)d)\exp\left(-i\omega\frac{d}{c_s(\omega)}\right), \tag{5}$$

$$T(\omega,d)=\frac{t_{w\to s}t_{s\to w}}{1-r_{s\to w}^2\exp(-2\alpha_s d)\exp(-i2\omega d/c_s)}, \tag{6}$$

where $\{t_{w\to s},t_{s\to w},r_{s\to w}\}$ are the transmission and reflection coefficients for either pressure or particle velocity valid for boundaries between infinite half-spaces, and $c_s(\omega)$ and $\alpha_s(\omega)$ are the phase velocity and attenuation coefficient of the sample. Explicitly writing this transform in terms of its magnitude and phase gives

$$F_s(\omega)=|T(\omega,d)||F(\omega)|\exp(-\alpha_w[L-d]-\alpha_s d)\times\exp\left(-i\omega\left[\frac{L-d}{c_w(\omega)}+\frac{d}{c_s(\omega)}\right]+i\phi_t+i\phi_0\right), \tag{7}$$

where $T(\omega,d)=|T(\omega,d)\exp(i\phi_t)$.

An expression for the phase velocity in the sample is obtained by subtracting the phase spectra of the through-sample and reference transforms. The resulting expression is

$$c_s(\omega)=\frac{c_w}{1-c_w\{[\Delta\phi(\omega)-\phi_t]/\omega d\}}, \tag{8}$$

where $\Delta\phi(\omega)\equiv\arg(F_s(\omega))-\arg(F_w(\omega))$. In this work the phase contribution of the transmission factor is negligible and for the remainder of the paper we will take $\phi_t=0$. Further comments on the size of the transmission phase factor are contained in Sec. III. This relation with $\phi_t=0$ has been employed by others using this same technique.[4,5] (Note that the attenuation coefficient of the sample can be derived by comparing the Fourier magnitude spectra of the same two signals. Attenuation measurements for Albunex suspensions similar to those studied in this work have been previously reported.[6])

In the laboratory, the signals deviate from the simple unidirectional plane-wave behavior assumed in the derivation of Eq. (8) due to the diffraction resulting from the finite size and detailed shape of the transmit and receive apertures. When Fourier amplitude spectroscopy is employed for the determination of attenuation coefficients, the impacts of these diffraction effects are estimated and corrections are applied as terms which multiply the linear domain Fourier magnitudes.[2,4] In the phase spectroscopic approach, the relevant corrections will come in as terms additive to the phase spectra [in the same manner as ϕ_t term in Eq. (7)],

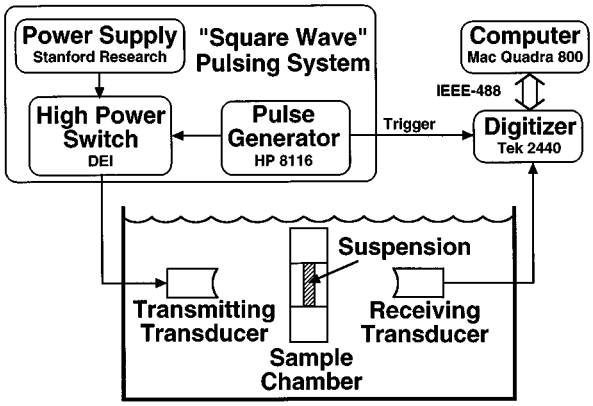


FIG. 2. Experimental apparatus used for acquiring rf. The transducers and sample chamber are completely immersed in a water bath. The suspensions are contained in a narrow section in the center of the sample chamber.

consistent with the magnitude corrections. These diffraction effects occur to some degree in all measurements whether the transducers are planar or focused. For the signal acquisitions performed in this work we believe diffraction effects have no significant impact on the measured phase differences. Further details regarding quantitative estimates of the diffraction corrections for these measurements, including the effects from using focused transducers, are left to Sec. III.

B. Measurement system

The data were acquired using a through-transmission broadband immersion system (see Fig. 2). The system employed two focused PVDF transducers (Panametrics ZF3002-SU, 3.25-in. nominal focus and 0.4-in. active diameter) immersed in a water bath and separated by 6.5-in. The transmitter was excited by a square-wave pulsing system consisting of a dc power supply (Stanford Research PS310, set to -200 V), a high power switch (DEI GRX 1.5K-E), and a pulse generator (Hewlett-Packard HP8116A). The received signals were captured by a digital oscilloscope (Tektronix 2440). The data were downloaded to a computer (Apple
Macintosh Quadra 800) for off-line storage and analysis. This acquisition system achieved a measurement spectrum from 1 to 15 MHz. The transmit pulse [see Fig. 6(a)] had one prominent zero crossing and 96% of its energy was contained in a 0.2-μs interval. The pressures near the focal region of the transmitting transducer were determined using a calibrated hydrophone (1 mm diameter, Force Institute). The peak positive pressure was 0.30 MPa and the peak negative pressure was -0.24 MPa.[6]

C. Data acquisition

Data were acquired for individual concentrations ranging from 1 μL of Albunex in 8 mL Isoton (0.089×10^6 microspheres/mL) up to 128 μL Albunex/8 mL Isoton (1.13×10^7 microspheres/mL). Stepping through from low concentrations to high, the amount of Albunex in the mixture was doubled for each new sample. Since the Albunex microbubbles are sensitive to environmental conditions such as temperature, gas balance, and hydrostatic pressure, a strict material handling protocol (see Marsh *et al.*[6]) was followed during the preparation of these suspensions to ensure their stability over the course of the measurements. The temperature of the water in the immersion tank ranged from 26.6°–26.8 °C during these data runs.

The suspensions were contained in a custom-designed sample chamber.[6] The chamber allowed for agitation during acquisition via pipette mixing to ensure homogeneity of the samples. This method of mixing has proven to be effective in obtaining stable and reliable measurements for both Albunex and polymer microsphere suspensions.[6,7] The walls of the sample chamber were oriented normal to the direction of propagation. To assess the impact of our experiments on the integrity of the suspensions, we performed a study to measure the distribution of microbubbles both before and after simultaneous insonification and agitation. These distributions were measured using an optical particle sizer (Accusizer 770, Particle Sizing Systems). Although it is known that Albunex can be unstable under insonification *in vivo*, these measurements did not detect any significant change in the microbubble distribution due to either insonification or mixing.[6]

For each suspension, 500 time traces of through-transmitted ultrasound were acquired. Additionally 500 water-path-only traces were captured to serve as the reference data. Each trace consisted of 1024 samples separated by intervals of 0.004 μs. The potential resolution of the phase velocity measurements are ultimately limited by the sampling interval (time resolution) of the digitizer and thus a narrow sampling interval lends precision to the measurement. Additionally, this sampling interval yields a Nyquist frequency of 125 MHz, thus ensuring that aliasing has negligible impact within the useable bandwidth of the system. The traces were acquired and downloaded to the computer at a rate of approximately 30 per second.

D. Data reduction and analysis

After being stored on computer disk, the rf waveform records for the through-transmitted signals were averaged in the time domain in sets of 50 before further processing. The waterpath reference signals were averaged in sets of ten.

The absolute phase spectra were calculated by unwrapping the phase derived from the discrete Fourier transforms (calculated via FFT) of the data and then compensating for time offsets such as the oscilloscope trigger delay using the Fourier shift theorem. The phase spectra for any given through-sample or waterpath acquisition were then averaged. With the phase of both through-sample and reference traces properly unwrapped and compensated, their difference is calculated. After this processing, there can still be an ambiguity in the phase difference of $m2\pi$, where m is an integer. The value for m is determined empirically since the phase velocity results exhibit large changes when m is changed by ± 1. For example, for the 16 μL/8 mL Albunex data choosing $m=1$ increased the outcome near 2.0 MHz by about 320 m/s over the $m=0$ value. The value of m was chosen in each case to yield the smallest dispersion. The phase differences are then used, along with the velocity in the host medium (water) and the thickness of the sample, in Eq. (8) to determine the phase velocities for the given sample. The velocity

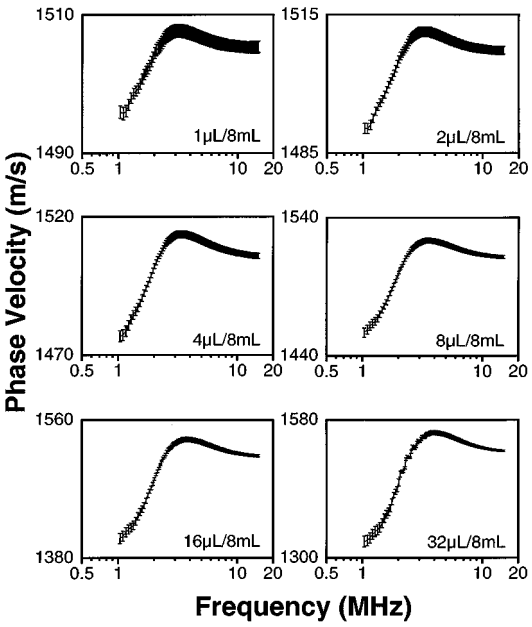


FIG. 3. The phase velocities for the specific concentrations of Albunex plotted versus a logarithmic frequency axis. Each graph has a unique vertical scale.

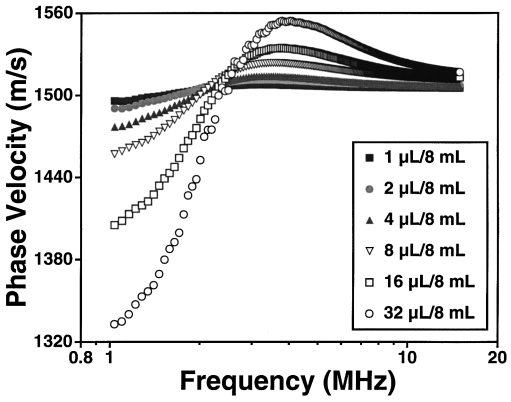


FIG. 4. The results of the phase velocity measurements for the Albunex concentrations from 1 μL/8 mL to 32 μL/8 mL.

in water is assumed to be dispersionless and determined by the temperature of the water using the formula of Del Grosso and Mader (Table III, combined fit).[8]

II. RESULTS

Figure 3 shows the phase velocity results from the Albunex suspensions for concentrations up to 32 μL/8 mL. Each individual graph in the figure has a unique vertical scale. At the 64 μL/8 mL and 128 μL/8 mL concentrations, the large attenuation through the suspensions overcame the dynamic range of our measurement system near the 2-MHz region of the spectrum. Therefore the results for these two concentrations could not be determined over the full bandwidth and are not presented here. Figure 4 displays the results for the all the concentrations from 1 μL/8 mL (0.089×10^6 microspheres/mL) to 32 μL/8 mL (2.8×10^6 microspheres/mL) in one plot so that their relative sizes are apparent. As can be seen, the most active regions of the dispersion curves occur below 4 MHz. Each curve starts near a minimum velocity at 1 MHz, climbs to a peak near 3.5 MHz, and then gradually falls toward a constant value. As a function of the concentration, the magnitude of dispersion (the difference between the maximum and minimum phase velocities) grows while the overall shapes of the dispersion curves remain similar. Figure 5 displays the magnitudes of dispersion as functions of concentration over a narrow frequency band and over the full bandwidth of the experiment. The narrow band of frequencies was chosen to coincide with the frequency spectrum used by some commercial echocardiographic systems for transthoracic imaging.

To estimate the error bounds of our results, the propagation of errors to first order in the measurement uncertainties has been performed. The velocity determination essentially requires the measurement of three quantities: temperature (determines the speed of sound in the water), length (sample chamber thickness), and time (delay interval as given by $\Delta\phi/\omega$). The fractional error equation [as derived from Eq. (8)] is

$$\frac{\delta c_s}{c_s}(T,d,t)=\sqrt{\frac{c_s^2}{c_w^4}\left(\frac{\partial c_w}{\partial T}\right)^2\delta T^2+\left(\frac{c_s}{c_w}-1\right)^2\left(\frac{\delta d}{d}\right)^2+\left(\frac{c_s}{d}\right)^2\left(\delta\frac{\Delta\phi}{\omega}(t)\right)^2}. \tag{9}$$

The temperature uncertainty, δT, was taken to be 0.2 °C. The fractional uncertainty in the thickness, $\delta d/d$, was conservatively estimated to be 0.07. The phase velocities were calculated using the average phase curves from the proper waterpath and through-sample acquisition sets; thus the phase uncertainties were calculated from the standard deviations of the two acquisition sets ($\delta\phi_w$ and $\delta\phi_s$) using $\delta\Delta\phi=\sqrt{\delta\phi_w^2+\delta\phi_s^2}$. The timing uncertainties, $\delta(\Delta\phi/\omega)$, were then obtained by dividing by the angular frequency. For the majority of data points, this calculated timing uncertainty is near 0.001 μs. The error bars displayed with the results in Fig. 3 were calculated from Eq. (9) using the uncertainties discussed above. The contribution of the fractional thickness uncertainty is scaled by the fractional deviation of the velocity in the sample from that of the water. For the lower concentrations, where the velocity deviations are small, all three uncertainties contribute similarly to the final error. For the lowest concentration, the fractional uncertainty is about 0.0006 (1 m/s) throughout the bandwidth. At the highest concentration, 32 μL/8 mL, the thickness uncertainty dominates throughout most of the bandwidth except near 2.3 MHz where the sample and water velocities were similar. The

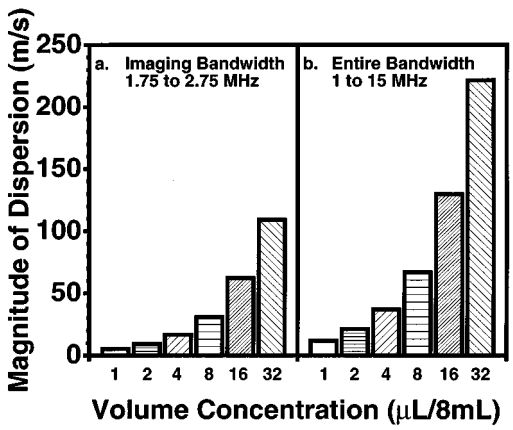


FIG. 5. The magnitude of the dispersion (difference between the maximum and minimum phase velocities) for the various concentrations of Albunex over (a) a limited band of frequencies between 1.75 and 2.75 MHz and (b) the entire bandwidth (1–15 MHz). A straight line through the origin (i.e., $y=m\cdot x$) was fit to both sets of values. For panel (a) $m=3.5$ with $r^2=0.99$; for panel (b) $m=7.2$ with $r^2=0.99$.

maximum fractional uncertainty for the 32 μL/8 mL case is 0.0083 (11 m/s) at the lowest frequency of 1.04 MHz. At the velocity maximum the uncertainty is about 0.003 (4 m/s) and falls to near 0.001 (1.5 m/s) at the high-frequency end of the data.

III. DISCUSSION

A. General features of the dispersion curves

The ultrasonic dispersion properties of Albunex suspensions can be understood by considering the compressibility of an encapsulated microbubble in the long wavelength limit. The linear acoustic properties of the Albunex microbubbles are consistent with a linear model developed by de Jong.[9,10] (A separate but related theoretical treatment of Albunex-like systems by Church deals with many of the issues of interest to this work including phase velocity in suspensions of encapsulated bubbles.[11]) The de Jong model modifies the equation of motion for the radial displacement of a bubble surface at long wavelengths by adding two terms which account for the Voigt-like viscoelastic properties of the shell. The specific values for the stiffness and friction constants used in these two additional terms are empirically determined. By linearizing and solving the equation of motion the displacement of the surface is found to behave as a damped harmonic oscillator. The resonant mode is a uniform radial oscillation of the surface and the frequency of this mode is a monotonically decreasing function of the microbubble radius. After accounting for the relevant energy-loss mechanisms, the damping term consists of contributions from the viscosity of the shell, the viscosity of the surrounding liquid, the acoustic radiation loss, and thermal effects. Of these four damping mechanisms, only the shell viscosity and radiative losses are important for the ranges of microbubble sizes (radii between 2 and 5 μm) and ultrasonic frequencies (<10 MHz) relevant to this work. Near the resonant frequencies for each respective size of microbubble under consideration, the shell viscosity term is at least an order of magnitude greater than the other contributions.

The compressibility of the microbubble is defined in terms of its volume V_b and the driving acoustic pressure P:

$$\kappa_b=-\frac{1}{V_b}\frac{dV_b}{dP}. \tag{10}$$

Using the results from the de Jong model, the compressibility can be expressed to first order in the surface displacement as follows:

$$\kappa_b=\frac{3}{\rho_l a^2}\frac{\omega_r^2(a)-\omega^2+i\,\delta(a,\omega)}{(\omega_r^2(a)-\omega^2)^2+\delta^2(a,\omega)}, \tag{11}$$

where ω_r is the frequency of the resonant mode, $\delta(a,\omega)$ is the damping term, ρ_l is the density of the surrounding liquid, and a is the radius of the microbubble at static equilibrium. The Albunex microbubbles are distributed such that approximately 80% of the total gas volume is encapsulated by shells with radii between 2 and 5 μm. Neglecting high-frequency modes of vibration, microbubbles within this range of sizes respond to impulses as underdamped oscillators with small quality factors [$Q(a=2\ \mu\text{m})=0.2$, $Q(a=5\ \mu\text{m})=1.2$]. The approximate condition for critical damping can be expressed in terms of damping term and resonance frequency as $\delta(a_{\text{crit}},\omega_r)=2\omega_r^2$ where a_{crit} is the radius of critically damped bubbles. Thus for the underdamped microbubbles, $\delta(a,\omega_r)<2\omega_r^2$ where $a>a_{\text{crit}}$. The response becomes overdamped as the bubble radius decreases below 2 μm. The overall properties of the Albunex suspensions are largely determined by the larger (2–5 μm) underdamped microbubbles and thus the rest of this treatment applies only to microbubbles in this size range. As stated above, near the resonant frequency of a given microbubble, the damping term is dominated by the contribution due to the shell viscosity. The elastic properties of the shell increase the resonance frequency by a factor of 3 for the 2-μm bubbles and a factor of about 2.5 for 5-μm bubbles over those for unencapsulated bubbles of the same size. When driven by an acoustic source, the microbubbles exhibit local extrema in their compressibilities just below and above their resonance frequency.

The velocity in the suspension can be calculated in terms of its compressibility κ and density ρ using the following relation:

$$c=\frac{1}{\sqrt{\rho\kappa}}. \tag{12}$$

The quantity c is a complex function that is related to the phase velocity c_s and attenuation coefficient α_s of the suspension by

$$c=c_s+i\,\frac{\omega}{\alpha_s}. \tag{13}$$

The compressibility and density for the suspension can be calculated using a law of mixtures,

$$\kappa=(1-\nu)\kappa_l+\nu\kappa_b, \tag{14}$$

$$\rho=(1-\nu)\rho_l+\nu\rho_b, \tag{15}$$

where ν is the volume concentration of encapsulated bubbles, κ_l is the compressibility of the liquid, and ρ_b is the density of the gas in the bubble. To first order in the volume concentration, the density–compressibility product can be written as

$$\rho\kappa = \rho_l\kappa_l + \nu\rho_l\kappa_b\left(1 + \frac{\rho_b\kappa_l}{\rho_l\kappa_b} - 2\frac{\kappa_l}{\kappa_b}\right). \tag{16}$$

The Albunex microbubbles are filled with air and the surrounding liquid is saline and thus we can estimate the sizes of the specific terms in the parentheses of Eq. (16) by considering the static compressibilities of air and water. The compressibility of air is on the order of 10^4 larger than that of water while the static compressibility of the encapsulated bubble is lower than that of the enclosed gas by a factor of about 10, primarily due to the reinforcing protein shell. Based on this consideration, the ratio of the compressibilities κ_l/κ_b is on the order of 10^{-3}. Thus the two ratios in the parentheses of Eq. (16) are estimated to be much less than unity and are thus discarded. Therefore the velocity in the suspension can be written as

$$c = \frac{1}{\sqrt{\rho_l\kappa_l + \nu\rho_l\kappa_b}} \tag{17a}$$

$$= \frac{1}{\sqrt{\rho_l\kappa_l(1 + \nu(\kappa_b/\kappa_l))}} \tag{17b}$$

$$= \frac{c_l}{\sqrt{1 + \nu(\kappa_b/\kappa_l)}}. \tag{17c}$$

Expanding the denominator to first order in the volume concentration,

$$c \cong c_l\left(1 - \frac{\nu}{2}\frac{\kappa_b}{\kappa_l}\right) \tag{18a}$$

$$= c_l\left(1 - \frac{\nu}{2}\frac{\kappa_b^{\text{real}} + i\kappa_b^{\text{imag}}}{\kappa_l}\right). \tag{18b}$$

(This expansion may strictly apply only to the smaller concentrations, given that the static compressibility of the encapsulated bubble is on the order of 10^3 greater than that of water. Combined with the fact that the volume fraction of microbubbles in undiluted Albunex is about 0.05, the term could lead to the expansion term being about 0.8 for the 32 μL/8 mL concentration. However, when the total volume of bubbles is distributed over of a range of sizes, the bubble compressibility averaged over the distribution is considerably smaller and thus the ratio is smaller.

The phase velocity is obtained by taking the real part of the above expression,

$$c_s = \text{Re}[c] = c_l\left(1 - \frac{\nu}{2}\frac{\kappa_b^{\text{real}}}{\kappa_l}\right) \tag{19a}$$

$$= c_l\left(1 - \nu\frac{3c_l^2}{2a^2}\frac{\omega_r^2(a) - \omega^2}{(\omega_r^2(a) - \omega^2)^2 + \delta^2(a,\omega)}\right). \tag{19b}$$

In the low-frequency limit ($\omega \ll \omega_r$), the microbubble compressibility is real and positive. Thus the volume of the bubble is at a minimum when the driving acoustic pressure is maximum, as one would expect under quasi-static conditions. In this region of the spectrum, the microbubbles increase the overall ''in-phase'' compressibility effectively softening the suspension. Above resonance, κ_b^{real} is negative and thus the ''in-phase'' compressibility is decreased and the microbubbles effectively stiffen the suspension. In the high-frequency limit, the compressibility goes to zero and the velocity in the suspension returns to that of the liquid alone. As ω_r is approached from below, κ_b^{real} is maximized just below resonance, is identically zero at resonance, and reaches a minimum just above resonance. Thus the phase velocity is minimum just below resonance, is equal to that of the liquid alone at resonance, and reaches a maximum just above resonance. Note that when κ_b^{real} goes to zero at ω_r, the microbubble compressibility is purely imaginary and is thus only responsible for dissipating and scattering acoustic energy in the suspension.

The maximum in phase velocity is clearly seen near 3.5 MHz in the results (see Figs. 3 and 4). However, the predicted minimum in phase velocity is not clearly resolved in our data. As a result of the microbubbles in Albunex being distributed over a range of sizes, the dispersion curves are smoothed and blunted through the resonance region compared to what one might expect for single-sized inclusions. The size distribution is such that the volume concentration is maximum for microbubbles near 4 μm in radius. Thus one might expect the shape and structure of the dispersion in the suspension to be similar to that of a monodisperse suspension of 4-μm microbubbles. For microbubbles of this size, using our determinations for the parameters of the de Jong model, the resonant frequency is predicted to occur near 2 MHz which is consistent with our results.

The overall magnitude of dispersion is proportional to the volume concentration of microbubbles. This data in both panels of Fig. 5 are consistent with a proportional relationship with concentration as both sets are well described by a straight-line fit through the origin. The model predicts this trend as can be seen using Eq. (19a):

$$\Delta c_s \equiv c_s^{\max} - c_s^{\min} = \nu\frac{c_l}{2}\frac{\kappa_b^{r,\max} - \kappa_b^{r,\min}}{\kappa_l}, \tag{20}$$

where $\kappa_b^{r,\min}$ and $\kappa_b^{r,\max}$ are the minimum and maximum values of the microbubble compressibility within the relevant spectral band. Equation (20) explicitly shows that the magnitude of dispersion in the suspension is proportional to the volume concentration ν when the first order expansion is valid.

Figure 5 displays the total magnitude of dispersion over the entire bandwidth (1–15 MHz) as well as the dispersion over a 1-MHz-wide bandwidth centered about 2.25 MHz. This narrow frequency band has been used in transthoracic cardiac imaging with ultrasound. Approximately half of the total magnitude of dispersion occurs over this narrow bandwidth.

It is not known in detail how the concentrations used in this work translate into what may be clinically relevant *in*

vivo (i.e., detectable in echocardiographic images). An idea of where these concentrations may fall relative to the spectrum of relevant concentrations can be gained by considering a standard dose of Albunex fully diluted in the circulation. The recommended dose for a 150-lb patient is from 5–15 mL of Albunex. Diluting these doses uniformly into 5 liters, an approximate human blood volume, would result in a range of concentrations from 8 μL/8 mL to 24 μL/8 mL. The filtering in the lungs and other mechanisms which remove or destroy the microbubbles before or during their initial entry into the left side of the heart will reduce these values. Even if 90% of the microbubble volume were removed, concentrations near 1 μL/8 mL to 2 μL/8 mL would result. The clinically relevant concentrations are most likely higher than these estimates since the Albunex is not uniformly distributed among the entire blood volume when it first becomes detectable in the left ventricle. More detailed estimates of the concentrations which may be encountered in the left ventricle are as high as 33 parts per thousand or 264 μL/8 mL.[12] Concentration itself is not the only relevant factor in comparing our suspensions to those *in vivo* since the relative size distributions will most likely be different. If the *in vivo* size distributions have a significantly lower fraction of larger microbubbles, the dispersion effects may be lessened for a given concentration.

B. Impact of Albunex on pulse propagation and broadband imaging

The dispersion in the suspensions of Albunex has a noticeable effect on the shape of through-transmitted ultrasonic pulses. To properly assess the influence of these suspensions on the propagation and evolution of broadband pulses, its attenuation properties must be considered in addition to the phase velocity. (The attenuation properties of Albunex over similar concentrations are reported in Marsh *et al.*[6]) The attenuation coefficient has a single peak with its maximum near 2 MHz. This region around 2 MHz is where the maximum slope is occurring in the phase velocity data. For a broadband pulse, one might expect significant broadening in time due to the dispersion. Additionally, however, the pulse can be separated into two parts due to the peak in attenuation. The Albunex acts like a band-reject filter, selectively attenuating away the intermediate frequencies and shifting the phase of the lower- and higher-frequency components. Thus a compact broadband pulse can be split into two packets by propagating through Albunex; one packet being a slower low-frequency segment, and the other a faster higher-frequency segment. Figure 6 shows this effect in some of the captured signals. Frames (a) and (b) of the figure show the waterpath only trace (representing the average of 500 captured waveforms) and its modulation envelope (magnitude of the analytic signal). Frames (c)–(h) show the waveforms (and their associated analytic signal magnitudes) for pulses transmitted through the Albunex filled chamber at concentrations of 32 μL/8 mL, 64 μL/8 mL, and 128 μL/8 mL, respectively. The chamber was 4.1 mm in thickness so these dispersion effects are occurring over a relatively short path length. As the concentration grows, the separation of the pulse into a narrow fast-moving packet and a broad slow-moving packet becomes more evident. The speeds of these two packets are perhaps better described by the group velocities at their respective center frequencies. The group velocity is related to the phase velocity by the following expression:

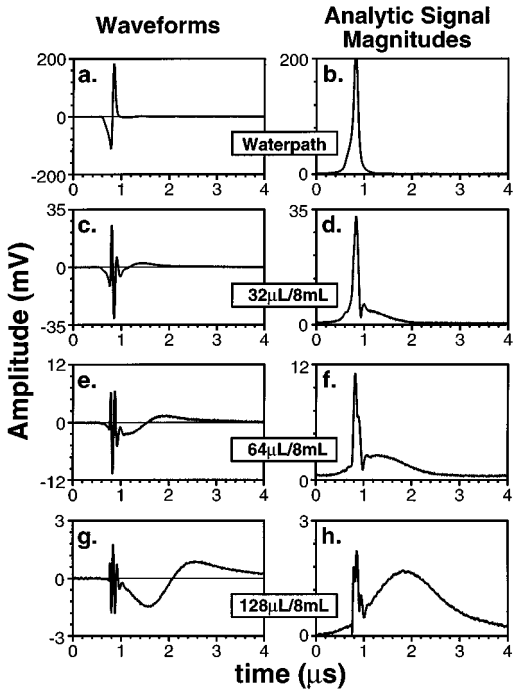


FIG. 6. Captured waveforms and their analytic signal envelopes for the waterpath-only propagation [(a) and (b)] and for propagation through the Albunex dilutions of 32 μL/8 mL [(c) and (d)], 64 μL/8 mL [(e) and (f)] and 128 μL/8 mL [(g) and (h)]. The separation of the signals into a fast high-frequency group and a slower low-frequency group becomes more evident with increasing concentration. The displayed rf waveforms are the average of 500 digitized raw signals.

$$c_{\text{group}} = \frac{c_{\text{phase}}}{1 - \dfrac{\omega}{c_{\text{phase}}} \dfrac{dc_{\text{phase}}}{d\omega}}. \tag{21}$$

Applying this relation to any of the Albunex phase velocity curves shows that the group velocities exhibit the same below-resonance, low-velocity/above-resonance, high-velocity trend. However, unlike the phase velocity, the group velocity has a relatively large maximum near the 2.0-MHz peak of the attenuation coefficient. This high-velocity group is not seen due to the large attenuation in this part of the spectrum.

In an imaging context, the packet separation can manifest itself both in propagation through a cavity containing a blood-contrast mixture and in the refraction across tissue–cavity boundaries. Considering image features that require propagation through a contrast-filled cavity, a sufficiently broadband pulse will be split into two pieces possibly causing a given scattering feature to misregister in the image. The first arriving peak in the backscatter will yield a well-localized image feature while its slower companion may result in a more diffuse signature offset from the first. For transmission across boundaries, the beam could be split into two pieces laterally as well. At non-normal angles of incidence, a narrow beam will be spread apart since a number of

modes in the bandwidth will have unique refraction angles. Again, this spatially broadened beam can be broken into separate pieces due to the strong attenuation of the intermediate frequencies around 2 MHz. Simple modeling of these transmission effects can be done with phase velocities from the 32 μL/8 mL dilution using Snell's law and the ray optics approximation. The values used correspond to the extrema in the phase velocities for the data set. For a plane wave in soft tissue ($c=1540$ m/s) incident at an angle 60° relative to the normal of a boundary between soft tissue and the 32 μL/8 mL Albunex suspension, the 1.0-MHz component ($c=1333$ m/s) would have an angle of refraction of 49° and would diverge from the 3.8-MHz component ($c=1554$ m/s) by 12°. If the two components were narrow beams whose axes roughly followed the plane wave trajectories, then for every 1 cm into the contrast (as measured normal to the boundary) the beams would diverge by 1.2 mm. In terms of separation over a given time interval, consider two points on the planes of constant phase of the respective components which were coincident at the boundary at time $t=0$. After 10 μs, the two would be separated by a total distance of 3.7 mm. Of this total separation 3.5 mm occur parallel to the boundary and 1.2 mm in the direction normal to the boundary. If the ultrasound were incident on the boundary from the contrast side of the interface, the 1.0-MHz component would not cross since 60° exceeds the critical angle for transmission of this component.

The manifestation of the beam-splitting effects can only occur for systems that employ broadband pulses for imaging. These pulses must have sufficient energy at frequencies both above and below the attenuation peak of Albunex for splitting to occur. However, significant refraction effects can still occur in narrow-band channels that use frequencies near 1 MHz.

C. Estimates of diffraction and transmission corrections

A procedure for calculating the diffraction corrections for focused transducer pairs is in development. Our approach involves the modeling of the fields using the angular spectrum technique. To account for the nonplanar shape of the face of the focused transducers, a phase-screen approximation is employed. Preliminary results suggest that the impact of diffraction on these velocity measurements is very small. With the possible exception of the 32 μL/8 mL concentration of Albunex, we estimate the systematic error introduced by neglecting these effects to be on the order of 10^{-2} m/s. For the 32 μL/8 mL Albunex results, the maximum correction is on the order of 10^{-1} m/s. Thus these corrections have a negligible impact on the measured dispersion curves. Diffraction impacts both the through-sample and waterpath pulses in a similar (although not identical) manner. These diffraction effects are minimized by employing the substitution method. For diffraction effects to have an impact in the context of the substitution technique, the sample has to significantly alter the diffraction pattern through the combination of the velocity contrast with the host medium and a sufficiently thick sample. For our system, we believe that the velocity mismatches are too small and the sample chambers are too thin for significant alteration of the diffraction patterns to occur.

As discussed in Sec. I A, transmission effects can contribute phase shifts to the signals. These shifts arise from two effects: (1) the generalization of the single-boundary transmission coefficients to complex quantities when attenuation is present and (2) multiple reverberations in the sample. Both of these transmission-related effects are also estimated to have little or no impact. For the single-boundary transmission coefficients, the ratio of the imaginary part to the real part is less than the ratio of the attenuation coefficient to the wavenumber (for whichever of the two media this ratio is greatest). For almost all measurements, this ratio is small. The exception again is the highest Albunex concentration of 32 μL/8 mL where this ratio is about 0.14 near 2 MHz. However, the estimated velocity correction is still less than 1 m/s and not significant on the scale of the dispersion. As for multiple reverberation effects, the first echo in the signal would arrive at the receiver at least 5 μs after the initial pulse due to the thickness of the sample chamber and velocities in the sample. Thus the reverberant echoes fell outside of our 4.092-μs acquisition window and no multiple reverberation effects were present.

D. Nonlinear effects

Nonlinear effects may occur in Albunex suspensions at pressure levels similar to those used in these experiments. However, we think that the linear response model (i.e., the linear phase spectroscopic approach; the attenuation coefficient and phase velocity as inherent properties of the suspension), although not the complete picture, remains a good approximation to the overall behavior of these suspensions. Thus we believe that the parametrization of the intrinsic properties of these microbubble suspensions in terms of the attenuation coefficient and phase velocity remains appropriate and thus still presents a valid framework for the reduction and interpretation of these data. In earlier work, we were able to get a consistent fit of the attenuation coefficient across concentrations for Albunex suspensions similar to the ones used here.[6] These fits were based on the linear de Jong model where the shell stiffness and viscosity were used as the fitting parameters. Although not reported here, the phase velocity data has also been successfully fit using the linear model as described in the discussion section. Thus the transmission properties of Albunex are consistent with a linear model used to interpret the encapsulated microbubble dynamics. At the same time, for a more complete understanding of Albunex suspensions, nonlinear effects (as well as multiple scattering effects which may be especially important near resonance) must be considered. The investigation of the nonlinear properties of Albunex microbubbles is an active area of research in our laboratory.

IV. CONCLUSION

The broadband phase spectroscopy technique has been applied to the measurement of phase velocities in dilutions of Albunex in Isoton. The dispersion measured for these sus-

pensions grows in proportion with Albunex concentration and may have a detectable impact on imaging when using Albunex during echocardiographic examinations. The dispersion and attenuation of Albunex suspensions can lead to the separation of a beam into low- and high-frequency groups both along the direction of propagation and laterally for pulses with the proper spectral content.

ACKNOWLEDGMENT

We would like to thank Mallinckrodt Inc. for its support of this work.

[1] W. Sachse and Y.-H. Pao, ''On the determination of phase and group velocities of dispersive waves in solids,'' J. Appl. Phys. **49**, 4320–4327 (1978).
[2] R. A. Kline, ''Measurement of attenuation and dispersion using an ultrasonic spectroscopy technique,'' J. Acoust. Soc. Am. **76**, 498–504 (1984).
[3] M. S. Hughes, ''Applications of Local Approximations to the Kramers-Kronig Relations in Ultrasonics,'' Ph.D. thesis, Washington University, 1987.
[4] J. Wu, ''Determination of velocity and attenuation of shear waves using ultrasonic spectroscopy,'' J. Acoust. Soc. Am. **99**, 2871–2875 (1996).
[5] D. Zellouf, Y. Jayet, N. Saint-Pierre, J. Tatibouet, and J. C. Baboux, ''Ultrasonic spectroscopy in polymeric materials. Application of the Kramers–Kronig relations,'' J. Appl. Phys. **80**, 2728–2732 (1996).
[6] J. N. Marsh, C. S. Hall, M. S. Hughes, J. Mobley, J. G. Miller, and G. H. Brandenburger, ''Broadband through-transmission signal loss measurements of Albunex suspensions at concentrations approaching *in vivo* doses,'' J. Acoust. Soc. Am. **101**, 1155–1161 (1997).
[7] C. S. Hall, J. N. Marsh, M. S. Hughes, J. Mobley, K. D. Wallace, J. G. Miller, and G. H. Brandenburger, ''Broadband measurements of the attenuation coefficient and backscatter coefficient for suspensions: A potential calibration tool,'' J. Acoust. Soc. Am. **101**, 1162–1171 (1997).
[8] V. A. Del Grosso and C. W. Mader, ''Speed of sound in pure water,'' J. Acoust. Soc. Am. **52**, 1442–1446 (1972).
[9] N. de Jong, L. Hoff, T. Skotland, and N. Bom, ''Absorption and scatter of encapsulated gas filled microspheres: theoretical considerations and some measurements,'' Ultrasonics **30**(2), 95–103 (1992).
[10] N. de Jong and L. Hoff, ''Ultrasound scattering properties of Albunex microspheres,'' Ultrasonics **31**(3), 175–181 (1993).
[11] C. C. Church, ''The effects of an elastic solid layer on the radial pulsations of gas bubbles,'' J. Acoust. Soc. Am. **97**, 1510–1521 (1995).
[12] M. W. Miller, M. Azadniv, Y. Doida, and A. A. Brayman, ''Effect of a stabilized microbubble contrast agent on CW ultrasound induced red blood cell lysis *in vitro*,'' Echocardiography **12**(1), 1–11 (1995).

Degradation of wren *Troglodytes troglodytes* song: Implications for information transfer and ranging

Jo Holland, Torben Dabelsteen, and Simon Boel Pedersen
Centre for Sound Communication, Zoological Institute, University of Copenhagen, Tagensvej 16, DK-2200 Copenhagen N, Denmark

Ole Næsbye Larsen
Centre for Sound Communication, Institute of Biology, Odense University, Campusvej 55, DK-5230 Odense M, Denmark

(Received 28 August 1997; revised 30 October 1997; accepted 10 December 1997)

The effects of bird song imply a transfer of information between conspecifics. This communication channel is constrained by habitat-induced degradation. Many studies suggest that birds can utilize features of degraded song to assess relative distance to the signaller (ranging). The degradation of transmitted song in the wren *Troglodytes troglodytes* is quantified to assess the opportunities offered in received song for both information transfer and ranging. This quantification incorporates three measurable aspects of degradation: signal-to-noise ratio; excess attenuation; blur ratio. Each aspect varies more-or-less predictably with transmission distance, i.e., a criterion for ranging. Significant effects of speaker and microphone elevation indicate a potential for birds to optimize both the opportunity for information transfer and ranging by considering perch location. Song elements are the smallest units of a song being defined as a continuous trace on a sonagram. Main and second-order effects of element type indicate element-specific patterns of degradation which could be a crucial factor in communication in this species. The element variation within a full song offers the potential for effective information transfer over a range of relevant distances and a variety of transmission pathways. It similarly offers highly flexible ranging opportunities. © *1998 Acoustical Society of America.* [S0001-4966(98)04603-7]

PACS numbers: 43.80.Ev, 43.80.Lb, 43.80.Nd [FD]

INTRODUCTION

The effects of long ranging bird song imply that conspecific receivers extract information from the song concerning the singer's identity, possible ensuing behavior, and location. This can involve the direct transfer of information which has been encoded into the song by aspects of the song features (e.g., Dabelsteen and Pedersen, 1992). The various selective pressures which shape such coding systems differ between species and environments. Song structure is unavoidably altered by the environment during transmission. The effects of this habitat-induced degradation are most serious when complex songs are broadcast in heterogeneous habitats. It has often been argued that selection should have favored long range signals which resist degradation and ensure effective transfer of encoded information over relevant distances, despite habitat differences (e.g., Robisson, 1987; Brémond and Kreutzer, 1986; Brémond and Aubin, 1990; Dabelsteen and Pedersen, 1992, 1993; Dabelsteen *et al.*, 1993; Mathevon *et al.*, 1996). However, information about the singer location can also be extracted from received song. This estimation has two components: distance and direction. While degradation may destroy directional cues (e.g., Michelsen and Rohrseitz, 1997), it seems to provide distance cues. Experiments have shown that birds can discriminate between degraded and nondegraded songs and this has been interpreted as an ability to use degradation cues to estimate the relative distance of the sender (ranging) (e.g., Richards, 1981; McGregor *et al.*, 1983; McGregor and Falls, 1984; McGregor and Krebs, 1984; Shy and Morton, 1986; Naguib, 1995a; Wiley and Godard, 1996). Communication by song therefore denotes possibly paradoxical problems to both senders and receivers. Two functions of singing behavior could be to effectively transfer encoded information despite degradation constraints and yet provide a ''rangeable'' signal which will facilitate an assessment of the relative distance of a singing conspecific using degradation cues. Conversely a receiver could need to both decode signals despite their being degraded and measure the level of degradation. Determining which of these are relevant involves finding out where the selective pressures lie. Unfortunately, the study of degradation has been hindered, first, by the word being used in various ways and, second, by difficulties associated with quantifying it.

We use degradation as a holistic term which incorporates all changes that occur during transmission through the habitat and identify three measurable aspects of degradation: (i) the signal-to-noise ratio (SNR), (ii) the level of attenuation in excess to that predicted by geometric spreading, and (iii) the blur ratio which expresses the distortions of the patterns of amplitude and frequency modulation. These distortions alone are often used synonymously with degradation.

A number of sources of degradation have been identified (e.g., Michelsen, 1978; Wiley and Richards, 1978, 1982). Sound is attenuated through processes of spherical spreading, absorption, scattering by heterogeneities in the atmosphere, turbulence and vegetation, and boundary interference. The latter three processes also result in frequency-

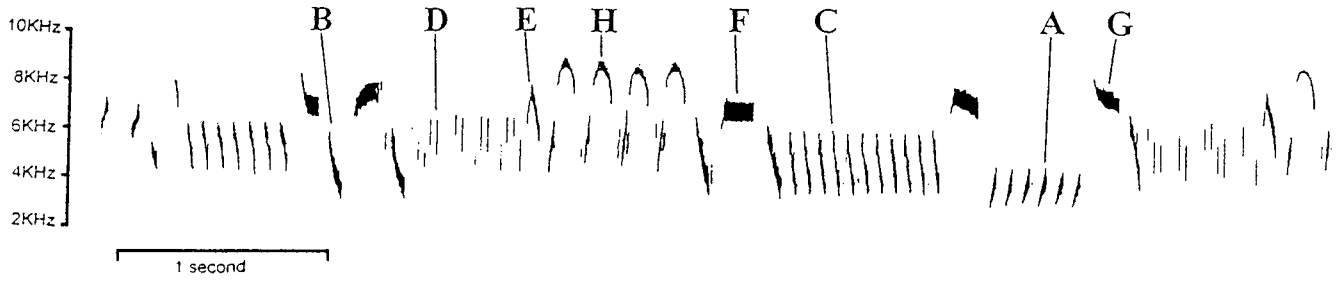


FIG. 1. Sonagram of whole wren song. Letters indicate those elements which were isolated and incorporated into the test series. Elements are lettered by increasing center frequency. Sonagram was created on a Kay Elemetrics Corp. DSP 5500 sonagraph (bandwidth 300 Hz, frequency range 0–16 kHz).

dependent attenuation by effectively filtering out specific frequencies. High frequencies in songs usually attenuate more rapidly than low frequencies (Morton, 1975; Marten and Marler, 1977; Piercy *et al.*, 1977; Wiley and Richards, 1982). Reverberations are a particularly serious consequence of communication in dense vegetation while communication in open habitats is constrained more by irregular amplitude fluctuations caused by atmospheric turbulence (Michelsen, 1978; Wiley and Richards, 1978; Richards and Wiley, 1980; Michelsen and Larsen, 1983; Wiley, 1991).

It has been argued that using patterns of habitat-induced song distortions (i.e., blurring) could offer greater reliability when ranging than attending to the overall song amplitude loss during transmission since they are less prone to irregular fluctuations, caused by unpredictable wind conditions for example (Wiley and Richards, 1978) and less susceptible to differences in the sender's arousal and orientation (Larsen and Dabelsteen, 1990). Birds are capable of ranging using distortion patterns independently of amplitude cues as experiments playing back degraded and undegraded songs were controlled for amplitude differences. Recent experiments indicate that Carolina wrens *Thryothorus ludoviciants* can use frequency-dependent attenuation and reverberation separately as ranging criteria (Naguib, 1995a), and that they can adjust their estimation according to the state of the foliage in a deciduous habitat (Naguib, 1996). Transmission studies have shown that the vertical position of sender and receiver can also influence the way degradation varies with distance (Lemon *et al.*, 1981; Dabelsteen *et al.*, 1993; Mathevon *et al.*, 1996).

This transmission study focuses on the song of the wren *Troglodytes troglodytes* which typically has series of rapidly modulated elements and has a centre frequency of about 6 kHz; features which should be particularly susceptible to degradation (Wiley and Richards, 1982; Mathevon *et al.*, 1996). In addition to this, wrens tend to inhabit acoustically complex environments, preferentially choosing the lowest layers of a deciduous or mixed forest with developed undergrowth (Cramp, 1988). Song perches are usually exposed with elevations often corresponding to the height of the upper part of the undergrowth, although it is not uncommon to see a wren using a perch near ground level (pers. obs.).

We investigate the possibilities offered for information transfer and simultaneous ranging in received song and the effects of perch height and morphology of song elements by quantifying the SNR, excess attenuation, and blurring of transmitted wren song.

I. METHODS

A. Study area

The experiment was performed in the mixed primeval forest at Strødam Biological Field Station, Denmark, early in December 1995. The weather conditions throughout the two experimental days were fairly constant: average (range) for temperature=0.7 °C (−1.2–2.2 °C), wind speed=0.4 m/s (0.1–1.3 m/s), and humidity=94% (83%–98%). At this time the physical conditions simulate those of early spring without the deleterious effects of masking by vocalizing birds. Transmissions were along a 100-m transect through habitat favoured by wrens (dominated by ~30-m-high beech *Fagus sylvatica* and smaller birch *Betula pendula* trees, occasional patches of dense bush and sapling undergrowth, and a ground covering of grass, leaf litter, and fallen trees. The canopy is approximately 10 m from ground level, mean trunk diameter is 0.24 m, and the average inter-tree spacing is 5–7 m).

B. Test sounds

Wren songs were recorded <3 m from an exposed individual in Strødam during spring 1995, using a Sony TCD-D10 PRO DAT recorder and a Sennheiser MKH 816 T directional microphone, and subsequently digitized (sampling frequency 20 kHz). Eight song elements were selected which were representative of the variation in element morphology and which covered the frequency range typically used by wrens (A–H, Fig. 1). The elements were extracted from three different renditions of a digitized song in order to ensure equivalent low levels of background noise for each element. The isolated elements were bandpass filtered using a digital linear phase FIR-filter of order 63 (element: frequency range kHz, center frequency kHz; A: 2.7–4.3, 3.5; B: 3.0–5.8, 4.4; C: 3.6–5.8, 4.7; D: 3.8–7.2, 5.5; E: 5.2–7.4, 6.3; F: 6.0–7.2, 6.6; G: 7.0–8.5, 7.75; H: 6.6–9.0, 7.8) and amplified to a uniform peak amplitude value (amplification range: 1.9–19.4 dB) (for details of the digital filter see Kaiser, 1974). These elements were concatenated to form a combined digital signal with the elements being presented in eight balanced random order sequences. Inter-element pauses

were 600 ms, thus minimizing the risk of forward masking. The test tape was produced by converting the complete series to analogue form and recording it onto a Sony TCD-D 10 PRO DAT. A 2-s period of silence started the test tape. The digitization, filtering, D/A-conversion, and also the later analyses of the transmitted song elements were done on a PC equipped with a Signal Data DSP SPB2 signal processor board, using the customised program DIGSIG ver. 1.2 (Pedersen, 1996).

C. Experiment design and field recordings

Two speaker elevations (3 and 6 m) and three microphone heights (6 m, 3 m, and 20 cm) were chosen as representative song perches and receiver locations, respectively, giving a total of six different transmission pathways for each of the propagation distances: 25, 50, 75, and 100 m. The latter were chosen to span the usual territory dimensions and distances between interacting neighbouring territorial wrens [territory area=0.5 ha (Cramp, 1988)]. The transect choice was restrained by two factors: (i) to avoid grossly uneven ground both along the transect line and surrounding area; (ii) to represent the typical variation in habitat characteristics. To form an impression of the degradation constancy through the habitat, transmissions were performed along both directions of the transect (TA and TB hereafter). A complete factorial design thus yielded a $2_{(\text{TA \& TB})}\times 4_{(\text{distance})}\times 2_{(\text{spk})}\times 3_{(\text{mic})}\times 8_{(\text{element})}$ recording routine. This design was followed with one exception; transmissions at 75 m, TB, were omitted due to a time limitation caused by changes in weather conditions. These recordings constituted the *observation sounds* which were compared with element *model sounds* during data analysis. The model sounds were derived from a recording of the test tape in TA with a transmission distance of 1.56 m and both speaker and microphone heights at 6 m. Under these circumstances the sounds had suffered negligible habitat-induced degradation while the effect of sound distortions caused by the broadcasting system was controlled for. Whenever a transmission was ''masked'' by transient fluctuations, loud noises or bird vocalizations, it was deemed unsuccessful and repeated. Transmissions were conducted over two consecutive days with the model sounds, 25 m and 50 m (TA) being recorded on day 1 and 100 m, 75 m (TA), 25 m, 50 m, and 100 m (TB) on day 2.

The test tape was broadcast from a Sony TCD-D10 PRO DAT recorder through a Vifa D26 NC-05-06 neodymium tweeter assembled with a 1.2-kHz ($f_{-3\text{ dB}}$) high-pass filter (Larsen and Dabelsteen, 1997). The signal to the tweeter was amplified with a Denon DCA-600 power amplifier and adjusted to a level where a broadcast wren song gave a maximum intensity of 65 dB(A) at a distance of 10 m in front of the speaker as measured with a Brüel & Kjær (type 2206) sound pressure level meter (e.g., Dabelsteen, 1981). The frequency response of the playback system was flat in the upper part of the frequency range of the test sounds (5–8.5 kHz) and then gradually reduced by 5 dB from 5 kHz down to 2.7 kHz. The broadcast song elements were re-recorded using a wind-shielded $\frac{1}{2}$-in. microphone (Brüel & Kjær, type 4176 and Brüel & Kjær, type UA 0237 wind screen) with built-in preamplifier and high-pass filter and connected with a 15-m cable to a measuring amplifier (Brüel & Kjær, type 2235). The received signals were then recorded on a Sony TCD-D 10 PRO DAT recorder. The Vifa and the microphone (in horizontal position) were fixed so that they respectively beamed and pointed directly toward each other with the Vifa on the top of a telescopic Clark Mast (type QT 15M/HP) and the microphone in a three-prong clamp which was mounted on the top of another Clark Mast or pushed into the ground.

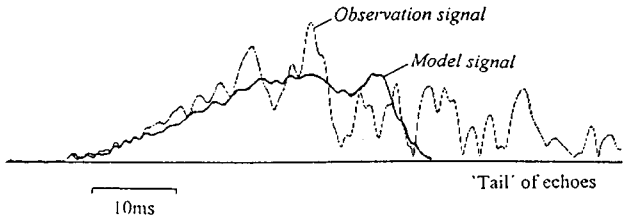


FIG. 2. An example of the alignment of a model and observation amplitude function (envelope) as a result of the cross-correlation process used to measure blur ratio [observation here is element A (TA) at 50 m]. Note that the energies of the model and the observation have been equated (by a factor k_{AF}, see text) and that the tail part of the reverberations is not included in the analysis.

D. Data analysis

The model sound series and each of the observation sound series including their introductory 2-s periods of ''silence'' were filtered (Stanford model SR650, LP 11.5 kHz; HP 2 kHz, 115 dB/octave), amplified when necessary and digitized (sampling frequency 20 kHz). The high upper limiting frequency of the LP filter was chosen to avoid additional attenuation of the high-frequency sounds. This is at the expense of a little aliasing of the noise signal above approximately 8.5 kHz. One second of uncorrupted, stationary (background) noise was sliced from the 2-s ''silence'' at start of each observation series. Out of the eight repetitions constituting each observation series the first five repetitions were selected which were not superimposed by transient noise fluctuations such as wind noise and bird vocalizations as detected through auditory and sonagram inspection. The models were also inspected and a single example nominated for each element on the basis of high cross-correlation factors in a pairwise correlation routine. For each of the elements the element-specific digital filter settings used at the production of the test tape were now applied to the respective model, observations, and background noise periods.

The data analysis involved comparisons between filtered, degraded observation sounds, and their respective undegraded model sounds. The outcome of the method used (for further details, see Dabelsteen *et al.*, 1993) is a *signal-to-noise ratio* (SNR) which is computed from the energy of the observation sound and the energy of the stationary background noise, both over the duration of the model sound, an *excess attenuation* (EA) which is the attenuation of the observation sound in excess of that due to spherical spreading, and a *blur ratio* (BR) which is computed from the amplitude function (AF) of the model sound and the AF of the observation sound (Fig. 2). The amplitude functions are envelopes

of the sound as computed through Hilbert transformations (for details see Dabelsteen and Pedersen, 1985).

The model sounds were of the same duration as the song element and during all cross correlations these were aligned with the start of the observation sound (Fig. 2). This restricted the analysis to the degradation throughout the duration of the element only and disregarded the additional tail part of the echoes that virtually always prolong transmitted sounds and may contain useful extra information to ranging birds (Wiley and Richards, 1982; Naguib, 1995a; Wiley and Godard, 1996).

The energy of the 1-s stationary noise was measured and this was used to estimate the noise component of the energy of the observation sound. This assumes that the stationary noise is the same throughout the observation recording. After that, the signal waveform of model and observation sounds were aligned in time by maximizing the cross correlation between them. In the aligned position, the energy of the observation sound (Ey) was determined and the SNR calculated; $\mathrm{SNR}=10\log[(Ey-En)/En]$. AF's of the model and observation sounds were then created through Hilbert transformations and these were aligned in time by maximizing the cross correlation between them. In the aligned position, the energy of the observation AF (Ey_{AF}) relative to the energy of the model AF was determined as a k_{AF} factor by attenuating the model to the same energy level as the observation. From the knowledge of k_{AF}, EA was calculated; $\mathrm{EA}=-20\log k_{\mathrm{AF}}-A$, where A=attenuation due to spherical spreading. The energy of the difference signal between the AF's of the observation and the k_{AF}-attenuated model was determined and named Ex_{AF}. The blur ratio (BR) is a ratio between the energy of this difference signal and the energy of the observation signal; $\mathrm{BR}=Ex_{\mathrm{AF}}/Ey_{\mathrm{AF}}$.

At relatively low SNRs (<12 dB) the determination of k_{AF} and Ex_{AF} is complicated by the fact that the AF of the noise combines with the AF of the transmitted wren sound in a nonlinear manner. The AF of a noise-corrupted observation sound is not equal to the AF of the sound plus the AF of the noise, resulting in biased values of k_{AF} and Ex_{AF} from the correlation process. The DIGSIG program has a built-in feature which allows compensation for this noise effect where nonbiased k_{AF} and Ex_{AF} values are obtained through a modification of the model AF on the basis of the SNR found. The model AF is changed in a nonlinear fashion over the duration of the model so that small values of the AF are affected most and large values least (Pedersen, 1996). This compensation procedure implies that the true model AF's are not used, but rather slightly modified versions of the model AF's. In practice, however, good SNRs made this effect negligible in nearly all cases (nonexistent in 88% and very small in the others).

Energies of all the signals were determined from the AF's by a summation of the squared AF values. All gains applied to the observation sounds, to enable a prudent analysis, were corrected for.

E. Statistics

TA and TB provided two separate data sets while data within transects were pooled for statistical analysis. There were no missing values (bar the aforementioned TB, 75 m) thus the following measures of SNR, excess attenuation, and blur came from 960 (TA) and 720 (TB) observations. These were subjected to a multifactor ANOVA (4 or $3_{(\mathrm{distance})}\times 2_{(\mathrm{spk})}\times 3_{(\mathrm{mic})}\times 8_{(\mathrm{element})}$) with five replications. Only main and two-factor interactions were considered. As the three measures represent different ''aspects'' of degradation, the significance of correlation coefficients between them were tested. Variations in background noise levels during each transmission were investigated for each element specific filter setting again as a multifactor ANOVA but with one measure per transmission therefore giving 192 (TA) and 144 (TB) observations. Multiple range tests were used to reveal where differences between groups lay. All statistic calculations were conducted with the Statgraphics package, STSC Inc. and all graphs created with Microsoft Excel Ver. 5.0a.

II. RESULTS

A. Variation in background noise

Before comparing levels of degradation in different situations, it is important to determine the variations in the background noise and its likely contribution to the overall variation.

The level of the stationary background noise varied with the different experimental configurations and the different filter settings. Both TA and TB showed significant main effects of microphone location, microphone height, and filter settings (Table I). The main and interaction effects explained 88.5% and 88.7% of the variation in TA and TB, respectively. There was no discernible regular pattern in the way the noise varied with either experimental day or time of day, although stationary noise levels did fluctuate between experiments as indicated by the ''speaker effect'' (TB) where the only logical explanation is an effect of the time interval between transmissions and thus it exposes underlying noise fluctuations. Much of the stationary background noise could be attributed to wind and traffic noise from distant roads around the study area.

In Table II the deviations between the means at each factor level are given and the significant differences shown by ■'s which are not vertically aligned. For example, considering the factor ''distance'' (TA), the lowest noise level was recorded at 100 m. Noise at 75 m was greater by +2.27 dB and not significantly different from that at 100 m or 50 m, whereas noise at 25 m was greater by +6.38 dB and this was significantly greater than the other distances. As expected, the greatest variation was caused by the filter setting factor with the level of background noise generally decreasing with increasing center frequency. Noise generally increased significantly with microphone height with the greatest increase being from 20 cm to 3 m. In both TA and TB it decreased as the microphone was located further along the transect, a variation that reflects the occurrence of nonuniform periods of noise from varying sources during the two experimental days since TB is the reversed TA. A significant interaction effect, particularly in TB, revealed that the effect of microphone height was most apparent when the microphone was in the 25-m location. This implies that levels of noise re-

TABLE I. Factorial ANOVA table for background noise: $4_{(\text{distance})}\times 2_{(\text{spk})}\times 3_{(\text{mic})}\times 8_{(\text{element-specific filter settings})}$(TA); $3_{(\text{distance})}\times 2_{(\text{spk})}\times 3_{(\text{mic})}\times 8_{(\text{element-specific filter settings})}$(TB); main and second-order interaction effects. Here ''distance'' refers to microphone location along the transect and the ''speaker height'' factor represents a time interval between transmissions when only speaker height was changed. Filter settings are the element-specific settings used during data processing (cf. Methods: test sounds).

	Transect A				Transect B			
Factor	*N*	d.f.	*F*-Ratio	*P*	*N*	d.f.	*F*-Ratio	*P*
Main effects:								
''distance''	48	3	138.5	<0.0001	48	2	71.2	<0.0001
''speaker height''	96	1	0.5	0.4944	72	1	11.9	0.0009
microphone height	64	2	13.5	<0.0001	48	2	3.6	0.0306
filter setting	24	7	61.1	<0.0001	18	7	46.3	<0.0001
Two-factor interactions:								
''distance'' by ''speaker height''		3	10.5	<0.0001		2	4.8	0.0101
''distance'' by microphone height		6	3.3	0.005		4	23.1	<0.0001
''distance'' by filter setting		21	0.6	0.8897		14	4.8	<0.0001
''speaker'' by microphone height		2	4.2	0.0173		2	5.3	0.0066
''speaker height'' by filter setting		7	1.3	0.2443		7	0.6	0.7656
microphone height by filter setting		14	0.6	0.8808		14	1.4	0.1867

ceived from a particularly exposed elevation or a location that is nearer to an extraneous source of noise, probably traffic in this case, are likely to affect the SNR more at a high perch elevation.

B. Degradation of the transmitted wren song elements

The three measures—SNR, excess attenuation, and blur ratio—were correlated showing that they measure related aspects of degradation (mutual correlation's between the three aspects of degradation were represented by the Pearson product-moment correlations (r) and a level of significance of $p<0.0001$ for each compared pair of measures. The level of excess attenuation increases as SNR increases (TA: $r=-0.712$; TB: $r=-0.772$). Likewise, at high SNRs the level of blurring was low (TA: $r=-0.439$; TB: $r=-0.635$). At high levels of excess attenuation a greater blurring occurred (TA: $r=0.456$; TB: $r=0.632$).

TABLE II. Deviations (in dB) of the mean from the mean with the lowest noise level are given to show variations in background noise levels for ''distance'' (microphone location along the transect), ''speaker height'' (time interval between transmissions), microphone height, and filter settings. The results of multiple range tests indicate homogeneities between groups within a factor where a nonaligned ■ denotes a statistically significant difference between groups (95% confidence limits using least significant differences). Vertically aligned ■s indicate that the groups are not significantly different whereas a ■ moved one position to the right indicates that this group is significantly different from the other groups within that factor. For example, the microphone height factor shows that the noise level is significantly lower at 0.2 m than at 3 m or 6 m and that there is no significant difference between 3 m and 6 m.

		Transect A		Transect B	
Factor		Dev. (dB)	homogeneous groups	Dev. (dB)	homogeneous groups
''distance'' (m)	100		■		■
	75	+0.27	■	-	-
	50	+0.24	■	+1.45	■
	25	+6.38	■	+3.87	■
''speaker	3		■		■
height'' (m)	6	+0.21	■	+0.89	■
microphone	0.2		■		■
height (m)	3	+1.33	■	+0.71	■
	6	+1.53	■	+0.86	■
element from	G 7.75		■		■
which the	F 6.6	+0.10	■	+0.10	■
element-	H 7.8	+1.6	■	+1.92	■
specific filter	E 6.3	+2.93	■	+2.52	■
settings were	C 4.7	+5.32	■	+4.80	■
designated+	A 3.5	+5.57	■	+4.92	■
center	D 5.5	+6.82	■	+6.03	■
frequency	B 4.4	+7.22	■	+6.42	■

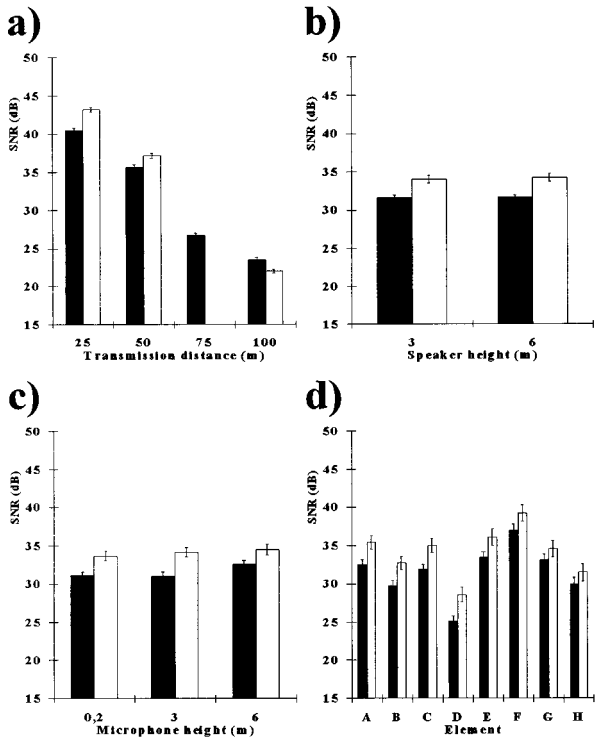


FIG. 3. Signal-to-noise ratio means (±s.e.) of main effect factors: (a) transmission distance; (b) speaker height; (c) microphone height; (d) element type. Black columns=TA and white columns=TB.

1. Signal-to-noise ratio (SNR)

SNR significantly decreased with transmission distance, increased with microphone height, and varied between elements, while there was no effect of speaker height for either TA nor TB (Fig. 3 and Table III). The effect of distance alone explained 67% (TA) and 82% (TB) of the variation in SNR. Element type had a greater effect on SNR than the microphone height (Fig. 3). The increase in SNR with microphone elevation varied for TA and TB. In TA 0.2 m and 3 m did not differ significantly, whereas the SNR at 6 m was significantly higher. In TB there was no difference between 3 m and 6 m and both were significantly greater than 0.2 m. All second-order interactions, except speaker by microphone height (TA), were significant (Table III) and these together with the main effects explained 93.5% (TA) and 96.4% (TB) of the variation. The only second-order interactions which consistently explained a substantial amount of variation in both transects were transmission distance by element type and transmission distance by microphone height. This indicates that the decrease in SNR with transmission distance depends on element type and that this measured decrease will depend on the microphone height. The distance by microphone height second order effect may explain the aforementioned difference between TA and TB in the main order effect of microphone height since there was a stepwise increase in SNR from 0.2 m to 3 m and further up to 6 m at three of the four distances in TA, but this pattern was not seen in TB.

2. Excess attenuation

The level of excess attenuation increased with transmission distance and element center frequency and decreased with speaker and microphone height (Fig. 4). In both TA and TB all of the four main factors significantly affected the excess attenuation (Table IV). Main and second-order interactions explained 91% of the variation in both TA and TB. Transmission distance had the greatest effect [alone explaining 50.5% (TA) and 62.1% (TB) of the variation], followed by element type. There was no significant difference in excess attenuation between transmission distances of 75 m and 100 m in TA (multiple range tests). Figure 5 shows an important second-order interaction between transmission distance and element type where the accumulation of excess attenuation with transmission distance increases with the center frequency of the element. Figure 5 shows this interaction effect in TB and the same tendency can be found in TA.

3. Blur ratio

The blur ratio was also significantly affected by all four main factors (Table V), and, as expected, the pattern of variation followed some of the same general tendencies as did SNR and excess attenuation. It increased with transmission distance, decreased with speaker and microphone height and varied with element (Fig. 6). The results were again

TABLE III. Factorial ANOVA table for signal-to-noise ratio: $4_{(distance)}\times2_{(spk)}\times3_{(mic)}\times8_{(element)}$(TA); $3_{(distance)}\times2_{(spk)}\times3_{(mic)}\times8_{(element)}$(TB); main and second-order interaction effects.

	Transect A				Transect B			
Factor	*N*	d.f.	*F*-Ratio	*P*	*N*	d.f.	*F*-Ratio	*P*
Main effects:								
distance	240	3	3074.7	<0.0001	240	2	7530.9	<0.0001
speaker height	480	1	0.3	0.6078	360	1	2.5	0.1120
microphone height	320	2	48.6	<0.0001	240	2	10.4	<0.0001
element	120	7	299.4	<0.0001	90	7	244.2	<0.0001
Two-factor interactions:								
distance by speaker height		3	101.5	<0.0001		2	8.4	0.0002
distance by microphone height		6	70.6	<0.0001		4	69.0	<0.0001
distance by element		21	19.5	<0.0001		14	28.7	<0.0001
speaker by microphone height		2	0.7	0.5100		2	17.0	<0.0001
speaker height by element		7	7.8	<0.0001		7	3.3	0.0016
microphone height by element		14	16.2	<0.0001		14	8.1	<0.0001

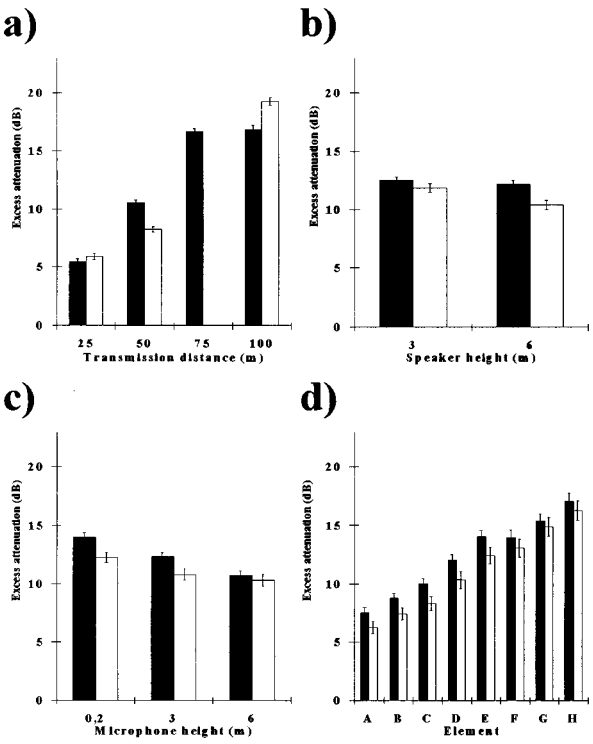


FIG. 4. Excess attenuation means (±s.e.) of main effect factors: (a) transmission distance; (b) speaker height; (c) microphone height; (d) element type. Black columns=TA and white columns=TB.

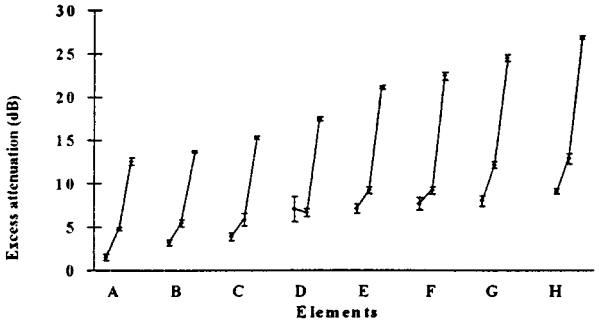


FIG. 5. Levels of excess attenuation showed a significant two-factor interaction effect of transmission distance by element type (TB illustrated here). Each element type has a cluster of three data points (mean ±s.e.) joined by straight lines representing distances 25 m, 50 m, and 100 m in order.

dominated by a main effect of transmission distance followed by element type, microphone height, and speaker height (Fig. 6). However, the main effects plus the second-order interactions only explained 58.3% (TA) and 60.1% (TB) of the total variation. The results of the range test are given in Table VI (constructed in the same way as Table II). Here it can be seen that, although every factor causes the elements to become blurred, the level of blurring at a transmission distance of 50 m is not significantly different from 100 m in TA, however this is not upheld in TB.

Although the overall level of blurring increases with transmission distance, this increase varies considerably between elements as evident in the second order interaction of these two factors. The significant interaction of distance by microphone elevation [Fig. 7(a) and (b)] suggests that a receiving wren has a better potential for ranging at heights of 3 m and 6 m than nearer the ground, although over different distances in the two transects.

4. Understanding the effect of perch elevation with regard to song degradation

It is possible to get an idea of the effect of increasing perch height on degradation by relating the measured difference in the degradation values to the horizontal distance that would have produced a comparable change. For each degradation measure linear regression lines were fitted to both sets of transmission distance data. A preliminary investigation of the data showed that regressions with distance predicted unrealistically high values when extrapolated to the source, whereas variation with log distance offered a more reasonable picture despite slightly lower coefficients of determination. Expressed in logarithmic terms the SNR is reduced by 8.8 dB/dd (dB per doubling of distance) in TA and 10.6 dB/dd in TB, the excess attenuation increases by 6.1 dB/dd (TA) and 6.7 dB/dd (TB), and the blur ratio increases by 0.04/dd (TA) and 0.05/dd (TB). The apparent discrepancy between the values for SNR and excess attenuation can be explained by the deviations in background noise (cf. Table

TABLE IV. Factorial ANOVA table for ecess attenuation: $4_{(distance)}\times2_{(spk)}\times3_{(mic)}\times8_{(element)}$(TA); $3_{(distance)}\times2_{(spk)}\times3_{(mic)}\times8_{(element)}$(TB); main and second-order interaction effects.

	Transect A				Transect B			
Factor	*N*	d.f.	*F*-Ratio	*P*	*N*	d.f.	*F*-Ratio	*P*
Main effects:								
distance	240	3	1165.2	<0.0001	240	2	2425.6	<0.0001
speaker height	480	1	4.7	0.0301	360	1	74.1	<0.0001
microphone height	320	2	203.6	<0.0001	240	2	48.7	<0.0001
element	120	7	311.5	<0.0001	90	7	236.2	<0.0001
Two-factor interactions:								
distance by speaker height		3	11.1	<0.0001		2	25.8	<0.0001
distance by microphone height		6	86.7	<0.0001		4	22.4	<0.0001
distance by element		21	24.6	<0.0001		14	15.8	<0.0001
speaker by microphone height		2	53.5	0.0008		2	0.4	0.6655
speaker height by element		7	3.6	<0.0001		7	1.3	0.2365
microphone height by element		14	14.5	<0.0001		14	7.1	<0.0001

TABLE V. Factorial ANOVA table for blur ratio: $4_{(distance)}\times2_{(spk)}\times3_{(mic)}\times8_{(element)}$(TA); $3_{(distance)}\times2_{(spk)}\times3_{(mic)}\times8_{(element)}$(TB); main and second-order interaction effects.

	Transect A				Transect B			
Factor	*N*	d.f.	*F*-Ratio	*P*	*N*	d.f.	*F*-Ratio	*P*
Main effects:								
distance	240	3	95.5	<0.0001	240	2	272.4	<0.0001
speaker height	480	1	7.0	0.0083	360	1	26.9	<0.0001
microphone height	320	2	68.1	<0.0001	240	2	9.1	<0.0001
element	120	7	37.9	<0.0001	90	7	23.8	<0.0001
Two-factor interactions:								
distance by speaker height		3	12.8	<0.0001		2	14.4	<0.0001
distance by microphone height		6	49.3	<0.0001		4	15.8	<0.0001
distance by element		21	5.2	<0.0001		14	2.1	0.0090
speaker by microphone height		2	32.1	<0.0001		2	1.1	0.3465
speaker height by element		7	2.1	0.0451		7	2.2	0.0309
microphone height by element		14	2.1	0.0080		14	7.5	<0.0001

II) which were particularly serious in TA. These regression lines were then used to translate the difference in degradation at the various speaker and microphone heights into horizontal approaches which would effect a comparable change in degradation. Table VII presents the equivalent horizontal approach distances effected by changes in perch height of wrens separated by 50 m. It suggests that a receiving bird may gain a small advantage in SNR by increasing perch elevation, despite the higher level of background noise there (Table II), whereas the advantage can be large with respect to both excess attenuation and blur ratio, particularly for receivers. By leaving the ground and flying up to a post only 6 m above ground level a receiving wren may sometimes obtain an improvement in blur ratio corresponding to a considerably longer horizontal movement (Table VII). Overall, a receiving wren seems to reduce degradation more by increasing the perch height from 3 to 6 m than does a singing wren since the average approach distances (both transects, all distances) equivalent to such an elevation are 5.8 m (excess attenuation) and 18.4 m (blur ratio) for microphone height and only 4.6 m (excess attenuation) and 13.1 m (blur ratio) for speaker height.

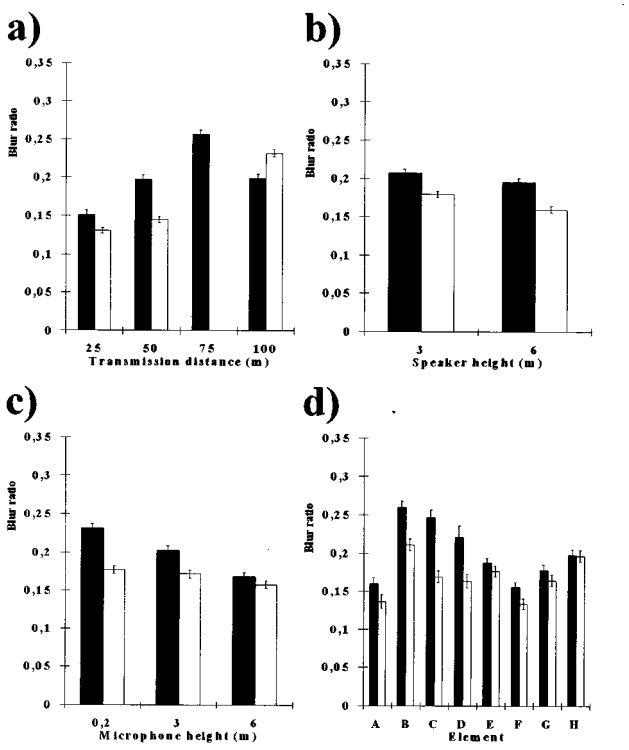


FIG. 6. Blur ratio means (±s.e.) of main effect factors: (a) transmission distance; (b) speaker height; (c) microphone height; (d) element type. Black columns=TA and white columns=TB.

III. DISCUSSION

These results confirm those of Dabelsteen Dabelsteen *et al.* (1993) in showing that it is possible to measure degradation variations in received song elements, within the original time frame of that element. This is true despite the high frequency and short durations of some of the natural wren song elements investigated. Levels of degradation were almost double that of the longer duration, lower-frequency blackbird motif song parts measured by Dabelsteen *et al.* (1993), being more comparable to the high-pitched twitter part of the blackbird song.

A. Degradation as a function of transmission distance

A general increase in the extent of degradation with transmission distance was measured [Figs. 3(a), 4(a), and 6(a)]. Any bird able to perceive these differences could potentially use them as ranging criteria.

In both transects the level of background noise is highest at the 25-m location and lowest at 100 m which, although the total variation is relatively low (<6.5 dB, TA; <4.0 dB, TB), influences the SNR to a greater extent at the smaller transmission distances. Without this variation in background noise the SNR would decrease with a steeper slope than that shown in Fig. 3(a). In reality, however, a successful communication and ranging competence depends on the ability to extract information under the usual noisy conditions of the environment where both noise and signal fluctuate irregularly. Often a song signal will be received having been

TABLE VI. Deviations of the mean from the mean with the lowest blur ratio are given to show variations in blur ratios for transmission distance, speaker height, microphone height, and filter settings. The results of multiple range tests indicate homogeneities between groups within a factor where a nonaligned ■ denotes a statistically significant difference between vertical groups (95% confidence limits using least significant differences). This table is constructed in the same way as Table II.

		Transect A	Transect B
Factor		Dev. Homogeneous groups	Dev. Homogeneous groups
distance (m)	25	■	■
	50	+0.046 ■	+0.014 ■
	75	+0.105 ■	- -
	100	+0.048 ■	+0.101 ■
speaker	6	■	■
height (m)	3	+0.012 ■	+0.020 ■
microphone	6	■	■
height (m)	3	+0.034 ■	+0.014 ■
	0.2	+0.063 ■	+0.019 ■
element	H6.6	■	■
+ center	A3.5	+0.004 ■	+0.003 ■
frequency	G7.75	+0.022 ■	+0.031 ■
	E6.3	+0.031 ■■	+0.043 ■
	H7.8	+0.041 ■	+0.059 ■
	D5.5	+0.065 ■	+0.030 ■
	C4.7	+0.090 ■	+0.036 ■
	B4.4	+0.103 ■	+0.077 ■

masked by conspecific song or songs of other species which utilize similar frequency and temporal parameters. An increase in excess attenuation mirrors the decrease in SNR as the amount of signal energy being absorbed and scattered accumulates with transmission distance and varies with element frequency. The SNRs greatly exceeded that necessary for signal detection and discrimination by birds at all distances (~3 dB) (Brenowitz, 1982; Klump, 1996). In this study, the elements were amplified to various extents to effect a blanket transmission sound pressure level, thus allowing inter-element comparisons (cf. methods: test sounds). However, the natural variation of element sound pressure levels and frequencies means that elements will fall below the detection threshold of the receiver at different distances. It is likely that some of the elements used in this study would not have been detectable had they not been amplified and could therefore represent ''gaps'' in natural received song that had been transmitted over 75 m. These obvious differences however do not suffice to explain a ranging ability in sougs with elements exceeding the critical masking ratio (e.g., Klump, 1996). To date, little is known about the perceptual sensitivity specifically for wrens. However, psychophysical experiments have shown that an intensity difference between two successive sounds must be greater than about 3 dB for a bird to detect the difference (Dooling, 1982). The decreasing SNR and the concerted increase in excess attenuation measured at the different transmission distances exceed 3 dB and should be discriminable to receiving wrens.

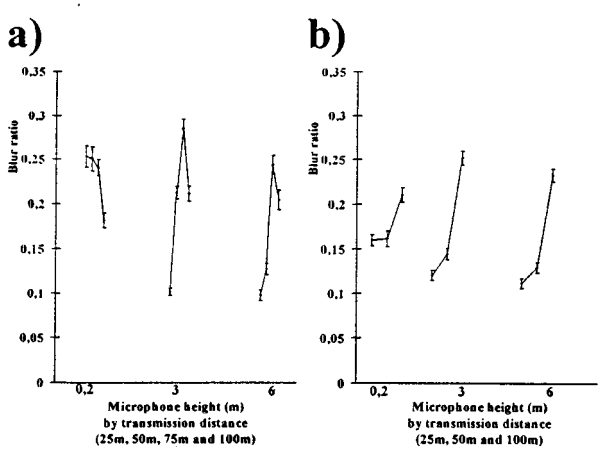


FIG. 7. Levels of blurring showed a significant two-factor interaction effect of transmission distance by microphone height: (a) =TA; (b) =TB. Each microphone height has a cluster of data points (means ±s.e.) joined by straight lines which represent distances 25 m, 50 m, 75 m, and 100 m (TA) and 25 m, 50 m, and 100 m (TB).

The SNR and the level of excess attenuation are covarying aspects of degradation which fulfil the major requirement of ranging criteria: They vary in a relatively predictable way with transmission distance, assuming some knowledge of the habitat type and broadcast intensity. One study has suggested that Carolina wrens can use intensity as a relative cue to deduce the approach or retreat of a conspecific (Naguib, 1995b). However, as mentioned in the Introduction, there is an error associated with using the overall amplitude loss during transmission. This aspect is prone to irregular amplitude fluctuations and manipulation by the sender. Wind gusts along the pathway and the orientation of the sender could severely affect ranging accuracy and could actually conceal the location of a territory intruder.

The intention of measuring blur ratio was to quantify differences in received song elements aside from the total sound intensity loss. The results demonstrate that these do

TABLE VII. SNR, excess attenuation, and blur ratio differences at speaker heights 3 m and 6 m and microphone heights 0.2 m, 3 m, and 6 m were translated into equivalent horizontal transmission range (signaller) or approach distances (receiver) which would effect the same change in the degradation measure. The values presented here indicate improvements measured with a transmission distance of 50 m. The logarithmic relationship with distance means that advantage of any such improvement increases with distance.

Change in elevation (m)	Equivalent horizontal distance (m)					
	SNR		Excess attenuation		Blur ratio	
	TA	TB	TA	TB	TA	TB
Speaker						
3 to 6			−6.9	5.3	60.9	1.0
Microphone						
0.2 to 3	8.8	5.4	27.9	6.3	77.7	9.3
0.2 to 6	11.7	−1.7	34.6	7.8	133.0	15.9
3 to 6	2.8	−7.1	6.8	1.4	55.3	6.6

change considerably during propagation. In general, the changes accumulate with transmission distance, thus fulfilling the requisite for ranging criterion. However, the data show the pattern of blur ratio variation to be more convoluted than the intensity loss previously discussed. In TA the average blur ratio increases steadily to 75 m but drops at 100 m [Fig. 6(a)] to a value that is not significantly different to that recorded at 50 m (Table VI). Since reverberations constitute a substantial part of a signal degraded in a forest habitat and are incorporated in blur ratio measures, reflective conditions local to the microphone may be an important influence on the blur ratio measure of degradation. Therefore, tree positions along and beside the transect were mapped. This revealed that the proximities of reflective surfaces at the 100-m microphone position (TA) and the 50-m position (TB) are lower than at the other positions. This could be part of the explanation for the relatively low blur ratios found at these locations.

The avian perception and discrimination of different levels of blurring has yet to be tested through psychophysical experimentation, but it appears that there could be cues represented in the blur ratio measure which have the potential to be used for ranging given a knowledge of the habitat through which the signal must have travelled. For a highly territorial songbird, such as the wren, an intimate knowledge of his territory does not seem too improbable. An obvious further question arising from this concerns its importance to any receiving wren attending to degradation features similar to the blur ratio. On average, wren territories have an area of 0.5 ha (Cramp, 1988) which crudely translates to radii of 40 m and could possibly suggest that accurate ranging is only crucial to approximately 80 m. Unrefined estimations of distant conspecifics could suffice using degradation features such as SNR and/or excess attenuation, while it may be adaptive to employ an additional, more accurate method when the integrity of the territory is under immediate threat and information about the acoustic properties of the transmission pathway is available. Dabelsteen *et al.* (1993) recorded a similar variation of blur with distance (over 150 m) for long ranging motif parts of the blackbird song. However, here the maximum blur was recorded at 100 m (although a transmission distance of 75 m was not measured) which would correspond to a greater average territory size in this species and could support this idea.

B. Degradation as a function of signaller and receiver elevations

Transmissions for both TA and TB show that an elevation of speaker height from 3 m to 6 m has a negligible effect on the SNR of the received signal [Fig. 3(b)] thus implying no increase in transmission range. The level of excess attenuation is slightly lower at the 6-m speaker location [Fig. 4(b)] presumably because the transmission pathway avoids the denser undergrowth layer to a large extent. Blur ratio is significantly greater when signals are transmitted from a 3-m elevation than at 6-m [Fig. 6(b)]. Interestingly, this suggests that an individual broadcasting from an elevated song post could optimize the transfer of encoded information in the song. In addition, a receiver using degradation cues to judge distance relative to the signaller could arrive at different estimations depending on the signaller's elevation (cf. Table VII for an example at 50 m). Is it possible that established territory holders sing from high perches specifically to issue vocalizations giving good distance estimations when ranged which they use to reiterate their location, while individuals communicating in sexual or conflict contexts might both benefit under the guise of a more degraded song and so use lower song perches?

In all three aspects of degradation the data suggest an advantage to receiving song from elevated perches, assuming that receiving a less degraded song is advantageous. SNR is slightly greater, excess attenuation lower, and blur ratio lower than when receiving from positions nearer the ground [Figs. 3(c), 4(c), and 6(c)]. This can be attributed to less complicated transmission pathways that avoid dense undergrowth and ground effects to an extent and endorses the importance of reflective conditions proximate to the microphone. Wrens increasing their perch height will generally receive songs with more of the original song integrity preserved and thus benefit from improved chances of encoded information transfer. If encoded information is lost in heavily degraded song, then receivers have two options: either to approach the sender to a distance where decoding is possible

or to increase perch elevation to a height where the level of degradation is lower. This strategy could conserve time and energy as the effect of receiver height can be equivalent to considerable approach distances (Table VII) especially when transmission distances are large. Figure 7 shows that the blur ratio varied with distance at microphone heights of 3 and 6 m. At 0.2 m the level of blurring either varied unpredictably with distance (TA) or covered a smaller extent (TB), both of which could make ranging difficult (Fig. 7). In the blackbird motif song transmission study this effect was only apparent at a 9-m microphone elevation, although degradation at the 6-m microphone height was not measured (Dabelsteen *et al.*, 1993). This correlates with a higher average song perch height in this species. The data suggest that, in the wren, the potential for ranging is approximately the same at 3 m and 6 m. Many territorial birds engaged in song interactions sing from high perches and can be followed moving up to these positions at the start of the interaction (pers. obs.). This behaviour changes when a bird intruding the territory proffers a threat necessitating an approach and possible attack. These results confirm a similar analysis with blackbird song (Dabelsteen *et al.*, 1993) in showing that an interacting bird located at 6 m above the ground benefits more, in terms of degradation, when adopting the receiver role than when signalling despite the higher level of background noise experienced at this elevation. This is also in accordance with the finding that territorial wrens use higher perches in response to degraded than undegraded song played back within their territory (Mathevon and Aubin, 1997).

C. Relative degradation of the eight different song elements

The results indicate that the level of degradation varies considerably between elements for all three measured aspects; SNR, excess attenuation, and blur ratio [Figs. 3(d), 4(d), and 6(d)]. This element factor accounts for a level of variation secondary only to that caused by transmission distance. Interestingly, the largest differences in the amount of blurring suffered in the two transects, is shown by elements with relatively low center frequencies [cf. Fig. 6(d) and Table VI]. Second-order interactions of element by transmission distance were significant for each aspect of degradation. The differential degrees of degradation suffered by elements must be an important property of received song containing multielement sequences. To range successfully receivers will either need an appreciation of element specific patterns of degradation expected in the habitat, or should restrict their estimations to using select elements offered in the song.

To investigate this differential degradation further, the measured level of degradation for each specific transmission pathway (every speaker and microphone configuration at each distance) was considered ($N=42$, TA and TB). The element means were ranked (1–8) for each setup with the element presenting as least degraded being assigned 1 with the most degraded being 8. The frequency of the occurrence of each rank value for the elements in the 42 setups allude to a probable pattern of differential degradation experienced by receiving wrens in the forest. Figure 8(a) suggests a dependency of SNR on element duration. The longer elements frequently held low rank positions and thus supports many previous studies where it has been found that long, tonal sounds transmit through habitat with less energy loss than sounds of short duration (e.g., Richards and Wiley, 1980). The total energy of a long received signal will incorporate some echoes from nearby reflective surfaces, whereas echoes from very short signals may appear outside the duration of the signal.

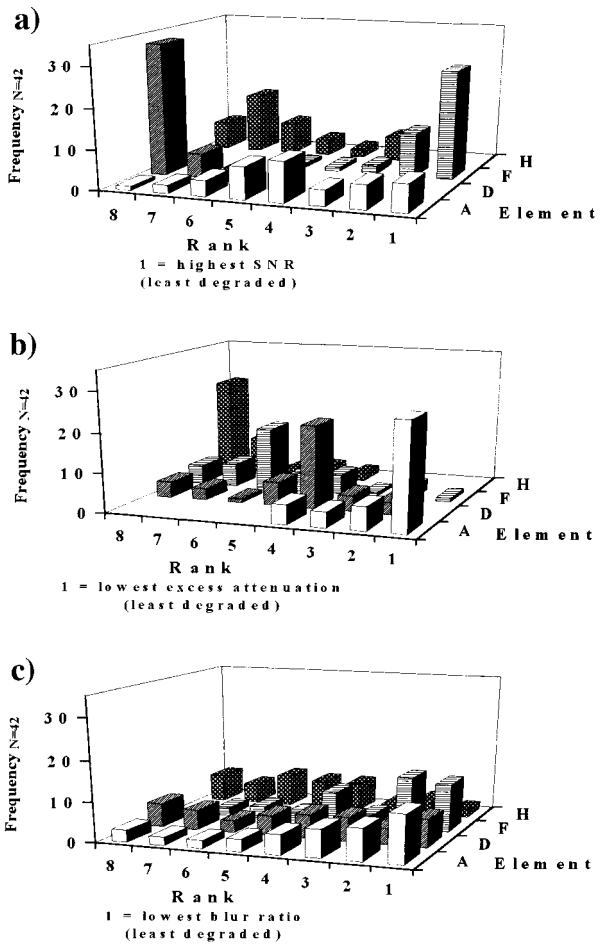


FIG. 8. Element rank frequencies: (a) SNR; (b) excess attenuation; (c) blur ratio. The relative degradation of the elements was considered for each transmission pathway ($N=42$). The element means were ranked (1–8) with the least degraded being assigned 1. These graphs show the frequency of each rank value for four of the elements; the longest element (F), the shortest (D), the lowest frequency (A), and the highest center frequency (H) (cf. Fig. 1 and methods). The other elements held intermediate positions but have been omitted here for presentation reasons.

In addition, Fig. 8(b) illustrates the much documented frequency dependency of the pattern of excess attenuation accumulation with transmission distance (e.g., Morton, 1975; Marten and Marler, 1977; Piercy *et al.*, 1977; Wiley and Richards, 1982; Naguib, 1995a). In contrast to these clear tendencies, Fig. 8(c) presents a confused picture and suggests that different elements suffer the least and most blurring during transmission through the different pathways. For example the least blurred at 25 m need not be the least blurred at 100 m and similarly the most blurred element received at 6-m elevation could differ from that received nearer the ground at the same location. The details of this have still to

be investigated but it remains an exciting possibility that the variation measured by the blur ratio could incorporate such fine detail concerning sender elevation.

Wren full song consists of sequences of elements which differ in duration, frequency and morphology (Fig. 1). In addition, each wren has a repertoire of approximately 20 song types of different sequences and elements (van Horne, 1995). The data suggest that a reason for this could lie in the differential degradation of elements depending on transmission pathway qualities; emitting a varied signal ensures that some elements will always reach receivers within detection range. If a number of elements of various characteristics convey the same encoded information, this redundancy would aid information transfer during long range communication in heterogeneous bioacoustic environments. The signaller also has the potential to exercise a degree of control over the transmission pathway it utilizes. Song post choice and orientation while singing affect the specific pathway and thus the level of degradation suffered.

IV. CONCLUSION

Receiving wrens experience songs that have passed through many different transmission pathways and will each be degraded in a pathway-specific fashion to some extent. Perch height and redundancy in the song are two factors which could aid the transfer of encoded information under these constraints. If a receiver's objective is to range conspecifics by attending to the degradation of a received song, wrens must first estimate the signaller's direction relative to themselves and then call upon knowledge of the acoustic properties of the probable pathway before assessing the level of degradation. This study indicates sufficient variation with transmission distance to range using any one of the three aspects of degradation; SNR, excess attenuation levels, and blur ratio. A striking feature of the results is the differential degradation between elements. A wren song consists of a sequence of elements with an array of characteristics which suffer different extents of habitat-induced degradation during transmission and as a whole could offer opportunities for information transfer and to range despite a heterogeneous forest environment.

ACKNOWLEDGMENTS

We thank Peter McGregor and Christine Holm for comments on the manuscript and Peter McGregor for earlier discussions. This study was funded by the Centre for Sound Communication which is financed by the Danish National Research Foundation.

Brémond, J. C., and Aubin, T. (**1990**). ''Responses to distress calls by black-headed gulls *Laris ridibundus*: The role of non-degraded features,'' Anim. Behav. **39**, 503–511.

Brémond, J. C., and Kreutzer, M. (**1986**). ''Comment le chant du troglodyte (Troglodytes troglodytes) émerge du bruit,'' Behaviour **98**, 361–370.

Brenowitz, E. A. (**1982**). ''The active space of Red-winged Blackbird song,'' J. Comp. Physiol. **147**, 511–522.

Cramp, S. (editor) (**1988**). *Handbook of the birds of Europe, the Middle East and North Africa. The birds of the Western Palearctic Vol. 5. Tyrant Flycatchers to Thrushes* (Oxford U.P., London), pp. 525–542.

Dabelsteen, T. (**1981**). ''The sound pressure level in the dawn song of the blackbird *Turdus merula* and a method for adjusting the level in experimental song to the level in natural song,'' Z. Tierpsychol. **56**, 137–149.

Dabelsteen, T., and Pedersen, S. B. (**1985**). ''A method for computerized modification of certain natural animal sounds for communication study purposes,'' Biol. Cybern. **52**, 399–404.

Dabelsteen, T., and Pedersen, S. B. (**1992**). ''Song features essential for species discrimination and behaviour assessment by male blackbirds (Turdus merula),'' Behaviour **121**, 259–287.

Dabelsteen, T., and Pedersen, S. B. (**1993**). ''Song-based species discrimination and behaviour assessment by female blackbirds (Turdus merula),'' Anim. Behav. **45**, 759–771.

Dabelsteen, T., Larsen, O. N., and Pedersen, S. B. (**1993**). ''Habitat-induced degradation of sound signals: Quantifying the effects of communication sounds and bird location on blur ratio, excess attenuation, and signal-to-noise ratio,'' J. Acoust. Soc. Am. **93**, 2206–2220.

Dooling, R. J. (**1982**). ''Auditory perception in birds,'' in *Acoustic Communication in birds, Vol. 1*, edited by D. E. Kroodsma and E. H. Miller (Academic, New York), pp. 131–181.

Kaiser, J. F. (**1974**). ''Nonrecursive digital filter design using the Io-sinh window function,'' in *Proceedings of the 1974 International Symposium on Circuits and Systems*, edited by S. R. Parker (IEEE, New York), pp. 20–23.

Klump, G. M. (**1996**). ''Bird communication in the noisy world,'' in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. M. Kroodsma and E. H. Miller (Cornell U.P., Ithaca, NY), pp. 321–328.

Larsen, O. N., and Dabelsteen, T. (**1990**). ''Directionality of blackbird vocalization. Implications for vocal communication and its further study,'' Ornis. Scand. **21**, 37–45.

Larsen, O. N., and Dabelsteen, T. (**1997**). ''The l''Vifa Neodymium Tweeter: a versatile speaker for playback experiments,'' Bioacoustics **8**, 323–326.

Lemon, R. E., Struger, M., Lechowicz, M. J., and Norman, R. F. (**1981**). ''Song features and singing heights of American warblers: Maximization or optimization of distance?'' J. Acoust. Soc. Am. **69**, 1169–1176.

Marten, K., and Marler, P. (**1977**). ''Sound transmission and its significance for animal vocalizations,'' Behav. Ecol. Sociobiol. **2**, 271–290.

Mathevon, N., and Aubin, T. (**1997**). ''Reaction to couspecific degraded song by the wren *Troglodytes troglodytes*: Territorial response and choice of song post,'' Behav. Process. **39**, 77–84.

Mathevon, N., Aubin, T., and Dabelsteen, T. (**1996**). ''Song degradation during propagation: Importance of song post for the wren *Troglodytes troglodytes*,'' Ethology **102**, 397–412.

McGregor, P. K., and Falls, J. B. (**1984**). ''The response of Western Meadowlarks (*Sturnella neglecta*) to the playback of undegraded and degraded songs,'' Can. J. Zool. **62**, 2125–2128.

McGregor, P. K., and Krebs, J. R. (**1984**). ''Sound degradation as a distance cue in great tit (*Parus major*) song,'' Behav. Ecol. Sociobiol. **16**, 49–56.

McGregor, P. K., Krebs, J. R., and Ratcliffe, L. M. (**1983**). ''The reaction of great tits (*Parus major*) to playback of degraded and undegraded songs: the effect of familiarity with the stimulus song,'' The Auk **100**, 898–906.

Michelsen, A. (**1978**). ''Sound reception in different environments,'' in *Sensory Ecology*, edited by M. A. Ali (Plenum, New York), pp. 345–373.

Michelsen, A., and Larsen, O. N. (**1983**). ''Strategies for acoustic communication in complex environments,'' in *Neuroethology and Behavioral Physiology*, edited by F. Huber and H. Markl (Springer-Verlag, Berlin), pp. 321–331.

Michelsen, A., and Rohrseitz, K. (**1997**). ''Sound localization in a habitat: an analytical approach to quantifying degradation of directional cues,'' Bioacoustics **7**, 291–313.

Morton, E. S. (**1975**). ''Ecological sources of selection of avian sounds,'' Am. Nat. **109**, 17–34.

Naguib, M. (**1995a**). ''Auditory distance estimation of singing conspecifics in Carolina wrens: The role of reverberation and frequency-dependent attenuation,'' Anim. Behav. **50**, 1297–1307.

Naguib, M. (**1995b**). ''Perception of auditory distance in song-birds and its implications for long-range communications,'' Ph.D. thesis, University of North Carolina, Chapel Hill.

Naguib, M. (**1996**). ''Ranging by song in Carolina wrens *Thryothorus ludovicianus*: Effects of environmental acoustics and strength of song degradation,'' Behaviour **133**, 541–559.

Pedersen, S. B. (**1996**). ''Preliminary operation manual for signal processor

SIGPRO,'' Center for Sound Communication, Odense University, DK.
Piercy, J. E., Embleton, T. F. W., and Sutherland, L. C. (**1977**). ''Review of noise propagation in the atmosphere,'' J. Acoust. Soc. Am. **61**, 1403–1418.
Richards, D. G. (**1981**). ''Estimation of distance of singing conspecifics by the carolina wren,'' The Auk **98**, 127–133.
Richards, D. G., and Wiley, R. H. (**1980**). ''Reverberations and amplitude fluctuations in the propagation of sound in a forest: Implications for animal communication,'' Am. Nat. **115**, 381–399.
Robisson, P. (**1987**). ''L'adaptation des règles de décodage des signaux acoustiques des oiseaux au canal de transmission. Etude appliquée aux cris de detresse du vanneau huppé *Vanellus vanellus*,'' C. R. Acad. Sci. Paris **304**, 275–278.
Shy, E., and Morton, E. S. (**1986**). ''The role of distance, familiarity, and time of day in Carolina wrens responses to conspecific songs,'' Behav. Ecol. Sociobiol. **19**, 393–400.
Van Horne, B. (**1995**). ''Assessing vocal variety in the winter wren, a bird with a complex repertoire,'' Condor **97**, 39–49.
Wiley, R. H. (**1991**). ''Associations of song properties with habitats for territorial oscine birds of eastern North America,'' Am. Nat. **138**, 973–993.
Wiley, R. H., and Godard, R. (**1996**). ''Ranging of conspecific songs by Kentucky warblers and its implications for interactions of territorial males,'' Behaviour **133**, 81–102.
Wiley, R. H., and Richards, D. G. (**1978**). ''Physical constraints on acoustic communication and the atmosphere: Implications for the evolution of animal vocalizations,'' Behav. Ecol. Sociobiol. **3**, 69–94.
Wiley, R. H., and Richards, D. G. (**1982**). ''Adaptations for acoustic communication in birds: Transmission and signal detection,'' in *Acoustic Comminucation in Birds, Vol. 1*, edited by D. E. Kroodsma and E. H. Miller (Academic, New York), pp. 131–181.

Target flutter rate discrimination by bats using frequency-modulated sonar sounds: Behavior and signal processing models

Anne Grossetête
Department of Psychology, Harvard University, Cambridge, Massachusetts 02138

Cynthia F. Moss[a)]
Department of Psychology, University of Maryland, College Park, Maryland 20742

(Received 2 July 1997; revised 20 October 1997; accepted 30 November 1997)

This study utilized psychophysical data and acoustical measurements of sonar echoes from artificial fluttering targets to develop insights to the information used by FM bats to discriminate the wingbeat rate of flying insects. Fluttering targets were produced by rotating blades that moved towards the bat, and the animal learned to discriminate between two rates of movement, a reference rate (30 or 50 Hz) and a slower, variable rate. Threshold discrimination performance depended on the rotation rate of the reference target, with a difference value of 9 Hz for the reference rate of 30 Hz and 14 Hz for the reference rate of 50 Hz. Control experiments demonstrated that the bats used sonar echoes from the moving targets to perform the discrimination task. Acoustical measurements showed that the moving target produced a Doppler shift in the echo and a concomitant change in the arrival time of each frequency in the linear period FM sweep. The difference in delay between echoes from moving and stationary parts varied linearly with flutter rate and depended on the characteristics of the bat's sonar sounds. Simulations also showed a reduction in average echo bandwidth with increasing flutter rate, which may account for a higher delay discrimination threshold using the 50-Hz reference rate. This work suggests that Doppler-induced changes in echo delays produced by fluttering targets may contribute to the FM bat's perception of flying insect prey.
 [S0001-4966(98)04203-9]

PACS numbers: 43.80.Jz, 43.80.Lb [FD]

INTRODUCTION

Echolocating bats orient in the environment and capture prey with the use of an active acoustic imaging system (Griffin, 1958). Bats using frequency-modulated (FM) sonar sounds produce vocalizations which are 1–5 ms long during the approach phase of insect pursuit (Kick and Simmons, 1984; Webster and Brazier, 1965), a duration far too short for a single echo to carry information encoding an insect's wingbeat over a complete cycle (periods ranging between 10 and 50 ms, e.g., Kober and Schnitzler, 1990; Moss and Zagaeski, 1994).

Nonetheless, FM bats can discriminate between artificial wingbeat targets that differ in their rate of flutter. Using an artificial wingbeat simulator, constructed from a rotating blade that was exposed through the window of a small box, Sum and Menne (1988) reported that the FM bat, *Pipistrellus stenopterus*, can discriminate between artificial targets moving at 41 and 50 Hz. In similar studies Roverud *et al.* (1991) and Moss *et al.* (1992) reported that the FM bat *Eptesicus fuscus* can also discriminate between rotating blades that differ in their rate of movement with a threshold close to that reported for *P. stenopterus*.

The low duty cycle of the FM bat's sonar emissions renders a reliable succession of strong echoes from a moving target improbable (Moss and Zagaeski, 1994). Sum and Menne (1988) speculated that the bat based its discrimination on sounds that hit the blade when it was positioned nearly perpendicular to the incident sound. Using a linear frequency model of the bat's sonar sound, they proposed that the bat heard a Doppler-shifted echo from the moving blade, together with an echo from the stationary part of the apparatus. The two surfaces would produce a two-wavefront echo with a time separation between the two components that depended on the velocity of the moving blade. Sum and Menne did not, however, test this hypothesis directly with an analysis of FM echoes taken from the fluttering target apparatus.

This paper presents behavioral data on moving target discrimination, acoustical measurements of the FM echoes from a fluttering target apparatus, and a theoretical analysis of the signals. Instead of a linear frequency model of the bat's sonar sound, we based our analysis on a linear period model. The FM sonar signal echoes returning from a wingbeat simulator showed that the delay of the echoes relative to the transmitted sounds changed linearly with the velocity of a moving target and that the duration of the echo decreased with increasing target velocity. Collectively, the behavioral and acoustical data contribute to a model of fluttering target perception and discrimination in bats that use brief FM sonar signals.

I. BEHAVIORAL STUDY

A. Materials and methods

1. Apparatus

Three big brown bats (*Eptesicus fuscus*) were initially trained to discriminate between two blades rotating at different rates [Fig. 1(A)]. Two animals became ill during the

[a)]Electronic mail: cmoss@bss3.umd.edu

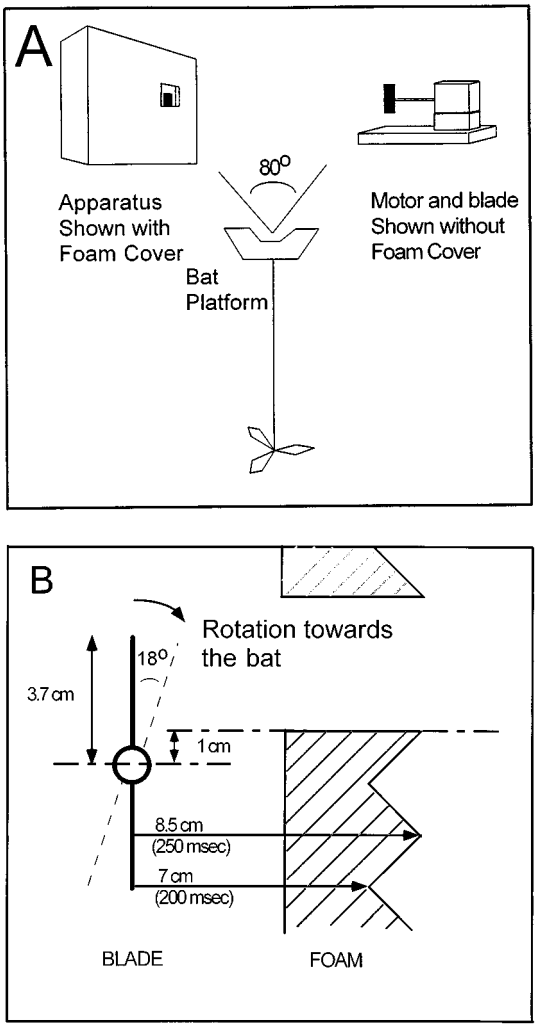


FIG. 1. (A) During behavioral experiments, bats were placed at the base of the Y platform, and were rewarded for crawling towards the rotating blade turning at the reference rate. To prevent passive listening, loudspeakers were placed behind the motors driving the rotating blades broadcasting low-pass filtered white noise which masked motor-generated sounds. (B) Cross section of rotating blade and foam. The rotating blade returned the strongest echo when it was near vertical.

course of training, and only one completed all of the discrimination tasks.

The two rotating rates simulated two different wing beat rates. In the present study, the blades always turned toward the bat. The two wingbeat simulators were positioned at an angular separation of approximately 80 degrees. In each simulator, two 1×3 cm brass blades were mounted on opposite sides of a 13 cm shaft of a dc motor. The shaft of the motor was oriented such that the flat surface of the blade would rotate in a plane perpendicular to the direction of the bat only once each revolution [Fig. 1(A)]. A foam cover was built over the apparatus, with a 2.7×3.0 cm window exposing the upper tip of the blade only. The bat received echoes from the moving blade tip and weaker echoes from the foam covering [Fig. 1(B)]. Acoustic measurements showed that mechanical noise produced by the dc motors driving the rotating blades was broadband, with energy below 16 kHz. To mask this mechanical noise, speakers broadcasting low-pass filtered (at 22 kHz) white noise (Stanford Research Systems function generator) were placed behind the motors.

In the behavioral experiments, the blades were 40 cm away from the bat's observing position on a discrimination platform. When the rotating blade was approximately perpendicular to the incident sound, the bat received an echo from the rotating blade with a delay of about 2.3 ms.

2. Behavioral procedures

The bat was trained to produce sonar vocalizations at the base of a Y-shaped platform and sample echoes from the two moving targets positioned to its left and right (Moss and Schnitzler, 1995). The left and right arms of the Y platform were directed at the two fluttering target simulators. The bat was rewarded for crawling towards the blade appearing at a reference rate of either 30 or 50 Hz. The blade on the other side always appeared at a slower rate. The reference rate was presented on the left or right side following a pseudorandom schedule (Gellerman, 1933) and was maintained at one of the set values of 30 or 50 Hz for several months at a time. The rates reported here refer to the rate of presentation of the blade in the window. The shaft rotation rate was half the presentation rate, since each blade appeared once per rotation.

The mean number of trials per day was 31, with over 20 trials per day on 90% of test days, up to a maximum of 65 trials per day. The mean number of trials per stimulus pair was 153 trials with a range of 18 to 671 trials per stimulus pair. The difference in rotation rates of the two artificial wing-beat stimuli was gradually decreased until the bat's performance fell to chance (50% correct selection of the reference stimulus). Discrimination threshold was arbitrarily defined as the difference value between the reference and comparison rotation rates producing 75% correct performance.

Control sessions in which the openings in the foam were closed with Plexiglas® were run up to a total of 65 trials at 30 Hz and 29 trials at 50 Hz. The bat's frustration limited the number of control trials that could be run on a given test day. Visual cues were still available, but the bat's sounds reflected off the foam and the Plexiglas®, not off the moving blade. If the bat were using passive listening or visual cues associated with motor speed, this information would still be available under these conditions.

3. Echo recordings from fluttering targets

For a half-dozen additional sessions, the bat's sounds were recorded with a microphone placed in front of the platform and approximately 8 cm below the bat, while the echoes were recorded with a microphone placed behind and approximately 8 cm above the bat's head. The sounds were recorded on a high-speed tape recorder at 30 ips (Racal Store 4), and later played back at $\frac{1}{32}$ of the recording speed and digitized at 50 kHz for an effective sample rate of 1.6 MHz on a 12-bit A/D board (Data translation model 2821G-DI). These signals were analyzed and displayed using Signal (Engineering Design) and Matlab (Mathworks) software.

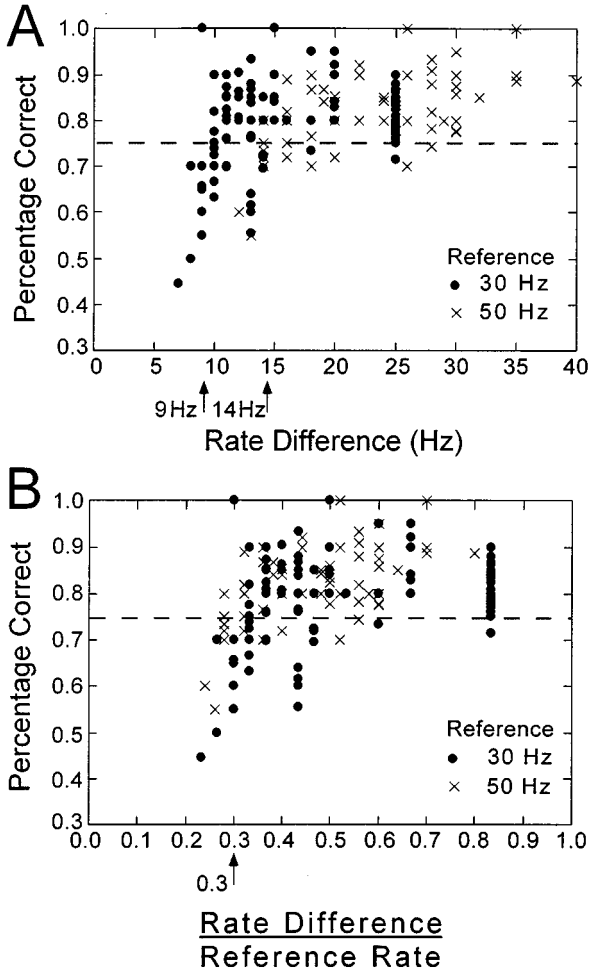


FIG. 2. (A) Each point corresponds to % correct performance for a session of 15–40 trials. Performance threshold of 75% correct occurred for a rate difference of 9 Hz for the 30-Hz reference, and 14 Hz for the 50-Hz reference. (B) Performance threshold of 75% correct occurred for a ratio of rotation rate difference to reference rate of 0.3 for both the 30- and 50-Hz reference rates.

B. Results

1. Behavioral performance

On average, the bat's performance fell to chance as the difference in rotation rate between the two blades was reduced [Fig. 2(A)]. For large differences, the bat's performance was stable at about 90% correct. When the reference rate was 30 Hz, the bat's performance deteriorated to 75% for a rate difference of 9 Hz. When the reference rate was 50 Hz, threshold performance was observed for a difference of 14 Hz.

The performance data was plotted as a ratio of rate difference to reference rate, and the two sets of data for 30 and 50 Hz reference rates overlapped [Fig. 2(B)]. The bat's performance fell to 75% correct when the ratio of the blade rotation rate difference to reference rate was about 0.3 ($\frac{9}{30}=0.3$, $\frac{14}{50}=0.28$).

In control trials with Plexiglas® placed in the opening to the moving blade, the bat's performance dropped to chance (58% and 48% correct performance for 30- and 50-Hz reference rates, respectively). This result confirmed that the bat was using echolocation and not visual discrimination or passive listening to sounds associated with blades rotating at different rates.

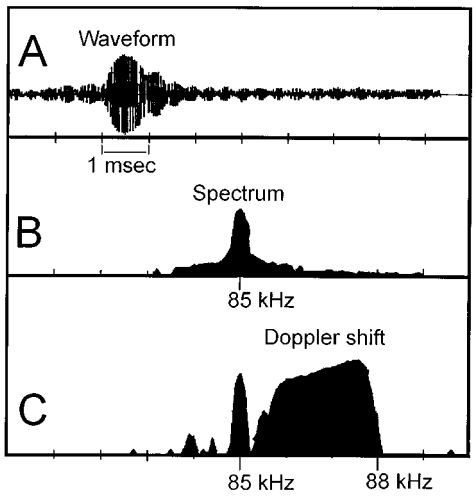


FIG. 3. (A) Echo from the continuous sine wave at 85 kHz. The blade moved towards the source at a presentation rate of 50 Hz. The echo is present for 2 ms out of a blade presentation period of 20 ms (10% of the time). (B) Spectrum of 85-kHz continuous sine-wave broadcast at the moving blade. (C) Spectrum of Doppler-shifted echo from blade rotating at 50 Hz. Energy spreads from 85.5 to 88 kHz, corresponding to the exposed blade from base to tip (1 to 3.7 cm) for this particular test.

2. Analysis of fluttering target echoes

In an attempt to identify the acoustic information the FM bat may have used to discriminate between blades rotating at different rates, we recorded sounds produced by the bat while it was performing the behavioral task, as well as the echoes reflected from the apparatus. When we analyzed these sounds, we found that a small proportion of echoes had a large amplitude. The moving blade only returned echoes to the bat with high signal-to-noise ratio (SNR) when the blade was nearly perpendicular to the incident sound (see Sum and Menne, 1988). When the sound struck the moving blade at a favorable angle, the echo from the moving blade combined with the echo from the stationary foam to form a composite echo with a SNR greater than that from the foam alone. The directionality of the reflecting target was confirmed by broadcasting a continuous sine wave at a rotating blade, and measuring the duration of the discrete signal returned. At 50 Hz, the moving blade returned a detectable echo for only 2 ms out of 20 ms, or 10% of the time (Fig. 3). This indicates that the angular range over which the moving blade returned a detectable echo was only 18 degrees. At these angular positions, the tip of the blade was moving at its maximum speed towards the bat, and the component of the acceleration of the blade was null in that direction. At 50 and 30 Hz, the tip of the blade moves at 6 and 4 mm/ms, respectively.

To examine the echoes from the moving blade in detail, we performed a spectrogram analysis which displayed the changes in spectral power distribution over time. Such an analysis is valuable in the study of FM bats because their echolocation sounds show a characteristic power distribution over time. *Eptesicus*'s sonar sound is composed of several

harmonics. The power in each harmonic sweeps from a high to a lower frequency, and each frequency is present only once per harmonic and only for a very brief instant in time. The most common method of generating spectrograms involves computing a series of spectra at equal time intervals. But this time sampling of the spectrogram may miss the peak power in a particular frequency band. In the present study, the spectrogram was generated by passing the sounds through a bank of digital bandpass filters 1000 points long ($\frac{1}{2}$ bandwidth of 3.2 kHz) and 3 kHz apart (see Appendices A and B). This method allowed us to directly measure the time at which power reached a local maximum in a given frequency band (see Saillant *et al.*, 1993).

We selected for analysis a set of bat sonar sounds recorded during the behavioral experiments that contained echoes from the moving blade. Figure 4(A) and (B) show conventional spectrograms of a sound and its echo. Figure 4(C) plots the frequency band center as a function of peak power times, showing a complex combination of several separable FM sweeps with two harmonics each. The first sweep (1) is the direct microphone pickup of the emitted sound. The second and third sweeps (2) and (3) are the main echoes from the apparatus. The fourth sweep (4) is an echo from unidentified clutter in the room. The time delays from the direct microphone pickup to the second and third sweeps (2.25 and 2.56 ms) were consistent with the echo delays recorded for the distance between the bat platform and the blade (around 2.3 ms for a distance of 40 cm, given the 58 μs/cm two-way travel time for sound in air).

When we plotted the signal period as a function of peak times [Fig. 4(D)], we found the emitted sound (1) could be reliably described by a linear period model (99% of the variance explained by the model). The main echoes (2) and (3) formed a pair of parallel lines, with the same slope as the sonar vocalization, a pattern that would be expected if the emitted sound reflected off two fixed surfaces. Further analyses confirmed that the power reflected from the stationary foam covering the apparatus and the power reflected from the moving surface of the blade appeared as parallel lines when plotted as signal period versus time.

II. THEORETICAL ANALYSIS

Agreement between our observations on the excellent fit of the linear period model and reports in the literature (Altes *et al.*, 1970; Masters *et al.*, 1991), combined with the ease of mathematically manipulating this model, led us to develop the following theoretical framework: Assume a linear period model for the emitted sound's first harmonic:

$$T = aP + b, \tag{1}$$

where T is the time of broadcast of period P, a is the slope, and b the offset. For simplicity, the following equations are limited to the first harmonic (see Appendix C for equations that include two harmonics).

When an emitted sound is reflected by a moving target, the sound's frequency content appears shifted to a fixed observer. This Doppler shift is illustrated in Fig. 3. If a portion of sound is broadcast from the transmitter with period P, it returns with period $P(1-2V/C)$, where V is the velocity of the moving target in the direction of the bat, C is the speed of sound in air, and the ratio (V/C) is small compared to 1. When only a small portion of the blade is exposed, V is the tip's average velocity.

A target moving with speed V towards the source is at a distance VD closer to the source at the offset of a sound of duration D than at the onset. Since the distance between source and target has shortened during the sound, the end of the sound travels $2(V/C)D$ less time than the beginning.

Combining the Doppler shift effect, and the shortening of the travel time (Fig. 5), it follows that the equation for the arrival time of the returning echo is (see Appendix C)

$$T = aP + b + L + 2a(V/C)P_0, \tag{2}$$

where T is the time of reception at a microphone of the signal with period P. The slope and offset a and b have the same values as in Eq. (1); L is a constant delay which is proportional to the distance between source and target at the onset of the sound, and P_0 is the period at onset. The equation for the echo from a moving target can be modeled by $T = aP + b'$. The slope a of the echo spectrogram is the same as the slope of the emitted sound. When the spectrogram is plotted in the period-time domain, the power reflected by the foam and the blade appear as parallel lines.

The change in offset, $b' - b$, is the apparent delay of the echo relative to the emitted sound. It is composed of a fixed delay L and of a variable delay $2\,a(V/C)P_0$. The slope $a = D/(P_D - P_0)$ is a function of the sound duration D, and the onset and offset periods, P_0 and P_D. The variable portion of the delay can be rewritten as $2[D/(P_D - P_0)] \times (V/C)P_0$. If beginning and ending periods P_0 and P_D are fixed, the variable portion of the delay is proportional to the sound duration D, and the target speed V. We thus formulate a testable theoretical prediction. When the upper and lower periods are fixed, the apparent change in delay should be proportional to the product of blade rotation rate by sound duration.

III. ARTIFICIAL SONAR SOUNDS ACOUSTIC STUDIES

A. Materials and methods

1. Apparatus

To obtain a larger acoustic data set with experimentally defined sonar sound parameters, we conducted further studies of echo reflections using computer-generated FM sounds characteristic of those produced by *Eptesicus fuscus* during the approach phase of insect pursuit (Kick and Simmons, 1984). The fluttering target apparatus used to collect acoustic data to test our theoretical prediction was the same as that used in the behavioral experiment. The bat was replaced by a loudspeaker and a microphone (Ultrasound Advice), both with relatively flat frequency responses (± 3 dB) between 20 and 90 kHz. A computer-generated linear period signal (a synthesized FM bat sonar sound) was digitally synthesized (RC electronics D/A board, sample rate 1 MHz) and broadcast through an electrostatic loudspeaker directed at the moving blade. This artificial bat sound was recorded directly from the D/A board at 30 ips on one channel of the Racal Store 4 recorder. The sound reflected off the apparatus was

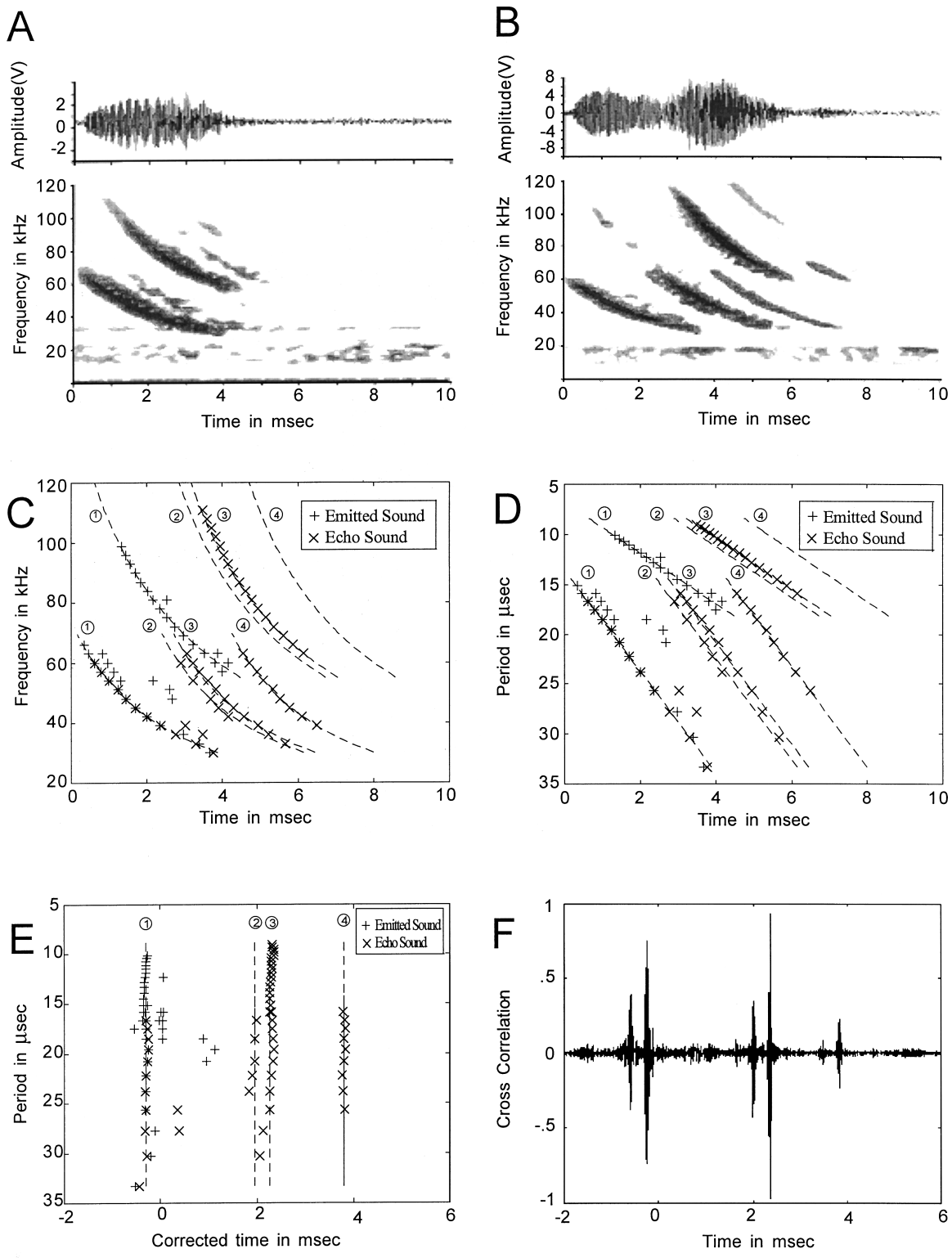


FIG. 4. (A) Waveform and spectrogram of FM sound emitted by bat during behavioral experiment. (B) Waveform and spectrogram of corresponding sonar echoes reflected from the foam and rotating blade. (C) Emitted sound and echoes peak power times in each frequency band, for peaks with power over −20 dB from maximum. Dashes represent linear period fit to data points. (D) Same as (C), but plotted with decreasing period on the *Y* axis. (E) Same as (D), but plotted with corrected time on the *X* axis (see text). (F) Example of cross correlation of emitted sound and echoes.

picked up by the microphone and recorded on a second channel of the Racal recorder at 30 ips. The magnitude of the echo changed with the position of the blade, and the input sensitivity of each Racal recorder channel was set to maximize the amplitude of the recorded signal without overloading.

We generated artificial sounds following a linear period model, and used a sine wave envelope to modulate the sound amplitude. In all cases, the first harmonic frequency swept from 50 to 25 kHz, while the second harmonic swept from 100 to 50 kHz. Recordings were made with 1-ms duration sounds for blade rotation rates of 5, 10, 20, 30, 40, and 50

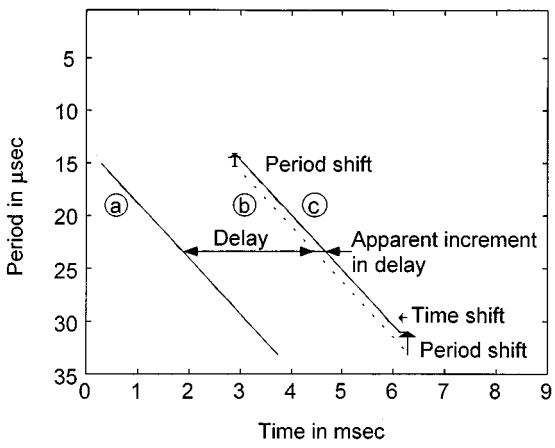


FIG. 5. Effect of a rotating blade, using a linear period model of the sonar signal. (a) Emitted sound. (b) Echo from fixed target. (c) Echo from moving target showing time shift and period shift due to the blade rotation toward the bat.

Hz, moving either towards or away from the speaker–microphone pair. In addition, recordings were made with 3- and 5-ms duration sounds for blade rotation rates of 5, 30, and 50 Hz, both towards and away from the speaker–microphone pair. For each condition, five echo sets were selected for digitization and analysis. The criterion for signal selection was an average echo amplitude of at least 80% of the maximum observed for that condition. This favored capturing sounds with strong echoes from the moving blade.

2. Data analysis

As reported above for the acoustical recordings in the behavioral experiments, each sonar transmission resulted in several overlapping sounds. Our analysis focused on echoes from the stationary foam and the moving blade, and the timing of these echoes was compared with that of the artificial bat sound recorded on a separate Racal channel. The sound data were digitized with a 12-bit Data Translation A/D board sampling at 50 kHz. Playing back the Racal recorder tape at $\frac{1}{32}$ of the original recording speed, we achieved an effective sampling rate of 1.6 MHz, which provides 16 points per 100-kHz cycle. We developed a spectral analysis to estimate the delays of the overlapping echoes from the foam and the moving blade.

The sound data were processed using a bank of digital bandpass filters set 3 kHz apart from 25 to 100 kHz and displayed as spectrograms. The filters were 600 points long for sounds of 1-ms duration (5.3-kHz half-bandwidth), and 1000 points long for sounds of 3- and 5-ms duration (3.2-kHz half-bandwidth; see Appendix B). To increase SNR, we measured the noise level at the beginning of the filter's output and only recorded the time and amplitude of peaks that were four times above that noise level. We started with the analysis of the echo recording. We assumed that in each echo, peak time was linearly related to period with an expression in the form $T=aP+b$, where a was the known quantity $D/(P_D-P_0)$, and P was the inverse of the filter's center frequency. We computed what we call a corrected time by subtracting aP (the time-of-occurrence offset of the period in the sweep) from the peak time. We expected that for each echo (returning from the stationary foam and from the moving blade), corrected time would be equal to the residual b, and would be independent of period. Indeed, plots of period with corrected time show corrected time to be constant with period for each echo. Figure 4(E) shows the relation between period and corrected time for the data of Fig. 4(D).

When we examined plots of peak amplitude versus corrected time, we typically found a pattern of one vertically grouped set of points corresponding to the emitted sound, and two vertically grouped sets of points separated by at least 100 μs corresponding to the reflected sounds from the foam and from the rotating blade. Five 1-ms-long sounds out of 60 were eliminated, because the echoes did not show this pattern (the blade echo was weak or absent). All 3- and 5-ms sounds were retained for further analysis. The timing of each vertical group of points was estimated from the median time of the nine data points with the largest amplitude. A similar analysis was conducted for the broadcast sound. We then subtracted the broadcast sound peak-time median from the two reflected sound peak-time medians, obtaining delay estimates for the foam echo and for the blade echo.

To cross-check our analysis, we created two-wavefront signals with sounds separated by the shortest expected delays, 366 μs for 1-ms sounds, 300 μs for 3-ms sounds, and 230 μs for 5-ms sounds. When the blade was moving away from the bat, the apparent delay decreased compared to the fixed blade delay. Our theoretical prediction leads us to expect the change in delay to be proportional to sound duration, and, therefore, sounds with longer duration should show a larger decrease in apparent delay. Plotting peak times against frequency for different filter lengths, we selected a filter length of 600 points for 1-ms sounds, and determined that the longer 3- and 5-ms sounds required a longer filter length of 1000 points (see Appendix B). We also created two-wavefront sounds with a range of expected delays, from 366–434 μs for 1-ms sounds, 300–500 μs for 3-ms sounds, and 230–570 μs for 5-ms sounds. We observed that for 3- and 5-ms sounds, the analysis showed more spurious peaks for the shorter delays, suggesting difficulties in separating the two-wavefronts near 300- and 230-μs delays, respectively. In general, separation of the two-wavefront was more reliable for delays larger than 400 μs than for smaller delays.

We also validated our analysis by measuring echo delay using a cross-correlation method. We used the broadcast sound to create a matched filter and ran the artificial sonar emissions and the echoes through the matched filter. The echo cross-correlation waveforms showed two separable peaks, corresponding to the echo arrival times from the stationary foam and from the moving blade. We observed that the maximum for the blade echo was always larger than the maximum for the foam echo. We then measured the time delay between the peak of the broadcast sound autocorrelation and the peak of the broadcast sound and moving blade echo cross-correlation [see echo cross correlation in Fig. 4(F)].

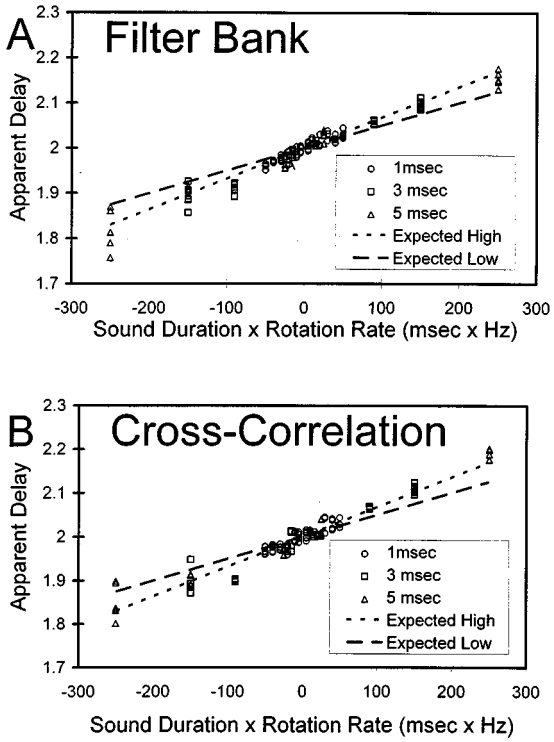


FIG. 6. (A) Delay of rotating blade echo obtained by using a bank of filters and computing the corrected time of each energy peak. (B) Delay of rotating blade echo obtained by detecting the peak of the cross correlation of echo and emitted sounds. Data are shown for 1-, 3-, and 5-ms sonar signals. Dashes indicate low and high range of apparent delay values expected from our theoretical model (see text).

3. Results

As predicted, the relationship between echo delay and the product of rotation rate and sound duration was linear [Fig. 6(A)]. The linear period spectrogram and cross-correlation methods produced very similar echo delay estimates and overall delay distributions [Fig. 6(B)]. The spread of data points was similar in Fig. 6(A) and (B), suggesting the two delay estimation methods employed had a similar accuracy.

For each sound duration, the linear characteristics correspond to what would be expected given the physical layout of the apparatus (Table I). For all three sound durations, the mean delay for the moving blade was 2 ms. This is close to the 2.05-ms delay expected with a distance of 35.5 cm between the loudspeaker and microphone. Given the range of radius of the exposed blade, the estimated slope of the spectrograms was on the order of magnitude expected of 0.5 to 0.7. The experimental slope estimates around 0.7 are near the high end. If we restrict the data to positive rotation rates (blade turning towards the bat), for which the delay difference between the foam and rotating blade is larger, and therefore it is easier to separate the two echoes in the analysis, the 3- and 5-ms sound data yield slope estimates of 0.67 and 0.56, respectively, which are well within the expected range.

TABLE II. Internal delay difference between comparison and reference.

	Sound duration		
	1 ms	3 ms	5 ms
5 vs 30 Hz	15 μs	45 μs	75 μs
21[a] vs 30 Hz	5.4 μs	16.2 μs	27 μs
5 vs 50 Hz	27 μs	81 μs	135 μs
36[a] vs 50 Hz	8.4 μs	25.2 μs	42 μs

[a]Threshold discrimination value from behavioral experiments. Numbers were derived assuming the apparent-delay change$=0.6*D*\phi$, where D is the sound duration in milliseconds and ϕ is the blade's rate of presentation in Hertz. The internal delay difference is $0.6*D*(\phi_{\text{ref}}-\phi_{\text{test}})$.

The internal delay difference between the foam and blade echoes depended on the moving blade's velocity and on the sound's duration. Table II illustrates differences in internal delays for each sound duration, for four stimulus pairs. If the bat was using 1-ms duration sounds, the smallest delay differences it discriminated was about 5.4 and 8.4 μs for the 30- and 50-Hz reference rates, respectively. This internal delay difference is consistent with behavioral data on electronically simulated two-wavefront echo discrimination by FM bats (e.g., Mogdans *et al.*, 1993; Schmidt, 1992). However, if the bat was using 3-ms duration sounds, its internal delay difference discrimination threshold was 16.2 and 25.2 μs for the 30- and 50-Hz reference rates, respectively, and notably larger than previously published two-wavefront discrimination data on FM bats (Mogdans *et al.*, 1993; Schmidt, 1992). Clearly, in order to derive an internal delay

TABLE I. Comparison of theoretical predictions and experimental results for the linear relation between apparent delay, and the product of sound duration and rotation rate.

	Theoretical limits		Least-square fitted parameters		
	Low ($r=2.75$ cm)	High ($r=3.75$ cm)	1 ms	3 ms	5 ms
Mean delay	2.05 ms[a]	2.05 ms[a]	2.0 ms	2.0 ms	2.0 ms
Linear slope	0.5[b]	0.68[b]	0.72 ($\pm$0.04)	0.73 ($\pm$0.03)	0.67 ($\pm$0.03)
% Variance explained			86%	96%	96%

[a]Mean delay$=2\times$distance$/C$, where $C=34$ cm/ms is the speed of sound in air, and the distance between broadcasting speaker and blade is 35.5 cm.

[b]Expected apparent-delay change$=2a(V/C)P_0$. $V=2\pi r(\phi/2)$, where r is the blade radius between 2.75 and 3.75 cm, ϕ is the blade's rate of presentation, and $a=D/(P_D-P_0)$, where $P_D=40$ μs and $P_0=20$ μs. The expected apparent-delay slope is $[2/(P_D-P_0)][2\pi r/(2C)]P_0=0.18r$, where r is expressed in centimeters. The slope coefficient is therefore between 0.5 and 0.68 with an average around 0.6.

difference threshold from the rotation rate difference discrimination threshold one must know the bat's sound characteristics (see Table II), and in our behavioral experiments these sound characteristics varied from trial to trial.

IV. DISCUSSION

The purpose of this study was to arrive at a better understanding of the information used by the FM bat to discriminate moving targets. Behavioral data showed that the FM bat can discriminate a target fluttering at a reference rate of either 30 or 50 Hz from a slower variable-rate moving target. While the bat's absolute discrimination threshold for the 30-Hz reference target was lower than that for the 50-Hz reference target, the ratio of the threshold flutter rate to reference rate was approximately the same (0.3) for the two reference rates.

In our acoustical analyses, we used a linear period model of sonar signals to develop a theoretical model which predicted the change in the apparent delay of the moving target's echo. Artificial sonar sounds were broadcast at a moving blade, and we measured the moving blade echo delay relative to the sound emission. Figure 6 displays the relation between delay, target velocity, and sound duration. As suggested by Sum and Menne (1988), the stationary apparatus and the moving blade produced a two-wavefront echo, with a time separation between the two components that varied linearly with the velocity of the moving blade. In addition, we found that this time separation also depended on sonar sound duration, which is under the bat's control.

The analysis presented in this paper leads us to formulate a hypothesis as to how bats discriminate between blades rotating at different rates. The bat rests on the platform and emits echolocation sounds towards the two moving blades. Most of the time, the blades deflect sounds away from the bat. Approximately once every ten sonar emissions, the bat's sound hits the moving blade when it is roughly vertical, and echoes return to the bat's ears. The bat hears a two-wavefront echo coming from the fixed acoustic foam and the moving blade, and it estimates the time delay of the echo from the blade relative to the echo from the foam (internal delay). The internal delays differ between the reference and comparison rates, and this may provide the foundation for the bat's moving target discrimination.

This can be illustrated with a 3-ms FM sound with the first harmonic sweeping from 50 to 25 kHz. If the left moving blade is turning towards the bat at 5 Hz, the internal delay between blade and foam covering is about 409 μs (400 μs$+0.6\times 3$ ms$\times 5$ Hz, where 400 μs is the round trip internal delay between foam and fixed blade, and 0.6 is a slope coefficient derived from Table I). If the right moving blade is turning at 50 Hz, the internal delay is about 490 μs. The difference between the internal delays of the left two-wavefront echo and the right two-wavefront echo is 81 μs. If on the other hand, the left moving blade is rotating at 36 Hz, the rotation rate at which the bat's performance fell to approximately 75% correct, the internal delay is 465 μs. The difference between the left and right internal delays is then 25 μs (see Table II).

The bat may compare the left and right internal delays, and choose the side with the larger internal delay. Or the bat may learn to recognize the internal delay of the reference stimulus, and use this information to perform the discrimination. The bat could also respond to spectral cues which are directly related to the internal delay difference of the reference stimulus (Mogdans *et al.*, 1993).

The calculated internal delay difference discrimination thresholds reported in this study are generally larger than those reported by Mogdans *et al.* (1993) for *Eptesicus fuscus* and by Schmidt (1992) for *Megaderma lyra*. Both Mogdans and Schmidt studied the bat's discrimination between electronically simulated two-wavefront targets that differed in internal delay. The electronically simulated components of the two-wavefront targets were replicas of the bat's sonar emissions and the two component echoes were equal in energy. In our study, the echo from the foam is complex, with energy smeared over time because the foam is not flat. In addition, the echo from the moving blade is intermittent, and echoes from those two reflecting surfaces (foam and blade) have unequal energy levels. Moreover, the internal delay between the two wavefronts jitters from sound to sound with changes in the bat sound characteristics, such as duration, starting frequency, and ending frequency (see Table II). These factors presumably contributed to the larger difference in internal delays associated with the moving targets at discrimination threshold in the present study.

Our hypothesis that the bat discriminated between rotating blade rates by comparing two-wavefront internal delays does not explain why the bat's discrimination threshold was 14 Hz with the 50-Hz reference rate, and 9 Hz with the 30-Hz reference rate. The reason may be that the rotation rate affects the sounds in ways other than those discussed up to this point. The blade rotating at 5 Hz moves slowly enough to return a complete echo of the bat's sonar emission. At higher rotation rates, the blade is more likely to return a truncated FM echo, which is either missing the beginning of the sound (upper frequencies), or the end of the sound (lower frequencies) [Fig. 7(A)]. Assuming that bats process sounds through a bank of filters, as in our analysis (see Menne, 1988), the bat would receive information from only approximately 50% of the filter banks at 50 Hz, compared to 100% at 5 Hz. The bat's accuracy would therefore be lowered when the reference rate was 50 Hz, compared to 30 Hz. Echoes from the moving blade are more frequent at 50 Hz [Fig. 7(B)], but receiving one echo every five sounds instead of one echo every ten sounds may not help the bat as much as the full use of all frequency channels for the echo analysis.

It is also important to highlight another feature of this study that may have influenced the bat's performance: The target velocity was controlled by the experimenter, but the duration of the sonar sounds was under the bat's control. Longer sonar sounds increased the slope a, and thus increased the apparent delay of the rotating blade echo. This would help the bat to discriminate small velocity differences between the reference and variable rates of target rotation. Indeed, bat Y56, whose performance is reported in this paper, did make unusually long sounds (5–7 ms) on one of the three taped sessions. Bat Y56 did not always use this strat-

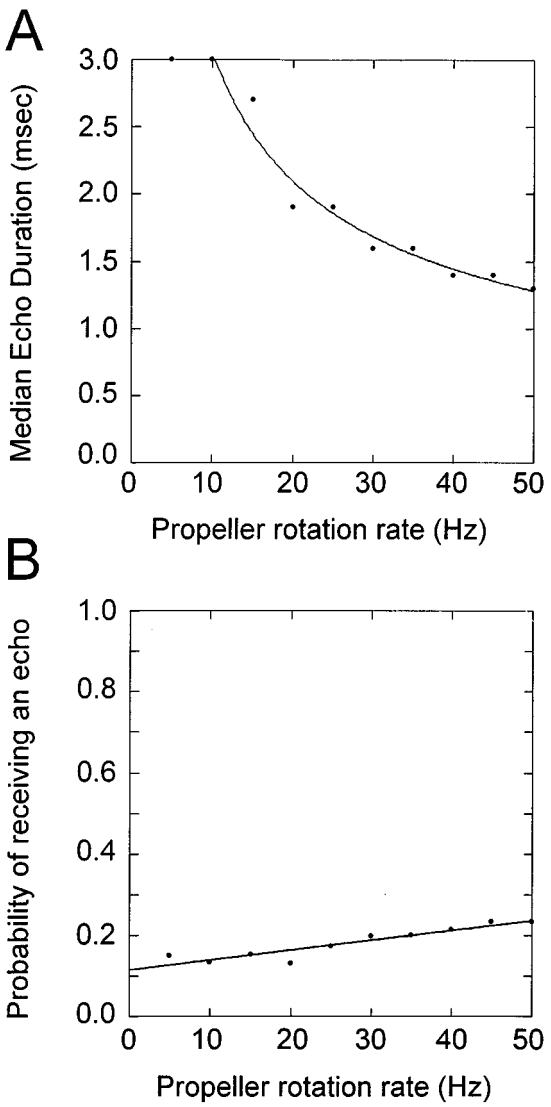


FIG. 7. Results from a simulation with 2000 artificial FM sounds. The sounds were 3 ms long, and emitted on average ten times per second. The simulation assumes the rotating blade reflects sounds over an 18-degree rotation range. (A) Median duration of the echo returned by the blade. (B) Probability of receiving an echo at least 0.1 ms long.

egy, perhaps because longer sounds are more frequently truncated, which may reduce the power of the analysis.

The work presented here raises some important considerations for developing ideas about the FM bat's perception of fluttering insects. In this study, only the tip of the artificial wing was exposed, limiting the velocity of the moving surface towards the bat to a narrow range. In the case of an insect wing, all of the wing reflects the sound towards the bat when it is roughly perpendicular to the impinging sonar sound (Moss and Zagaeski, 1994). The velocity of the moving surface would spread between zero at the body's wing attachment point, to a maximum at the tip of the insect's wing. Instead of a discrete two-wavefront echo, the bat would receive an echo with energy spread over a range of delays. That range would be proportional to the maximum wing velocity and might be fairly stable from echo to echo. In addition, the bat may use other cues, such as interference patterns between the insect's wings and body, to discriminate insect prey. Further studies of moving target discrimination that include echoes from fluttering insects will help to elucidate the acoustic cues that FM bats may use to identify airborne prey items.

ACKNOWLEDGMENTS

We wish to thank Colin Gounden and Doreen Valentine for their assistance in collecting the behavioral data and Itiel Dror for his assistance with echo recordings. We also thank Doreen Valentine, Jim Wadsworth, and Mark Zagaeski for their many contributions to the experimental and theoretical work. This work was supported by the Whitehall Foundation and by an NSF Young Investigator award to CFM.

APPENDIX A: FILTERING METHOD

The method used to synthesize synthesis a single channel's short-term power spectrum amplitude $|X_n(e^{jw}k^n)|$ with a band pass filter centered at the frequency w_k is shown in Fig. A1. The window W has the properties of a low-pass filter (Rabiner and Schafer, 1978).

APPENDIX B: SPECTRAL ANALYSIS

As in a conventional spectrogram computation, time resolution is achieved at the expense of frequency resolution. The length of the filter was chosen to achieve a good compromise between frequency and time resolution.

For this work, we ran simulations to chose an appropriate filter length for each of the three types of sounds used in the artificial data (1, 3, and 5 ms). We constructed a two-wavefront sound, with echo onsets separated by the average expected apparent delay, given the spatial positions of the acoustic foam covering and the blade. We compared the spectrograms obtained for a number of filter lengths in 100-point increments. For the shorter filters we considered, we observed three peaks or more in the filter's output, while for the longer filters considered we observed a single peak. We chose the shortest filter showing two peaks (one for each wavefront) on average in most frequency bands: For 1-ms sounds, the filter length selected was 600 points and for 3- and 5-ms sounds the filter length selected was 1000 points.

For 3- and 5-ms sounds, we observed that the 1000 points filter length did not cleanly separate the two wavefronts offset by delays as short as 300 μs and 250 μs, respectively, suggesting that delay estimates would have a larger error for rotation rates away from the platform than for rotation rates towards the platform.

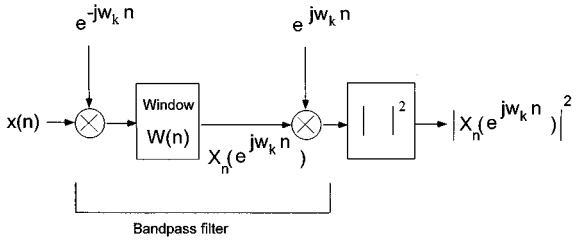


FIG. A1. Filtering method.

APPENDIX C: EQUATIONS

1. Formulation of the linear period model for the emitted sound

One way to formulate the LPM relation between period and time is that the time of emission of a given period is a linear function of period:

$$t_e = a^* P_{e1} + b, \tag{C1}$$

where t_e is the time of emission of a segment with the first harmonic period P_{e1}. Since the first harmonic period is twice the second harmonic period P_{e2},

$$t_e = a^*(2P_{e2}) + b.$$

Therefore, for both first and second harmonics, the relation between time of emission and period emitted can be formulated as

$$t_e = a^* P_{e\_\mathrm{adj}} + b, \tag{C2}$$

where $P_{e\_\mathrm{adj}} = P_{e1}$ for the first harmonic, and $P_{e\_\mathrm{adj}} = 2P_{e2}$ for the second harmonic.

2. Model for the echo returned by a moving target

Assume a sound is reflected by a plate moving with velocity V towards the bat. Each segment of sound arriving at the microphone has been shifted up in frequency by the movement of the plate (Doppler effect). The echo energy centered around the period P actually started as a larger period $P^*(1+2V/C)$, where C is the speed of sound in the air. Each segment of sound also traveled a shorter distance than the segment before. The first segment emitted covered the distance to the target and back to the microphone in L ms. But a segment emitted dt ms later traveled $2(V\,dt)/C$ ms less, where $dt = (t_e - t_{es})$ and t_{es} is the onset segment emission time. A segment emitted at time t_e reaches the microphone at time t_m:

$$t_m = t_e + L - 2(V(t_e - t_{es}))/C. \tag{C3}$$

According to the LPM, $t_e = aP_{e\_\mathrm{adj}} + b$. The period emitted P_{e_adj} can in turned be expressed as a function of the Doppler-shifted period of the echo at the microphone P_{m_adj} by the relation $P_{e\_\mathrm{adj}} = P_{m\_\mathrm{adj}}{}^*(1+2V/C)$.

Equation (C3) can therefore be developed as Eq. (C4) and rearranged as Eq. (C5):

$$t_m = aP_{m\_\mathrm{adj}}(1+2V/C) + b + L - 2(V/C)(aP_{m\_\mathrm{adj}}(1+2V/C) - aP_{es\_\mathrm{adj}}), \tag{C4}$$

$$t_m = a(1-4(V/C)^2)P_{m\_\mathrm{adj}} + (b + L + 2(V/C)aP_{es\_\mathrm{adj}}). \tag{C5}$$

Assuming that the velocity V is small compared to the speed of sound C and that second orders of (V/C) can be neglected, the equation becomes

$$t_m = aP_{m\_\mathrm{adj}} + b + L + 2(V/C)aP_{es\_\mathrm{adj}}. \tag{C6}$$

This equation shows that the relation between time and adjusted period is the same for the echo and the emitted sound, except for a change in time delay. The effect of the moving target is to increase the delay L by a factor which is proportional to the target velocity. The increase in delay is also a function of the emitted sound characteristics, a and P_{es_adj}.

Altes, R. A., and Titlebaum, E. L. (**1970**). ''Bat signals as optimally Doppler tolerant waveforms,'' J. Acoust. Soc. Am. **48**, 1014–1020.

Gellerman, L. W. (**1933**). ''Chance orders of alternating stimuli in visual discrimination experiments,'' J. Gen. Psychol. **42**, 206–208.

Griffin, D. (**1958**). *Listening in the Dark* (Yale U.P., New Haven) (reprinted by Cornell U.P., Ithaca, NY, 1986).

Kick, S. A., and Simmons, J. A. (**1984**). ''Automatic gain control in the bat's sonar receiver and the neuroethology of echolocation,'' J. Neurosci. **4**, 2725–2737.

Kober, R., and Schnitzler, H.-U. (**1990**). ''Information in sonar echoes of fluttering insects available for echolocating bats,'' J. Acoust. Soc. Am. **87**, 882–896.

Masters, W. M., Jacobs, S. C., and Simmons, J. A. (**1991**). ''The structure of echolocation sounds used by the big brown bat *Eptesicus fuscus*: Some consequences for echo processing,'' J. Acoust. Soc. Am. **89**, 1402–1413.

Menne, D. (**1988**). ''A matched filter bank for time delay estimation in bats,'' in *Animal Sonar Processes and Performance*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum, New York), pp. 835–842.

Mogdans, J., Schnitzler, H.-U., and Ostwald, J. (**1993**). ''Discrimination of 2-wavefront echoes by the big brown bat, *Eptesicus fuscus*: behavioral experiments and receiver simulations,'' J. Comp. Physiol. A **172**, 309–323.

Moss, C. F., and Schnitzler, H.-U. (**1995**). ''Behavioral studies of auditory information processing,'' in *Hearing by Bats*, edited by A. N. Popper and R. R. Fay (Springer-Verlag, New York), pp. 87–145.

Moss, C. F., and Zagaeski, M. (**1994**). ''Acoustic information available to bats using frequency-modulated echolocation sounds for the perception of insect prey,'' J. Acoust. Soc. Am. **95**, 2745–2756.

Moss, C. F., Gounden, C., Booms, J., and Roach, J. (**1992**). ''Discrimination of target movement by the FM-bat, *Eptesicus fuscus*,'' Midwinter Meeting of the Association for Research in otolaryngology, St. Petersburg, FL.

Rabiner, L. R., and Schafer, R. W. (**1978**). *Digital Processing of Speech Signals* (Prentice-Hall, Englewood Cliffs, NJ).

Roverud, R. C., Nitsche, V., and Neuweiler, G. (**1991**). ''Discrimination of wingbeat motion by bats correlated with echolocation sound pattern,'' J. Comp. Physiol. A **168**, 259–263.

Saillant, P. A., Simmons, J. A., Dear, S. P., and McMullen, T. A. (**1993**). ''A computational model of echo processing and acoustic imaging in frequency-modulated echolocating bats: The spectrogram correlation and transformation receiver,'' J. Acoust. Soc. Am. **94**, 2691–2712.

Schmidt, S. (**1992**). ''Perception of structured phantom targets in the echolocating bat, *Megaderma lyra*,'' J. Acoust. Soc. Am. **91**, 2203–2223.

Sum, Y. W., and Menne, D. (**1988**). ''Discrimination of Fluttering Targets by the FM-Bat *Pipistrellus stenopterus*,'' J. Comp. Physiol. A **163**, 349–354.

Webster, F. A., and Brazier, O. B. (**1965**). ''Experimental studies on target detection, evaluation and interception by echolocating bats,'' Aerospace Medical Research lab, Wright-Patterson Air Force Base, Ohio, AD 673373.

Broadband spectra of seismic survey air-gun emissions, with reference to dolphin auditory thresholds

John C. Goold
University of Wales Bangor, School of Ocean Sciences, Menai Bridge, Anglesey LL59 5EY, United Kingdom

Peter J. Fish
University of Wales Bangor, School of Electronic Engineering and Computer Systems, Dean Street, Bangor, Gwynedd LL57 1UT, United Kingdom

(Received 19 May 1997; revised 15 September 1997; accepted 17 January 1998)

Acoustic emissions from a 2120 cubic in. air-gun array were recorded through a towed hydrophone assembly during an oil industry 2-D seismic survey off the West Wales Coast of the British Isles. Recorded seismic pulses were sampled, calibrated, and analyzed post-survey to investigate power levels of the pulses in the band 200 Hz–22 kHz at 750-m, 1-km, 2.2-km, and 8-km range from source. At 750-m range from source, seismic pulse power at the 200-Hz end of the spectrum was 140 dB *re*: 1 μPa^2/Hz, and at the 20-kHz end of the spectrum seismic pulse power was 90 dB *re*: 1 μPa^2/Hz. Although the background noise levels of the seismic recordings were far in excess of ambient, due to the proximity of engine, propeller, and flow sources of the ship towing the hydrophone, seismic power dominated the entire recorded bandwidth of 200 Hz–22 kHz at ranges of up to 2 km from the air-gun source. Even at 8-km range seismic power was still clearly in excess of the high background noise levels up to 8 kHz. Acoustic observations of common dolphins during preceding seismic surveys suggest that these animals avoided the immediate vicinity of the air-gun array while firing was in progress, i.e., localized disturbance occurred during seismic surveying. Although a general pattern of localized disturbance is suggested, one specific observation revealed that common dolphins were able to tolerate the seismic pulses at 1-km range from the air-gun array. Given the high broadband seismic pulse power levels across the entire recorded bandwidth, and known auditory thresholds for several dolphin species, we consider such seismic emissions to be clearly audible to dolphins across a bandwidth of tens on kilohertz, and at least out to 8-km range.
 [S0001-4966(98)04004-1]

PACS numbers: 43.80.Ka, 43.80.Nd, 43.30.Nb, 43.50.Lj [FD]

LIST OF SYMBOLS

DAT	Digital Audio Tape
GPS	Global Positioning System
CTD	Conductivity, Temperature, and Depth
HPF	High Pass Filter
2-D	two dimensional

INTRODUCTION

One of the ''noise pollution'' issues of great concern in recent years is that of high levels of manmade noise introduced into the marine environment, and the effects this may have on marine life, especially higher organisms such as marine mammals. Much concern has focused upon the possible effect that increased noise levels may have on cetaceans (whales and dolphins). These concerns range from impairment of cetaceans' ability to communicate acoustically, due to overall increases in background noise levels, to the possibility of physiological damage to sensory and other body organs induced by the proximity of high energy sound sources. Maritime seismic surveying is perceived as one particular threat, the effects of which may span this range of possibilities.

Seismic surveys in the marine environment are generally one of the first stages in the exploration for oil and gas reserves beneath the seabed. Operational modes may vary, but the general methodology involves a large survey ship towing the sound source and receiver equipment. The sound source is generally an array of air guns which produce loud impulsive bursts of underwater sound from the release of compressed air, forming a rapidly expanding and contracting bubble. The bubble collapse causes a transient, of relatively slow rise time compared to chemical explosives, which is geared to the production of primarily low-frequency energy in the subkilohertz region. The characteristics of the sound pulse produced can be modified to a certain extent by the composition and geometry of the gun array. The receiver equipment is generally a long streamer behind the ship (often several kilometers in length) which houses an array of hydrophone elements. Sound energy radiating outward from the gun array propagates downward through the seabed and reflects from discontinuities in the underlying rock strata, such as oil and gas pockets. The reflected sound waves are received by the trailing hydrophone array and the data processed and archived aboard ship, ultimately to be used for mapping the subsurface topography.

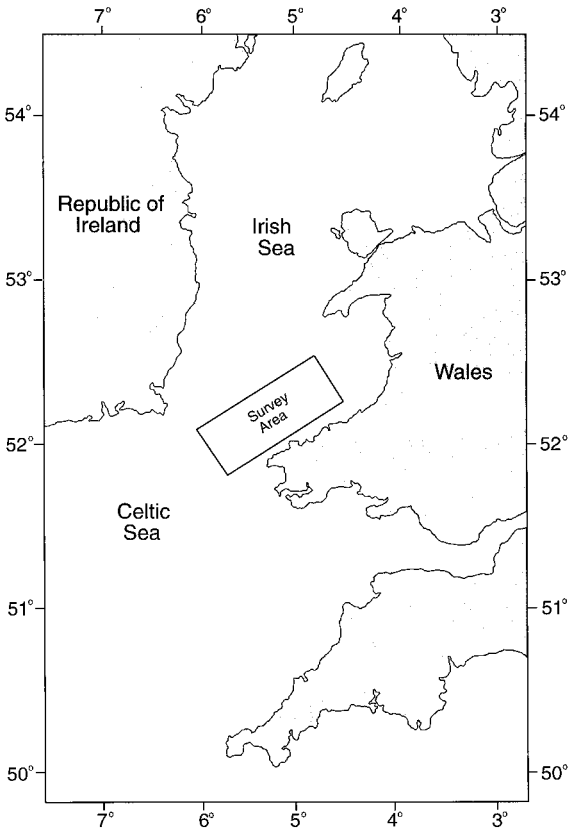


FIG. 1. Chart showing location of seismic/dolphin survey area in relation to the British Coastline. Degrees of latitude and longitude are north of the equator and west of Greenwich meridian, respectively.

Although the output of air gun arrays is usually ''tuned'' to produce a concentration of low-frequency energy, the impulsive nature of the bubble collapse inevitably results in a broadband sound characteristic. In addition there is considerable lateral spreading of sound energy from the gun array into the wider marine environment. Air gun array source levels at the low-frequency end of the spectrum below 100 Hz, are high, typically in the order of 220–240 dB *re*: 1 μPa @1 m (Barger and Hamblen, 1980; Greene and Richardson, 1988; Richardson *et al.*, 1995; Richardson and Wursig, 1997). Seismic pulses are often detectable >100 km from source (R. J. Morris, personal communication; Richardson and Wursig, 1997), and in deep water may be audible >1000 km from source (Richardson and Wursig, 1997). Indeed our own observations during Irish Sea seismic surveys revealed that pulses were audible on hydrophone recordings above the highly elevated background ship noise at least up to 20-km range. High-frequency energy in the seismic pulse spectrum (i.e., above 1 kHz) has generally been ignored in the literature, and in any case is of little interest to the seismic industry, being filtered from survey recordings as ''unwanted clutter.'' In this paper we examine the broadband output from a 2120 cubic in. air-gun array, used at sea during an oil industry 2-D seismic survey off the West Wales Coast of the British Isles in March 1996.

I. METHODS

A. Field recording

Seismic surveying was conducted in the southern Irish Sea off the West Wales Coast (Fig. 1), a shallow shelf area of between 50 and 100 m depth. Recordings of seismic emissions were made from the guard ship, a smaller vessel used in conjunction with the main seismic ship to keep other shipping traffic clear of the long survey streamers. The guard ship generally steamed just ahead of the seismic ship while ''shooting seismic'' was underway. Shooting along a single seismic survey line typically lasted some 2–3 h, with air-gun shots approximately every 10 s; the seismic survey lasted the entire month of March. The guard ship recording apparatus consisted of a 250-m towed stereo hydrophone assembly, with transducer–preamplifier elements positioned 180-m and 230-m aft of the guard ship during tow. The tow depth of the elements was undetermined, but probably within 10 m of the surface and definitely within 20 m as shallow obstacles at 20-m depth were cleared by the tow. Signal output from the rearmost hydrophone element was taken directly to the left channel input of a SONY TCD-D10 DAT recorder, and was used to record seismic pulses flat ± 2.5 dB *re*: 1 μPa in the band 200 Hz–22 kHz. The hydrophone–preamplifier units had integral first order 200-Hz HPF circuits attenuating the input signal below the 200-Hz threshold frequency. Signal output from the forwardmost hydrophone was fed through an additional external HPF stage, with filter break at approximately 6 kHz, and passed to the right channel input of the DAT recorder. The right channel recording (filtered at 6 kHz) was used for manual listening post survey to evaluate dolphin sounds in the area (see Goold, 1996). The left channel recording (filtered at 200 Hz) was used for the analysis of seismic pulses.

The seismic ship was fitted with a differential GPS navigation unit for accurate determination of position to approximately ± 2 m. The guard ship was fitted with nondifferential civilian GPS, and could fix its geographic position to approximately ± 100 m. Both vessels stored absolute time and position continuously from their GPS units on computer disk and these data were used post survey to determine source–receiver ranges. With the air-gun array (geometry illustrated in Fig. 2) towed 90 m behind the large seismic vessel, and the hydrophone receive element towed 230 m behind the much smaller guard ship, GPS center positions for both vessels were used to calculate the distance between sound source and receiver to a precision of ± 100 m.

Although the two vessels generally remained close together for operational reasons, with the guard ship directly ahead of the seismic ship, the occasional opportunity was taken to perform ''walk-away'' tests where the guard ship increased speed and steamed ahead of the seismic ship. This procedure was used to increase the distance between the two ships during the 2–3 h of continuous seismic shooting per survey line. Assessment of received seismic power at extended distance from source was thereby possible, with GPS positions used to determine source–receiver ranges. Seismic pulses analyzed in this work were all recorded at bow aspect

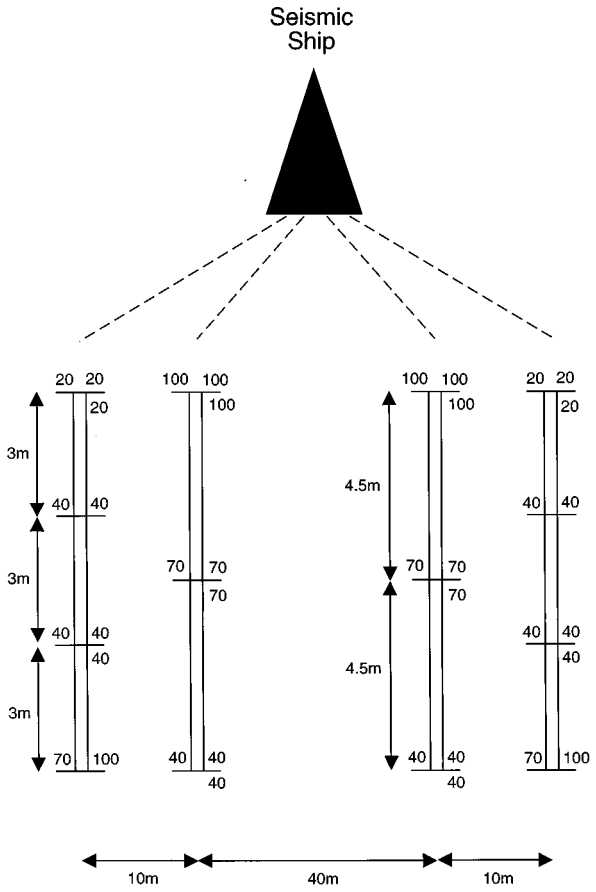


FIG. 2. Geometry of 2120 cubic in. air-gun array used during the seismic surveys. The survey vessel towed four elongated floats, shown as vertical bars, from which 14 gun clusters were suspended to 5 m depth. Numbers along the floats indicate position and capacity (in cubic inches) of each individual sleeve gun, totaling an array capacity of 2120 cubic in. Arrowed lines indicate the geometrical spacing of the various elements in meters.

to the seismic vessel, i.e., with the guard ship steaming directly ahead of, and on the same course as, the seismic ship.

B. Equipment calibrations

1. Hydrophone

The transducers used in the towed hydrophone were Benthos AQ-4 elements, and the preamplifiers were miniaturized transistor units with a gain of 30 dB. The composite hydrophone–preamplifier unit was calibrated against a precise reference transducer at the Sonar and Signal Processing Unit of the Loughborough University of Technology and shown to have a response of -172 dB *re*: 1 V/μPa$\pm$1.5 dB across the frequency range 4 kHz–24 kHz. The result is consistent with Benthos specifications for the AQ-4 element (plus 30-dB preamplifier gain). Although calibration could not be made below 4 kHz due to reverberation within the test tank, Benthos specifications indicate the element to have a flat response at least down to 200 Hz. As the preamplifiers had built in 200-Hz high pass filter circuits, the -172 dB *re*: 1 V/μPa voltage response of the hydrophone was taken to apply across the recorded bandwidth between 200 Hz and 22 kHz.

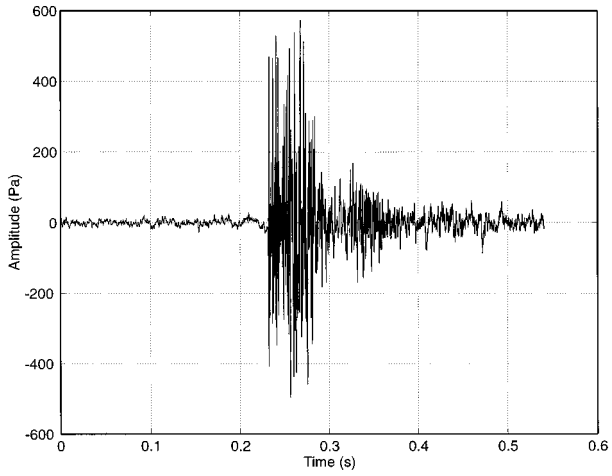


FIG. 3. Waveform trace showing a typical seismic pulse recorded through a 200-Hz High Pass Filter 750 m from the air-gun source at bow aspect.

2. Recording/sampling system

Recorded sounds were sampled from the DAT recorder analog line output onto computer disk using a Cambridge Electronic Design (CED) 1401 laboratory interface. These data were then imported into MATLAB version 4.2c for further analysis. The conversion factor from DAT recorder input to MATLAB data was measured by recording and replaying a series of 10-kHz sinewaves (a frequency in the center of the DAT recorder band) from 1 to 128 mV amplitude, with all recorder gain settings fixed from their fieldwork levels. A linear regression analysis on the MATLAB data versus DAT input voltage yielded the required factor. This factor together with the hydrophone calibration enabled the conversion of MATLAB data into acoustic pressure.

A cursory check was also made of the DAT recorder frequency response, which was specified by the manufacturers to be flat ± 1 dB between 20 Hz and 22 kHz. A synthesized composite signal of eight sinewaves, each of equal amplitude and spanning the DAT recorder frequency range, was recorded through the line input and subsequently replayed and sampled to disk from the line output. The composite signal was imported to MATLAB, rescaled to true voltage using the DAT to MATLAB voltage conversion factor, and subjected to power spectral density analysis through a sliding 2048 point Hanning window. rms amplitude values from integration around the eight frequency peaks were found to be within ± 0.4 dB of the rms sinewave amplitudes, and the manufacturers specification of ± 1 dB was therefore considered adequate as a measure of frequency response flatness. The overall spectral analysis of seismic data, taking into account the hydrophone response, is considered to be flat ± 2.5 dB.

C. Seismic analysis

Seismic sounds were sampled to disk from the DAT recordings, with a sampling rate of 45.455 kHz, and the wave data units converted to sound pressure amplitude by application of the DAT to MATLAB conversion factor. Seismic pulse waveforms were selected from recordings taken at

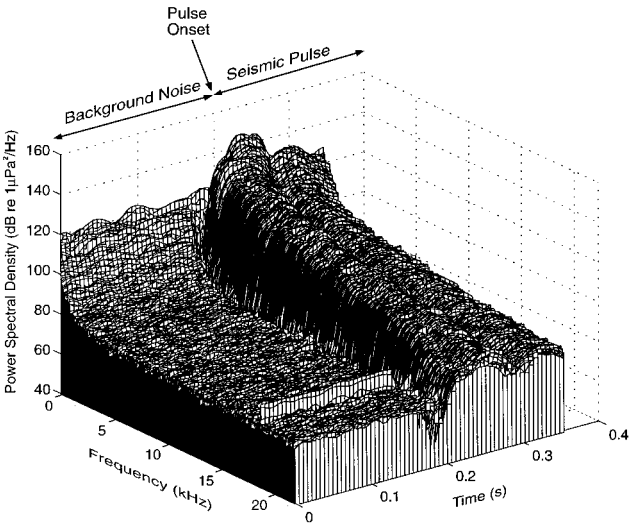


FIG. 4. Power-frequency-time matrix plot of average background noise preceding average maximum seismic pulse power at 750-m range from the 2120 cubic in. air-gun array. The seismic pulse onset occurs at 0.2 s in the plot, and there is clearly a surge in power across the entire bandwidth of 200 Hz–22 kHz.

source–receiver ranges of 750 m, 1 km, 2.2 km, and 8 km. Figure 3 shows a typical seismic pulse recorded 750 m from source. The waveform is a complex of high and low frequencies of uncertain duration, and a short-time Fourier transform method was employed to analyze the time-frequency structure of sequences of pulses. In analyzing any signal with a time-varying frequency content the choice of analysis window duration is important. Time and frequency resolution are inversely proportional, short windows giving good temporal resolution and poor frequency resolution and *vice versa* (Papoulis, 1984). To ascertain the window length giving the least spectral power estimate reduction resulting from time-frequency analysis in the onset region of the pulses, power spectral density analysis was performed on the pulse onset region (i.e., maximum energy region) through Hanning windows of various length, based on 2^n for computational efficiency. Power maximum occurred from application of a 2048 point (45 ms) windows across the initial section of the pulse, and analysis was therefore continued with this window. It should be noted that although the window duration is 45 ms the effective time resolution is better as a result of the tapering nature of the window. By simple calculation the 3-dB width of a 45-ms Hanning window is 16.4 ms, and 80% of the windowed signal energy is contained within this width.

An average power-frequency-time matrix of background noise power and seismic pulse power was constructed at each of the four source–receiver ranges. Power spectral density analysis was performed across 8192 waveform points, in advancing steps of 256 points, through 2048 point Hanning windows on (i) 11 sequential background noise waveform sections immediately preceding seismic pulses and (ii) the seismic pulses themselves. Background noise analysis, and construction of an average noise matrix at each source–receiver range, was performed prior to seismic pulse analysis. The average background noise matrix was then subtracted from each individual seismic pulse matrix prior to pulse power averaging, in order to remove the contribution of incoherent background noise from the pulse power. Figure 4 is a composite plot of the average background noise matrix preceding the average seismic pulse power matrix 750 m from source, and illustrates the huge cross bandwidth power surge above the already elevated background noise at the seismic pulse onset.

Finally, for each source–receiver range, the maximum pulse power spectrum slice was extracted from the average matrix, and each of the average background noise matrices for the four different source–receiver ranges were averaged internally to a single power-frequency spectrum. Average background noise and average maximum seismic pulse power spectra were then smoothed using a 10 point running average (resulting in a 222-Hz resolution spectrum) to further reduce estimation noise, and plotted together on a logarithmic (dB) scale for each source receiver range.

II. RESULTS

The solid curves in Fig. 5(a)–(d) illustrate maximum seismic pulse power at the four source–receiver ranges of 750 m, 1 km, 2.2 km, and 8 km. It can be seen from Fig. 5(a) and (b) (750-m and 1-km range, respectively) that seismic pulse power is far in excess of even the highly elevated background ship noise levels across the entire recorded bandwidth 200 Hz–22 kHz. Seismic pulse power ranges from approximately 140 dB *re*: 1 μPa^2/Hz at the 200-Hz end of the spectrum to approximately 90 dB *re*: 1 μPa^2/Hz at the 20-kHz end. High-frequency sound is attenuated by absorption through molecular resonance in seawater more rapidly with distance than low-frequency sound, and at 2.2-km range [Fig. 5(c)] seismic pulse power merges with the background noise at approximately 20 kHz. It should be stressed that the background noise spectrum is highly elevated from ambient by mechanical and flow sources of the guard ship (and at close source–receiver ranges the seismic ship), and water flow over the towed hydrophone. A ''quiet'' ambient noise environment would produce a curve several tens of dB below those obtained here. Even given the elevated background noise levels, it is clear that even 2.2 km from source, seismic pulse power dominates virtually the entire spectrum. At 8-km source–receiver range [Fig. 5(d)] seismic pulse power merges with background noise at approximately 8 kHz. If similar recordings were made in a quiet ambient environment (e.g., such as a fixed listening post far removed from engine machinery, or perhaps the ear of a swimming dolphin) the gap between seismic power curve and noise power curve would be increased from the illustrated examples at equivalent ranges from the air gun array. Both high- and low-frequency energy is present in the pulses at considerable magnitude and will certainly be detectable tens of kilometers from source.

Since the spectra of the seismic pulses cover a wide frequency range over which the Dolphin's hearing sensitivity varies very significantly it was thought worthwhile to make use of the same kind of procedure used in human environmental noise assessment whereby the frequency components of the noise (in this case the seismic pulses) are weighted

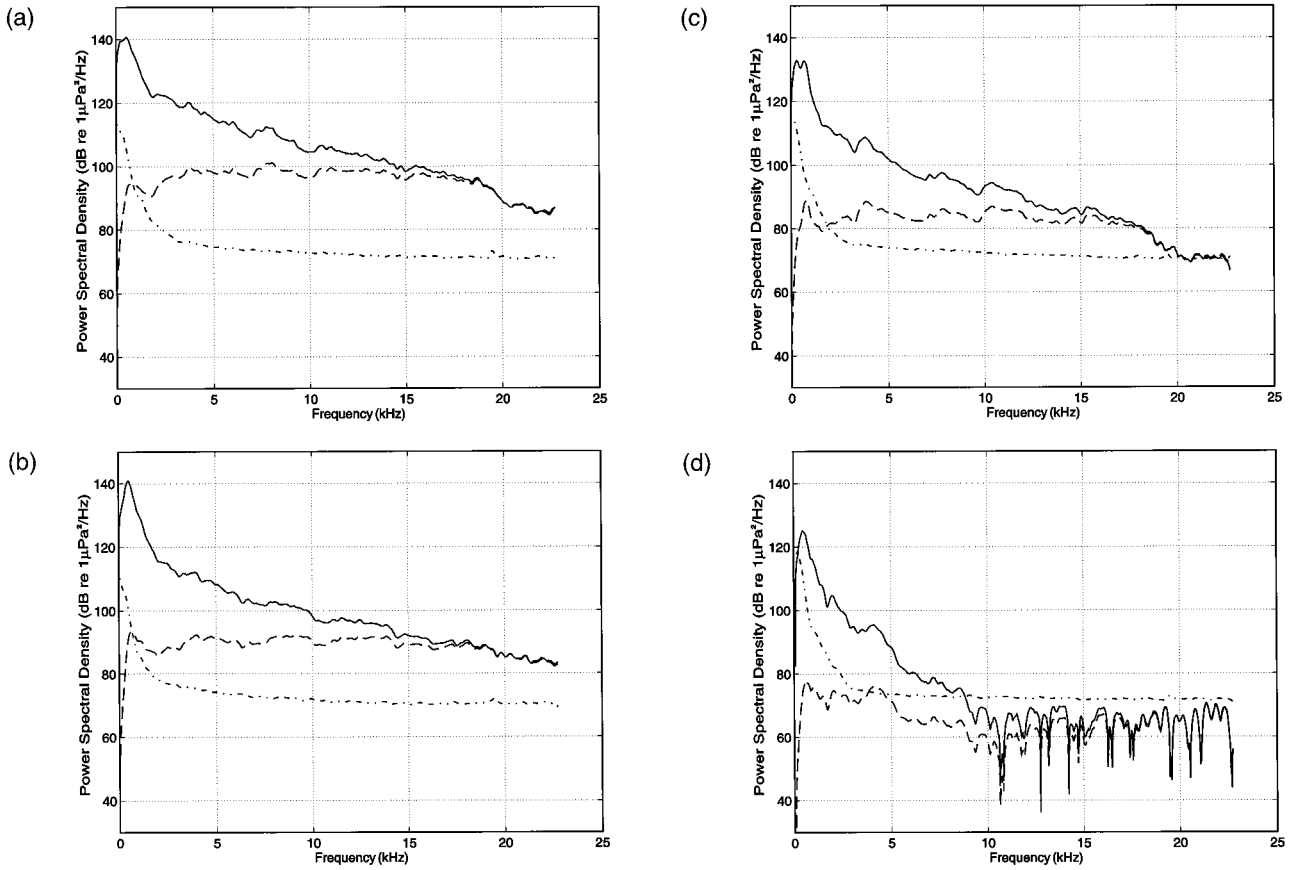


FIG. 5. Plots of maximum seismic power spectral density (solid line), spectral density normalized using dolphin threshold curve (dashed line), and average background noise, including high ship noise (dot-dash line) at four different source–receiver ranges: (a) 750 m; (b) 1 km; (c) 2.2 km; and (d) 8 km. Plots are shown on a logarithmic scale of power spectral density in dB *re*: 1 μPa^2/Hz.

according to the listener's hearing sensitivity at that frequency (Kinsler *et al.*, 1982). In the human case the standard A-weighting (relative sensitivity) curve approximately mirrors the human ten phon equal loudness contour which gives the level of sound at each frequency having the same loudness as a 10 dB (*re*: 10^{-12} W m^{-2}) signal at 1 kHz. The signal, after such weighting, should have approximately the same loudness as a 1-kHz tone having the same power.

In order to gain further insight into the likely impression of the seismic pulses on dolphins the seismic pulse spectra were weighted using published data on the variation of auditory threshold with frequency in the bottlenose dolphin (Johnson, 1967; Au, 1993). The dolphin auditory threshold curve for continuous tones (Fig. 3.2 in Au, 1993) was increased by 5 dB to allow for the increase with reducing tone burst duration (Johnson, 1967) and is shown in Fig. 6(a). It is recognized that pulse duration tends to increase during propagation and that the threshold shift with pulse duration is slightly frequency dependent. However, for simplicity a fixed figure of 5 dB (figure for 50 ms at 20 kHz) was used but the sensitivity to duration and frequency in this tone duration region is small (within 1.8 dB of this figure for a pulse duration range 50–100 ms and frequency range 1–20 kHz). A weighting (relative sensitivity) curve appropriate for dolphins using a reference frequency of 20 kHz is shown in Fig. 6(b). Note that this has been obtained, as in the case of the human A-weighting curve, by mirroring the sensitivity curve [Fig. 6(a)] in the sensitivity level at the reference frequency and setting this level to 0 dB. The choice of reference frequency is arbitrary and has been set, as in the human case, at approximately one-third of the frequency of maximum sensitivity. The dashed lines in the Fig. 5 plots are the seismic spectra with the weighting curve added. As in the human case the integral of each weighted spectrum (expressed in

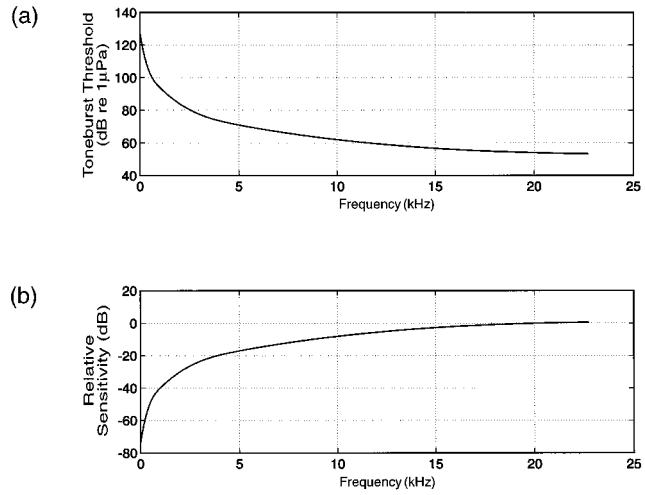


FIG. 6. (a) Dolphin hearing threshold curve for 50-ms tone bursts. (b) Relative sensitivity or weighting curve derived from (a).

TABLE I. Seismic pulse power at four source–receiver ranges corrected for sensitivity variation with frequency.

Range (km)	Power (dB_{dt})
0.75	140
1.0	133
2.2	127
8	112

$\mu Pa^2/Hz$) should have approximately the same loudness as a tone at the reference frequency having the same power and allows comparison with the dolphin hearing threshold at 20 kHz.

Table I shows this integral, expressed in dB *re*: 1 μPa. Continuing the comparison with the methodology of human noise measurement and noting that this figure is equivalent to the dB figure used for A-weighted sound levels adjusted for the human frequency response we term this the dolphin dB figure for sensitivity-weighted transient signals or dB_{dt}. It is recognized that there may be a slight underestimate in this figure for the 750-m, 1-km, and 2.2-km ranges as a result of the exclusion of signal components above the upper cutoff frequency of 22 kHz and below the lower cutoff frequency of 200 Hz. However, the error resulting from the exclusion of frequencies above 22 kHz will not be greater than 1 or 2 dB since the spectral density is falling and the dolphin sensitivity variation is leveling off. The spectral density estimation at the 8-km range is unreliable above 10 kHz, the variation shown resulting from random errors in the background noise subtraction. At this range absorption is an important factor, ensuring a rapid decrease in signal power at high frequencies and the true spectrum will continue to decrease above 10 kHz. Integrating the spectrum up to 10 kHz leads to a figure within 1 dB of that quoted. The exclusion of frequencies below 200 Hz is potentially more serious since the spectral density is still rising with reducing frequency at this point and will reach a peak below 200 Hz. However the dolphin's hearing sensitivity is still reducing and the excluded bandwidth is small. Assuming that the tendency for the reduction in sensitivity to compensate for the increase in spectral power resulting in the approximately flat adjusted spectra seen in Fig. 5(a)–(d) continues below 200 Hz, then the excluded frequencies below 200 Hz lead to errors of less than 1 dB at all distances. An alternative approach to error estimation is to take the upper figure of the typical range for air-gun outputs (240 dB *re*: 1 μPa @ 1 m), assume inverse square law power reduction (note that the figures in Table I follow this law approximately), and that all this power is concentrated below 200 Hz. Using the hearing threshold figure at 100 Hz to adjust for hearing sensitivity again leads to errors in the dB_{dt} figures due to this excluded power of less than 1 dB at all distances.

III. DISCUSSION

It was mentioned in the Introduction to this paper that seismic survey recordings were conducted in conjunction with an acoustic dolphin survey, using passive listening for dolphin vocalizations. During the March 1996 seismic survey, on which the sound analyses in this paper are based, natural dolphin population densities were so low that it was impossible to assess any effect on their distribution. The paucity of dolphin encounters was likely the result of a seasonal offshore migration the previous Autumn (Goold, in press), which meant that there were very few dolphins in the area at the time of the seismic survey. However some 18 months previous to the seismic survey described in this paper, in the Autumn of 1994, an identical seismic operation had been conducted by the same operators in adjacent prospect blocks, and using the same air-gun array. The 1994 seismic operation was conducted at a time when common dolphin population density was much higher (Goold, 1996), and acoustic dolphin surveying carried out in conjunction with seismic operations at that time suggested a localized disturbance of dolphins. For example, during the first week of seismic surveying only 4% of dolphin contacts occurred when air-gun shots were in progress, the remaining 96% occurring when the guns were silent between shot lines. However, there was also an observation during the 1994 survey of common dolphins close to the guard ship hydrophone for approximately 2 h while air-gun shots were in progress at an average distance of 1.34 km. The closest approach between the two ships during this encounter was 1070 m (both vessels fitted with differential GPS) and common dolphin sonar clicks could be heard very clearly through the guard ship hydrophone, causing intermittent record level overload. The inference was therefore drawn that (i) common dolphins were close to the guard ship hydrophone and (ii) that seismic energy from this particular air gun array under the prevailing conditions was therefore tolerable to common dolphins at a radius from the gun array of 1 km. No observations were made of dolphins at closer range to the gun array while firing was in progress. Figure 5(b) shows seismic pulse power at 1 km from the gun array in the same waters, and this plot may well be representative of a broadband threshold at which dolphins will tolerate seismic pulses.

Although it would be useful to present analyses of the seismic pulses recorded through the guard ship hydrophone during the aforementioned encounter, such analyses are precluded due to the recorder setup at the time. During the 1994 seismic survey, seismic pulses were recorded merely as a by-product of the acoustic dolphin survey, which was a new venture in itself. Gain levels were set high to capture faint dolphin whistles, and both recorder channels were 6-kHz high pass filtered. As a result of the high gain setting, seismic pulses recorded at source–receiver ranges of less than 2.5 km exceeded the DAT recorder dynamic range as configured at the time, resulting in clipped signals on tape. We can therefore only assume that the seismic power level presented in Fig. 5(b) is closely representative of that to which dolphins were exposed during the aforementioned encounter. This is a reasonable assumption given that both seismic operations were conducted in essentially the same shallow sea area, and that the same air-gun array (indeed the same contractor) was employed in both instances. Although propagation conditions in the sea may vary considerably, both surveys were conducted in a perpetually mixed shelf sea region, thereby eliminating any complications of thermal stratifica-

tion. Thermal stratification would not have occurred at the time of year in which these surveys were conducted in any case, and a CTD cast during the 1996 seismic survey confirmed a homogeneous vertical temperature-salinity profile. It should be noted at this point that sound pressure level estimates for the air-gun array at 5 km, quoted merely as ''rough cut estimates'' in Goold (1996), were based on an erroneous assumption of the high pass filter performance particularly at the low-frequency end of the spectrum. Some time after the 1994 survey work the filter unit was tested and found to have a much shallower attenuation slope than the design specified. Sound pressure level measures below 6 kHz quoted in Goold (1996) were overcompensated, resulting in estimates that were too high. The power spectral density measures made in the current work, however, are based essentially on a flat ± 2.5-dB bandpass recording between 200 Hz and 22 kHz, and a total system through-put calibration.

There is some difficulty in determining how loud the seismic pulse seems to a dolphin at different ranges. This stems from the fact that hearing threshold measurements have been carried out using long (quasi-continuous) tones and tone bursts of varying duration but not for more complex signals such as the seismic pulses and isophon (equal loudness) curves have not been measured at all. Noise masking measurements have been used to determine the bandwidth of the constant Q filters by which it is suggested the dolphin's auditory system can be modeled. The critical ratio and critical bandwidth methods indicate the filter Q to be 12.3 and 2.2, respectively (Au and Moore, 1990; Johnson, 1968). It is fairly clear that the dolphin can hear the seismic pulse even at the maximum range (8 km) considered. Taking the narrowest bandwidth case ($Q=12.3$) at the frequency of peak spectral density 466 Hz, the power within the filter bandwidth is [from Fig. 5(d)] approximately $125+10\log_{10}(466/12.3)=141$ dB *re*: 1 μPa, well above the threshold of 105 dB *re*: 1 μPa at this frequency and this is not even taking into account the contributions from other spectral components. This conclusion is also clearly not critically dependent on the bandwidth estimation used. The question of how the seismic pulses affect the dolphin at different ranges is less easily answered. It is known that the dolphin adapts to environmental noise during echo location by increasing its emitted click amplitude. Peak energy in the dolphin click spectrum in a quiet environment occurs around 50 kHz, however this peak shifts to higher frequency along with increasing click amplitude in response to noise (Au, 1980, 1993) and is in the range at which the seismic spectral density is very low and decreasing and the dolphin hearing threshold is beginning to increase again. It is likely therefore that the predominant interference is with the dolphins' communication which takes place at lower frequencies (5–30 kHz). In order to have some idea of how loud the seismic signal might appear to the dolphin, it is necessary to take into account their dependence of sensitivity on frequency and to weight the spectrum before calculating an effective power.

Use of the low level threshold curve is likely to lead to an underestimate of the dolphin dB_{dt} figure since it will reduce the contribution of low-frequency components if the equal loudness contours become flatter with increasing acoustic pressure as in the human case. Table I shows the ''dB_{dt}'' figures for the four source distances, all clearly exceeding the threshold at 20 kHz.

In order to determine the potential for interference with dolphin communication typical levels for these signals are required. Unfortunately the literature is sparse on this subject; although a figure of 155 dB *re*: 1 μPa at 9 kHz at source (1-m range) has been ascertained for common dolphin (Hall, 1997). At a range of 10 m (assuming predominately inverse square law attenuation) the power level would be 135 dB *re*: 1 μPa which is equivalent to 125 dB at 20 kHz using the same normalization curve. The range of 10 m has been used as a typical distance between dolphins in the same pod. Comparing this figure of 125 dB with those in Table I it is seen that the seismic pulse has comparable response-adjusted level (perceived as comparably loud) as this typical communication signal at an air-gun-to-dolphin distance of 2.2 km. At the closer distances of 750 m and 1 km the pulse will appear much louder. Bearing in mind the assumptions and approximations made the figures in Table I are not inconsistent with the observation that dolphins find the seismic signal levels at 1 km and closer distressing and that this distress may be due to interference with communication. It should be remembered, however, that these power levels occur within only a 50–100 ms period repeated at 10-s intervals. This low duty-cycle signal would not seem to be a serious potential interruption of echo location or communications. However, it is known that, in humans, loud sounds cause a temporary threshold shift (permanent for very loud and sustained sounds) within approximately 0.5 ms of the start of the sound (acoustic reflex) and recovery may take several minutes (Kinsler *et al.*, 1982). This protection mechanism may well exist in dolphins and, if it is present, would lead to reduced hearing sensitivity even during the interpulse intervals.

A recent publication (Ridgway *et al.*, 1997) describes experiments in which the thresholds for behavioral changes and temporary threshold shifts in captive dolphins were measured at sound levels of 181 dB and 193–196 dB *re*: 1 μPa, respectively, for a 1-s tone burst at 20 kHz. Even allowing for uncertainty in our distance estimation it is clear that our observation of dolphins exhibiting behavioral changes (avoidance) at distances less than 1 km together with our estimate of a sensitivity-weighted sound level referenced to 20 kHz of 133 dB at this distance suggests a far greater sensitivity to environmental noise than that presented in the Ridgway report. One possible explanation is that the threshold for the observed behavioral changes in captive dolphins is far higher than that for avoidance behavior in free ranging dolphins. Captive dolphins cannot swim away and may be more tolerant through necessity!

The dolphin is a inquisitive animal and will tolerate fairly loud engine and propeller noise in order to investigate ships at close quarters. However, there must be a distance from a source of loud noise at which the desire to investigate is balanced by auditory distress. Further work needs to be done to determine ''safe'' distances at which high source level sounds are tolerable to cetaceans. Our best estimate at present for dolphins and the 2120 cubic in. air-gun array in a shelf sea environment is 1 km.

ACKNOWLEDGMENTS

The fieldwork described in this paper was funded by Chevron UK Limited, Aran Energy Exploration Ltd., and Repsol Exploration (UK) Ltd. Thanks are also expressed to David Goodson of the Loughborough University of Technology for assistance with the hydrophone calibration. Additional acknowledgment is also due to operations geophysicist Norman White, for coordinating walk-away tests aboard the seismic vessel PACIFIC HORIZON. Similarly the skipper and crew of the guard ship EJP are thanked for their efforts and cooperation.

Au, W. W. L. (**1980**). ''Echolocation signals of the Atlantic Bottlenose Dolphin, *Tursiops truncatus*, in open waters,'' in *Animal Sonar Systems*, edited by R. G. Busnel and J. F. Fish (Plenum, New York), pp. 251–282.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, New York).

Au, W. W. L., and Moore, P. W. B. (**1990**). ''Critical ratio and critical bandwidth for the Atlantic bottlenose dolphin,'' J. Acoust. Soc. Am. **88**, 1635–1638.

Barger, J. E., and Hamblen, W. R. (**1980**). ''The air gun impulsive underwater transducer,'' J. Acoust. Soc. Am. **68**, 1038–1045.

Goold, J. C. (**1996**). ''Acoustic assessment of populations of common dolphin, *Delphinus delphis*, in conjunction with seismic surveying,'' J. Mar. Biol. Assoc. UK **76**, 811–820.

Goold, J. C. (**in press**). ''Acoustic assessment of populations of common dolphin, *Delphinus delphis*, off the West Wales Coast, with perspectives from satellite infra-red imagery,'' J. Mar. Biol. Assoc. UK.

Greene, C. R., and Richardson, W. J. (**1988**). ''Characteristics of marine seismic survey sounds in the Beaufort Sea,'' J. Acoust. Soc. Am. **83**, 2246–2254.

Hall, J. D. (**1997**). Private communication.

Johnson, C. S. (**1967**). ''Relation between absolute threshold and duration-of-tone pulses in the bottlenose porpoise,'' J. Acoust. Soc. Am. **43**, 757–763.

Johnson, C. S. (**1968**). ''Masked tonal thresholds in the bottlenose porpoise,'' J. Acoust. Soc. Am. **44**, 965–967.

Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**1982**). *Fundamentals of Acoustics* (Wiley, New York), 3rd ed.

Papoulis, A. (**1984**). *Signal Analysis* (McGraw-Hill, New York).

Richardson, W. J., Greene, C. R., Malme, C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, San Diego).

Richardson, W. J., and Wursig, B. (**1997**). ''Influences of man-made noise and other human actions on cetacean behaviour,'' Mar. Freshwater Behav. Physiol. **29**, 183–209.

Ridgway, S. H., Carder, D. A., Smith, R. R., Kamolnick, T., Schlundt, C. E., and Elsberry, W. R. (**1997**). ''Behavioral responses and temporary shift in masked hearing threshold of bottlenose dolphins *Tursiops truncatus*, to 1-second tones of 141 to 201 dB *re*: 1 μ Pa,'' Tech. Rep. 1751. NRAD, RDT&RE Div., Naval Command, Control & Ocean Surveillance Center, San Diego, CA.

Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study

Joseph A. Kogan and Daniel Margoliash
Department of Organismal Biology and Anatomy, 1027 East 57th Street, The University of Chicago, Chicago, Illinois 60637

(Received 4 August 1997; revised 5 November 1997; accepted 2 January 1998)

The performance of two techniques is compared for automated recognition of bird song units from continuous recordings. The advantages and limitations of dynamic time warping (DTW) and hidden Markov models (HMMs) are evaluated on a large database of male songs of zebra finches (*Taeniopygia guttata*) and indigo buntings (*Passerina cyanea*), which have different types of vocalizations and have been recorded under different laboratory conditions. Depending on the quality of recordings and complexity of song, the DTW-based technique gives excellent to satisfactory performance. Under challenging conditions such as noisy recordings or presence of confusing short-duration calls, good performance of the DTW-based technique requires careful selection of templates that may demand expert knowledge. Because HMMs are trained, equivalent or even better performance of HMMs can be achieved based only on segmentation and labeling of constituent vocalizations, albeit with many more training examples than DTW templates. One weakness in HMM performance is the misclassification of short-duration vocalizations or song units with more variable structure (e.g., some calls, and syllables of plastic songs). To address these and other limitations, new approaches for analyzing bird vocalizations are discussed. [S0001-4966(98)02004-9]

PACS numbers: 43.80.Ka, 43.72.Ne, 43.60.Lq [FD]

INTRODUCTION

Many biological studies require identification of constituents of animal vocalizations from continuous recordings obtained in the field or in the laboratory (e.g., Payne *et al.*, 1981; Marler and Peters, 1982a). Most of these studies are based on manual inspection and labeling of sound spectrographs, which relies on agreement between human experts, often not explicitly described, for reproducible results, rather than a quantitative approach. In addition, many biological studies involve a large corpus of vocalizations and therefore are extremely labor extensive, yet substantial databases may be biologically important, for example, in identifying rarely observed if highly significant behaviors (e.g., Margoliash *et al.*, 1991, 1994), or in assessing variability of brain activity in relation to behavior (e.g., Yu and Margoliash, 1996).

Manual inspection of multiple vocalizations is prone to errors, which can be minimized by cross checking (duplicate scoring) but only at the cost of additional effort. On the other hand, the ability of a human expert in visual analysis of sound spectrographs supported by auditory playback cannot be outperformed on a small task. Since automated analysis techniques tend to reliably generate correct and erroneous constituent identification (labels), this suggests that a practical strategy for maximizing performance while minimizing manual effort should be based on training on a small data set manually prepared by experts, then automated recognition followed by limited inspection by human experts to correct obvious errors.

Recent progress in automated speech recognition (Makhoul and Schwartz, 1995) encourages expectations for achieving the goal of reliable automated recognition of animal vocalizations. Although the differences between human speech and animal vocalizations, and the different conditions under which they are recorded, are significant and have to be taken into account, the relative simplicity of animal vocalizations compared to speech can facilitate recognition of animal vocalizations. For example, most animal vocalizations consist of discrete subunits organized in stereotyped hierarchies. This is true for bird songs, which can be viewed as comprising notes, syllables (or figures), phrases (or motifs), and songs (Catchpole and Slater, 1995). On the other hand, many biological studies require recording animal vocalizations under adverse conditions (e.g., field recordings with contaminant vocalizations and nonhomogeneous noise backgrounds), whereas speech recognition is almost always focused on low-clutter high signal-to-noise (S/N) ratio conditions. Here we evaluate automated recognition techniques as applied to bird songs recorded under a variety of noise conditions in the laboratory. This may also be a first step towards developing suitable analysis techniques for field recordings (Larkin *et al.*, 1996).

A dynamic time-warping (DTW) template-based approach is potentially attractive for analyzing stereotyped vocalizations (Silverman and Morgan, 1990) such as the songs of adult birds. Thus we developed the DTW-based Long Continuous Song Recognition (LCSR) system (Anderson *et al.*, 1996). Similar dynamic programming techniques have also proven useful in analyzing and comparing manually extracted single animal vocalizations, for example, dolphin signature whistles (Buck and Tyack, 1993) and bird calls (Ito *et al.*, 1996).

In this study, we continue investigation of LCSR and compare its performance with hidden Markov models (HMMs) on a large database of songs. As would be expected with a template-based system, good LCSR performance depends on the degree of variability within and separability between classes represented by the templates. Thus LCSR performance can depend on the often quite subtle tradeoff between the effectiveness of templates in representing constituent vocalization variability while maintaining separability between vocalization classes (Anderson *et al.*, 1996).

In contrast to template matching, HMMs statistically represent and estimate the constituent vocalizations to be recognized. As a result of such estimation (training), these models can accumulate more information and generalize better than can template-based techniques [for recent reviews, see Rabiner and Juang (1993); Makhoul and Schwartz (1995)]. To explore the application of HMMs to bird song, we employed the Hidden Markov Model Toolkit (HTK) (Entropic Research Laboratory) developed by Young *et al.* (1995). HTK exhibits excellent performance for continuous speech recognition applications compared to other systems in this area (Kubala, 1995).

We compared the performance of HTK and LCSR on a large database of songs of two species: stereotyped songs of zebra finches (*Taeniopygia guttata*), which have broadband vocalizations, and stereotyped and plastic songs of indigo buntings (*Passerina cyanea*), which have relatively spectrally compact vocalizations. Both techniques were tested over a broad range of regimes, and we attempted to make balanced comparisons. Our results show that HTK usually outperforms LCSR when tested with relatively noisy recordings or variable bird vocalizations with confusing calls and notes. For optimal performance we find HTK in general requires considerably more training examples than LCSR requires templates, while LCSR requires greater expert knowledge and repeated trials in selecting an optimal set of templates, especially for noisy recordings. For practical laboratory usage, an efficient approach may be to use LCSR to prepare a training set, correct that training set manually, then train HTK and use it for subsequent recognition.

I. METHODS AND TECHNIQUES

Here we emphasize specific differences in the two techniques employed for the recognition task. Technical details can be found in the cited references.

A. Long continuous song recognition

The LCSR system segments and labels an acoustic input (bird songs) using a prespecified set of template patterns of bird song constituents (Anderson *et al.*, 1996). In our application we have used digital spectrographs to represent the signals.

LCSR is based on the one-stage synchronized search which is known as one of the most computationally efficient DTW algorithms. This algorithm was first introduced by Vintsyuk (1971) for automated word recognition and subsequently extended by Bridle *et al.* (1982) and Ney (1984) for connected word recognition. For this search, the time frames of the input acoustic signal and the time frames of the templates are organized in a lattice (i,j,k), where i and j are the indexes of the time frames of the input song and each individual template, respectively, and k is the template counter. The quality of the match is measured through the sum of local metrics $d(i,j,k)$ which represent the distances between the two multidimensional vectors of the input signal at time frame i and template k at time frame j. In Anderson *et al.* (1996), $d(i,j,k)$ was computed as the Euclidean distance between the log magnitudes of the fast Fourier transform (FFT) bins of those vectors. Based on Bellman's principle of optimality, the one-stage algorithm ''recognizes'' the continuous input as the optimal sequence of selected template patterns. This minimizes the total distance between the sequence of warped templates and the continuous input under the allowable nonlinear compression–dilation transformations (warping) of the templates. In the present realization of LCSR, the range of such warping is constrained by a factor of 2. Note that in LCSR compression–dilation of the templates is the only mechanism to compensate for the variability of sounds (see Anderson *et al.*, 1996).

B. Hidden Markov toolkit

In HTK, the process of recognition of the input signal can be divided into several steps including preprocessing and representation, modeling the acoustic patterns present in the input signals, training these models, and ultimately, recognition. Here we focus on the HTK statistical models for acoustic patterns (HMMs) and their training, which distinguishes HTK from LCSR and other template based approaches in automated recognition. Though HTK is primary designed for speech recognition, its 18 major tools and programs can be adapted and used for analysis and recognition of any acoustic time series, including bird vocalizations. During this work, all major tools and most auxiliary programs of HTK were extensively tested under a variety of regimes.

1. Modeling

Using HTK, each sound category (calls, syllables, cage noises) was modeled by an HMM. The HMM is a finite state machine consisting of two levels. The hidden level operates in time, t $(1 \leqslant t \leqslant T)$ as a Markov chain with a finite number of states $X=\{1,\ldots,n\}$ and the transitions from state i to state j occur with a transition probability a_{ij}. The level of observations of the acoustical signals are represented by a sequence of vectors $\mathbf{O}=\{o_t\}$ which are assumed to be emitted from the hidden state j with some probability density $b_j(o_t)$. If the transition and the output probabilities of the HMM, **H**, are known, then the hidden state sequence $X=\{x(1),x(2),\ldots,x(T)\}$ of interest can be estimated as the most likely sequence that has generated the observed output **O**:

$$\hat{X}=\mathrm{argmax}_X\left\{a_{x(0)x(1)}\prod_{t=1}^{T} b_{x(t)}(o_t)a_{x(t)x(t+1)}\right\}, \qquad (1)$$

where $x(0)$ is the model entry state and $x(T+1)$ is the model exit state. In our typical model, the HMM was a left-to-right model $(a_{ij}=0, j<i)$ with five states, with only the middle three states {2,3,4} emitting signals. The first and the

last states {1,5} served to glue these elementary HMMs into a compound model.

The densities of observation probabilities in the emitting states are usually modeled as mixtures of multidimensional Gaussian distributions:

$$b_j(o_t) = \sum_{m=1}^{M} c_{jm} N(o_t; \mu_{jm}, \Sigma_{jm}), \quad (2)$$

where M is the number of mixtures, c_{jm} is the weight of the mth component, $\Sigma_{m=1}^{M} c_{jm} = 1$, and $N(o_t; \mu, \Sigma)$ is some n-dimensional Gaussian density with mean vector μ and covariance matrix Σ where n is the dimensionality of observed vector o_t.

In practice, the parameters of HMMs are unknown and have to be estimated from data during the training phase. The choice of model parameters is important for reliable performance, and will reflect the data to be modeled. For example, the number of mixtures should reflect, roughly speaking, the number of different subcategories of acoustic patterns combined within the same signal class. This interacts with the amount of data available for reliable estimation (training) of the model parameters. In our experiments we used from 2 to 4 mixtures with a diagonal covariance matrix for modeling syllables and calls.

2. Training

Although direct computation by Eq. (1) is intractable, the efficient and well-known Baum–Welch algorithm (see Rabiner and Juang, 1993; Young *et al.*, 1995) implemented in HTK finds the maximum-likelihood estimates (MLE) of unknown parameters (see Appendix A). In HTK the training of HMM prototypes is subdivided into several stages including initialization of the models, isolated training for each vocalization class, and finally the embedded training when the compound HMMs are trained on the set of entire songs (Appendix A). In addition, we divided the intersyllable intervals into five duration ranges (see Williams, 1990), then derived a simple probabilistic bigrammar by counting the frequency of co-occurrences of each vocalization class with each intersyllable interval range.

In our experiments, songs for training were selected at random, but such that the training set contained exemplars of all vocalization classes. Since some calls or introductory notes occurred very infrequently, and because compound HMMs are capable of accumulating more information from the entire songs (one or more of which were contained in each file), we selected the training set on a file level, specifically including files with examples of infrequently encountered elements. These procedures tended to increase the size of the HTK training sets. Because HTK and LCSR do not have a mechanism for selecting an optimal training (or template) set, we did not obtain precise estimates for the size of the training (template) set or other parameters that may influence recognition accuracy. This would require a more statistically oriented approach such as cross validation (e.g., McLachlan, 1992). Thus direct comparison of the size of the training sets we used for HTK and number of templates we used for LCSR is somewhat difficult to interpret, although general trends were observed (see Sec. III).

3. Recognition

Trained HMMs along with the song bigrammars were used for recognition. In HTK the recognition is performed by a dynamic programming algorithm called token passing (Young *et al.*, 1989) which represents an adaptation of the Viterbi algorithm (Viterbi, 1967; Forney, 1973; Kogan, 1996) for continuous speech recognition. Interestingly, the computational scheme employed in the token passing algorithm is also the one-stage synchronized search used in LCSR, but modified for the HMM setting.

C. Database

Oscine passerine birds learn their vocalizations by reference to auditory feedback (Thorpe, 1958; Konishi, 1965). This accounts for the complexity and individual variety in many species' songs (Catchpole and Slater, 1995), and provides examples of complex bioacoustic signals to be recognized. Both LCSR and HTK were evaluated with continuous recordings of stereotyped songs of four zebra finches (bu41, gr43, gr46, yl49) as well as stereotyped and plastic songs of four indigo buntings (ib7, ib5, ib13, ib20). These recordings are representative of different types of vocalizations (broad/narrow band, stereotyped/plastic) and different conditions (low noise/noisy). To facilitate assessment of the performance of the recognizers, each call and syllable in the data files was labeled (i.e., assigned to a vocalization class and assigned an onset and offset). Some data files were scored completely manually, others were prepared with LCSR then manually reviewed. Many of the labeled vocalizations were carefully segmented and extensively analyzed in the context of behavioral or neurophysiological experiments. In some cases, labels were cross verified in blind conditions. Thus the data base is likely to be highly reliable. The overall number of syllables and calls analyzed as well as approximate number of songs per animal can be found in Table I.

Zebra finches produce a repeated sequence of broadband vocalizations often with a prominent harmonic structure. Their songs consist of introductory notes/calls followed by one or more motifs (Sossinka and Böhner, 1980). Motifs tend to be highly stereotyped, consisting of ordered sequences of two or more syllables which typically are 20–200 ms in duration separated by brief (5–50 ms) silent intervals. Zebra finch calls may be learned or innate (Zann, 1996). Compared to syllables, calls may have a more variable structure and amplitude, be shorter in duration, and can occur in a broader range of contexts (temporal sequences of vocal elements). The number of different vocalization classes in the database were ten for zebra finch bu41, 14 for gr46, 13 for gr43, and 17 for yl49.

Indigo buntings typically produce sequences of repeated (doublets, triplets, etc.) narrow-band syllables often followed by one or several spectrally broadband syllables in their advertisement songs (Thompson, 1970). Indigo buntings sing two types of songs, stereotyped songs and plastic songs (Margoliash *et al.*, 1991). The plastic songs can be distin-

TABLE I. Database content for evaluation of LCSR and HTK.

Zebra finch				Indigo bunting			
Bird	Type[a]	Level[b]	Total (N)	Bird	Type[a]	Level[b]	Total (N)
bu41	S	file	59	ib7	S	song	154
		syl	1115			syl	1061
gr43	S	file	54		P	song	61
		syl	1131			syl	582
gr46	S	file	56	ib5	S	song	100
		syl	1700			syl	975
yl49	S	file	114		P	song	70
		syl	3732			syl	1619
				ib13	S	song	100
						syl	1199
					P	song	49
						syl	833
				ib20	S	song	100
						syl	1209
					P	song	56
						syl	692

[a]S stands for stereotyped song, P stands for plastic song.
[b]A ''file'' consists of 1–5 zebra finch's songs and usually corresponds to a recording of a song bout; ''syl'' stands for syllable.

guished from stereotyped songs by their low amplitude, larger syllable repertoire, more variable syllable structure and order, and inclusion of high-frequency ''squeaky'' notes and other transients, incompletely formed sounds and indistinct sounds typically not found in the advertisement songs (Margoliash *et al.*, 1991, 1994). The indigo buntings vocalizations consisted of 13 song element classes for ib7, 13 for ib5, 22 for ib13, and 16 for ib20, for both stereotyped and plastic songs.

For each bird, there was a large range in the frequency of expression of different elements of the vocal repertoire. Commonly, there were hundreds of exemplars of each stereotyped syllable and call class, but occasionally a syllable or call class was only represented by tens of exemplars. In addition, some call/note classes occurred even less frequently. For example, element $\{s\}$ of gr43 was recognized only once, and calls $\{k\}$ and $\{m\}$ of yl49 were recognized only once and twice, respectively. Automated recognition of such infrequently occurring elements was difficult for both recognizers. For HTK, the small sample size tended to preclude adequate training. For LCSR, just a few templates were often sufficient to recognize a class (avoiding deletion errors), but the inclusion of such templates also increased the

TABLE II. Dependence of HTK performance on parametrization for zebra finch gr43. (target-rate=6 ms, window-size=18 ms, 10 coefficients).

Parametrization[a]	% Correct	% Accuracy	S[b]	D[b]	I[b]
FBANK_E	40.4	36.5	738	708	95
FBANK_EDA	89.7	76.6	192	59	316
	92.2	80.6	75	13	131
LPC_EDA	82.3	74.9	280	156	172
	81.3	74.0	149	63	83
LPCEPST_EDA	88.6	82.5	164	112	149
	91.4	86.1	48	49	60
LPREFC_EDA	81.3	78.8	233	220	62
	82.8	80.2	103	92	29
MELSPEC_EDA	81.0	68.3	317	145	308
	89.7	62.9	97	20	303
MFCC_EDA	90.4	83.9	108	124	160
	92.3	88.7	30	57	41
MFCC_EDAZ	81.6	45.6	432	15	874

[a]First row tests were conducted on the set of vocalizations as manually scored; second row tests with the combined set of calls or syllables (see text).
[b]D≡deletion error, S≡substitution error, I≡insertion error. % correct ignores syllable level insertions, and song level insertions and deletions. % accuracy accounts for all types of errors.

TABLE III. HTK and LCSR recognition performance on zebra finches vocalizations (LCSR, window-size/step=25.6/12.8[a] ms, HTK, target-rate=6 ms, window-size=18 ms).

Tool/Bird	Level	% Correct[b]	% Accuracy[b]	% S[b]	% D[b]	% I[b]	Train/Test[c]
HTK/gr43	file	25.9, 53.7§		74.0, 46.2			27/54
HTK	syl	92.4, 97.9§	90.1, 96.1§	6.4, 1.4§	1.2, 0.7§	2.3, 1.8§	668/1131
HTK	file	35.2§		64.8§			14/54
HTK	syl	89.4, 97.0§	85.8, 94.6§	2.1§	0.8§	2.4§	383/1131
LCSR	syl	75.8, 92.7§	62.3, 80.0§	21.0, 4.2§	3.2, 3.2§	13.4, 12.6§	5(70)
LCSR	syl	84.6, 95.8§	81.8, 92.5§	11.9, 0.8§	3.4, 3.4§	2.8, 2.8§	(130)
HTK/gr46	file	17.9		83.1			16/56
HTK	syl	96.6	92.4	3.2	0.2	4.1	513/1700
HTK	file	16.0		84.0			13/56
HTK	syl	95.9	92.4	3.7	0.3	3.4	485/1700
LCSR 12.8/3.2	syl	92.5	82.1	5.5	2.0	10.3	5(75)
HTK/bu41	file	49.1		50.9			20/59
HTK	syl	96.5, 97.5§	95.0, 95.5§	3.1, 1.0§	0.4, 1.5§	1.5, 2.0§	439/1115
LCSR	syl	97.4, 98.6§	79.7, 95.2§	2.5, 0.7§	0.1, 0.7§	17.7, 3.4§	5(60)
LCSR	syl	98.8	95.5	0.9	0.3	3.3	(93)
HTK/yl49	file	4.4		95.6			18/114
HTK	syl	87.1	83.6	10.8	2.0	3.5	653/3732
LCSR 12.8/12.8	syl	89.7	45.4	8.3	2.0	44.3	5(78)
LCSR 12.8/12.8	syl	92.9	78.2	5.8	1.2	14.7	(156)

[a]The default FFT window-size/step, otherwise shown in the first column.
[b]§ is for combined set of calls or syllables (see text).
[c]HTK: Number of vocalizations in training/test set. Number in training set does not include following ''silent'' intervals. LCSR: first number is typical number of templates per syllable; number in parentheses is the total number of templates.

frequency of substitution and especially insertion errors elsewhere.

D. Representation and preprocessing

1. Template creation

As described by Anderson *et al.* (1996), the most significant preprocessing step in using LCSR is the creation of templates for the song units, background noises, and a variety of ''silent'' intervals to be identified from the input stream. In the current study, templates were manually selected based on visual inspection of the spectrograms of individual bird songs with some attempt made to represent the variation within each vocalization class (call or syllable), as well as cage noises and ''silent'' intervals. As a rule, the templates were chosen from the beginning, middle, and end of an entire set of recordings. In initial experiments the number of templates within each class was limited to 3–7, which in some cases has provided accurate recognition while limiting computational overhead (see Anderson *et al.*, 1996). Under challenging conditions (short confusing calls, cage noises, noisy recordings), this number of templates was insufficient for accurate recognition and we increased the number of templates by choosing confusing elements as additional templates (see below).

2. Preprocessing and parametrization

In our experiments with LCSR presented in Tables III and IV, the bird song signals were represented by discrete Fourier transforms. Frequencies below 500 Hz, where vocalizations have little power, were omitted to remove fan and other low-frequency noise evident in much of the data. Feature vectors are thus the log magnitude FFT bins from 0.5 to 10 kHz. To optimize the LCSR performance, the parameters of the FFT size and frame rate, thresholds for minimum durations of elements to be recognized, etc., were experimentally selected for each bird individually. The amplitude of the acoustic signals was not normalized, because it was thought that the signal amplitude might provide valuable information for separation of song elements. To achieve better LCSR performance on indigo buntings' stereotyped songs, the input representation was enhanced by zeroing all feature vector components of magnitude <1 s.d. above the mean value (see Anderson *et al.*, 1996). This procedure, however, requires prior knowledge regarding these songs.

HTK facilitated analysis of the applicability to bird songs of different types of representations common in speech recognition work. Thus we compared the influence of six different types of parametrizations on HTK performance. These included linear predictive coding (LPC), LPCepstral, LPCreflection, mel-frequency cepstral, log mel-filter bank channel, and linear mel-filter bank channel (Rabiner and Juang, 1993; Deller *et al.*, 1993). These parametrizations were also evaluated with different classifiers: _E (energy), _D and _A (first and second derivatives), and _Z (cepstral mean normalization). The results of these experiments on songs of zebra finch gr43 are presented in Table II. As one can see, the performance of HTK for linear parametrizations (LPC, LPCreflection, linear mel-filter bank), decreased in all cases but by a variable amount compared to nonlinear cepstral parametrizations. Among the latter, the best results were achieved for mel-frequency cepstral coefficients (MFCC). For this reason we used the MFCC parametrization for all other experiments (e.g., Tables III and IV). Classifiers _E, _D, and _A improved the performance but _Z did not.

TABLE IV. HTK and LCSR recognition performance on indigo bunting vocalizations (LCSR, window-size=25.6/3.2, HTK, target-rate=10 ms, window-size=25 ms).

Tool	Bird/Type	Level	% Correct	% Accuracy	% S	% D	% I	Train/Test[a]
HTK	ib7/S	song	88.3		11.7			26/154
HTK		syllable	98.4	97.8	1.4	0.2	0.6	217/1061
LCSR		syllable	99.2	97.8	0.8	0.0	1.4	3(45)
HTK	P	song	65.6		35.4			29/61
HTK		syllable	98.5	94.76	0.5	1	3,7	341/582
LCSR		syllable	94.5	83.7	3.1	2.4	10.8	3(45)
HTK	ib5/S	song	67.7		26			32/68
HTK		syllable	93.2	92.7	2.8	4	0.4	318/975
LCSR	12.8/3.2	syllable	97.4	97.3	0.9	1.6	0.1	5(113)
HTK	P	song	21.3		78.7			23/47
HTK		syllable	83.4, 95.0§	76.5, 90.7§	3.2§	1.7§	4.4§	644/1619
LCSR	12.8/3.2	syllable	75.4, 86.5§	66.6, 77.7§	3.4§	10.1§	8.8§	5(113)
HTK	ib13/S	song	72, 76§		27,34§			15/100
HTK		syllable	98.8, 99.4§	97.2, 97.5§	0.1§	0.4§	1.9§	192/1199
LCSR		syllable	80.2	22.6, 78.8§	12.1	7.8	57.6,2.2§	7(161)
HTK	P	song	24.5, 44.9§		55.1§			26/49
HTK		syllable	90.9, 95.6§	89.1, 94§	3.3§	1.1§	1.5§	460/833
LCSR		syllable	62.0	−29.6, 53.3§	20.0	18.0	91.6,8.7§	7(161)
HTK	ib20/S	song	55.0		45.0			26/100
HTK		syllable	96.3	93.8	2.9	0.8	2.4	402/1209
LCSR	12.8/3.2	syllable	92.9	87.0	1.8	5.3	5.9	5(52)
HTK	P	song	64.3, 83.9§		16.1§			18/56
HTK		syllable	98.7, 99.2§	95.8, 98.6§	0.5§	0.3§	0.7§	251/692
LCSR		syllable	90.9	85.0	0.6	8.5	5.9	5(52)

[a]For LCSR, the same templates were used for analysis of stereotyped and plastic song. See Tables II and III for all symbols.

Hence, we also used _EDA as well for all other experiments.

The LPC representation was also explored for LCSR. We tested the short-time modified representation based on linear predictive coding (LPC) applied to the double autocorrelation function which de-emphasizes the amplitude of the signal. We also tested modification of the Euclidean distance for the FFT feature vectors to LPC cepstrum, lifted cepstrum, weighted likelihood ratio distances, as well as their linear combinations (Shikano and Itakura, 1991). Those methods were suggested by Kobatake and Matsunoo (1994) for improving recognition of words degraded by white noise. However, as with the results for HTK, in general the alternate representations also decreased the performance of the LCSR system.

We investigated the influence of HTK performance of the number of coefficients in the parametrizations, and the influence of target-rate (t-r) (the output sample rate which determines the period of the parameter vector) and window size (w-s) (segment of waveform used to determine each parameter vector). Our experiments have shown that the use of more than ten coefficients does not improve HTK performance. We also observed that the influence of target-rate and window-size was species dependent. The best results for zebra finches were achieved at t-r=6 ms and w-s=18 ms, whereas for indigo buntings best results were achieved at t-r=10 ms and w-s=25 ms. These differences probably arise because zebra finch vocalizations are broadband whereas indigo bunting vocalizations are narrow band. Based on these results, for most of our experiments we have used t-r=6/10 ms and w-s=18/25, depending on species, with the number of cepstral coefficients equal to 10.

E. Scoring

The performance of each recognizer was compared against a baseline of manually labeled vocalizations drawn from the database using auxiliary programs for scoring evaluation. The programs match each of the recognized and reference label sequences. Once the alignment of the two label files was found, the number of substitution (misrecognition) (S), deletion (D), and insertion (I) errors was calculated. The percentage correct (% correct) was calculated as 100% times the ratio of correctly recognized elements to a total number of song elements. This measure ignores insertion errors. To take insertion errors into account, the percentage accuracy (% accuracy) was computed as % correct minus percentage of insertion errors. The % correct on the song/file level was computed only for HTK as the percentage of completely correct recognized songs/files.

Although the main principles of scoring as applied to both recognizers are the same, they differ somewhat in detail. LCSR scoring allows one to constrain the discrepancy in the boundaries of the matched segments. In the LCSR tests reported here, the maximum time discrepancy (δ) was 100 ms. HTK scoring achieves a similar effect by performing optimal string matching (Sankoff and Kruskal, 1983) making the penalty for a substitution error (10) higher than penalties for insertion and deletion errors (7). We noted increasing δ beyond 100 ms hardly affected the numbers or distribution of LCSR errors. This implies that for δ=100 ms, the LCSR

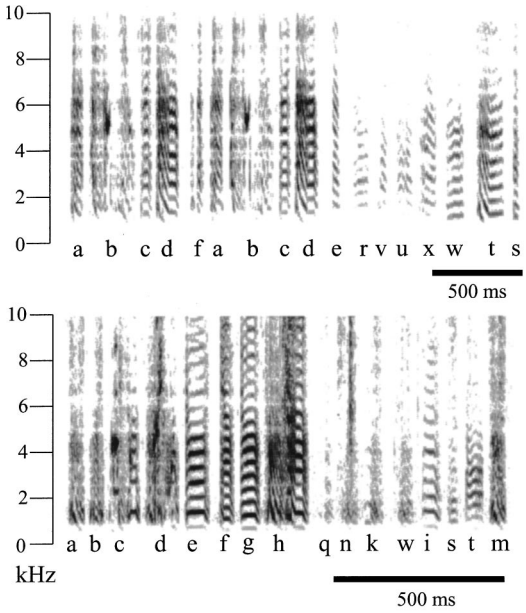


FIG. 1. Stereotyped song and additional vocalizations of zebra finch gr43 (top panel) and of zebra finch yl49 (bottom panel). The vocalization classes are represented for gr43 and all but two for yl49. ''*n*'' for yl49 is a cage noise.

score was determined by the sequence of elements and was therefore equivalent to the HTK score.

II. RESULTS OF EVALUATION

Both techniques were tested against continuous, unsegmented recordings of all animals in the database. Summaries of these tests are presented in Table III (for zebra finches) and Table IV (for indigo buntings). Unless otherwise noted, test sets included training (template) sets.

A. Comparison of recognition for zebra finch songs

Both LCSR and HTK exhibited almost perfect performance in recognition for relatively clean recordings of highly stereotyped parts of the vocalizations (e.g., major syllables of song motifs), but were vulnerable to noise and to variability in vocalizations. Performance also suffered during separation of short duration or low amplitude calls from noises such as cage hopping, bill wiping, etc. We observed that for adequate performance under noisy conditions, LCSR required many more templates for representing the variety of different cage noises and short calls which are not the part of the song than templates for stereotyped notes and syllables that comprise the bulk of song.

In general, a template set with five elements per class was sufficient to achieve high % correct scores of LCSR performance, but this measure ignores the relatively large number of insertion errors typically caused by mislabeling cage noises as short calls. Thus such small template sets were typically inadequate for achieving high accuracy. For example, for gr43 vocalizations, which included five short duration calls that were relatively confusing (see Fig. 1), LCSR accuracy was only 62.3% (accuracy) when tested with five templates for each class of elements. Increasing to 26 the number of templates for noises which were confused with calls or other vocalizations helped to reduce the number of insertion errors by a factor of 5. To reduce the number of substitution errors caused by misclassified elements of the motif, the number of templates for the introductory note $\{a\}$ was increased to 18 and for syllable $\{d\}$ to 9, and the number of templates for short confusing calls $\{r\}$ and $\{u\}$ were increased to 26 and 14, respectively. These changes completely eliminated the 35 substitution errors for $\{a\}$, eliminated two of three substitution errors for $\{d\}$, and reduced substitution errors for $\{r\}$ from 107 to 50, and for $\{u\}$ from 59 to 41. These changes also slightly increased the substitution errors for other calls, however. Another approach to resolving this problem which may have applicability for some experimental designs is to combine the confusing calls into one class. For gr43, combining confusing calls into one class improved LCSR accuracy to 80.0% for five templates per class and to 92.5% for the larger set of templates.

The same strategy was used to improve LCSR performance for zebra finches bu41 and yl49. For example, by selecting nine additional templates for confusing cage noises, and 24 additional templates for two confusing calls and one syllable, we dramatically improved LCSR performance for bu41, achieving a better score than achieved by HTK. For yl49, selecting 16 additional templates for confusing cage noises and 62 additional templates for the most confusing syllables $\{a,b,f,g\}$ and calls $\{i,s,w\}$, we dramatically improved the accuracy from 45.4% to 78.2%, with only modest increases in correctness (89.7% to 92.9%). The decreased insertion error was largely attributed to a drop in the number of insertions (from 1450 to 190) for a low-amplitude, short duration call $\{q\}$ that was manually labeled only three times out of 3732 vocalizations. The additional templates also helped to reduce the number of substitution and deletion errors for almost all elements. However, the number of insertions for some elements even increased, for example, for confusing calls $\{s\}$ (from 0 to 195), and $\{w\}$ (from 97 to 119). Further attempts to improve the performance for yl49 by increasing the template size for confusing elements did not give uniform results.

In most experiments, HTK exhibited better performance in accuracy than did LCSR, although the correctness of LCSR was higher for bu41 and yl49 using the expanded template sets (Table III). The primary reason for HTK's superior performance was that typically it produced fewer insertion errors than did LCSR. This effect was most noticeable for noisy recording as for yl49. On the other hand, LCSR could make fewer substitution and deletion errors, (e.g., bu41 and yl49), which partially offset the improvement with HTK. For comparison on the most challenging set of recordings (yl49) the number of errors for LCSR/HTK were as follows: 218/406 (substitution), 44/74 (deletion), 550/113 (insertion). The impact of the song elements on the number and type of errors was also quite different. For example, most of the substitution (and therefore deletion) errors for HTK were caused by confusing an introductory note $\{a\}$ and syllables of the motif with each other, and to a lesser extent from misclassifying calls within call classes, whereas for LCSR errors were mostly caused by confusing call $\{i\}$ with syllable $\{g\}$, and to a lesser extent by confusing introductory

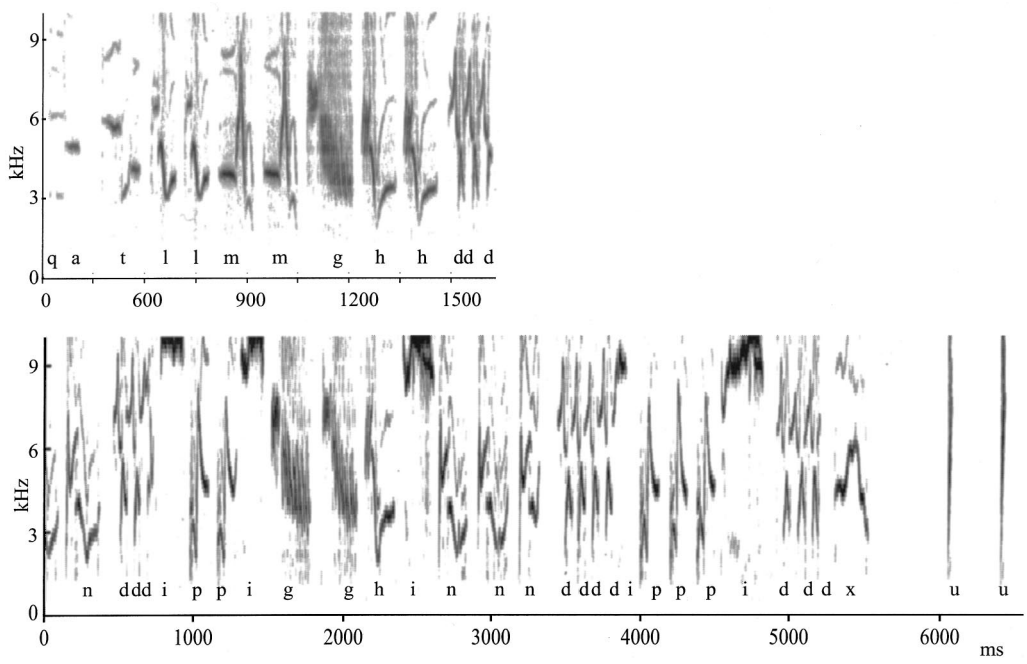


FIG. 2. Stereotyped song (top panel) and plastic song (bottom panel) for ib5, showing all 13 vocalization classes. Notes ''i'' of plastic songs are ''squeaky'' notes (see Margoliash *et al.*, 1991).

note $\{a\}$ with calls $\{w\}$, $\{s\}$, and syllable $\{b\}$. The percentage of insertion errors for LCSR/HTK were distributed across the song elements as follows: low amplitude calls $\{s\}$, $\{q\}$, and $\{w\}$: 35.5/34.6, 35.8/0.0, and 21.6/0.0, respectively; introductory note $\{a\}$: 1.3/10.5; and syllables $\{e\}$: 0.4/16.5, and $\{h\}$: 0.0/9.7.

The reason for the high level of substitution errors, which also often involved insertions and deletions, is that HTK (and LCSR) does not have a good mechanism for discriminating similar sounds. Therefore use of a larger number of exemplars per syllable in the training set did not guarantee better performance on the test set. For example, for gr43, only 16 exemplars for syllable $\{f\}$ were sufficient to obtain 100% accuracy for that syllable. However, 135 exemplars for syllable $\{a\}$ still gave five substitution and two deletion errors, and increasing the training set for $\{a\}$ to 241 exemplars still gave two substitution and one deletion errors. The same trend was noted in other experiments as well, where a large number of training exemplars did not necessarily guarantee good performance, and syllables with larger numbers of training exemplars did not necessarily have higher performance than did syllables with fewer numbers of training exemplars. For example, for yl49, 95/62 exemplars for syllables $\{a\}/\{d\}$ gave 180/18 substitution, 8/1 deletion, 2/0 insertion errors, respectively. Good performance depended more on homogeneity within the category. We did not investigate the use of discriminative procedures, but these may significantly improve error performance, and to some extent automate the selection of a representative set of templates for LCSR, or training set for HTK (see Sec. III).

B. Comparison of recognition for indigo bunting songs

The summary of LCSR and HTK performance for indigo buntings is given in Table IV. We experimented with both techniques using different numbers of templates or songs in the training sets. Both plastic and stereotyped songs were included in all training sets, and both types of songs were tested using the same training set. In general, performance at the syllable level was comparable comparing zebra finches and indigo buntings but was better for the latter at the song/file level because the finch recordings typically contained many songs per file whereas the bunting recordings typically contained only one song per file (Table I).

In general, both LCSR and HTK exhibited excellent performance for indigo bunting stereotyped songs (Table IV). LCSR exhibited marginally better performance than did HTK for stereotyped songs of one bird (ib5) (see Fig. 2). LCSR also exhibited poor performance on stereotyped songs of one bird (ib13). LCSR performance on plastic songs was also somewhat variable, ranging from excellent to poor depending on the bird (Table IV). In contrast, HTK exhibited excellent performance for stereotyped songs for all birds but one, and very good to excellent performance on plastic songs, significantly better performance than LCSR, for all birds (Table IV). The performance of HTK on plastic songs is a result of potential importance to biologists, because it demonstrates the potential of HMM-based approaches to properly represent variable vocalizations in juvenile and adult birds that learn songs (e.g., see Marler and Peters, 1982b; Margoliash *et al.*, 1994). It is noteworthy also that the robust performance of HTK was accomplished without enhancement of the input representation.

Excellent LCSR performance for bird ib7 has been previously reported by Anderson *et al.* (1996). We also achieved excellent LCSR performance for ib7 using a different set of templates (Table IV). As was noted by Anderson *et al.* (1996), LCSR performance on stereotyped and plastic indigo buntings songs was mixed, and depended on enhancement of the signal. This resulted in dramatic improvement of

performance on stereotyped songs, for example, for ib7 (from 54.6% to 97.3% in accuracy), but with some decrease in performance for plastic songs (87.6% to 82.6% accuracy) (Anderson *et al.*, 1996). We noted the same impact of enhancement on LCSR performance for other indigo bunting recordings (not shown), and therefore enhancement was used for all LCSR results given in Table IV. The opposite effects of enhancement on accuracy of recognition of high-amplitude stereotyped songs and low-amplitude plastic songs is another example of LCSR sensitivity to amplitude (hence S/N ratio) in the input representation. The slight decrease in LCSR performance on plastic songs probably occurs because the enhancement reduces not only the noise signal but the low-amplitude plastic song signal as well.

The poor performance of LCSR on ib13 resulted from a large number of insertion errors. Analysis of the confusion matrix (not shown) revealed that 665 out of 682 insertion errors for stereotyped songs, and 663 out of 733 insertion errors for plastic songs, were caused by a buzzy syllable $\{z\}$ which was present only four times in the manually scored plastic songs and absent from the stereotyped songs. As shown in Table IV, elimination of this sound, which could be considered an indistinct vocalization in the sense of Margoliash *et al.* (1991), improves accuracy of LCSR from 22.6% to 78.8%. HTK still performed better on ib13, however, without eliminating syllable $\{z\}$. In contrast, on stereotyped songs of ib5, HTK exhibited slightly worse performance than the excellent performance (97.3% accuracy) of LCSR. Interestingly, the high-amplitude stereotyped song (but not the low-amplitude plastic song) recordings of this bird were badly contaminated by aliasing noise, which created some discrepancies within the signal classes that trained HTK. In contrast, aliasing was removed when the stereotyped songs were enhanced prior to processing by LCSR. For ib5, most of the HTK errors resulted from a deletion (22 of 26 deletion errors) of part of a triplet or quadruplet of narrow-band syllables $\{d\}$. The typical error arose from substitution of a pair of $\{d\}$ syllables by another narrow-band syllable. Thus HTK is also susceptible to producing a relatively large number of insertion or deletion errors caused by one or more syllables recorded under noisy conditions, but on a much smaller scale than is LCSR.

To understand how the overlapping of the training and testing sets may influence the results of recognition, we also tested the performance of HTK under two extreme cases, when the training and testing sets were identical and when they did not overlap. For example, for identical sets HTK performed for ib5 as follows: for stereotyped songs 93.4% correct and 93.4% accuracy; for plastic songs 87.7 (95.2)% correct and 84.0 (94.1)% accuracy on the entire (or combined) set of syllables. Results of recognition for the same bird for nonoverlapping training and testing sets are slightly lower but generally very similar (Table IV). Thus since HTK learns a statistical model of the data, the absence of overlapping training and testing sets does not significantly degrade HTK performance, especially for stereotyped songs. One should note the differences when comparing the effects on LCSR performance of including templates in the test set. If a constituent in the test set which is also a template has not been improperly segmented by a prior error, LCSR is guaranteed to match that constituent to the template.

III. DISCUSSION

Bird songs represent relatively complex animal vocalizations, but such complexity is also known for the vocalizations of nonavian species. Our results highlight the main differences, strengths, and limitations of DTW-based and HMM-based approaches to recognition of bird song elements, and by extension to animal vocalizations in general. These conclusions serve to outline some important features which reliable animal recognition systems should inherit from the current recognizers.

A. Sensitivity and discriminative power of LCSR and HMMs

1. Template and training set sizes

As a rule, LCSR performance was determined by the discriminative power of the templates. In LCSR, since all features of sound spectrograms are weighted equally, more variable sounds required more templates for recognition. Thus a mere five templates per class were in general sufficient for reliable recognition of stereotyped syllables and calls of zebra finches, and only three templates sufficed for reliable recognition of stereotyped and even many of the plastic syllables of indigo buntings. On the other hand, further increase of the template size for these types of sounds did not give significant improvement in performance (see also Anderson *et al.*, 1996). The more variable types of sounds such as some calls and syllables, especially those of short duration and low amplitude, and cage noises, often required tens of carefully selected templates for proper discrimination. This limitation of LCSR has been previously noted by Anderson *et al.* (1996), who obtained an accuracy of 98.1% for the songs of a zebra finch by combining confusing calls into one class, increasing the number of templates for those calls to 13, and by using 27 templates for noises and nine for silent intervals.

The size of the training sets we used for HTK was usually much larger. However, as with the template set, simply increasing the training set size did not reliably lead to better performance. From 20 to 40 exemplars per syllable or call class were sufficient to obtain near-maximal performance, and subsequent increases in the training set resulted in only very modest improvements in recognition accuracy. Our actual training set was typically large, because we trained HTK on entire files (continuous recordings), and selected several files to provide sufficient exemplars for infrequently encountered vocalization classes. Training on continuous records is a natural approach for an HMM-based system.

2. Parametrization and normalization

The performance of the recognizers depends on the parametrization and amplitude normalization of the signals. Parameters which can affect LCSR performance are the resolution of FFT which determines the window and temporal step sizes. The choice of parameters depends on the specifics of a bird's vocalizations. In general, adjacent time windows

have to overlap at least 50%. A smaller step size (especially for stereotyped vocalization of indigo buntings) may reduce the number of deletions but increase the number of insertions (Anderson *et al.*, 1996). Conflicting effects on performance are also obtained in some cases when manipulating the minimum permitted syllable duration. In considering the number of templates chosen or the degree of window overlap, it is also important to remember that computational load increases as a square of the number of input patterns and template time frames.

The problem of efficient and parsimonious representation of acoustic signals has not received much attention in bird song research. Our experiments with various representations of bird songs have shown that nonlinear parametrizations are better than linear ones (e.g., LPC). Because linear representations take into account only the second-order dependencies in time (correlations or covariances) of the signal, they give smoothed averages of the signals. Therefore linear parametrizations are better suited for quasilinear signals such as in speech (Gidas and Murua, 1996), but are less appropriate for the rapid transients and nonlinearities which are common in bird songs. In contrast, nonlinear parametrizations are better at modeling bird songs because they can better represent the higher-order dependences of the signals (see Bell and Sejenowski, 1996). It is reasonable to expect that nonlinear parametrizations derived specifically for bird songs will give even better results.

Amplitude normalization of signals produced mixed results. On one hand, LCSR without normalization often gave results that were more correct but less accurate than HTK with normalization. This means that LCSR made less substitution and deletion errors which in HTK were caused by misrecognition of motif elements of relatively high amplitude. On the other hand, HTK made much fewer insertion errors than LCSR, especially for noisy recordings which caused LCSR to confuse (insert) low-amplitude short calls instead of noises. Interestingly, our experiments with LCSR using the double autocorrelation function (for LPC), which attempts to de-emphasize signal amplitude, often resulted in fewer insertion errors but at a cost of greatly increased substitution and deletion errors. Therefore normalization may be useful for recognition of low-amplitude sounds in relatively noisy recordings (e.g., as with field recordings), but may worsen the recognition of prominent song elements (as often achieved in laboratory recordings).

3. HMM structure

We evaluated increasing or decreasing the number of states and number of mixtures, and changing left-to-right models to recurrent models. In general, we observed that reduction in the number of states, for example, from five to four, improved the performance of recognition for shorter duration sounds. Reduction of the number of mixtures of distributions per state, for example, from four to two, typically worsened performance. Thus this procedure could potentially be considered in order to reduce the number of estimated parameters only in cases where a very limited amount of training data were available. We observed insignificant changes in performance with recurrent models. We conclude that using HTK, the default structure is best suited for modeling the longer duration (more stationary) bird song syllables and calls. For the shorter duration calls and notes it is often useful to reduce the number of states (which also has the effect of reducing the number of exemplars required for training). There is no mechanism within HTK, however, to select an optimal structure for a given acoustical pattern.

We also attempted to improve the discrimination of confusing syllables and calls using models having different number of states, as well as using recurrent models. These experiments did not give consistent results. It is noteworthy that easily confused sounds shared similar morphologies, that is they appeared as classes within a larger group. This can explain why improvement in discrimination of some sounds within such a group also resulted in an increase in the number of substitution errors for other sounds of the group. This represents another significant limitation of HTK. For HTK, training the statistical models for each song element is based on the maximum-likelihood estimation within the group but not across the groups, and therefore often lacks sufficient discriminative power to distinguish confusing transient calls and introductory notes of bird songs. In addition, in traditional HMMs the transition probabilities of the underlying Markov chains have smaller effects on the model than do the observation probabilities. Taking into account that the observation probabilities are assumed to be conditionally independent explains our result that HTK models are much better suited for representing more stationary rather than more transient bird song sounds (cf. Bourlard *et al.*, 1994; Gidas and Murua, 1996).

B. Further development of LCSR and HMMs for automated analyses of bird vocalizations

Both LCSR and HTK can be successfully used to automate segmentation (labeling) of bird song elements of nonoverlapping continuous vocalizations recorded in the laboratory. However, both techniques produce sufficient errors that may limit their utility in daily laboratory use. Here we suggest that extracting more biologically meaningful information from bird song recordings may significantly improve performance.

1. Discriminative features extraction

One major drawback of both techniques is the lack of a method to find discriminative features from the given signals. This limitation results in many obvious errors of song element misrecognition. For both LCSR and HTK, this limitation can be addressed by the discriminative training method (Chang and Juang, 1993; Juang *et al.*, 1997). This approach weighs more heavily those features which distinguish between classes. For example, subtly distinct budgerigar calls were separated using DTW applied to the frequencies of the two most intense peaks of each call with a specially designed scoring function of frequency tolerance (Ito *et al.*, 1996). The specific model used by Ito *et al.* (1996) was designed for budgerigar calls, however, and presumably is not generally valid for all birds. Therefore it would be valuable to have a technique that finds such features, which can potentially be of significant biological interest, with a

minimum of prior knowledge. For HMMs, one attractive approach may be to use the so-called hybrid HMM/ANN models, which switch the emphasis from observation probabilities to transition probabilities (Bourlard *et al.*, 1994). These hybrid models use an artificial neural network as an input to an HMM which are then also discriminatively trained.

2. Bird song structure

Additional improvement in performance can be gained by learning more of the structure of the bird songs of the training set. For example, even the simplest bigrammar model significantly improved HTK performance for both stereotyped and plastic songs. This implies that the probabilistic song structure can be learned based on Markov dependencies (co-occurrences) and that longer dependencies of song elements can be approximated as a product of those Markov transitions. It would therefore be valuable to automate the learning of the song probabilistic structure, for example, using probabilistic automata with a fixed length memory such as *n*-grammars (Rabiner and Juang, 1993), or more biologically meaningful ones with a variable length memory (Ron *et al.*, 1996).

More generally, hierarchical dependences of bird songs are well established, even at the physiological level (e.g., see Yu and Margoliash, 1996). This potent source of information regarding song structure is ignored in the current models, however, often causing obvious errors (e.g., song level recognition for HTK, see Tables III and IV). This limitation can be addressed by embracing appropriate hierarchical HMMs (hHMMs) (Hihi and Bengio, 1995; Fine *et al.*, 1998), or possibly by applying dynamic programming in multiple layers (see Ney *et al.*, 1992). hHMMs would learn (and hence, describe) the hierarchy from the training examples.

3. Self-learning

Finally, a significant limitation of LCSR and HTK is that the training and template creation stages are separated from recognition. As a result, classes of elements which are not present in the training/template set will be misrecognized. Presently there are no mechanisms to automatically identify new classes and thereafter include them in the training/template set. As well as a technical problem (missing classes in a fixed repertoire), this is also a biologically important problem (as when birds acquire new vocalizations). It would be very valuable to modify the one stage DTW (token passing) algorithm in such a way as to include adaptive thresholds on the discrepancies in measures for matching the input stream against the reference (training) patterns. Such thresholds will not only prevent these algorithms from often obvious errors, but will also allow a closed form of self-learning: first, by incorporating unrecognized elements in a new training/template set; second, by adapting the thresholds to the modified training/template set. Such an algorithm would have useful applications in many areas.

IV. SUMMARY

Massive databases are common in bioacoustic studies of animal vocalizations. Automated analyses may significantly facilitate such studies. We explored the application of DTW and HMM recognition techniques to a large database of continuous bird song recordings. Our experiments highlight the advantages and limitations of the two approaches. Excellent performance was achieved under some conditions. Nevertheless, further development is needed to improve the automated recognition of short duration bird song sounds, and to improve performance under adverse conditions.

ACKNOWLEDGMENTS

We thank Sven E. Anderson and Alexander Peryshkin for help in using and modifying LCSR, Amish Dave for modifying some auxiliary code, and Han Y. Kim for extensive work preparing the vocalization database. S.E.A. provided valuable comments on the manuscript. This work was supported in part by U.S. Army Research Office (DACA88-95-C-0016) and the NIH (NS25677).

APPENDIX A: BAUM–WELCH REESTIMATION

At the initialization stage, the observation vectors are equally subdivided among the model states for each label, then the initial MLE of the mean and variance for each state j are computed as averages

$$\hat{\mu}_j=\frac{1}{T}\sum_{t=1}^{T} o_t, \quad \hat{\Sigma}_j=\frac{1}{T}\sum_{t=1}^{T}(o_t-\mu_j)(o_t-\mu_j)', \qquad \text{(A1)}$$

where the prime denotes the vector transpose. Then, the ML state sequence is found by using the Viterbi algorithm (see Appendix B), the observation vectors are reassigned to the states that were found, and Eq. (A1) is used again to get better estimates. This process is repeated until the estimates converge below some threshold value.

Because each observation vector o_t may contribute to the ML parameter values for each state j, the initial estimates are then improved by assigning each observation to every state in proportion to the probability $L_j(t)$ of the model being in that state j when the vector is observed at time t

$$\begin{aligned}\hat{\mu}_j&=\sum_{t=1}^{T} L_j(t)o_t \Big/ \sum_{t=1}^{T} L_j(t),\\ \hat{\Sigma}_{\mathbf{j}}&=\sum_{t=1}^{T} L_j(t)(o_t-\mu_j)(o_t-\mu_j)' \Big/ \sum_{t=1}^{T} L_j(t),\end{aligned} \qquad \text{(A2)}$$

where $L_j(t)=P(x(t)=j/\mathbf{O},\mathbf{H})$ and is computed by using the Forward–Backward algorithm (Rabiner and Juang, 1993). Equations (A2) are the Baum–Welch reestimation formulas for the means and covariances of an HMM. Similar formulas can be derived for reestimation of the coefficients of the mixtures and the transition probabilities. These Baum–Welch formulas are used in HTK first for reestimation of the parameters of each single HMM by training isolated labels, and then for embedded training when all HMMs are trained in parallel using the entire training set.

APPENDIX B: VITERBI ALGORITHM

Given an HMM, let $\pi_j(t)$ represent the maximum log likelihood of observing acoustic vectors $\{o_1,\ldots,o_t\}$ and being in state j at time t. Then, $\pi_j(t)$ can be efficiently computed using the following dynamic programming equation:

$$\pi_j(t)=\max_i\{\pi_i(t-1)+\log(a_{ij})\}+\log\, b_j(o_t), \qquad \text{(B1)}$$

where $\pi_1(1)=0$, and $\pi_j(1)=\log(a_{1j})+\log b_j(o_1)$, for $j\neq 1$. The MLE state sequence $\hat{X}$ is then retrieved by keeping track of the argument which maximizes Eq. (B1) for each j and $1<t\leqslant T$

$$f_j(t)=\operatorname{argmax}_i\{\pi_i(t-1)+\log(a_{ij})\}$$

by backtracking, $\hat{\imath}_t=f_{t+1}(\hat{\imath}_{t+1})$,

$$t=T-1,\ldots,1, \text{ starting from } \hat{\imath}_T=\operatorname{argmax}_i\{\pi_T(i)\}.$$

Anderson, S. E., Dave, A. S., and Margoliash, D. (**1996**). ''Automated recognition and analysis of birdsong syllables from continuous recordings,'' J. Acoust. Soc. Am. **100**, 1209–1219.

Bell, A. J., and Sejnowski, T. J. (**1996**). ''Learning the higher-order structure of a natural sound,'' Comp. Neural Sys. **7**, 261–267.

Bourlard, H., Konig, Y., and Morgan, N. (**1994**). ''REMAP: Recursive estimation and maximization of a posteriori probabilities, application to transition-based connectionist speech recognition,'' Technical Report TR-94-064, International Computer Science Institute, Berkeley, CA.

Bridle, J. S., Chamberlain, R. M., and Brown, M. D. (**1982**). ''An algorithm for connected word recognition,'' in *Proceedings of IEEE Conference on Acoustics, Speech, and Signal Processing, Paris, France* (IEEE, New York), pp. 899–902.

Buck, J. R., and Tyack, P. L. (**1993**). ''A quantitative measure of similarity for *Tursiops truncatus* signature whistles,'' J. Acoust. Soc. Am. **94**, 2497–2506.

Catchpole, C. K., and Slater, P. J. B. (**1995**). *Bird Song: Biological Themes and Variations* (Cambridge U.P., Cambridge, England).

Chang, P. C., and Juang, B. H. (**1993**). ''Discriminative training of dynamic programming based speech recognizers,'' IEEE Trans. Speech Audio Process. **1**, 135–143.

Deller, J. R., Proakis J. G., and Hansen J. H. L. (**1993**). *Discrete-Time Processing of Speech Signals* (Prentice-Hall, Upper Saddle River, NJ).

Fine, S., Singer, Y., and Tishby (**1998**). ''The hierarchical hidden Markov model: analyses and applications,'' Machine Learning (to be published).

Forney, Jr., G. D. (**1973**). ''The Viterbi algorithm,'' Proc. IEEE **61**, 268–278.

Gidas, B., and Murua, A. (**1996**). ''Stop consonant discrimination and clustering using nonlinear transformations and wavelets,'' in *Image Models (and Their Speech Model Cousins)*, edited by S. E. Levinson and L. Shepp (Springer-Verlag, New York), pp. 13–62.

Hihi, S. E., and Bengio, Y. (**1995**). ''Hierarchical recurrent neural networks for long-term dependencies,'' in *Advances in Neural Information Processing Systems 7*, edited by G. Tesauro, D. Touretzky, and T. Leen (Morgan Kaufmann), pp. 493–499.

Ito, K., Mori, K., and Iwasaki, S.-i. (**1996**). ''Application of dynamic programming matching to classification of budgerigar contact calls,'' J. Acoust. Soc. Am. **100**, 3947–3956.

Juang, B. H., Chou, W., and Lee, C. H. (**1997**). ''Minimum classification error rate methods for speech recognition,'' IEEE Trans. Speech Audio Process. **5**, 257–265.

Kobatake, H., and Matsunoo, S. (**1994**). ''Degraded word recognition based on segmental signal-to-noise ratio weighting,'' Proc. IEEE ICASSP I-425–I-428.

Kogan, J. A. (**1996**). ''Hidden Markov models estimation via the most informative stopping times for the Viterbi algorithm,'' in *Image Models (and Their Speech Model Cousins)*, edited by S. E. Levinson and L. Shepp (Springer-Verlag, New York), pp. 115–130.

Konishi, M. (**1965**). ''The role of auditory feedback in the control of vocalization in the white-crowned sparrow,'' Z. Tierpsychol. **22**, 770–783.

Kubala, J. (**1995**). ''Design of the 1994 CSR Benchmark Tests,'' in *Spoken Language Systems Technology*, Workshop (ARPA), pp. 41–46.

Larkin, R. P., Margoliash, D., Kogan, J. A., and Pater, L. L. (**1996**). ''Recognition of the utterances of terrestrial wildlife: A new approach,'' J. Acoust. Soc. Am. **99**, 2532.

Makhoul, J., and Schwartz, R. (**1995**). ''State of the art in continuous speech recognition,'' Proc. Natl. Acad. Sci. USA **92**, 9956–9963.

Margoliash, D., Staicer, C., and Inoue, S. (**1991**). ''Stereotyped and plastic song in adult indigo buntings, *Passerina cyanea*,'' Animal Beh. **42**, 367–388.

Margoliash, D., Staicer, C., and Inoue, S. (**1994**). ''The process of syllable acquisition in adult indigo buntings, *Passerina cyanea*,'' Behavior **131**, 39–64.

Marler, P., and Peters, S. (**1982a**). ''Developmental overproduction and selective attrition: New processes in the epigenesis of birdsong,'' Dev. Psychobiol. **15**, 369–78.

Marler, P., and Peters, S. (**1982b**). ''Subsong and plastic song: Their role in the vocal learning process,'' in *Acoustic Communication in Birds, Vol. 2, Song Learning and Its Consequences*, edited by D. E. Kroodsma and E. H. Miller (Academic, New York), pp. 25–50.

McLachlan, G. (**1992**). *Discriminative Analysis and Statistical Pattern Recognition* (Wiley, New York).

Ney, H. (**1984**). ''The use of a one-stage dynamic programming algorithm for connected word recognition,'' IEEE Trans. Acoust., Speech, Signal Process. **29**, 284–297.

Ney, H., Mergel, D., Noll, A., Paeseler, A. (**1992**). ''Data driven search organization for continuous speech recognition,'' IEEE Trans. Signal Process. **40**, 272–281.

Payne, R. B., Thompson, W. L., Fiala, K. L., and Sweany, L. L. (**1981**). ''Local song traditions in indigo buntings: cultural transmission of behavior patterns across generations,'' Behaviour **77**, 199–221.

Rabiner, L. R., and Juang, B. H. (**1993**). *Fundamentals of Speech Recognition* (Prentice-Hall, Englewood Cliffs, NJ).

Ron, D., Singer, Y., and Tishby, B. (**1996**). ''The power of amnesia: learning probabilistic automata with variable memory length,'' Machine Learning **25**, 117–149.

Sankoff, D., and Kruskal, J. B. (**1983**). *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison* (Addison-Wesley, Reading, MA).

Shikano, K., and Itakura, F. (**1991**). ''Spectrum distance measures for speech recognition,'' in *Advances in Speech Signal Processing*, edited by S. Furui and M. Sondhi (Marcel Dekker, New York), pp. 419–452.

Silverman, H. F., and Morgan, D. P. (**1990**). ''The application of dynamic programming to connected speech recognition,'' IEEE ASSP Mag. **7**, 7–24 (July).

Sossinka, R., and Böhner, J. (**1980**). ''Song types in the zebra finch (*Poephila guttata castanotis*),'' Z. Tierpsychol. **53**, 123–132.

Thompson, W. L. (**1970**). ''Song variation in a population of indigo buntings,'' Auk. **87**, 58–71.

Thorpe, W. H. (**1958**). ''The learning of song patterns by birds, with especial reference to the song of the chaffinch, Fingilla coelebs,'' Ibis **100**, 535–570.

Vintsyuk, T. K. (**1971**). ''Element-wise recognition of continuous speech composed of words from a specified dictionary,'' Kibernetika **7**, 133–143.

Viterbi, A. J. (**1967**). ''Error bounds for convolutional codes and an asymptotically optimal decoding algorithm,'' IEEE Trans. Inf. Theory **IT-13**, 260–269.

Williams, H. (**1990**). ''Bird song,'' in *Neurobiogy of Comparative Cognition*, edited by R. C. Kesner and D. C. Olton (Erlbaum, Hillsdale, NJ), pp. 77–125.

Young, S., Odel, J., Ollason, D., Valtchev, V., and Woodland, P. (**1995**). *The HTK Book* (Cambridge University Technical Services Ltd., Cambridge, England).

Young, S., Russel, N. H., and Thorton, N. H. (**1989**). ''Token passing: a simple conceptual model for connected speech recognition system,'' Technical Report, Cambridge University Engineering Department.

Yu, A. C., and Margoliash, D. (**1996**). ''Temporal hierarchical control of singing in birds,'' Science **287**, 1871–1875.

Zann, R. A. (**1996**). *The Zebra Finch* (Oxford U.P., Oxford).

Acoustic and articulatory correlates of stop consonants in a parrot and a human subject

D. K. Patterson
Department of Psychology, University of Arizona, Tucson, Arizona 85721

I. M. Pepperberg
Department of Ecology and Evolutionary Biology and Department of Psychology, University of Arizona, Tucson, Arizona 85721

(Received 14 June 1997; accepted for publication 2 January 1998)

General acoustic patterns in productions of American English stops /p,b,t,d,k,g/ by a Grey parrot (*Psittacus erithacus*) and his primary human trainer are identified. Comparisons demonstrated both differences (coherence of voicing [/p,t,k/ and /b,d,g/] and place [/p,b/, /t,d/, and /k,g/] subsets, predictive power of measures related to F_1 and F_3) and similarities (categorical distinctions among stop consonants) in acoustic properties of psittacine and human speech. The authors suggest how acoustic data might be correlated with articulatory events in the parrot. Our data are also compared with speech from a mynah (*Gracula religiosa*). Because humans perceive and correctly label psittacine stop consonants despite differences in avian and human articulatory and acoustic parameters, our findings (a) further expand our initial work with vowels [Patterson and Pepperberg, J. Acoust. Soc. Am. **96**, 634–648 (1994)], (b) corroborate findings suggesting that the perceptual discontinuities exploited by speech may be accessible to various sensory systems, and (c) challenge the notion that human speech production is a unique process. [S0001-4966(98)01904-3]

PACS numbers: 43.80.Ka, 43.70.Fq [FD]

INTRODUCTION

How mechanisms of ''speech'' production by psittacids compare with those of humans poses questions to several disciplines. Physiologists, anatomists, and phoneticians must explain how perceivable speech is produced by a biological system with neural control centers and articulatory apparatuses that differ significantly from those of humans. Only a few studies provide avian data to compare with human data. Researchers have identified neuroanatomical structures responsible for conspecific vocalizations in budgerigars (*Melopsittacus undulatus*, Durand *et al.*, 1997; Striedter, 1994) but are only now examining how these structures may be used for production, learning, or storage of human speech (Banta, in prep). As for other species, we know only that anatomical structures similar to those in budgerigars exist in Grey parrots (*Psittacus erithacus*, Banta, personal communication). Nottebohm (1976) and Homberger (1979, 1986) described parts of the physical system that parrots use to produce ''speech,'' but not the relationships between anatomy and speech acts that are needed for comparison with humans. By examining cineradiographic data of the psittacine vocal tract in action, Scanlan (1988) and Warren *et al.* (1996) described physical changes correlated with speech, but could not determine how these changes are responsible for production.

To characterize psittacine production mechanisms, we compare acoustic and articulatory parameters of avian and human utterances. In Patterson and Pepperberg (1994), we found both parallels (e.g., separation of vowels into front and back categories) and striking differences (e.g., a relatively invariant psittacine F_1) between Grey parrot and human production of American English vowels. The next step has been to examine Grey parrot stop production.

To produce stops in meaningful minimal pairs,[1] birds must first discriminate among and appropriately categorize such stops. Studies of nonhuman perception of stops show that some aspects of speech processing are not uniquely human. Chinchillas (*Chinchilla laniger*) taught to respond to the end points of synthetic voice onset timing (VOT) continua (voiced /b,d,g/versus voiceless /p,t,k/) then divide intermediate VOT continua in a manner comparable to English-speaking human listeners (Kuhl and Miller, 1975, 1978). Dooling *et al.* (1989) report similar abilities in budgerigars. Macaques (*Macaca fuscata*, *Macaca nemistrina*) tested on F_2 onset continua corresponding to place distinctions (/b,d,g/) exhibit roughly human phonetic boundaries (Kuhl and Padden, 1982; Morse and Snowdon, 1975; Waters and Wilson, 1976). Japanese quail (*Coturnix coturnix japonica*) taught to categorize natural syllables (/b,d,g/ plus four different vowels) correctly categorize the same stops preceding eight novel vowels (Kluender *et al.*, 1987). No single invariant feature could explain the quails' performance, suggesting that their success was based on formation of polymorphous acoustic categories. Moreover, when tested for effects of F_1 onset frequency on stop discrimination, quail performed at about human levels (Kluender, 1991; Kluender and Lotto, 1994); the effects on perception of the voiced–voiceless distinction was explained by general auditory processes.

Studies comparing human and nonhuman stop *production* are limited to mimetic birds. Klatt and Stefanski (1974) report some stop data for a mynah (*Gracula religiosa*); Scanlan (1988), using radiologic techniques on a Grey parrot,

described how laryngeal and syringeal valve closure may contribute to sublaryngeal pressure needed for a plosive-like /k/. The limited lexicons of birds in these studies (see also, Silaeva, 1995) prevented identification of more general acoustic and articulatory patterns needed for cross-species comparisons. Our Grey parrot subject, Alex, however, produces over 100 referential English utterances (Pepperberg, 1990a); we thus could examine a large corpus of psittacine word- and syllable-initial stops.

We studied voiced stops /b,d,g/ and voiceless stops /p,t,k/. We primarily examined word-initial stops, which have been studied extensively (Lauter *et al.*, 1986; Lisker and Abramson, 1991; Pickett, 1991a, b) because of their well-defined acoustic characteristics. Word-internal stops, in contrast, are more likely than word-initial stops to be reduced to flaps or glottal stops. Word-final stops often lack burst releases and are thus difficult to characterize. By examining sonagrams and waveforms of Alex's utterances and comparing them statistically with each other and with equivalent utterances by his primary trainer, we describe key features that differentiate Alex's stops and identify similarities and differences between parrot and human speech.

Our study not only describes Alex's stops but may also provide insights into acoustic correlates of human stop perception. To produce stops that humans can recognize, parrots must duplicate essential characteristics of these sounds, but not necessarily those commonly used to characterize human speech (Ohala, personal communication). Extrapolating from data on Alex's vowels (Patterson and Pepperberg, 1994), we expected each of his stops to be acoustically distinct, and context to be more important for characterizing Grey parrot than human stops.

We expected to see our conjectures borne out in two ways. First, we anticipated that traditional acoustic measures used to characterize stops would be similar for Alex and his principal human trainer, but that, overall, relationships between these measurements and stops would be more robust for the trainer. Second, we expected correlations between stops and measurements associated with context (e.g., formant targets) to be stronger for Alex.[2]

Based on acoustic evidence and published models (e.g., Kuhn, 1975; Dorman *et al.*, 1977), we also propose correlations between acoustic and articulatory events during Alex's stop production. For humans, energy peaks associated with each stop (stop loci) are used to estimate length (i.e., resonant properties) of the upper (oral) cavity in a two-chamber model (Kuhn, 1975; Stevens, 1993). Warren *et al.* (1996) proposed a similar two-chamber model for Alex's vowel production. Sound generated in the avian syrinx resonates in the lower tube, the trachea (Fig. 1). At the top of the trachea is the larynx. The glottis, a slit in the larynx, is not a sound source but can be constricted to stop air flow; it forms the border with the upper chamber, i.e., the oropharyngeal cavity, hinged beak, flexible tongue, and nasal passage (choana) that connects to sinuses above the palate. To produce stops, Alex may alter upper chamber resonances by opening or closing glottis and beak, moving tongue and larynx, and possibly using the esophagus as a side branch resonator (Patterson *et al.*, 1997). Calculations similar to those used for humans, based on stop loci, might thus predict Alex's upper cavity lengths.

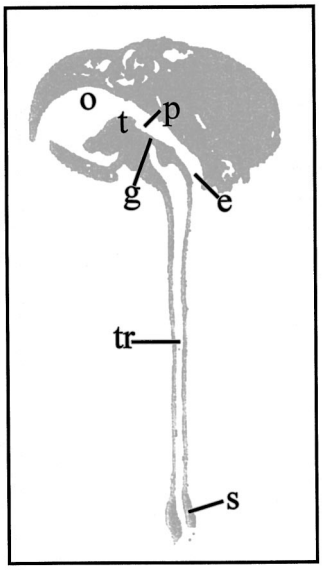


FIG. 1. Lateral view of *Psittacus erithacus* head and vocal tract. e=esophagus, g=glottis, o=oral cavity, p=pharynx, s=syrinx, t=tongue, tr=trachea; o+p=oropharyngeal cavity.

Finally, we compare Alex's utterances to those of a mynah, Ig-wog. Although limited by Ig-wog's small repertoire (Klatt and Stefanski, 1974), we contrast frequency and timing data for these subjects. We suggest that the observed acoustic differences may relate to their particular anatomies.

I. EXPERIMENT

A. Subjects

The primary avian subject, Alex, a Grey parrot (*Psittacus erithacus*), uses English speech to identify, request, refuse, categorize, quantify, or comment upon >100 objects; his repertoire includes labels for colors, shapes, materials, numbers, and categories (Pepperberg, 1981, 1990a). He also uses functional phrases (e.g., ''look,'' ''I want X,'' ''Wanna go Y''; X and Y, respectively, are object or locale labels, Pepperberg, 1988). He reliably produces minimal word pairs under distinct conditions (/pi/ ''pea,'' /ki/ ''key,'' etc.; Fig. 2). We compared his utterances with equivalent samples by his primary trainer, Irene M. Pepperberg (IMP), a native speaker of American English (Northeast dialect).

B. Methods

1. Recording and analysis equipment

Alex and IMP were audiotaped with a Sony TCM 5000 recorder, AKG CK8 microphone, and Maxell XL-UDII tapes. Alex's speech samples were obtained by questioning (e.g., Pepperberg, 1990a) or from spontaneous utterances both in the presence and absence of trainers. The latter material was documented via a voice-activated tape recorder (Pepperberg *et al.*, 1991).[3] To perform acoustic analyses on a Kay 5500 DSP Sona-Graph, we transferred signals from the Sony or a Marantz PMD 201 recorder. We noted no difference in sound quality in the two recorders.

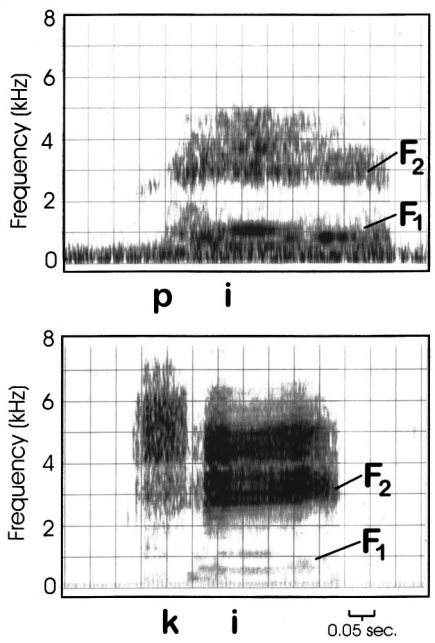


FIG. 2. The minimal pair /pi/ and /ki/ produced by *Psittacus erithacus* (Alex).

We videotaped Alex's speech acts with a Panasonic SVHS AG-450 camera and Maxell XR-S120 SVHS tape (30 frames/s). We analyzed a 20 min tape frame-by-frame and obtained stills using a Panasonic AG 1960 proline deck and CT-2082y color monitor. From the stills, we traced externally visible articulatory correlates for each vocalization.

2. Transcription for analysis and interobserver agreement of categorization

One of us (DKP) chose examples of stops from the audiotapes and transcribed samples using the International Phonetic Alphabet. DKP had considerable experience transcribing human and psittacine utterances. Standard American English equivalents of this alphabet are /p/=*p*at, /b/=*b*at, /t/=*t*ap, /d/=*d*ab, /k/=*c*ap, /g/=*g*ap. For vowels that preceded or followed our stops, equivalents are /i/=h*ee*d, /ɪ/=h*i*d, /ɛ/=h*ea*d, /æ/=h*a*d, /ɒ/=h*o*t, /ə/=h*u*t, /ɚ/=h*ur*t, /o/=h*oe*, /ʊ/=h*oo*d, /u/=wh*o*'d.

We performed two checks on the transcription and categorization process. In one check, questionable samples of each stop were classified independently by other trainers; samples were deleted if DKP and these trainers disagreed. As a separate test of inter/intraobserver reliability (CATMOD,[4] SAS, 1989), we made a tape (Signal,[5] Version 2.2, Beeman, 1993) of six samples of each stop for both IMP and Alex; we isolated samples of each stop from a variety of word-initial contexts (equivalent for both subjects). Samples were five repetitions of the utterance separated by 0.75 s; 5 s elapsed between samples. Six listeners (four for Alex's tape; two for IMP's) identified stops from relevant choices on a data sheet. Listeners could stop the tape or replay samples. Listener familiarity with our subjects' speech varied. Intraobserver reliability scores for listeners (CB, IMP) with >2-yr experience with Alex were 100% and 97%, respectively; interobserver agreement was 97%; their respective agreement with Alex's actual utterances was 100% and 97%.[6] Intraobserver reliability scores for listeners (MM, AP) with <2-yr experience with Alex were 84% and 88%, respectively; interobserver agreement was 71%; their respective agreement with Alex's actual utterances was 87% and 85%. Intraobserver reliability scores for listeners (PB, DM) with >2-yr experience with IMP were both 93%, respectively; interobserver agreement was 89%; their respective agreement with IMP's actual utterances was 95% and 91%.

3. Data

We collected at least 30 samples of each of the six stops for Alex and IMP (Table I). For two reasons, we collected extra samples of certain stops. First, Alex produced some stops (e.g., /k/) in several contexts and, by including multiple samples of each, we characterized each stop's full range of variation. Second, not all information could be read from sonagrams of all samples; by using more samples, we obtained adequate data. Note that despite Alex's extensive repertoire, he did not produce all stop/vowel combinations represented in a ''standard minimal data'' set.[7] For purposes of comparison, we collected the same number of equivalent utterances for both subjects.[8]

Most studies of human stops exclusively use word-initial stops in single word samples. All our samples were word-initial, except for 11 /p/ samples (22% of /p/, 3% of total samples), which were taken from the second syllable of ''paper'' because Alex's initial stop in this label was never clearly identified by listeners or its acoustic characteristics. Because he embedded some words in phrases (e.g., ''calm *down*''), half our samples were phrase-internal.

Videotapes provided data on Alex's possible coarticulation. We made three tracings of /k/ in back vowel (/kork/, ''cork'') versus front vowel (/ki/, ''key'') contexts; we chose /k/ because Alex's /k/ had a high amplitude and we could easily elicit samples. For analysis, we selected tracings of stills made as close to the release of the stop as possible.

II. PROCEDURES OF ANALYSIS

A. Digital signal processing

We analyzed stop signals using real-time fast Fourier transforms. We displayed signals in three formats: (1) broadband (300 Hz/100 point) sonagrams, (2) zoomed waveform displays (31.2 ms), and (3) between-cursors power spectra (300 Hz/100 points, flat shaping). Sonagrams provide data on voicing, location and number of bursts, stop loci, and formant transitions and targets. Zoomed time waves provide convergent evidence about the beginning of voicing and location and number of bursts, with more precise temporal resolution than sonagrams. Between-cursors power spectra enabled us to corroborate frequency information in the sonagrams.

We chose the specific sonagraph settings for five reasons. First, these settings were previously used to analyze F_1 and F_2 for vowels (Patterson and Pepperberg, 1994) and thus provided consistency in our analyses. Second, although Halle *et al.* (1957) used a 150-Hz bandwidth to determine stop loci, this choice obscured temporal information in our son-

TABLE I. Distribution of consonant contexts for Alex and IMP. Number in parenthesis refers to total number of instances for that category. Specific tokens are reported in footnote g.

	Consonant					
	Voiced			Voiceless		
Vowel	[b][a]	[d][b]	[g][c]	[p][d]	[t][e]	[k][f]
Front	(31)	(30)	(0)	(8)	(18)	(26)
i	7	⋯	⋯	1	1	22
I	5	30	⋯	2	4	⋯
ɛ	10	⋯	⋯	5	13	4
æ	9	⋯	⋯	⋯	⋯	⋯
Back	(18)	(8)	(32)	(30)	(19)	(39)
ɒ	10	8	⋯	30	1	2
o	4	⋯	6	⋯	⋯	12
ə	4	⋯	7	⋯	⋯	25
U	⋯	⋯	19	⋯	⋯	⋯
u	⋯	⋯	⋯	⋯	18	⋯
Other	(15)	(0)	(2)	(11)	(18)	(25)
r	⋯	⋯	2	⋯	14	⋯
l	1	⋯	⋯	⋯	⋯	10
m	⋯	⋯	⋯	⋯	⋯	13
ɚ	14	⋯	⋯	11	⋯	⋯
ø	⋯	⋯	⋯	⋯	4	2

[a]Word/phrase initial, 32 samples; word initial but phrase internal, 32 samples; word internal, 0 samples.
[b]Word/phrase initial, 0 samples; word initial but phrase internal, 38 samples; word internal, 0 samples.
[c]Word/phrase initial, 26 samples; word initial but phrase internal, 8 samples; word internal, 0 samples.
[d]Word/phrase initial, 8 samples; word initial but phrase internal, 30 samples; word internal, 11 samples.
[e]Word/phrase initial, 30 samples; word initial but phrase internal, 25 samples; word internal, 0 samples.
[f]Word/phrase initial, 58 samples; word initial but phrase internal, 32 samples; word internal, 0 samples.
[g]**bi**: ''bean'' /bin/ 1; /biwi/ 1; ''beeper'' /bipɚ/ 5; **bI**: /bIr/ 2; /bIri/ 3; **b**ɛ: ''bear'' /bɛr/ 2; ''better'' /bɛtɚ/ 1; **b**æ: ''back'' /bæk/ 9; /bæm/ 8; **b**ɒ: /bɒ/ 6; ''box'' /bɒks/ 2; ''bye'' /bɒi/ 1; **bo**: ''boy'' /boi/ 4; **bə**: /bənɛl/ 1; /bənɛri/ 3; **bl**: ''block'' /blɒk/ 1; **b**ɚ: ''bread'' /bɚɛd/ 12; ''bird'' /bɚd/ 1; ''birdy'' /bɚdi/ 1. **dI**: ''dinner'' /dInɚ/ 30; **d**ɒ: ''down'' /dɒun/ 8. **go**: ''go'' /go/ 6; **gə**: ''gonna'' /gənə/ 7; **gU**: ''good'' /gUd/ 19; **gr**: ''grain'' /grɛn/ 1; ''green'' /grin/ 1. **pi**: ''pea'' /p^{h}i/ 1; **pI**: ''pick'' /p^{h}Ik/ 2; **p**ɛ: ''parrot'' /p^{h}ɛrIt/ 5; **p**ɒ: /p^{h}ɒ/ 21; ''popcorn'' /p^{h}ɒk^{h}orn/ 7; ''pot'' /p^{h}ɒt/ 2; **p**ɚ: ''paper''/ʔep^{h}ɚ/ 11. **ti**: ''tea'' /t^{h}i/ 1; **tI**: ''tickle'' /t^{h}Ikə1/ 4; **t**ɛ: ''tell'' /t^{h}ɛ1/ 13; **t**ɒ: ''talk'' /t^{h}ɒk/ 1; **tu**: ''two'' /t^{h}u/ 18; **tr**: ''training'' t^{h}renin/ 1; ''tray'' /t^{h}re/ 6; ''truck'' /t^{h}rək/ 3; ''try''/ t^{h}rai/ 3; **t**: /t^{h}/ 5. **ki**: ''key'' /k^{h}i/ 21; ''kiwi'' /k^{h}iwi/ 1; **k**ɛ: ''carrot'' /k^{h}ɛrIt/ 3; ''carry'' /k^{h}ɛri/ 1; **k**ɒ: ''calm'' /kɒlm/ 1; /k^{h}ɒ/ 1; **ko**: ''cork'' /k^{h}ork/ 9; ''corn'' /k^{h}orn/ 3; **k**ə: ''color'' /k^{h}ələ/ 20; /k^{h}əmIr/ 4; **kl**: ''clear''/k^{h}lIr/ 1; ''clearly'' /k^{h}lIrli/ 1; ''climb'' /k^{h}laim/ 8; **km**: /k^{h}mɒn/ 7; /k^{h}mIr/ 6; **k**: /k^{h}/ 2. Note that some utterances are the result of sound play and therefore are not standard English words.

agrams that was necessary for locating bursts and voicing bars. We used a 300-Hz bandwidth instead. Third, we used flat input shaping because high shaping can distort the formant/peak relationship between the bird's power spectra and the sonagram. These formant/peak relationships are important for analyses of formant transitions and stop loci. Flat shaping also provided consistency with our vowel analyses and ''A-weighting'' and ''high-shaping'' were unnecessary to resolve the signals of either subject. Fourth, we used a Hamming window (traditional in human speech analysis) because its wide main lobe is appropriate for temporal measurement and complements the broadband sonagram. Neither the Hamming window nor flat shaping produced marked differences in gross acoustic parameters for representative stop samples for our subjects compared with other alternatives. Fifth, the frequency range we used, 8000 Hz, is typical for speech analysis and captures the important components of Alex's utterances. Note that anti-aliasing is automatic for the Kay 5500.

Where possible, for each sample we obtained 14 measures: VOT, number of burst releases (i.e., bursts), two stop loci, duration, three formant onsets, three formant targets, and three slopes. We followed previous research to define and identify these measures. We measured VOT in ms from the beginning of the first burst to the beginning of periodic vibration (Fant, 1991; Olive *et al.*, 1993). Bursts are discretely marked shifts in the instantaneous amplitude of a signal that are associated with stop releases. To determine number of bursts, we counted occurrences before vowel onset because only in that region could bursts clearly be identified. Using between-cursors power spectra, we selected the region of the signal defined by VOT for analysis of stop loci. Following Cole and Scott (1974), Dorman *et al.* (1977), Just *et al.* (1978), Kuhn (1975), and Winitz *et al.* (1972), we include the burst and aspiration. Human front cavity resonance estimates rely on burst and aspiration data (Kuhn, 1975; Dorman *et al.*, 1977), and we wished to obtain similar estimates for Alex. After discounting constant energy $<$500 Hz, we designated the two first (lowest frequency) high-energy peaks in the signal as stop loci ''dB1'' and ''dB2.''[9] Duration was the time from beginning of voicing to the first place where formants assume a steady-state pattern. For the first three formants (F_1, F_2, F_3), we measured F_x onset (frequencies during the first 10 ms of voicing following the stop)

and F_x target (average frequencies during the steady-state portion of the vowel following the stop). We then calculated SLOPE*x* (difference in Hz between F_x onset and target divided by duration in ms).

B. Statistical analyses

Statistical tests characterized our data in meaningful ways. We examined the relationship between Alex's (or IMP's) utterances and our measures of those utterances with Tukey–Schleffe tests (SAS, 1989) and GLMs (multiple regression analyses; General Linear Model; PROC GLM, SAS, 1989). We compared Alex's with IMP's utterances via bivariate correlations. To see if Alex's and IMP's stops grouped differently with respect to a common measure, we performed hierarchical cluster analyses (Aldenderfer and Blashfield, 1984; Mezzich and Solomon, 1980) in SYSTAT.

One set of Tukey–Schleffe tests distinguish between stops with respect to various measures, e.g., distinguish between /b/ and /g/ with respect to VOT for each subject;[10] additional tests within voiced/voiceless or place of constriction subsets (e.g., /b,d,g/, /p,b/), provide some information about subset cohesiveness. Tukey–Schleffe tests, however, are less sophisticated than GLMs: We could not use Tukey–Schleffe tests, for example, to examine the effects of acoustic context.

GLMs show the extent to which acoustic context affects each measure (e.g., VOT, burst, etc.) taken on an utterance, and thus allow us to compare context effects across subjects. For GLMs we first define an utterance as STOP (identity of stop /p,t,k,b,d,g/) plus its acoustic context. We then divide context into three components: (1) C1, the two phonemes[11] (if present) preceding the stop; (2) C2, the two phonemes (if present) following the stop; and (3) WORD, the entire label from which a stop was taken. (We could not exclude C1 *a priori* because samples could be phrase-internal and context effects may cross word boundaries.) These components, along with STOP, are called *variables* in GLM analyses. GLMs show how each of our four variables, singly and in various combinations, affect each of our 14 measures, e.g., how STOP versus C1+C2 affects VOT. Each single or combined variable set is called a model. One variable may be better than another at accounting for variance in a given measure; for other measures, some combination or perhaps all are important.

Although limiting IMP's data set to utterances that matched Alex's showed that differences in GLMs between our subjects were independent of our data set, we still needed to examine how GLMs were themselves affected by our inability to include all stop/vowel combinations represented in a ''standard'' data set (see footnote 7). A factor like following vowel (i.e., *C2, a reduced C2 that includes only the phoneme immediately following a stop), for example, is not correlated to STOP in a standard set (e.g., one with /pa,ba,ta,da,ka,ga/), but could be correlated in Alex's restricted data set. Such correlations could affect the relative extent to which particular variables (e.g., STOP versus *C2) affected particular measures (e.g., VOT). We thus used two statistical techniques, CATMOD (SAS Institute, 1989) and Nested Model Comparisons (NMCs), to examine these correlations. CATMOD determines the extent of interdependence in our variables (WORD, STOP, C1, C2, and *C2) and thus which models should be omitted from GLMs.[12] By running CATMOD on voiced/voiceless and place of constriction subsets as well as on the full set of stops, we could see if context effects were stronger or weaker for any particular set. NMCs compare GLMs and tell which models (a) explain exactly the same variance and are thus redundant and (b) optimally account for variance in any given measure.[13]

Specific GLMs also enabled us to examine whether certain intrasubject phonetic relationships known for humans held for Alex. We thereby examined whether relationships, such as those between VOT and *C2, might occur in our data. We also used GLMs to determine the relative coherence of the voicing and place subsets for Alex and IMP.

GLMs and Tukey–Schleffe tests provide information about grouping by place or voicing, but fail to reveal possible additional or alternative groupings in our data. We thus performed hierarchical cluster analyses. Cluster analyses sort stimuli into hierarchical dendrograms using average linkages on Euclidean distances; resulting subsets are divided according to similarities and differences that should correspond to stop categories (see Dooling *et al.*, 1987). For reasons discussed below, we ran two cluster analyses, one on a representative set and one on a more limited set of VOT data.

III. RESULTS

Alex's stops were generally distinguishable by ear, particularly during successive productions of minimal pairs (e.g., ''pea'' /pi/, ''key'' /ki/; Fig. 2); not surprisingly, so were IMP's. Significant intrasubject differences existed among stops in the acoustic parameters described below. Also, the particular parameters that were distinctive sometimes differed for the parrot and human.

A. Acoustic parameters

1. VOT data

We provide VOT for each stop for Alex and IMP and published human values for comparison (Table II). We compare our subjects' voiced/voiceless ranges to learn whether Alex's stops occupy the equivalent (relative) amount of acoustic space as do IMP's. Comparisons of *means* across subsets tell us if Alex's stops occupy the same (absolute) acoustic space as do IMP's.

For both subjects, the range of values is roughly similar to reported human values. Human VOTs for voiceless initial stops /p,t,k/ are generally 44–115 ms (Lauter *et al.*, 1986), although Fant (1991) reports 125 ms. Voiced stops /b,d,g/ are shorter, 4–30 ms (Lauter *et al.*, 1986; Fant, 1991). Ranges and standard deviations (s.d.) for /p,t,k/ for Alex and IMP are remarkably similar. Alex's range for /b,d,g/ is greater than IMP's and his s.d. for /b/ is considerably greater than hers.

Some mean VOTs are noteworthy. Alex's /b/ value is strikingly shorter (negative) than IMP's; his /d/ is considerably longer than hers. Unlike some humans (Lauter *et al.*, 1986), Alex's VOTs for velars are shorter than for alveolars; the same is true for IMP's /t/ and /k/ (see Sec. IV A 1).

TABLE II. Mean VOT values (first lines)[a] and standard deviations (second lines) in ms and number of samples (third line) for humans, *Psittacus erithacus*, and IMP.

	Consonant					
	Voiced			Voiceless		
Subject	[b]	[d]	[g]	[p]	[t]	[k]
M	8.80	14.25	24.06	55.33	68.47	77.04
F	8.55	17.28	24.05	73.37	82.57	83.93
Alex	−14.64	28.45	22.08	51.47	113.90	89.23
	42.79	9.59	12.41	26.58	24.52	18.76
	54	38	31	48	50	85
IMP	5.88	17.11	24.71	63.11	103.21	89.61
	24.13	7.03	7.54	30.34	25.42	22.90
	64	38	32	49	55	89

[a]Values for M (man) and W (woman) are derived from Lauter *et al.* (1986); no standard deviations or number of samples are given. Alex provided the values for *Psittacus erithacus*, and his principal trainer, Irene M. Pepperberg (IMP), provided data for comparisons.

2. Number of bursts

Mean number of bursts and s.d.'s are in Table III. Alex has fewer bursts than IMP for /p,b,t/ but a comparable number for /d,k,g/. Both subjects display considerable variability, but, unlike IMP, Alex's s.d. often is equal or double the mean.

IMP's data, but not Alex's, generally conform to human norms. IMP, like other humans, often has double bursts for /k/ (Olive *et al.*, 1993) and occasionally for /g/ (Fischer-Jorgensen, 1954). Alex follows the human pattern for /k/, but has more double bursts for /d/ than /g/. We found occasional double bursts in all samples except IMP's /d/. Voiced and voiceless subsets differ significantly between and within subjects (t tests, $p \leqslant 0.041$) with respect to number of bursts. We examine possible causes of Alex's multiple bursts and high variability in Sec. IV A 1.

3. Stop loci

We report mean stop loci (dB1,dB2) in Table IV. Our subjects' absolute values differ, and Alex's dB1s have somewhat more overlap across stops and his dB2s somewhat less than IMP's. Nevertheless, our data generally follow human patterns (Halle *et al.*, 1957; Kuhn, 1975; Dorman *et al.*, 1977; Stevens, 1993) in that stop loci differ more with respect to place of constriction than the voiced/voiceless distinction; also, as for humans (Olive *et al.*, 1993), our /p,t,k/ stop loci are of greater intensity and temporal length and are more distinguishable from each other than are /b,d,g/ stop loci.

We compare our subjects' mean /p,b/ stop loci to those of humans. For humans, Halle *et al.* (1957) found values of 500–1500 Hz, but /b/ could be 2000 Hz (Dorman *et al.*, 1977; Kuhn, 1975). Only our subjects' dB1s and Alex's /b/ dB2 correspond to those values. For our subjects as for humans (Dorman *et al.*, 1977; Kuhn, 1975), amplitude of /b/ is absent, or weak compared to that of /p/.

We also compare our /t,d/ data to human values. Human stop loci for /t/ have the highest frequencies (>3000 Hz, Liberman *et al.*, 1952; Dorman *et al.*, 1977; >4000 Hz, Halle *et al.*, 1957) and vary least with context across stops (Kuhn, 1975). For /d/, researchers report energy >2000 Hz (Dorman *et al.*, 1977) or >4000 Hz (Halle *et al.*, 1957). For Alex, dB1 for /t/ displays the expected high frequency, acoustically stable values, although contexts of our /t/ samples vary as much as those for all other stops. Alex's dB2 for /t/ and /d/ are relatively high and stable compared to /p,b,k,g/, but his stability for /d/ may be a consequence of limited context, i.e., of values only for ''dinner.'' For IMP, in contrast, dB1 and dB2 for /d/ are of higher frequency and more stable than for /t/ (also possibly the effect of limited context). Alex's dB1s for /d/ are consistent with Dorman *et al.*'s (1977) report that /d/ has higher frequency stop loci before unrounded vowels /i,I,e,ɛ,ɘ,/: his /d/ in ''dinner'' is higher ($\bar{x}=2146$ Hz, $N=29$) than in ''down'' ($\bar{x}=1173$ Hz, $N=3$). IMP's dB2 but not dB1 values are consistent with Dorman's claim: For dB1, her /d/ in ''dinner'' is lower ($\bar{x}=2888$ Hz, $N=30$) than in ''down'' ($\bar{x}=3086$ Hz, $N=7$) but her dB2 is higher ($\bar{x}=5101$ Hz, $N=30$) than in ''down'' ($\bar{x}=4512$ Hz, $N=5$).

For /k,g/ we again compare our data to published human values. Our dB1s resemble Halle *et al.*'s (1957) 1500–4000 Hz for humans. Kuhn (1975) and Dorman *et al.* (1977) suggest that /k,g/ values are highly dependent on following vowel; for Alex's /k,g/, statistical tests (see below) show that C2 accounts for significant variance (80.5%) in dB1. Our

TABLE III. Mean number of burst releases (first line), standard deviation (second line), and number of samples (third line) for *Psittacus erithacus* and IMP.

	Consonant					
	Voiced			Voiceless		
Subject	[b]	[d]	[g]	[p]	[t]	[k]
Alex	0.57	1.11	0.84	0.38	0.58	1.65
	0.61	0.45	0.78	0.53	0.57	0.98
	53	38	31	47	52	88
IMP	0.92	1.00	1.03	0.92	0.96	1.60
	0.38	0	0.61	0.49	0.51	0.72
	62	38	31	49	54	89

TABLE IV. Mean stop loci values (first line) and standard deviation (second line) in Hz, number of samples (third line) and (under dB1) estimated place of articulation in cm (fourth line) *for Psittacus erithacus* and IMP.

	Consonant					
	Voiced			Voiceless		
Subject	[b]	[d]	[g]	[p]	[t]	[k]
Alex dB1	1349	2055	1329	1054	2292	2013
	743.20	913.82	402.65	506.59	222.40	829.26
	11	32	24	39	55	85
	6.38	4.18	6.47	8.16	3.75	4.27
dB2	1928	2973	2991	2608	4146	3115
	987.28	414.73	672.24	884.41	663.55	1283.00
	5	9	9	29	20	25
IMP dB1	1549	2925	1649	1675	2285	2453
	869.35	290.92	332.77	390.61	413.50	1355.00
	40	37	32	38	52	83
	5.55	2.94	5.22	5.13	3.76	3.51[a]
dB2	3048	5017	4602	3060	4613	4953
	1155.00	745.26	978.60	412.93	1033.00	992.06
	21	35	25	8	46	30

[a]This is the mean value for /k/. The median stop locus value for IMP (1720 Hz) predicts an estimated place of articulation at 5 cm.

data may also reflect other context effects. Our /g/ samples predominantly involve low back vowels (e.g., /o,ə/); thus Alex's (and IMP's slightly) low dB1 /g/ values may reflect this bias. In particular, Alex's /o/ generally has a single, low broad formant (Patterson and Pepperberg, 1994). Alex's dB2s for /g/, however, correspond to the reported human range; IMP's dB1s but not dB2s for /g/ match published human data.

Values that other researchers designate ''stop loci'' mostly correspond to our dB1. For /p,b,k,g/, IMP's dB1s but not dB2s correspond to reported values; neither her dB1s nor dB2s match reported /d/ values, but her dB2s for /t/ match published values. For Alex, dB1s match human values for /p,b,d,k/, but his dB2s match published values for /t,g/. Note, however, our discussion of context above.

Researchers have used stop loci to predict the place of constriction between cavities in a two-chamber model for humans (Ladefoged, 1982). We use dB1s, which generally correspond to published stop loci, to predict places of constriction for our subjects' stops.[14] Following Kuhn (1975), we use the formula $l = c/4f$ (l=length of cavity; $c = 344$ m/s, speed of sound in air, 25 °C; f=mean frequency in Hz) for a tube open at one end. Results are in Table IV; predictions for place of constriction within human voiced/voiceless subsets respectively are, in front-to-back order, /b,d,g/ and /p,t,k/ (Ladefoged, 1982).

According to our calculations, order of stop occurrence differs for Alex and IMP within their respective upper cavities. Humans produce /p,b/ at the lips and have no cavity forward of the constriction; thus stop loci cannot be used to calculate front cavity resonances of /p,b/. Alex, however, need not produce /p,b/ at the front of the cavity; e.g., he can produce /b/ with beak open and tongue down (Patterson, unpublished). Calculating places of constriction for all his stops may provide insights into his underlying production mechanisms. For Alex, calculations indicate a model in which /d,k/ occur at the back of a ~4-cm tube, /t/ is produced forward of /k/, /b,g/ occur at the back of a tube ≳6 cm, and /p/ occurs at the back of an 8-cm tube. For IMP voiced stops occur in the expected order /d,g/; voiceless stops appear to occur in the order /k,t/. This finding suggests that IMP's /k/ is fronted, but her /k/ values may be skewed by the nature of our data set: Her mean dB1 is 2453 Hz, but her median is 1720 Hz; i.e., many /k/ samples must occur at the back of a tube longer than the mean dB1 suggests. Her median dB1 places the point of constriction at 5 cm, which preserves the reported human order. Alex's (and IMP's other) medians and means do not differ enough to affect other stop orderings; using Alex's medians for /p,b/ puts production further back in the vocal tract (>9 cm). Alex's oropharyngeal cavity is ~3.7 cm; thus his predicted constriction sites do not make biological sense. We discuss other likely mechanisms for his /b,p/ (and possibly /g/) production below (Sec. IV B 2).

4. Mean onset and target frequencies and transition durations

Formant onsets, targets (Table V), and transition durations are usually analyzed as slope, where SLOPE$x = [F_x\text{target} - F_x\text{onset}]/\text{DUR}$. A slope is positive if it rises from a stop into a following vowel, and negative if it falls. Two formants' slopes converge if they shift toward each other as they move from a stop into a following vowel (Stevens and House, 1956). Slopes tend to rise out of labials, converge out of velars, and remain constant (horizontal and parallel) for alveolars (e.g., Liberman *et al.*, 1954). In perceptual experiments, slopes provide almost as much information on stop recognition as the ''full complement of cues'' before most vowels (Dorman *et al.*, 1977, p. 116); furthermore, SLOPE2 and SLOPE3 were sufficient to cue place of articulation.

TABLE V. Mean onset and target frequencies (first line) and standard deviation (second line) in Hz; number of samples (third line), *Psittacus erithacus* and IMP.

Subject	Measure	[b]	[d]	[g]	[p]	[t]	[k]
Alex	$F_{1\text{onset}}$	836.23	788.42	847.27	830.00	756.36	738.71
		834.72	184.34	189.49	236.07	176.27	782.05
		53	38	22	48	22	62
	$F_{1\text{target}}$	816.15	803.16	874.78	822.50	821.82	674.75
		160.83	164.63	167.68	148.73	131.06	193.28
		52	38	23	48	22	61
	$F_{2\text{onset}}$	1789.00	2488.00	1593.00	1636.00	2222.00	2088.00
		416.54	133.53	272.74	414.54	273.18	750.99
		51	38	24	47	22	56
	$F_{2\text{target}}$	2102.00	2513.00	1663.00	1660.00	2100.00	2075.00
		511.37	118.68	283.82	403.88	378.22	822.86
		50	38	23	47	22	55
	$F_{3\text{onset}}$	2669.00	3180.00	⋯	3400.00	3360.00	2783.00
		445.45	254.03	⋯	⋯	⋯	695.27
		22	4	⋯	1	1	16
	$F_{3\text{target}}$	3019.00	3200.00	⋯	3160.00	3960.00	2560.00
		672.69	340.98	⋯	⋯	⋯	731.63
		19	4	⋯	1	1	19
IMP	$F_{1\text{onset}}$	364.38	345.26	338.06	602.76	299.13	277.33
		98.53	46.89	47.15	272.12	79.44	79.27
		64	38	31	29	46	60
	$F_{1\text{target}}$	445.63	413.68	385.81	635.15	402.67	454.00
		177.58	70.99	119.35	259.18	119.51	186.29
		64	38	31	33	45	60
	$F_{2\text{onset}}$	1806.00	2345.00	1723.00	1655.00	2013.00	1966.00
		498.60	102.48	178.57	413.36	423.84	798.48
		56	38	31	30	45	57
	$F_{2\text{target}}$	2008.00	2344.00	1720.00	1612.00	1916.00	1982.00
		629.52	121.09	231.40	474.16	403.10	810.96
		59	38	31	34	45	57
	$F_{3\text{onset}}$	2855.00	3036.00	2763.00	3160.00	2956.00	3409.00
		325.67	60.83	128.16	⋯	341.32	415.04
		41	38	23	1	34	47
	$F_{3\text{target}}$	3122.00	3002.00	2677.00	⋯	2914.00	3409.00
		393.75	67.67	88.99	⋯	356.57	437.46
		41	38	24	⋯	34	47

SLOPE, although highly context dependent, is an important correlate of stops for humans; thus, even though our data set is skewed with respect to certain contexts, we examine SLOPE*x* for our subjects (Table VI) and find some agreement with published descriptions (Liberman *et al.*, 1954). Alex's SLOPE2 and SLOPE3 rise as expected for /b/, but SLOPE1 is neutral. As expected, his three /d/ slopes are relatively constant and his first and second /g/ slopes converge, but we have no F_3 data for SLOPE3 for /g/. For /d/ and /g/, context in our data set is very limited (see Table I) and we interpret our results cautiously. Nevertheless, Alex's /b,d,g/ values do correspond roughly to those predicted for humans. For his voiceless stops, SLOPE1 and SLOPE2 rise for /p/ as predicted, but SLOPE3 falls; we found no other predicted patterns (note that /p/ contexts are limited). For IMP, voiced stops follow predicted patterns. For her voiceless stops, SLOPE1 and SLOPE2 for /k/ converge but SLOPE3 is constant; we found no other predicted patterns. For both subjects, but especially IMP, voiced stops better represent predicted slope patterns than voiceless stops (see Halle *et al.*, 1957).

Following Halle *et al.* (1957), we calculated SLOPE*x* independently for each *C2. Even then, our subjects' data did not fully match reported values. We discuss our results below (Sec. IV A 2).

B. Statistical analyses

1. Tukey–Schleffe tests

Stops are distinguishable from each other if their values on some measure (e.g., VOT, F_x onset) differ significantly ($p<0.05$). Tukey–Schleffe tests indicate that within subjects (Alex and IMP), all stop pairs for the full set are distinguishable for some measure (Table VII). Within voice and place subsets, stops for both subjects, except /d,g/ for Alex's voiced subset, differ with respect to the same or more measures than when analyzed as part of the full set (Table VII). Such findings contrast with our vowel data, for which vowels within front/back categories are less distinguishable from one another than vowels across such categories (Patterson and Pepperberg, 1994). Because IMP's voice and place subsets are indeed coherent, we suggest that Tukey–Schleffe tests are not sensitive to these differences.

TABLE VI. Mean slope values (first line) and standard deviation (second line); number of samples (third line), *Psittacus erithacus* and IMP.

Subject	Measure	[b]	[d]	[g]	[p]	[t]	[k]
Alex	SLOPE1	−0.01	−0.63	1.59	1.08	2.82	−3.18
		23.06	5.76	10.24	12.09	4.92	28.87
		47	36	19	41	20	51
	SLOPE2	12.29	1.29	−0.98	1.41	−5.42	−0.29
		10.71	6.46	8.21	12.68	12.92	7.13
		46	36	19	41	20	50
	SLOPE3	13.89	1.34	⋯	−11.82	22.59	−0.70
		12.73	6.05	⋯	⋯	⋯	10.27
		19	4	⋯	1	1	14
IMP	SLOPE1	3.10	2.27	1.74	0.27	3.25	6.11
		8.50	2.79	2.81	5.08	3.67	7.62
		55	37	15	19	24	58
	SLOPE2	7.66	−0.31	−1.81	−0.25	−4.78	−0.92
		7.69	5.08	8.59	7.19	5.74	15.07
		53	37	15	21	24	57
	SLOPE3	8.02	−1.03	−6.39	⋯	1.01	−0.25
		5.99	2.99	5.12	⋯	3.65	4.40
		38	37	10	⋯	13	47

2. CATMOD results

Before performing further statistical tests to account for sources of variance in each measure, we ran CATMOD to see if WORD, STOP, C1 and C2 were independent of each other. Such findings also suggest how the restricted nature of our data set may affect GLMs and NMCs: If, for example, STOP is 100% predictable from C2 but C2 is only 14% predictable from STOP (e.g., if given /u/, stop ''must'' be /t/, but given /t/, C2 could be any of seven possibilities), then C2 is more likely than STOP to account for variance in measures in our data set than in a standard data set (see footnote 7).

CATMODs do reveal interdependencies among variables; we report the major findings. Not surprisingly, STOP is 100% predictable from WORD across the full set and all subsets (i.e., if word is ''key,'' stop must be /k/). WORD is 30%–50% predictable from STOP (i.e., if stop is /k/, word could be ''key,'' ''cork,'' ''kiwi,'' etc.), and, as expected, predictability corresponds to the number of contexts in our data set. STOP does account for a certain amount of variance, but this variance is subsumed by WORD; thus we omit WORD+STOP in GLM models. WORD is the best predictor of other variables (80%–100%), followed closely by C2 (56%–100%); thus WORD and C2 subsume most of the variance of other variables and will likely be optimal at accounting for variance in measures. Both C1 (44%–76%, except for the alveolar and labial subsets) and STOP (29%–68%) are poor predictors of other variables and are least likely to be optimal in accounting for variance in measures.

The relationships among variables and measures are, however, more complex than might be expected from this analysis. Whereas CATMODs tell us how our results are affected by the restricted nature of our data set and the general extent to which a variable can predict variance in a measure, GLMs and NMCs enable us to quantify variable/measure relationships, i.e., show the exact amount of variance in a measure that is the consequence of phonetic context of the vocalizations involved.

3. Nested model comparisons: Full data set

NMCs enabled us to contrast inclusive GLMs (WORD +C1+C2) and more restricted models (e.g., WORD, WORD +C2) with respect to their ability to account for variance in each measure. We thereby determined optimal models and models equal to or not significantly different from optimal models for the full set for our subjects (Table VIII). For comparison, we examined individual variables (WORD,

TABLE VII. Results of Tukey tests. Number of comparisons that were significant at ≤0.05 level for the full, voiced/voiceless, and place sets. ''na'' indicates a nonapplicable comparison.

(a) Full set, voiced and voiceless.

	Alex			IMP		
	Full	Voiced	Voiceless	Full	Voiced	Voiceless
p–t	6	na	6	5	na	5
p–k	7	na	8	6	na	6
p–b	4	na	na	3	na	na
p–d	7	na	na	5	na	na
p–g	4	na	na	1	na	na
t–k	4	na	3	4	na	5
t–b	6	na	na	5	na	na
t–d	4	na	na	4	na	na
t–g	3	na	na	5	na	na
k–b	9	na	na	7	na	na
k–d	6	na	na	5	na	na
k–g	6	na	na	6	na	na
b–d	6	8	na	6	8	na
b–g	4	6	na	3	5	na
d–g	6	6	na	3	3	na

(b) Place of articulation.

	Alex	IMP
p–b (labial)	4	5
t–d (alveolar)	6	6
k–g (velar)	6	9

TABLE VIII. Nested model comparisons. Full set: Optimal model (model), percentage of variance explained by optimal models (%O), and other models using solely stop (%S), word (%W), context (%C1, %C2), place of constriction (%P), and voicing (%V) factors.

	Alex								IMP							
Measure	Model	%O	%W	%S	%C1	%C2	%P	%V	Model	%O	%W	%S	%C1	%C2	%P	%V
VOT	W+C1+C2	88.9	88.1	75.2	24.5	73.3	27.9	56.7	W+C1+C2	89.6	89.0	74.2	31.8	78.2	18.0	66.2
Bursts	W+C1	68.9	65.0	31.9	40.5	53.5	22.3	1.6	W+C1+C2	51.1	47.1	22.5	19.0	35.1	16.6	4.8
dB1	W	72.8	72.8	31.5	32.8	68.0	24.4	ns	W+C1+C2	83.4	80.2	23.8	25.3	75.8	14.3	ns
dB2	W+C1	77.5	71.7	31.4	28.6	56.0	22.2	ns	W+C1+C2	69.2	55.8	34.6	36.5	55.5	32.4	ns
$F_{1\text{onset}}$	ns	ns	ns	ns	ns	ns	ns	ns	W+C1	75.5	73.7	39.6	39.4	55.3	17.8	ns
$F_{1\text{target}}$	W+C1+C2	47.1	43.9	13.6	14.8	31.4	5.6	3.7	W+C1+C2	81.5	76.8	16.0	52.7	67.3	6.0	2.6
$F_{2\text{onset}}$	W+C1	85.5	84.0	30.6	42.5	80.6	22.9	ns	W+C1	83.6	83.2	15.5	32.1	81.9	10.1	ns
$F_{2\text{target}}$	W+C1	89.1	88.5	23.8	45.5	85.6	11.2	3.4	W+C1+C2	96.7	95.4	13.8	41.0	95.5	3.8	1.9
$F_{3\text{onset}}$	W	80.5	80.5	ns	ns	75.6	8.2	ns	W+C1	94.9	91.1	37.3	ns	82.6	12.1	17.3
$F_{3\text{target}}$	W	91.1	91.1	ns	33.5	88.8	15.7	ns	W+C1	94.4	89.6	34.3	ns	84.7	5.2	7.9
SLOPE1	ns	ns	ns	ns	ns	ns	ns	ns	W+C1+C2	76.6	73.9	7.8	63.6	48.9	3.6	ns
SLOPE2	W+C1	49.8	47.7	23.7	26.5	38.6	11.6	9.4	W	38.0	38.0	14.7	29.6	33.6	8.3	4.0
SLOPE3	W	72.9	72.9	35.8	35.3	61.2	21.4	18.7	W+C1+C2	65.2	60.0	47.2	29.7	55.4	4.1	3.5
DUR	ns	ns	ns	ns	ns	ns	ns	ns	W+C1+C2	41.5	35.4	ns	13.5	ns	ns	ns

STOP, C1, or C2) to determine whether they accounted for significant variance in each measure (e.g., VOT, dB1, etc.). We could account for significant variance in only 11 measures for Alex, whereas we could account for significant variance in all 14 of IMP's measures. Here we discuss our most salient findings. The results must, however, be interpreted in light of our restricted data set.

STOP accounted for significant variance in fewer measures for Alex than IMP. For Alex, however, STOP accounted for a greater proportion of variance in VOT, number of bursts, dB1, F_2 onset, F_2 target, and SLOPE2. We examine these differences in Sec. IV A 3.

Our models also demonstrate context effects for both subjects. As in standard data sets, our samples are mostly word-initial, and C2 and WORD explain considerable variance in the measures. Our data set, however, may be affected by C1: 50% of our samples are phrase-internal (e.g., ''go eat *d*inner'') and 3% (i.e., 22% of /p/) are word-internal. Although C1 alone cannot optimally explain variance for any measure, it accounts for significant variance in ten measures for Alex and 12 for IMP.

In the full set, WORD[15] is the only single variable optimally to account for variance in some measures (Table VIII). WORD, even when not optimal, accounts for 92.5%–99% of the variance explained by optimal models for both subjects. STOP, which is subsumed by WORD, typically accounts for $<$50% of the variance accounted for by WORD. Even for VOT, for which STOP accounts for $\sim$75% of the variance, WORD accounts for an additional 13%–14% for both subjects.

4. Nested Model Comparisons: Voiced/voiceless and place subsets

Based on data from Patterson and Pepperberg (1994), we wondered if context effects were greater for categorically similar stops; i.e., within subsets. For both subjects, *t* tests revealed that the mean variance across all measures explained by each variable does not differ significantly across voiced, voiceless, and full sets, but may be greater in the full set than in place subsets. Each variable may account for significantly different amounts of variance for individual measures across sets but, overall, individual variables account for as much or more variance in the full set than in all subsets.

5. Coherence of subsets

Our data suggest that Alex's stops fall into the same subsets as those of humans and we had thus performed calculations based on this premise; we had not, however, tested the extent of coherence of such subsets. Tukey–Schleffe tests did not provide adequate information about subset coherence; we thus used GLMs to learn how well human voicing and place divisions characterized Alex's data.

Alex's stops, like IMP's, separate into voicing and place subsets, but his subsets are less coherent than IMP's (Table VIII): Both voicing and place distinctions account for significant variance in fewer measures for Alex than for IMP. For both subjects, place accounts for significant variance in more measures than does voicing, but voicing and place distinctions differ considerably in their power to account for variance for individual measures. Interestingly, voicing accounts for more variance than does place in VOT for both subjects and in IMP's F_3 onset and target.

6. Correlations between subjects

Positive correlations between our subjects' measures reveal how similar Alex's speech is to IMP's. The same nine measures for our subjects are significantly correlated (Table IX) for the full set and voiceless subset; five of these nine are correlated for the voiced subset. Correlations are somewhat stronger for voiceless than voiced stops. As in our vowel data (Patterson and Pepperberg, 1994), F_1 correlations either fail to exist or are the weakest significant correlations between subjects.

We wonder, of course, whether Alex (a) specifically attempts to reproduce IMP's vocal patterns because she was his principal trainer; (b) simply uses the female register because trainers were mostly female and a higher register is more easily produced by a short (compared to human) psittacine vocal tract; or (c) uses some combination of these

TABLE IX. Correlations between Alex and IMP for measures of stop consonants in order of descending value.

	Measure	Pearson's r	Signficance (p)	No. of observations
All stops	$F_{2\text{target}}$	0.85	0.0001	210
	VOT	0.85	0.0001	307
	$F_{2\text{onset}}$	0.82	0.0001	205
	dB1	0.68	0.0001	228
	SLOPE2	0.58	0.0001	168
	$F_{3\text{target}}$	0.55	0.0026	28
	$F_{3\text{onset}}$	0.50	0.0050	30
	bursts	0.32	0.0001	306
	$F_{1\text{target}}$	0.15	0.0294	219
[b,d,g]	$F_{2\text{target}}$	0.79	0.0001	111
	$F_{2\text{onset}}$	0.77	0.0001	109
	VOT	0.59	0.0001	125
	dB1	0.55	0.0001	65
	SLOPE2	0.46	0.0001	87
[p,t,k]	$F_{2\text{target}}$	0.90	0.0001	99
	$F_{2\text{onset}}$	0.88	0.0001	96
	dB1	0.73	0.0001	163
	$F_{3\text{onset}}$	0.73	0.0076	12
	$F_{3\text{target}}$	0.71	0.0065	13
	VOT	0.65	0.0001	182
	SLOPE2	0.60	0.0001	81
	bursts	0.40	0.0001	185
	$F_{1\text{target}}$	0.22	0.0228	105

strategies. We compare Alex to IMP, however, because she was his most frequent acoustic model and we wanted to determine how well their acoustic measures match, and, by implication which features he might best be able to reproduce.

7. VOT and preceding context (C1)

NMCs reveal that C1 accounts for significant variance in VOT for both subjects (Table VIII). Studies of C1 effects are uncommon, but C1 can affect VOT perception (Repp, 1991). If VOT stimuli are phrase-initial (e.g., ''**B**ills are here'') or immediately preceded by a voiced sound (''The **b**ills are here''), listeners demonstrate category boundaries at the same place as for isolated samples. A short VOT preceded by a voiceless context (''Most **b**ills are here''), however, may be interpreted as belonging to a stop associated with longer values. Voiceless C1s could thus cause short VOTs normally indicative of /b,d,g/ to be interpreted as longer VOTs indicative of /p,t,k/; long VOTs should still be interpreted as /p,t,k/. These C1 effects can persist over several seconds and across word boundaries; because some of our samples were phrase-internal, C1 might have affected our data.

Specifically, this perceptual phenomenon may have affected our subjects' stop *productions*. To be understood, Alex (and perhaps IMP) might produce only very short /b,d,g/ VOTs following voiceless C1s; /p,t,k/ VOTs would not be affected. VOTs following voiceless C1s might thus be significantly different from VOTs following voiced or absent contexts. Of our 330 data samples, only those for /d/ included both these contexts, although for different C2s (''go eat **d**inner,'' ''calm **d**own''). Alex's VOT for /d/ in ''dinner'' is significantly shorter than in ''down'' ($t=8.39$, $p<0.001$).[16] The reverse is true for IMP ($t=-5.11$, $p<0.001$). IMP's results might be a consequence of C2 effects (see below).

8. VOT and following context (*C2)

We performed statistical tests to learn how VOT in our data is affected by the immediately following phoneme, *C2. For humans, VOT is generally longer before sonorants /l,m,r/ and high vowels /i,I,ɛ,U,u/ than before mid and low vowels /æ,ɒ,ə,o/ (Klatt, 1991);[17] /i/, in particular, is associated with long VOTs (Repp, 1986). As noted above, IMP's VOT for /d/ in ''dinner'' (i.e., before /I/) was significantly longer than in ''down'' (i.e., before /ɒ/ in /ɒu/). While examining other patterns in our data set, we encountered difficulties with /r/ in our subjects' productions of ''bread'' and ''paper.''

To address this problem, we classified the data three different ways and tested each set. (1) We first classified /r/ in ''bread'' as a sonorant/high vowel. (2) We reclassified /r/ because our subjects' manner of production (/bəˠɛd/) suggested that it likely was a mid/low vowel, /əˠ/. (3) We omitted all ''bread'' and ''paper'' samples because /əˠ/ in both cases was associated with unusual VOTs.[18]

When we treated /r/ in ''bread'' as a sonorant/high vowel, our independent sample t tests (Norušis, 1993) revealed no significant differences in VOT values associated with sonorants/high vowels versus mid/low vowels (for Alex, $t=-0.52$, $p=0.605$; for IMP, $t=-0.84$, $p=0.399$). These relationships were not likely affected, moreover, by the distribution of samples or place of constriction: We had (a) almost as many voiced as voiceless stops in the sonorant/high vowel and mid/low vowel categories; (b) almost equal numbers of sonorants and high vowels following /t/, /k/, and /d/ (i.e., across place and voicing boundaries); and (c) almost as many mid and low vowels following /p/, /b/, and /k/.

As noted above, however, /r/ might be a mid/low syllabic /əˠ/. Given such a classification, independent sample t tests showed that *C2 affected VOTs for Alex ($t=-3.87$, $p<0.001$) and IMP ($t=-3.15$, $p=0.002$): VOTs were longer before sonorants/high vowels, consistent with Klatt's findings.

The reversal suggested that the restricted nature of our data set might cause /əˠ/ to influence VOT to an unusual extent. We thus performed paired comparisons (Tukey–Schleffe tests) both within and across stops to identify *C2s associated with significantly different VOTs. /əˠ/ affected Alex's VOT more than any other *C2. Moreover, for the two stops in the data set with a following /əˠ/, /p/ and /b/, /əˠ/ was associated with the greatest number of significantly different VOTs for both stops for Alex, but only for /p/ for IMP. /əˠ/ did not predict VOT for IMP's /b/ but no other *C2 did so either.

Because of /əˠ/'s significant effect on preceding stop, we removed all /əˠ/ samples (''bread,'' ''paper'') from the data and reran our t tests. Significant differences still existed in VOTs for following sonorants/high versus mid/low vowels for Alex ($t=-2.13$, $p=0.034$), but not for IMP ($t=-1.78$, $p=0.076$). Removing /əˠ/ samples did not affect

our results for IMP as much as did reclassifying /r/ in ''bread.''

We also used mean VOTs and paired Tukey–Schleffe tests to learn if /i/ was associated with long VOTs (Repp, 1986). For our /k/s, /i/ was associated with longer VOTs than were /ə,ɛ,ɒ,l,m,o/ ($p<0.05$). For /p/, for which we have one /pi/ sample, /i/ was associated with a longer VOT than were /ɚ,ɛ,l,ɒ/, but VOT for /bi/ for both subjects was the third shortest of nine /b/ contexts. In all other cases for our subjects, VOTs associated with /i/ were as long as or significantly longer than other VOTs.

Our analyses thus support Klatt's and Repp's reports only partially, but agree with the general claim that *C2 can predict VOT. *C2 accounts for significant variance in VOT for the full set for Alex ($R^2=0.51$, $F[12,295]=25.67$, $p=0.0001$) and IMP ($R^2=0.48$, $F[13,315]=22.24$, $p=0.0001$), and for each stop except Alex's /g/ and IMP's /b/ (Tukey–Schleffe test, $p<0.05$). The relationship between VOT and *C2 was not significantly affected by removal of /ɚ/ ($t=-1.63$; $p=0.104$) and remained significant for the full set for Alex ($R^2=0.44$, $F[11,273]=19.71$, $p=0.0001$) and IMP ($R^2=0.47$, $F[11,285]=20.11$, $p=0.0001$). Removal of /ɚ/, however, significantly affected Alex's mean VOT for /b/ ($t=-2.55$, $p=0.014$); this result was important for our VOT-based cluster analysis.

9. Cluster analyses

We based cluster analyses only on VOT for three reasons: (1) STOP accounts for relatively more variance in VOT than in other measures for both subjects. (2) Other measures similarly related to STOP (e.g., F_x target) are generally contextual and CATMOD calculations show that context terms alone can predict STOP in our data set; thus cluster analyses based on context-related measures would not provide an independent test of stop distinctiveness (Sec. III B 2). (3) STOP accounted for more variance in measures other than VOT for IMP's voiced set, but our focus was Alex, and for him VOT was the noncontextual measure most consistently related to STOP.

Cluster analyses (Fig. 3) based on a representative sample of VOT data[19] revealed different patterns for our subjects. Overall Alex's voiced/voiceless distinction [Fig. 3(A)] was less coherent than IMP's [Fig. 3(B)]. His /p/ clustered more closely to /d,g/ than to /t,k/ and /b/ clustered less closely to /d,g/ than did /p/. IMP's /b,d,g/ clustered closely and were separate from the less closely clustered /p,t,k/.

Given /ɚ/'s effect on VOT, especially for Alex, we repeated the cluster analyses, omitting ''bread'' and ''paper'' samples. Alex's general pattern did not change, but /b/ was less of an outlier [Fig. 3(C)]. This second set of analyses thus provides additional support for our concerns about the effects of /ɚ/ on our data set, particularly for Alex's /b/.

C. Video analyses

Tracings of SVHS video stills of Alex's utterances reveal different beak openings during the /k/ burst for different C2s, i.e., ''key'' /ki/ versus ''cork'' /kork/. The /k/ burst is associated with rapid retraction of Alex's slightly protracted trachea. For ''key,'' his beak is open as much as 300 ms

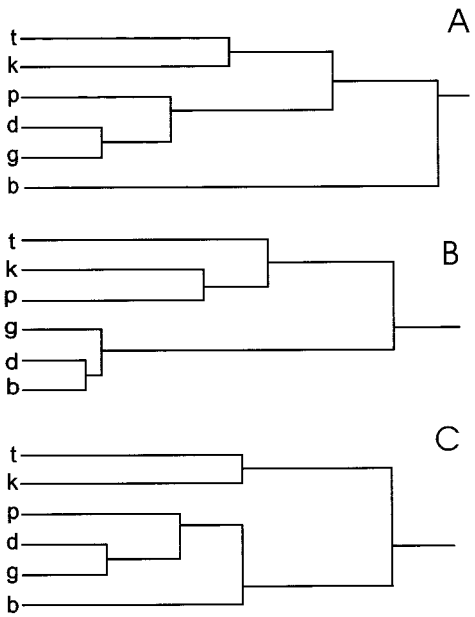


FIG. 3. Hierarchical cluster analyses showing separation of stops into VOT-based voiced and voiceless groups: (A) Alex's stops including /ɚ/ data; (B) IMP's stops; (C) Alex's stops without the /ɚ/ data.

before /k/ release [Fig. 4(A)], and continues to open during production [Fig. 4(B)]. In contrast, his beak is completely closed before he produces ''cork'' and opens only slightly during production [Fig. 4(C)]. These data suggest that he prepares his vocal tract to produce particular sounds, and are consistent with the possibility of anticipatory coarticulation (see Sec. IV B 1). Such data are not available for his voiced stops: Voiced VOTs are short and insufficient time exists to

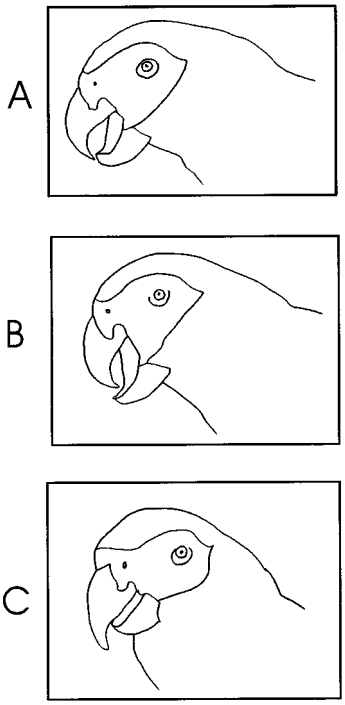


FIG. 4. Anticipatory coarticulation of vowels with stops: (A) more than 300 ms before /k/ release in /ki/; (B) after the burst release of /k/ in /ki/ but before the aspiration; (C) immediately preceding burst release for /kork/.

identify the effect of C2 before it is actually produced. Moreover, voiced bursts are weak and we cannot see the tracheal protraction associated with pressure buildup.

IV. DISCUSSION

We divide our discussion into three main sections. First, we compare how well our acoustic measures characterize our subjects' stops, and describe how Alex may compensate for lack of information in some acoustic measures with increased information in others. Second, we discuss articulatory correlates of stops with an emphasis on Alex, i.e., the possibility of psittacine coarticulation, how acoustic data may predict places of constriction for his stops, and how he might produce stops. Third, we examine how Alex's stop production compares with that of a mynah.

A. Acoustic measures

1. VOT, number of burst releases, stop loci

VOT is one of the best descriptors of stops across languages (Pisoni, 1977). VOT is primarily associated with voiced/voiceless distinctions (Kuhl and Miller, 1975) but may also relate to place of constriction (Ladefoged, 1982). For our full data set, VOT is the single most robust indicator of our subjects' stops and varies as a function of the voiced/voiceless distinction and place of constriction (Sec. III B 5; Table VIII). Both subjects generally follow expected VOT distribution patterns (Lauter *et al.*, 1986; Fant, 1991), with three exceptions. First, Alex's /p/ clusters with /d,g/ rather than /t,k/ but it has (like IMP's) the shortest VOT of the voiceless stops and a longer VOT than any voiced stop. Second, Alex's but not IMP's VOT for /d/ is longer than for /g/; his result is unexpected from context effects. Alex's long VOT for /d/ may be related either to his specific vocal mechanisms or to the phrase-internal nature of the most common /d/ sample, ''Go ea*t d*inner'': His samples were noisier than IMP's, and we may have counted extra bursts for t–d sequences, and thus included some /t/ voice-offset timing in VOT for /d/ (see below). Third, for both subjects, VOT for /k/ is shorter than for /t/. Typically, human VOT for /t/ is shorter than for /k/ (Ladefoged, 1982), but Lauter *et al.* (1986) report five of 16 cases in which the reverse is true. Our subjects' k–t ordering may correspond to Lauter *et al.*'s cases. Alternatively, our data may originate from two distinct but not mutually exclusive factors. First, our data set contains more mid/low vowels following /k/ than /t/, this weighting might shorten VOTs for /k/ relative to those for /t/ (Klatt, 1991); also, most /t/ samples involve high vowels. Alex may not have the high/low vowel distinction (Patterson and Pepperberg, 1994), but his VOTs are affected by the mid/low versus sonorant/high distinction; thus the different distributions of contexts in /k/ and /t/ samples may also affect his VOTs. Second, although IMP produces most vowels at higher frequencies than many women (Ladefoged, 1982; Borden and Harris, 1984), she also produces these vowels more toward the front of her mouth (Patterson and Pepperberg, 1994), and this tendency might be (a) transferred to stops and (b) imitated in some way by Alex.

Number of bursts, although not usually an important descriptor of stops (Olive *et al.*, 1993), differentiates some stops in our data set, particularly for Alex. We occasionally found multiple bursts for stops that are not associated with multiple bursts by other researchers (Table III). These multiple bursts may derive from two difficulties in reading sonagrams, particularly Alex's: (1) 50% of our stops were phrase-internal and a burst in C1 could be interpreted as belonging to the following stop; e.g., Alex's /d/ in ''go ea*t d*inner'' has multiple bursts in several instances. Perhaps we counted the preceding /t/ as an extra /d/ burst. In contrast, /m/ in Alex's ''cal*m d*own'' was unlikely to have a burst and that /d/ had only one questionable instance of a double burst. (2) Failure to count bursts correctly is more likely in a noisy signal. Alex's signals were often noisy and had more variable counts than IMP's. Thus noise may have contributed to problems in identifying the /t–d/ boundary in Alex's ''go ea*t d*inner.''

Stop loci, used by Kuhn (1975) to predict a stop's place of constriction, were difficult to identify. Stop loci occurred at several frequencies; we recorded the lowest and next highest as, respectively, dB1 and dB2. In $<1\%$ of our samples, the only stop locus was so high that we treated dB1 as missing; 60% of Alex's samples and 39% of IMP's samples lacked dB2. dB1 and dB2 could overlap or include outlier values, but both measures were coherent enough (s.d.s of 10%–55% for both subjects) to have predictive value. Statistical analyses suggest that our dB1 values correspond to loci described by other researchers; we thus use mean dB1s to predict a stop's place of constriction for both subjects (see Sec. IV B 2).

2. Formant onsets and targets, and transition durations

Onsets, targets, and transition durations can be combined into SLOPE, a well-researched correlate of stop (Dorman *et al.*, 1977). Slope values represent the relationship between STOP and *C2. In general, slopes are more easily discerned if preceded by voiced stops (with short VOTs) than by voiceless stops (with long VOTs) because a formant slope is not clearly visible until voicing begins (Olive *et al.*, 1993). The angle of the slope varies primarily with STOP (Olive *et al.*, 1993), but may also be affected by *C2. We find several predicted correlations in our data between SLOPE and STOP but fewer between SLOPE and *C2. Also, consistent with our vowel study (Patterson and Pepperberg, 1994), Alex's SLOPE1 (based on F_1) does not vary significantly with either STOP or *C2. For both subjects, most significant correlations occurred for SLOPE2, which is where predictive power primarily resides for humans (Olive *et al.*, 1993). In our study, however, SLOPEx (Table VIII) is a poorer predictor of stop identity than in that of Dorman *et al.* (1977).

3. GLMS and NMCs

GLMs and NMCs give us a structured view of how speech information is encoded differently for Alex and IMP, i.e., allow us to compare how much information is carried by each portion of the speech signal for each subject. GLMs allow us to determine if variance in measures is explained more fully for Alex or IMP. If more variance in VOT, for example, is explained by STOP for Alex than for IMP, then

his /p,b,t,d,k,g/ VOT distinctions are more sharply defined than hers. In general, variables explained only slightly more variance in IMP's measures than in Alex's, which suggest that his stop categories are nearly as distinct as hers. We expected NMCs to show whether context effects were more important in subsets than in the full set, but our calculations did not yield decisive results. In part, difficulties were a consequence of our restricted data set.

4. Compensatory acoustic strategies

Significant differences exist in how Alex's and IMP's F_x values vary across stops. STOP fails to account for significant variance in Alex's F_3 target, probably because he rarely produces F_3. Alex's stops, like his vowels (Patterson and Pepperberg, 1994), differ less with respect to F_1 and more with respect to F_2 than do IMP's. In humans, F_2 is important for vowel perception but F_1 and F_3 also carry relevant information (Olive *et al.*, 1993); for Alex, most information about vowel and preceding stop identity is in F_2. He seems to use F_2 to compensate for lack of information in F_1 and F_3. If redundancy in the human system is critical for disambiguating speech signals in noisy environments (Repp, 1983), then Alex's intelligibility should drop off more sharply than that of humans in such situations.

B. Articulatory correlates of stops

1. Coarticulation

Anticipatory coarticulation, which occurs when a phoneme is produced so as to configure the vocal tract for the subsequent sound, may exist in Alex's speech. Anticipatory coarticulation effects have been regarded as evidence for preparatory strategies and even for top-down processing (Ladefoged, 1982). We examine two instances of possible psittacine anticipatory coarticulation, note a related observation, and discuss whether such material is evidence for top-down processing in Alex.

Analyses of SVHS video of Alex's /k/ productions in /kork/ and /ki/ reveal differences that appear contingent upon C2 (Sec. III C). Alex varies the amount and timing of his beak opening in these two cases. Such data suggest the possibility of choice in the type of /k/ that is produced.[20]

Observations of Alex and another Grey, Griffin, and x-ray videos of Alex (Warren *et al.*, unpublished) also suggest that our birds position their vocal tracts for the second sound in an utterance: They both protract the trachea, for example, before producing /f/ in anticipation of /or/ in ''four.'' Alex engages in this anticipatory positioning even if he is interrupted and does not immediately produce ''four.'' Interpreting this behavior can be problematic because /f/ is occasionally too weak in amplitude to be identified sonagraphically. Nevertheless, the behavior is evidence for anticipatory positioning, if not for anticipatory coarticulation. This configuring for /or/ may be unique to parrots (or at least Greys); human anatomy prevents use of the same articulatory strategy.

Statistical patterns in the acoustic data are consistent with anticipatory coarticulation for voiced stops, despite lack of physical evidence for this phenomenon. For Alex and IMP, VOT is significantly affected by *C2 (see Sec. III B 8), which suggests that VOT is adjusted in preparation for *C2. For both subjects, this effect is strongest for voiceless stops; however, the presence of a following /ɔ/ extensively shortens Alex's VOT for /b/, but not IMP's. Although likely explicable in terms of Alex's anatomy and imitative goals, his acoustic patterns are also consistent with a mechanism involving anticipatory coarticulation for voiced stops.

Interpretation of anticipatory coarticulation in a parrot, however, is difficult. Because the phenomenon presupposes knowledge of sounds before they are produced, Ladefoged (1982) treats this behavior as evidence of top-down processing. Repp (1986), however, suggests that ''phenomena commonly lumped together under the heading of 'coarticulation' may have diverse origins and hence different roles in speech development. Some forms of coarticulation are an indication of advanced speech production skills whereas others may be a sign of articulatory immaturity, and yet others are neither because they simply cannot be avoided'' (p. 1618). Alex's vocal tract is not immature, but his nonhuman anatomy might make anticipatory coarticulation unavoidable (see footnote 20). Thus whether Alex's behavior reflects top-down processing or simply anatomical constraints cannot be determined until we know more about the mechanisms he uses for such productions.

2. Place of constriction

Analyses of SVHS, x-ray, and infrared videos provide limited data on Alex's articulatory mechanisms (Warren *et al.*, 1996); to predict how he might use his vocal tract to produce human speech, we thus use acoustic data and equations for calculating resonances of a tube closed at one end (Dorman *et al.*, 1977; Kuhn, 1975; Ladefoged, 1982). As in the vowel study (Patterson and Pepperberg, 1994), we regard kinematic inferences from Alex's acoustic data with caution given the differences in psittacine and human vocal tracts and the human-based assumptions of various models. By estimating the place of constriction based on stop loci, however, we propose mechanisms Alex might use to produce stops. We apply similar techniques to IMP's data; differences between IMP and published results on other humans may identify effects related to the restricted nature of our data set.

Calculations based on stop loci provide intriguing comparisons between our subjects. As noted earlier, calculation of place of constriction does not apply to IMP's /p,b/, but might apply to Alex. For Alex, stop loci predict the order for voiced and voiceless stops as, respectively, /d,b≈g/ and /t,k,p/; for IMP, the predicted order is /d,g/ and /t,k/, using a median dB1 value for /k/. If we use a standard human model of a tube closed at one end for Alex's upper cavity (upper tube of a two-tube model), stop loci predict /b,p,g/ production at places of constriction that fail to make biological sense. Possibly Alex produces stops via a tube that approaches closure at both ends (Brittan-Powell *et al.*, 1997; Nottebohm, 1976; Westneat *et al.*, 1993). The end of a tube that approaches closure resembles the neck of a Helmholtz resonator (Rossing, 1982); the effect is similar to lengthening the tube. Alex's data are also consistent with a perturbation

theory model (Chiba and Kajiyama, 1941, discussed in Johnson, 1997, p. 100), which predicts that ''constriction of the vocal tract near a point of maximum velocity...*lowers* the formant frequency.'' Details of such calculations, however, are beyond the scope of this paper.

Alex's /p,b/ production is intriguing because he lacks lips. Ventriloquists produce /p,b/ with tongue taps and laryngeal and pharyngeal constrictions instead of lips (Semok, personal communication); Alex's anatomy also likely requires alternative strategies. Possibly, he lengthens the upper cavity by involving his esophagus, which is directly caudal to the trachea; both emerge at the pharynx (Fig. 1). If the esophagus is included as part of the upper tube of a two-tube model, the lower tube (trachea) might connect to the middle of the upper tube rather than to one end, effectively lengthening the upper tube to produce the observed low frequencies. Alex's dB1s for /b/ actually split into lower ($\bar{x}$ = 747 Hz) and upper ($\bar{x}$ = 2072 Hz) groups that, according to cluster analysis, correlate, respectively, with F_1 and F_2; two groups also exist for /p/ (lower, $\bar{x}$ = 820 Hz; upper, $\bar{x}$ = 1578 Hz). Given that we previously associated F_2 with Alex's upper cavity (Warren *et al.*, 1996), calculations based on upper values place /b/ at 4.15 cm and /p/ at 5.45 cm. Esophageal involvement is consistent with these values (Demers, personal communication) and, depending upon tongue placement and degree of beak openness, with models of a tube closed at one end or approaching closure at both ends. Alex may also lower resonances by involving nasal cavities or limiting his glottal aperture; Nottebohm (1976) proposes a glottal mechanism for an Amazon parrot.

All these mechanisms are consistent with Alex's other /p,b/ data. The low loci and burst amplitudes for /p,b/ and many missing cases associated with both stops but primarily /b/ are consistent with lack of pressure buildup in the trachea prior to the burst and a glottis that need not be maximally opened at release. On video, /b/ has the least tracheal protraction of any stop, also suggesting minimal pressure buildup. Such a glottal mechanism would allow Alex to produce low frequencies, but at the expense of loudness.[21] Conceivably, Alex combines esophageal and glottal mechanisms or uses some others. /p,b/ may be produced by related mechanisms but /p/ appears in few contexts and always has a burst; also /p/'s lower frequencies may require more glottal or esophageal involvement. The low mean values for /p/'s upper dB1 may be related to having more /p/ than /b/ samples that are followed by vowels with low F_2s: For a stop to be perceived as labial, stop loci must be lower than the following F_2 target. Alex's strategy for /b/ may also depend on context: e.g., /b/ in ''boy'' (in ''good boy'') often has strong /w/-like slopes but no burst, and may be produced by a different mechanism.

Calculations from our subjects' stop loci suggest that /g/ occurs further back than /k/, although both are produced at the same place of constriction in most humans. The observed differences may be caused by two confounds in our data. First, our /k/ samples are followed by a normal distribution of front and back vowels, but g/s are followed primarily by back rounded vowels. This restriction likely accounts for /k/'s frontedness compared to /g/. Second, the mean dB1 for /k/ may not be representative of IMP's speech (Sec. III A 3): If we use her median dB1 instead, /k/ and /g/ occur at roughly the same place of constriction. Alex's mean dB1 /k/ value is likely representative of his speech: His mean and median /k/ values are within ~200 Hz. Thus the difference in /k,g/ place of constriction appears to be real for Alex, but an artifact for IMP. Given the predicted place of constriction for Alex's /g/ and its relatively low dB1s, he may produce /g/ via mechanisms similar to those proposed for /p,b/. Data for Alex's /k/ suggest that it might be included in the /t,d/ category we describe next.

Calculations from stop loci support the hypothesis that Alex produces /t,d,k/ with a tube open at one end. Such a configuration implies an open glottis or open oropharyngeal cavity. Unlike the low frequencies of /p,b/ that may also involve the esophagus, the high frequencies of /t,d,k/ do not seem to require esophageal involvement (Patterson *et al.*, 1997).

Warren *et al.* (1996) proposed that Alex may use laryngeal positioning to alter the resonant characteristics of his front cavity to produce vowels; for three reasons we suggest that stop production involves additional mechanisms. First, stops involve voiced/voiceless sound production, pressure buildup, bursts, and aspiration, and thus likely entail added physical movements. Second, predictions of Alex's place of constriction, based on stop loci and the physical model proposed for both human vowel and stop production, are not consistent with what we know of his biology. Specifically, based on the premise of a tube open at one end and a simple two-tube model, Alex would have to double the natural length of the upper cavity to produce low-frequency stop loci [the Grey parrot oropharyngeal cavity is ~3.7 cm (Warren *et al.*, 1996)]. Whether such a maneuver is possible solely via laryngeal retraction is unknown. Third, recent data (Patterson *et al.*, 1997) suggest that Alex may not only manipulate beak, tongue, and larynx to produce stops, but also use esophageal involvement, changes in glottal opening, and/or changes in sublaryngeal pressure (see Scanlan, 1988) not used by humans, or not used in the same ways.

In summary, Alex's stop production appears consistent with a modified version of the two-tube model generally used for humans (e.g., Kuhn, 1975). He likely closes the lower (tracheal) tube to build pressure (see Scanlan, 1988); when pressure is released, upper tube length determines the stop locus. To produce /p,b/ and possibly /g/, Alex seems to differ from the human model in two ways. First, he may extend his upper tube by esophageal involvement. Second, he may constrict his glottis (i.e., produce a small aperture between upper and lower tubes) to control output frequencies: His /p,b/ data are consistent with a mechanism in which his upper tube approaches closure at both ends and lower tube approaches closure at one end, i.e., a mechanism that involves a Helmholtz resonator (Rossing, 1982) or perturbation theory (Johnson, 1997).

C. Comparison with Mynah data

For two reasons we compare Alex's data to that of Ig-wog, a mynah whose utterances were described by Klatt and Stefanski (1974): (1) Ig-wog's vocabulary consisted of only

seven phrases, but provides the only available detailed information about word-initial /t,g/, syllable initial /b,k/, and word final /p,t,d,k,g/ in a nonpsittacine talking bird. (2) Differences in analyzable characteristics of Alex's and Ig-wog's utterances may be correlated with different articulatory and brain structures. Specifically, the mynah, an oscine songbird [order Passeriformes], has a complex syrinx and thin triangular tongue; the Grey parrot [order Psittaciformes] has a simple syrinx and fleshy tongue. Mynahs lose their ability to acquire new vocalizations at maturity (likely a consequence of reduction of dendritic spine density and enlargement of the remaining spines, Rausch and Scheich, 1982); Grey parrots not only continue to learn but, unlike mynahs, also vary their speech patterns throughout life (see Pepperberg, 1990c; Todt, 1975; neuroanatomic correlations of such flexibility have yet to be studied).

VOTs for Ig-wog and Alex differ considerably. For /t,d/, Ig-wog's VOTs are both ~40 ms whereas Alex's vary predictably across stops and are comparable to IMP's (Table II). For Ig-wog, unlike Alex, VOT does not indicate stop identity. VOT data thus suggest species differences in mechanisms for stop production. Spectral changes in Ig-wog's speech were associated with rapid articulator motion ''preceding closure and at the release of stop consonants'' (Klatt and Stefanski, 1974, p. 825). Klatt and Stefanski (1974) argue that Ig-wog's short VOTs are a consequence of his small, light articulators (e.g., small thin tongue, tiny tympanic membranes in the syrinx; Brackenbury, 1982, 1989), but humans (and parrots), with more massive articulators, produce short and even negative VOTs (Borden and Harris, 1984; Table II). Thus other physical structures (and mechanisms) are likely responsible for VOT differences: e.g., mynah external labia (syringeal membranes; Klatt and Stefanski, 1974) do not exist in Grey parrots; mynah and Grey parrot syringes also differ in other respects.

We compare stop loci for the two birds. Ig-wog's /t/ and /d/ have energy between 2 and 7 kHz, the fronted /k/ is primarily between 2 and 4 kHz (Klatt and Stefanski, 1974). Both birds' stop loci are within the range of human values but Alex displays few values >5 kHz perhaps because his articulators are not optimal for producing the high frequencies possible for the mynah. We do not know if place of constriction estimates would be appropriate for Ig-Wog's stops (Klatt and Stefanski, 1974).

Different slopes for Ig-wog and Alex may be related either to their different anatomies or trainers. Ig-wog's slopes, which are steeper than Alex's, are not likely a consequence of small fast-moving articulators: His slopes resemble those of his trainer, whose articulators are large compared to a mynah or parrot. Alex's more gradual slopes also somewhat resemble those of IMP, his primary trainer. Our data do not allow us to determine the origins of the slopes, but bird-to-trainer similarities suggest that slope patterns may be independent of anatomy and acquired by learning.

Ig-wog and Alex also differ in the plasticity of their speech. Ig-wog's utterances, especially his stops, are consistent ''from repetition to repetition'' (Klatt and Stefanski, 1974, p. 825). In contrast, Alex varies production of individual labels and phonemes. Intrasong variability exists in some songbird repertoires (e.g., song sparrows [*Melospiza melodia*], Stoddard *et al.*, 1988), but, to our knowledge, variability has not been reported for oscine mimetic speech. Ig-wog's consistency also suggests that, unlike humans, he does not parse utterances into smaller units. Alex, however, displays at least two patterns in his vocal behavior: Some phrases, like Ig-wog's, may be undivided unanalyzed chunks, but much of Alex's speech, especially referential words and phrases, is parsed into component parts, as reflected in his scaffolding and sound play (Pepperberg, 1990b; Pepperberg *et al.*, 1991).[22] His parsing ability suggests a more complex program or programs for producing speech than Ig-wog's. Although the training techniques Alex experienced (e.g., Pepperberg, 1981), which likely differed from Ig-wog's, may be responsible for his parsing ability, significant recombinatory ability has been noted in the speech of other psittacids (e.g., Banta, in preparation; Patterson and Pepperberg, 1994; Todt, 1975). Given that brain structures that support vocal learning in oscines and psittacids appear anatomically different (see Striedter, 1994), our data suggest that these structures may also be functionally different, not only with respect to open versus closed learning, but also possibly in terms of parsing ability. Clearly, we need additional studies on the acquisition of human speech by mynahs.

V. GENERAL DISCUSSION

As for vowels (Patterson and Pepperberg, 1994), similarities and differences exist in our subjects' stop productions. General similarities involved acoustic characteristics and separation of stops into sets based on voicing and place of constriction. Differences involved specific ways stops separated into these sets and, not surprisingly, the mechanisms used to produce speech. Overall, our findings suggest that perceptual discontinuities exploited by speech may be basic to vertebrate systems.

Alex's stops have statistically distinct acoustic characteristics. His stops were generally characterized by the same acoustic measures used for humans, but relationships between stops and measures were more robust for IMP than for Alex. Also, Alex's stops divided into voicing and place subsets that were not identical to IMP's: His /b/ was somewhat of an outlier, and /p/ clustered nearer /d,g/ than /k,t/. Such findings are not surprising given that Alex produces /b,p/ without lips or teeth; he may compensate by using specific glottal and esophageal mechanisms unavailable to humans.

We initially predicted stronger correlations for Alex than IMP between stops and measures closely associated with context effects, e.g., F_x. Assessing the relative importance of C2 in the two subjects, however, was difficult: Alex's F_2 is more closely correlated to stop identity than IMP's, but F_1 and F_3 are less closely correlated. Thus we could neither support nor reject our prediction.

Our findings allow us to suggest possible strategies for Alex's speech production and make comparisons with those proposed for humans and a mynah. Our results are consistent with both a modified version of the two-tube model generally accepted for human speech (e.g., Borden and Harris, 1984) and proposed for Alex's vowels (Patterson and Pepperberg, 1994; Warren *et al.*, 1996), and with a form of esophageal

and/or glottal involvement (Patterson *et al.*, 1997) unlike that in humans.[23] Alex's and a mynah's speech differ in acoustic characteristics; in conjunction with known species' differences in vocal tract anatomy and neuroanatomy, these findings suggest that not all avian imitators of human speech use similar mechanisms. Differences between species likely extend beyond production mechanisms to involve neurological mechanisms that underlie vocal plasticity.

Our findings also have implications concerning the uniqueness of human speech. Specifically, Alex does not seem to mimic either IMP's articulatory motions or acoustic idiosyncracies; his use of F_2 to do more work than IMP's and possible use of glottis and esophagus suggest that he constructs his own articulatory and acoustic solutions to the task of producing speech with a psittacine anatomy (e.g., lacking teeth and lips, using syrinx and beak). That he maintains, to some extent, distinctions between voiced/voiceless stops and among labial, alveolar, and velar stops implies that these distinctions may be basic to vertebrates rather than to mammals. In addition, his recombination of human phonemes to produce new speech patterns from existent ones (solitary sound play, Pepperberg *et al.*, 1991) suggests that he maintains a complex acoustic map between the human system and his own. To understand Alex, humans must likewise maintain some mapping between the two systems. These observations not only provide evidence for flexibility in human speech perception, but also suggest that such flexibility is shared with other species and that phonetic categories of considerable generality can develop in nonhumans. Also, Alex may exploit anticipatory coarticulation, which, in conjunction with his solitary sound play (Pepperberg *et al.*, 1991), is consistent with possible top-down processing; i.e., Alex not only appears to separate specific phonemes from the flow of speech, but also to produce these sounds so as to facilitate production of upcoming phonemes. Specifically, our data suggest, contra Lieberman (1996), that voluntary control of portions of the vocal tract for linguistic ends is not uniquely human, and that Grey parrots, like humans, use phonetic distinctions to produce meaningful sounds that are ''largely independent of affect and psychological state'' (p. 57).

ACKNOWLEDGMENTS

We thank Steven Hopp for access to and training on Signal, and Pamela Banta and particularly Dick Demers and Kerry Green for comments on the original manuscript. We thank numerous students for eliciting Alex's speech samples. This study was supported by NSF Grant IBN 92-21941 to I.M.P. and funds provided by the Alex Foundation, the van Cleave family, Gary and Susan Biggs, Sally Blanchard, the Gateway Parrot Club (St. Louis), Avicultural Society of Tucson, Midwest Bird Club of Omaha, Greater Rochester Hookbill Association, South Jersey Bird Club, Long Island Feather Enthusiasts, Bird Clubs of Virginia, Alabama Bird Clubs, Central Coast Avicultural Society, Northern Illinois Parrot Society, Midwest Avian Research Expo, Boston Exotics, Tennessee Valley Exotic Bird Club, and the Wasatch (Salt Lake City) Avian Education Society.

[1]At least one Grey parrot, the subject of this study, does indeed produce *meaningful* minimal pairs (e.g., *tea*, *pea*): He understands that requesting the former provides a sip of liquid and the latter a vegetable (Pepperberg, 1990a).

[2]Specifically, if Alex's isolated phonemes are more difficult to distinguish auditorially than those of his human trainers, context effects may be more important for helping us understand his speech overall. We have some preliminary data that suggest this possibility (Patterson, unpublished).

[3]This tape recorder uses a ''bucket brigade device'' that ensures recording without cutting off the starting point of an utterance. The sensitivity was set so that recording began at sound levels comparable to human conversation.

[4]CATMOD (CATegorical data MODeling) is a procedure for running categorical data in SAS; CATMOD, for example, streamlines the process of performing correlations on categorical data.

[5]Signal is a computer program specifically designed for acoustic analysis; one of its features enables researchers to construct tape loops from individual sound elements.

[6]Scores might in fact have been depressed by our taking samples out of their phrase context: Some of Alex's samples, for example, come from rote phrases (e.g., ''go back'') whose familiarity, if kept intact, would assist in identification of the phrase-internal stop.

[7]Most speech research involves adult human subjects, and phoneticians typically control for a factor like following vowel by eliciting only samples of interest. Alex, however, like a small child, has a limited repertoire and we had to use available speech samples rather than a standard set. Some combinations (e.g., /pi/; Table I) were difficult to elicit or produced too rarely for us to collect many samples; other combinations (e.g., /tu/) provided a high proportion of samples for a particular stop.

[8]All of IMP's samples matched Alex's; when Alex's samples were extracted from phrases, IMP reproduced the same phrases to provide equivalent samples.

[9]Some disagreement exists as to how terms like stop loci (our dBx) and F_x onset are to be defined and measured. Sussman and Shore (1996), for example, recently proposed to begin measurement of onset during the aspiration period following release of the stop; thus their formant onset begins at the point that we define as a stop locus. Other researchers, such as Olive *et al.* (1993), do not begin to measure formant onset until onset of voicing, i.e., the first glottal pulse of the vowel. We followed the latter convention.

[10]Although VOT is expected to correlate with the voiced/voiceless distinction for humans, we did not know what to expect for Alex with respect to either the voiced/voiceless distinction or the place of constriction distinction; we thus tested all possible correlations.

[11]Again, we use the human term ''phoneme'' for convenience when describing Alex's productions.

[12]For example, if CATMODs show that STOP and WORD are strongly interdependent in our data set, we can use the models STOP and WORD individually, but not the model WORD+STOP.

[13]NMCs perform pairwise comparisons of regression coefficients, R^2s, of the various models (NB: the number of variables must differ between models and a model may have a single variable), using an F test to determine if the models account for significantly different amounts of variance. Optimal models produce the highest R^2s, relative to the number of terms in that model, for the measure under consideration. R^2 either increases or remains constant as the number of terms in a model increases. Thus the optimal model does not necessarily provide the highest R^2, but the highest R^2 relative to the number of terms included; i.e., if R^2s from two models do not differ significantly (based on an F test), the model with fewer terms is optimal. Thus a model with one term that explains 75% of the variance in VOT (i.e., has an $R^2=0.75$) is better than a model with two terms that explains 75.5%.

[14]We do not mean to imply here that Alex constricts his vocal tract in the same manner as do humans; clearly, he lacks lips and his tongue may or may not function as does the human tongue. We wish, however, simply to determine the size of Alex's chamber above the place of constriction from the acoustic data for comparison with humans.

[15]Much of the variance in measures can be attributed to WORD or C2 (Table VIII). Extensive multicollinearity occurs for STOP+C2, C2, and WORD, but these categories are not isomorphic; in some cases, measures are clearly better predicted on the basis of one than another. In ''I wanna go eat'' /ai wanə go it/, ''go'' is a WORD, but STOP+C2 is /goi/ and C2 is /oi/. In ''I wanna go back'' /ai wanə go bæk/, WORD would be the same /go/, but STOP+C2 is /gob/ and C2 is /ob/. In such a case, C2 or

STOP+C2 would have extra predictive power. Alternatively, ''birdy'' /bɚrdi/ and ''bird'' /bɚrd/ are distinguished from each other at the word level, but not by STOP+C2 or C2. Finally, ''parrot'' /pɛrlt/ and ''carrot'' /kɛrIt/ are distinguished at the word level and by STOP+C2, but not by C2.

[16]Note that C1 effects may be particularly strong for these phrases for Alex: The words in these phrases, unlike words in most of his other utterances (and all of IMP's), are never produced independently; that is, he seems to produce these phrases as unanalyzed wholes. Thus the stops in these phrases might be considered word-internal.

[17]We note that although Alex technically does not have the high/low vowel distinction (Patterson and Pepperberg, 1994), his VOTs may still be differentially correlated to vowels in these groups, for example, because of anatomical constraints.

[18]/p/ in ''paper'' was word-internal, which may also have been responsible for unusual VOTs for those samples.

[19]SYSTAT will not process more than 25 samples for the cluster analysis; we thus took representative samples of each word type.

[20]Conceivably, Alex might treat /k/ in /ki/ and /k/ in /kork/ as two separate phonemes (e.g., somewhat like /hw/ versus /w/ in dialects that separate ''which'' from ''witch''), because we cannot know what, for Alex, ''...has the capacity to distinguish words or morphemes under normal conditions'' (Wolfram and Johnson, 1982, p. 42). If this assumption were true, coarticulation would not necessarily be involved with his production. We believe, however, that his data on sound play (e.g., variants of ''nail'': /bæneIl/, /bəneIl/, /bleIl/, /ʧeIl/, /jəmeIl/, /ləmeIl/, /lobaneIl/, /lubaneIl/, /meIl/, /ʃeIl/, /wgeIl/; Pepperberg *et al.*, 1991) and minimal pairs, e.g., his productions of /ki/, /pi/, /kɛrIt/, /pɛrIt/, etc., argue against this interpretation.

[21]We emphasize that Alex's glottis, unlike the human glottis, is not his sound source. Rather, Alex uses his glottis like the neck of a Helmholtz resonator to alter the resonant properties of his vocal tract.

[22]We believe that Alex's behavior is not simply a consequence of his larger repertoire, because our juvenile parrots, with small repertoires (3–5 labels) engage in similar behavior (Pepperberg *et al.*, personal observation). Note, too, that Alex engaged in this behavior even at the age of 16: When given peaches for the first time in several years and told their label, he produced, in the course of approximately 10 min, /ʧ/, /ʧIt/, /piʧ/ (Patterson *et al.*, personal observation).

[23]In humans, the esophagus is not used as a resonant chamber as it can be in Alex because in humans the esophagus is below the sound source (larynx). Even for human esophageal speech, which is generally used when the larynx is missing, the esophagus is below the pharyngo-esophageal segment which becomes the sound source (Boone and McFarlane, 1994). Alex's esophagus is above his sound source, his syrinx. We propose that the role of the syrinx is not supplanted when Alex incorporates his esophagus; rather the esophagus is used to extend the oropharyngeal cavity and thus lower values of F_2 and F_3.

Aldenderfer, M. S., and Blashfield, R. K. (**1984**). *Cluster Analysis* (Sage, Newbury Park, CA).

Banta, P.A. (in preparation). ''Lesions in the budgerigar correlate of HVc affect specific motor aspects of the production of English words and natural vocalizations.''

Banta, P.A. (personal communication).

Beeman, K. (**1993**). Signal; RTS version 1.2; October (Engineering Design, Belmont, MA).

Boone, D. R., and McFarlane, S. C. (**1994**). *The Voice and Voice Therapy* (Prentice-Hall, Englewood Cliffs, NJ).

Borden, G. J., and Harris, K. S. (**1984**). *Speech Science Primer: Physiology, Acoustics, and Perception of Speech* (Williams and Wilkins, Baltimore, MD).

Brackenbury, J. H. (**1982**). ''The structural basis of voice production and its relationship to sound characteristics,'' in *Acoustic Communication in Birds, Vol. 1: Production, Perception and Design Features of Sound*, edited by D. E. Kroodsma, E. H. Miller, and H. Ouellet (Academic, New York), pp. 53–73.

Brackenbury, J. H. (**1989**). ''Functions of the syrinx and the control of sound production,'' in *Form and Function in Birds*, edited by A. J. King and J. McLelland (Academic, London), pp. 193–220.

Brittan-Powell, E. F., Dooling, R. J., Larsen, O. N., and Heaton, J. T. (**1997**). ''Mechanisms of vocal production in budgerigars (*Melopsittacus undulatus*),'' J. Acoust. Soc. Am. **101**, 578–589.

Chiba, T., and Kajiyama, M. (**1941**). *The Vowel: Its Nature and Structure* (Kaiseikan, Tokyo).

Cole, R. A., and Scott, B. (**1974**). ''The phantom in the phoneme: Invariant cues for stop consonants,'' Percept. Psychophys. **15**, 101–107.

Demers, R. A. (personal communication).

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.

Dooling, R. J., Park, T. J., Brown, S. D., and Okanoya, K. (**1987**). ''Perceptual organization of acoustic stimuli by budgerigars (*Melopsittacus undulatus*): II. Vocal signals,'' J. Comp. Psych. **101**, 367–381.

Dorman, M. F., Studdert-Kennedy, M., and Raphael, L. J. (**1977**). ''Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues,'' Percept. Psychophys. **22**, 109–122.

Durand, S. E., Heaton, J. T., Amateau, S. K., and Brauth, S. E. (**1997**). ''Vocal control pathways through the anterior forebrain of a parrot (*Melopsittacus undulatus*),'' J. Comp. Neurol. **377**, 179–206.

Fant, G. (**1991**). ''Stops in CV syllables,'' in *Readings in Clinical Spectrography of Speech*, edited by R. J. Baken and R. G. Daniloff (Singular Publishing Group and KAY Elemetrics), pp. 179–215.

Fischer-Jorgensen, E. (**1954**). ''Acoustic analysis of stop consonants,'' Miscellanea Phonetica **2**, 42–49.

Halle, M., Hughes, G. W., and Radley, J.-P. A. (**1957**). ''Acoustic properties of stop consonants,'' J. Acoust. Soc. Am. **29**, 107–116.

Homberger, D. G. (**1979**). ''Functional morphology of the larynx in the parrot *Psittacus erithacus*,'' Am. Zool. **19**, 988.

Homberger, D. G. (**1986**). *The Lingual Apparatus of the African Grey Parrot*, Psittacus erithacus Linne *(Aves: Psittacidae) Description and Theoretical Mechanical Analysis. Ornithological Monographs, No. 39* (The American Ornithologists' Union, Washington, DC).

Johnson, K. (**1997**). *Acoustic & Auditory Phonetics* (Blackwell, Oxford).

Just, M. A., Suslick, R. L., Michaels, S., and Shocky, L. (**1978**). ''Acoustic cues and psychological processes in the perception of natural stop consonants,'' Percept. Psychophys. **24**, 327–336.

Klatt, D. H. (**1991**). ''Voice onset time, frication, and aspiration in word-initial consonant clusters,'' in *Readings in Clinical Spectrography of Speech*, edited by R. J. Baken and R. G. Daniloff (Singular Publishing Group and KAY Elemetrics), pp. 226–246.

Klatt, D. H., and Stefanski, R. A. (**1974**). ''How does a mynah bird imitate human speech?'' J. Acoust. Soc. Am. **55**, 822–832.

Kluender, K. R. (**1991**). ''Effects of first formant frequency on VOT judgments result from processes not specific to humans,'' J. Acoust. Soc. Am. **67**, 971–995.

Kluender, K. R., Diehl, R., and Killeen, P. R. (**1987**). ''Japanese quail can learn phonetic categories,'' Science **237**, 1195–1197.

Kluender, K. R., and Lotto, A. J. (**1994**). ''Effects of first formant onset frequency on [-voice] judgments result for auditory processes not specific to humans,'' J. Acoust. Soc. Am. **95**, 1044–1052.

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.

Kuhn, G. M. (**1975**). ''On the front cavity resonance and its possible role in speech perception,'' J. Acoust. Soc. Am. **58**, 428–433.

Ladefoged, P. (**1982**). *A Course in Phonetics* (HBJ, San Diego).

Lauter, J. L., Pearl, N., and Baldwin, C. M. (**1986**). ''VOT variability: Within-subject and between subject measurements of stop-consonant productions by female speakers of English, Japanese, Navajo, and Spanish,'' J. Acoust. Soc. Am. Suppl. 1 **80**, S62.

Liberman, A. M., Delattre, P., and Cooper, F. S. (**1952**). ''The role of selected stimulus-variables in the perception of stop and nasal consonants,'' Psychol. Monogr. **68**, 1–13.

Liberman, A. M., Delattre, P., Cooper, F. S., and Gerstman, L. J. (**1954**). ''The role of consonant-vowel in the perception of the unvoiced stop consonants,'' Am. J. Psychol. **65**, 497–516.

Lieberman, P. (**1996**). ''Some biological constraints on the analysis of prosody,'' in *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, edited by J. L. Morgan and K. Demuth (Erlbaum, Hillsdale, NJ), pp. 55–65.

Lisker, L., and Abramson, A. S. (**1991**). ''A cross-language study of voicing in initial stops: Acoustical measurements,'' in *Readings in Clinical Spectrography of Speech*, edited by R. J. Baken and R. G. Daniloff (Singular Publishing Group and KAY Elemetrics), pp. 247–285.
Mezzich, J., and Solomon, H. (**1980**). *Taxonomy and Behavioral Science: Comparative Performance of Grouping Methods* (Academic, New York).
Morse, P. A., and Snowdon, C. T. (**1975**). ''An investigation of categorical speech discrimination by rhesus monkeys,'' Percept. Psychophys. **17**, 9–16.
Norušis, M. J. (**1993**). *SPSS for Windows. Base System User's Guide, Release 6.0* (SPSS, Chicago, IL).
Nottebohm, F. (**1976**). ''Phonation in the Orange-winged Amazon parrot, *Amazona amazonica*,'' J. Comp. Physiol. **108**, 157–170.
Ohala, J. (personal communication).
Patterson, D. K. (unpublished).
Patterson, D. K., Fountain, A., and Pepperberg, I. M. (personal observation).
Olive, J. P., Greenwood, A., and Coleman, J. (**1993**). *Acoustics of American English Speech: A Dynamic Approach* (Springer-Verlag, New York).
Patterson, D. K., and Pepperberg, I. M. (**1994**). ''A comparative study of human and parrot phonation: Acoustic and articulatory correlates of vowels,'' J. Acoust. Soc. Am. **96**, 634–648.
Patterson, D. K., Pepperberg, I. M., Story, B. H., and Hoffman, E. (**1997**). ''How parrots talk: Insights based on CT scans, image processing and mathematical models,'' SPIE Proc. Physiology and Function from Multidimensional Images, 3033-02, Newport Beach, CA, pp. 14–24.
Pepperberg, I. M. (**1981**). ''Functional vocalizations by an African Grey parrot (*Psittacus erithacus*),'' Z. Tierpsychol. **55**, 139–160.
Pepperberg, I. M. (**1988**). ''An interactive modeling technique for acquisition of communication skills: Separation of 'labeling' and 'requesting' in a psittacine subject,'' Appl. Psycholing. **9**, 31–56.
Pepperberg, I. M., Neal, K., and Patterson, D. K. (personal observation).
Pepperberg, I. M. (**1990a**). ''Cognition in an African Grey parrot (*Psittacus erithacus*): Further evidence for comprehension of categories and labels,'' J. Comp. Psych. **104**, 41–52.
Pepperberg, I. M. (**1990b**). ''Referential mapping: A technique for attaching functional significance to the innovative utterances of an African Grey parrot,'' Appl. Psycholing. **11**, 23–44.
Pepperberg, I. M. (**1990c**). ''Some cognitive capacities of an African Grey parrot (*Psittacus erithacus*),'' in *Advances in the Study of Behavior, Vol. 19*, edited by P. J. B. Slater, J. S. Rosenblatt, and C. Beer (Academic, New York), pp. 357–409.
Pepperberg, I. M., Brese, K. J., and Harris, B. (**1991**). ''Solitary sound play during acquisition of English vocalizations by an African Grey parrot (*Psittacus erithacus*): Possible parallels with children's monologue speech,'' Appl. Psycholing. **12**, 151–178.
Pickett, J. M. (**1991a**). ''Consonant features, glides and stops,'' in *Readings in Clinical Spectrography of Speech*, edited by R. J. Baken and R. G. Daniloff (Singular Publishing Group and KAY Elemetrics), pp. 96–112.
Pickett, J. M. (**1991b**). ''Consonants: nasal, stop, and fricative manners of articulation,'' in *Readings in Clinical Spectrography of Speech*, edited by R. J. Baken and R. G. Daniloff (Singular Publishing Group and KAY Elemetrics), pp. 113–123.
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.
Rausch, G., and Scheich, H. (**1982**). ''Dendritic spine loss and enlargement during maturation of the speech control system in the mynah bird (*Gracula religiosa*),'' Neurosci. Lett. **29**, 129–133.
Repp, B. H. (**1983**). ''Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization,'' Speech Commun. **2**, 341–361.
Repp, B. H. (**1986**). ''Some observations on the development of anticipatory coarticulation,'' J. Acoust. Soc. Am. **79**, 1616–1619.
Repp, B. H. (**1991**). ''Effects of preceding context on the Voice-Onset-Time category boundary,'' J. Exp. Psychol: Human Percept. Perform. **17**, 289–302.
Rossing, T. D. (**1982**). *The Science of Sound* (Addison-Wesley, Phillippines).
SAS Institute (**1989**). *SAS/STAT User's guide, Version 6, Fourth ed.* (SAS Inst., Inc., Cary, NC).
Scanlan, J. (**1988**) ''Analysis of avian 'speech': Patterns and production,'' Ph.D. dissertation, University College, London.
Semok, A. (personal communication).
Silaeva, O. L. (**1995**). ''Bioacoustic types of avian vocal imitations,'' Biol. Bulletin **22**, 608–615.
Stevens, K. N. (**1993**). ''Models for the production and acoustics of stop consonants,'' Speech Commun. **13**, 367–375.
Stevens, K. N., and House, A. S. (**1956**). ''Studies of formant transitions using a vocal-tract analog,'' J. Acoust. Soc. Am. **28**, 578–585.
Striedter, G. (**1994**). ''The vocal control pathways in budgerigars differ from those in songbirds,'' J. Comp. Neurol. **343**, 35–56.
Stoddard, P. K., Beecher, M. D., and Willis, M. S. (**1988**). ''Response of territorial male song sparrows to song types and variations,'' Behav. Ecol. Sociobiol. **22**, 125–130.
Sussman, H. M., and Shore, J. (**1996**). ''Locus equations as phonetic descriptors of consonant place of articulation,'' Percept. Psychophys. **58**, 936–946.
Todt, D. (**1975**). ''Spontaneous recombinations of vocal patterns in parrots,'' Naturwissenschaften **62**, 399–400.
Warren, D. K., Patterson, D. K., and Pepperberg, I. M. (unpublished).
Warren, D. K., Patterson, D. K., and Pepperberg, I. M. (**1996**). ''Mechanisms of American English vowel production in an African Grey parrot,'' Auk **113**, 41–58.
Waters, R. S., and Wilson, W. A., Jr. (**1976**). ''Speech perception by rhesus monkeys: The voicing distinction in synthesized labial and velar stop consonants,'' Percept. Psychophys. **19**, 285–289.
Westneat, M., Long, J. H. L., Hoese, W., and Nowicki, S. (**1993**). ''Kinematics of birdsong: Functional correlation of cranial movements and acoustic features in sparrows,'' J. Exp. Biol. **182**, 147–171.
Winitz, H., Scheib, M. E., and Reeds, J. A. (**1972**). ''Identification of stops and vowels for the burst portion of /p,t,k/ isolated from conversational speech,'' J. Acoust. Soc. Am. **51**, 1309–1317.
Wolfram, W., and Johnson, R. (**1982**). *Phonological Analysis: Focus on American English* (Center for Applied Linguistics, Washington, DC).

Low-frequency amphibious hearing in pinnipeds: Methods, measurements, noise, and ecology

David Kastak and Ronald J. Schusterman
Long Marine Laboratory, 100 Shaffer Road, Santa Cruz, California 95060

(Received 4 August 1997; revised 20 October 1997; accepted 3 January 1998)

Aerial low-frequency (100–6400 Hz) hearing thresholds were obtained for one California sea lion (*Zalophus californianus*), one harbor seal (*Phoca vitulina*), and one northern elephant seal (*Mirounga angustirostris*). Underwater thresholds over a similar frequency range (75–6300 or 6400 Hz) were obtained for these three animals in addition to another California sea lion. Such data are critical, not only for understanding mechanisms about amphibious hearing and relating them to pinniped ecology and evolution, but also for identifying species at risk to man-made noise in the marine environment. Under water, the elephant seal was most sensitive, followed by the harbor seal and the sea lions. In air, the harbor seal was most sensitive, followed by the older of the two sea lions and the elephant seal. The following trends emerged from comparisons of each subject's aerial and underwater thresholds: (a) the sea lion (although possessing some aquatic modifications) is adapted to hear best in air; (b) the harbor seal hears almost equally well in air and under water; and (c) the elephant seal's auditory system is adapted for underwater functioning at the expense of aerial hearing sensitivity. These differences became evident only when aerial and underwater thresholds were compared with respect to sound pressure rather than intensity. When such biologically relevant comparisons are made, differences in auditory sensitivity can be shown to relate directly to ecology and life history. © *1998 Acoustical Society of America.* [S0001-4966(98)01804-9]

PACS numbers: 43.80.Lb [FD]

INTRODUCTION

Richardson *et al.* (1995) outlined the current state of knowledge with respect to the hearing abilities of marine mammals. These authors noted a distinct lack of data on low-frequency hearing, specifically frequencies below 1000 Hz. These data are crucial, considering recent concerns about anthropogenic noise in general, and specifically, low-frequency sound produced by oil drilling, seismic exploration, shipping, and oceanographic experiments. There are little available data on the low-frequency hearing abilities of pinnipeds, including the Phocidae (true seals) and Otariidae (sea lions and fur seals). The northern elephant seal is a species of particular concern because of its geographic range (proximity to shipping lanes as well as ATOC transmission paths) and its deep-diving ability. This study examines the low-frequency (below 6400 Hz) hearing of three species of pinniped, the harbor seal (*Phoca vitulina*), the California sea lion (*Zalophus californianus*), and the northern elephant seal (*Mirounga angustirostris*) using behavioral psychophysics. The absolute auditory thresholds obtained are placed in a comparative framework which relates hearing sensitivity to components of pinniped biology. These factors are divided into several categories: pinniped vocal behavior; auditory anatomy; amphibious hearing, including a consideration of the ''adequate'' acoustic stimulus with respect to sound source specification; the relationship between low-frequency hearing and the effects of noise; and finally, evolutionary and ecological considerations.

A. Pinniped vocal behavior and hearing

Airborne vocalizations by pinnipeds play important roles in social functions, including the delineation of territory, advertisements of dominance status, and female attendance behavior [for example, see Bartholomew and Collias (1962); Peterson and Bartholomew (1969); Schusterman *et al.* (1992)]. Some underwater vocalizations are also related to social interactions, particularly among breeding males [see Schusterman (1978) for review]. However, despite the initial claim by Poulter (1963) and later revived by Renouf and Davis (1982), that many pinniped sounds were biological sonar signals, experimental evidence for such claims has come under heavy criticism (Evans and Haugen, 1963; Schusterman, 1967; Schusterman, 1981; Wartzok *et al.*, 1984). In general, phocids are far more vocal under water than are otariids, although evidence for underwater sound production in elephant seals and monk seals is sparse [for a review, see Watkins and Wartzok (1985)].

Descriptions of the aerial and underwater sound production by pinnipeds suggest that these animals ought to possess amphibious hearing, and indeed, available evidence confirms this hypothesis. The hearing sensitivity of the ringed seal (Terhune and Ronald, 1975), harp seal (Terhune and Ronald, 1971, 1972), harbor seal (Møhl, 1968a; Turnbull and Terhune, 1990; Terhune, 1991), California sea lion (Schusterman *et al.*, 1972; Schusterman, 1974), and northern fur seal (Moore and Schusterman, 1987; Babushina *et al.*, 1991) have been behaviorally measured at frequencies above 1 kHz. Sensitivity to underwater sounds has generally been reported to be as good as or better than sensitivity to aerial sounds in all these species. The superiority of underwater sensitivity was first noted by Møhl (1968a) and later corroborated by Moore and Schusterman (1987), Terhune (1989) and Babushina *et al.* (1991).

Pinnipeds evolved from terrestrial carnivores, whose

outer and middle ears are adapted to hear airborne sound. Such an ear should be ineffective under water because of the impedance mismatch between air and water and therefore the ability of a pinniped to hear well under water is intriguing (Repenning, 1972). The 30 or so dB hearing loss in submerged humans (Wainwright, 1958) is thought to occur because under water, the sound transmission path to the inner ear shifts from middle ear conduction to bone conduction through the skull (Hollien and Brandt, 1969). Repenning (1972) hypothesized that bone conduction also occurs under water in the pinnipeds, but is enhanced relative to humans by modifications in the bones of the skull. However, he and others (Møhl, 1968b; Ramprashad, 1975; Moore and Schusterman, 1987; Terhune, 1989) also suggested that changes in the middle ear that occur upon submersion may allow it to function in a conventional fashion during diving; that is, to transmit acoustic energy to the inner ear via the tympanic route. The role of the middle ear as a variable transformer (i.e., similarity of function in air and under water accomplished by impedance matching to both media) received some experimental support from Møhl and Ronald (1975) who measured the cochlear microphonic response of a harp seal and determined that under water, sound was most efficiently transmitted to the inner ear along the external meatus. The results suggest that acoustic energy may be preferentially channeled through the lumen, walls, and/or surrounding tissue of the ear canal when the seal is under water. Additional data from the same animal suggest that the effective interaural distance corresponds to the separation of the external auditory meatal openings, and not the interbullar distance, which would be expected if massive bone conduction was the predominant mode of hearing (Møhl and Ronald, 1975). This suggests that pinnipeds are likely to rely upon ''conventional'' sound conduction pathways as well as bone conduction to detect underwater acoustic signals (see, e.g., Ramprashad, 1975).

B. Underwater functioning of the pinniped ear

The gross anatomical characteristics of the pinniped ear are suggestive of adaptation for aerial hearing, but some features are presumed to enable rather efficient underwater hearing (King, 1983). Detailed anatomical observations of the pinniped ear can be found in Repenning (1972) and Ramprashad *et al.* (1972). There are several notable differences between the two major pinniped families.

Phocids lack an external ear pinna. The external meatus is long, narrow, and filled with cerumen and hairs. It is supported by cartilage throughout most of its length, and is flexible and easily collapsed. Muscular attachments allow closure of the meatal opening, although it is unclear whether the canal is air-filled under water. The middle ear bones are larger and more dense than those of terrestrial mammals. These bones are loosely attached to the inner wall of the middle ear, and surrounded by a highly vascularized corpus cavernosum or cavernous tissue (Møhl, 1967, 1968b; Ramprashad, 1975), which is also found in the external meatus. The round window is partly or entirely shielded from the middle ear space in both the phocids and the otariids. In the elephant seal, the round window opens at the junction of the bulla and the mastoid, to the exterior of the skull (Repenning, 1972). In contrast to the phocid ear, the ear of otariids has a pinna, which is greatly reduced when compared to the pinnae of terrestrial carnivores and is often considered vestigial. The middle ear bones of otariids are less massive than those of phocids, but cavernous tissue is present and highly developed in both the meatus and middle ear space (Odend'hal and Poulter, 1966).

Møhl (1968b), as well as Odend'hal and Poulter (1966), Repenning (1972), and later Moore and Schusterman (1987) and Terhune (1989) suggested that the pinniped middle ear might function as a variable transformer, allowing these animals to hear well in air and with better than expected sensitivity in water. According to this hypothesis, as a pinniped dives, the cavernous tissue of the middle ear and external meatus engorges with blood. This may be an active mechanism, accompanying changes in blood circulation that take place during the dive response (Castellini *et al.*, 1994), or a passive response to pressure changes at depth. The swelling of cavernous tissue collapses the lumen of the external meatus as well as the middle ear air space, increasing its acoustic impedance to more closely match that of the surrounding water. This impedance change ostensibly reduces the reflection of sound energy from the tympanic membrane. Under extreme pressures, both sides of the tympanic membrane will contact fluid (tissue in the meatus and in the middle ear), allowing sound to be transmitted efficiently to the cochlea through the ossicular chain because of the minimized impedance difference across the tympanic membrane. Such hearing in the ''conventional'' sense ought to occur at depths below about 80–100 m (Repenning, 1972), because at these depths pressure-induced volume changes in the middle ear should cause cavernous tissue to contact both sides of the tympanic membrane.

C. Pressure versus intensity

It has become standard practice, when comparing aerial and underwater hearing, to convert sound pressure into sound intensity, correcting for the impedance difference between air and water [see Carey (1995) for a discussion of suggested reference units for underwater sound]. This practice has been justified on the presumption that underwater hearing takes place via bone conduction, rather than ''conventional'' pressure transduction. Sound energy reaching the inner ear in this fashion is often assumed to be transferred via particle displacement, or one of its derivatives, velocity or acceleration (Lombard and Hetherington, 1993). While both the inner ear and lateral line systems of many fish are sensitive to acoustic particle motion rather than pressure (Harris and van Bergeik, 1962; Kalmijn, 1989), many aquatic or amphibious animals, such as otophysine fish (Popper and Fay, 1973, 1993) and certain frogs (Hetherington and Lombard, 1982) have been shown to be sensitive to sound pressure rather than, or in addition to, particle motion. These animals make use of functional analogs to the mammalian middle ear to detect pressure fluctuations related to acoustic phenomena. It is reasonable to suggest that the pinniped ear responds to acoustic pressure (rather than particle motion) when submerged, since it possesses the typical mammalian

middle ear structures. However, partly because of a lack of understanding of the ''adequate stimulus'' for underwater hearing in mammals, and partly because of recent inderdisciplinary crossing-over among biologists and ocean engineers, measures of intensity have been adopted as a means of comparing aerial and underwater sound, and consequently, hearing thresholds in amphibious animals.

Intensity is strictly a measure of the rate of energy flow in the sound field. Therefore it can be used as a basis of comparison without considering the transduction of sound energy from the environment to the receiver's auditory system. In addition, its use is already accepted by ocean engineers and acousticians (Carey, 1995). The use of intensity is problematic, however, since it is not directly measured but rather estimated from pressure measurements using the equation $I=p^2/\rho c$. This equation describes the relationship between pressure and intensity in a plane progressive wave, free from boundaries and reflection. In practice this relationship rarely holds, especially in test tanks (mainly due to the presence of reflective boundaries), where intensity can be underestimated, sometimes by one or more orders of magnitude (Lombard *et al.*, 1981). Further, it is extremely difficult to directly measure intensity under these conditions, especially when the acoustic stimuli consist of pure tones (Fahy, 1989). Test environments which are far from ideal in these terms include shallow tanks, ocean pens in shallow water, and the hydrodynamic near field of a projector. All data published on underwater hearing in marine mammals were obtained under at least one of these three conditions. Although sound-pressure thresholds obtained in such studies are accurate, intensity calculations based solely on these pressure thresholds are likely to be in error.

In this experiment, our approach was to compare aerial and underwater thresholds with respect to pressure (in addition to intensity) for three reasons: (1) the difficulties in estimating intensity in our testing situation precluded the determination of accurate intensity thresholds; (2) there is a body of data (see discussion) implicating pressure as the primary stimulus to which the pinniped ear is sensitive under water; and (3) comparisons using pressure rather than intensity illustrate the relative amphibious adaptations of the three species we tested, because the pinniped ear appears to have retained its pressure transducing capabilities in air (Møhl, 1968b; Repenning, 1972; Ramprashad, 1975).

D. Man-made noise

Even though we know that some species of pinnipeds are quite sensitive to frequencies ranging from 1 to about 50 kHz, virtually nothing is known about their low-frequency hearing sensitivity. Since most man-made sounds are low in frequency, and the levels of such noise in the ocean are increasing (Urick, 1986), the impact of anthropogenic sound in the ocean on pinnipeds is largely unknown. In a band between about 20 and 200 Hz, ambient noise in the ocean is dominated by shipping noise. This noise is higher in shipping lanes, along coasts, and in bays and harbors; however, sounds can propagate great distances with little loss, depending on oceanographic conditions (see Urick, 1983; Richardson *et al.*, 1995, for reviews). Other sources of anthropogenic noise in the ocean are sounds from construction, ice-breaking, oceanographic experiments, and oil drilling operations. All of these sounds have the potential to interfere with acoustic signal detection in aquatic animals, but only if (1) the animals are sensitive to sounds similar in frequency, and/or (2) the sounds are loud enough to mask biologically relevant acoustic signals or to cause temporary or permanent damage to the auditory system. In light of recent increases in anthropogenic noise and concerns about acoustic habitat degradation, low-frequency hearing thresholds for a variety of species should be considered critical data.

The goals of the experiments described in this paper are to characterize the low-frequency aerial and underwater hearing sensitivity of individual animals representing three pinniped species, two relatively shallow divers (*Zalophus* and *Phoca*) and one deep diver (*Mirounga*). The procedures utilize operant conditioning of behavioral responses to pure tones to determine auditory thresholds (Stebbins, 1970). The resulting data are compared with respect to medium (air or water) and species and related to ecology, evolution, life history, and auditory anatomy.

I. METHODS

A. Acoustic mapping

The acoustic response of a 7.6-m test pool to low-frequency pure tones was mapped prior to the audiometric tests on three pinnipeds. Pure tones at frequencies ranging from 100 to 1000 Hz were projected from a J9 underwater transducer. Measurement was made with an H56 calibrated hydrophone, a Tektronix oscilloscope, and a General Radio Company type 1554A vibration and noise analyzer equipped with third-octave filters to measure ambient noise. Once a region of sufficient and constant signal intensity was located, its position was recorded. Away from its walls, the tank had good response i.e., a signal with minimal fluctuations in amplitude was measured) to all frequencies below 1 kHz with variable reliability (some amplitude variation) at frequencies between 1600 and 6400 Hz. Although higher-frequency signals could be produced and received, frequency-dependent amplitude fluctuations of up to 10 dB were apparent. Reliability at frequencies much higher than 6400 Hz was generally poor. Once an area of maximal intensity was mapped, the experimental apparatus was arranged so that the subjects could be trained to station in that position for threshold testing.

B. Subjects

The subjects of aerial threshold testing were Rocky, a female California sea lion (*Zalophus californianus*), aged 17–19 years during the course of the experiments, Sprouts, a male harbor seal (*Phoca vitulina*), aged 5–7 years during testing, and Burnyce, a female northern elephant seal (*Mirounga angustirostris*), aged 1–3 years during testing. These three animals, in addition to Rio, a female *Zalophus* aged 7–9 years, were the subjects of the underwater hearing experiments. All three animals were kept in free-flow saltwater pools and adjacent haulout areas at Long Marine Laboratory

in Santa Cruz, California. They were fed a mixed diet of herring and capelin (4–10 kg per day), and usually consumed 20%–30% of a daily ration during test sessions.

Neither of the sea lions nor the harbor seal had ever been treated for disorders involving the ear. Although these subjects were occasionally treated with antibiotics, aminoglycosidic compounds were never used. The elephant seal, however, developed a chronic otitis externa prior to being tested. The infection was confined to the right external meatus, and occurred following an intrusive procedure conducted at Hubbs-Sea World Research Institute (Yochem, personal communication). The ear was treated both systemically and topically. Hair cell damage caused by topical administration of neomycin was determined to be extremely unlikely given the external localization of the infection. Aminoglycosidic compounds were not administered systemically.

C. Apparatus

1. Aerial

In-air threshold measurements were obtained outdoors, on a haulout space adjacent to the test pool. The aerial response apparatus was an approximately cubic PVC box measuring 45×45×63 cm. Two metal slots were attached to the inside front face of the apparatus. An opaque Plexiglas door (39×45 cm) was mounted in these slots and connected to a rope and pulley assembly. The door served to separate a chin station and a response paddle.

2. Underwater

The response apparatus was a larger version of the one used in air. The dimensions were 43×135×100 cm. The relative positions of projector and receiver were modified to maintain a uniform sound field at the chin station during all phases of testing.

D. Stimuli

1. Aerial

Pure tones were produced by a Stanford Research Systems DS345 function generator and SRS Arbitrary Waveform Composer software run on a 486-based PC. All waveforms were 500 ms in duration, and shaped with rise–fall times of 40 ms to eliminate switching transients. Waveforms were monitored on a Hitachi V202 oscilloscope. Signals were triggered manually from the function generator. The output of the DS345 was fed to an H-P 350C stepwise attenuator, then to a Realistic MPA-20 power amplifier. Signals were presented to the subjects through Telephonics TDH-39 earphones that were secured to pockets of specially designed neoprene harnesses. The earphone openings were placed directly over the ears of the subjects. Signal measurements at the opening of the external meatus were made with an Etymotic Research ER-7C clinical probe microphone system. Ambient noise levels under the earphones were determined in third-octave bands using the probe microphone, a PC sound card (22-kHz sampling rate, 16-bit recording) and Spectra Plus software (Pioneer Hill). At frequencies below 1 kHz, placement of the earphones decreased ambient noise at the meatus by approximately 7–15 dB.

2. Underwater

Pure tones for underwater testing were generated in the same manner as the in-air signals. All waveforms had durations of 500 ms and rise–fall times of 40 ms. The signals were projected by a J9 underwater transducer. The J9 was placed 135 cm away from the pool wall and 157 cm below the pool rim, on the horizontal axis shared by the stationing arm. The distance between the J9 and the end of the stationing arm was approximately 5 m. Sound-pressure levels were recorded at the stationing device by a calibrated H56 hydrophone. Signal waveforms were monitored for distortion on the V202 oscilloscope during all phases of testing the three animals. Additionally, the PC-based real-time spectrum analyzer was used to monitor signal and noise levels on a trial to trial basis.

E. Procedure

1. Aerial

Prior to the start of a session, the subject was called out from the pool and the headphones were fitted into position by a trainer. Before each trial, the placement of the headphones was checked and the subject was required to place its nose on the station. When the subject was stationed properly, a concealed assistant raised the door to expose the response paddle. The opening of the door served as a ''ready'' signal for the animal. The door remained open between 5 and 7 s. The experimenter observed the session on closed-circuit video, and could communicate with the trainer through a microphone attached to an in-air speaker. For a signal trial, the stimulus was triggered by the experimenter between 2 and 4 s after the door opened. The trainer had no knowledge of the presence or absence of a signal. A correct detection occurred if the animal pressed the paddle. If the trial was a catch trial (no signal), a correct rejection occurred if the animal remained stationed until the door was closed, signifying the end of the trial. All correct responses were confirmed by the experimenter and the information was relayed to the trainer. Subsequently, the trainer would reinforce the response with a piece of fish. Incorrect responses were not reinforced, and the animal was simply restationed following these trials.

2. Underwater

Prior to the start of each underwater trial, the subject was instructed to swim down to the station by a trainer seated at the side of the pool. After the subject was properly stationed, a trial began when an assistant opened the door to expose the response paddle. Trial durations were similar to those used in air (5–7 s). Reinforcement for correct responses was delivered by the trainer upon instruction by the experimenter; incorrect responses were not reinforced.

3. Psychophysical techniques

The experimental protocols for sessions and threshold determination were set up the same way for both aerial and underwater experiments. Signal and catch trials were presented quasirandomly, with a conditional probability of 0.50 for either trial type (Moore and Schusterman, 1987). A series

of two psychophysical methods was used to obtain thresholds. The first was a tracking or ''staircase'' method (Cornsweet, 1962), in which the signal intensity was decreased by 4 dB for each correct detection (hit). Following the first failure to detect a signal (miss), the increments were changed to 2 dB (increased for misses, decreased for correct detections). The sound level was not altered after catch trials. After three to five sessions in which consistent reversals occurred, a threshold was estimated as the average between the upper and lower limits of the reversals.

Following preliminary threshold estimation, a final threshold was obtained using a method of constant stimuli (Stebbins, 1970). A series of five or six sound levels (separated by 2 or 4 dB) was chosen from a 12–20-dB range surrounding the estimated threshold. Determinations of the number of signal levels and their separation were made following the staircase phase. These determinations were based on the range of variability in the subject's staircase performance. Four or five trials of each signal level were arranged randomly within each session so that the subject could not respond on the basis of systematic changes in signal intensity. Sessions were composed of either 40 or 60 trials, depending on the number of stimulus levels used. As in the staircase sessions, 50 percent of the trials in each session were catch trials.

Threshold determination using the method of constants was based on Finney's (1971) probit analysis. Most thresholds were determined using probit methods in the following fashion: a minimum of two sessions using the method of constants were run for each frequency. If, after two sessions, the 95% confidence limits of the threshold estimate fell within ± 3 dB, that estimate was used as the final threshold. If the confidence limits fell outside ± 3 dB, threshold was recalculated following sessions 3, 4, and 5, if necessary. The maximum number of sessions of this type for any frequency was five.

Aerial and underwater thresholds were obtained at each of the frequencies for each of the animals shown in Tables I and II. Occasionally, data from the initial method of constants sessions were considered unreliable because of unacceptable variability in testing behavior. In these cases, thresholds were determined following a minimum of 5 days testing, using the staircase method. Calculations of threshold and standard deviation were made using the method of Dixon and Mood (1948) for analyzing staircase data. Criteria for acceptable variability were the same as for the method of constant stimuli.

II. RESULTS

Aerial threshold estimates and false alarm rates (during sessions used for threshold determination) for each subject are shown in Table I, and these thresholds are graphically displayed in Fig. 1. In air, the harbor seal was most sensitive at all frequencies, followed by the older sea lion and the elephant seal.

Underwater thresholds and false alarm data for all subjects are shown in Table II. These data are graphically displayed in Fig. 2. The underwater curve for the sea lion represents average thresholds obtained from both animals tested. In general, the elephant seal was most sensitive to underwater sound, followed by the harbor seal and the California sea lion. The elephant seal and harbor seal curves cross at just over 1000 Hz, but thresholds at the highest frequency tested (6300 Hz for the elephant seal and 6400 Hz for the harbor seal) are similar. Each curve shows a trend of generally increasing sensitivity from lower to higher frequencies. The two phocid curves are similar except in the vicinity of 200 to 400 Hz, where the elephant seal is about 10 dB more sensitive. The curve for the sea lion shows relatively poor hearing at frequencies below 1000 Hz. All three species exhibited similar sensitivity between 1000 and 6400 Hz.

TABLE I. Aerial sound detection thresholds (in dB *re*: 20 μPa) and false alarms (in percent of catch trials) for Rocky (*Zalophus californianus*), Sprouts (*Phoca vitulina*), and Burnyce (*Mirounga angustirostris*).

Frequency (Hz)		Rocky (*Zalophus*)	Sprouts (*Phoca*)	Burnyce (*Mirounga*)
100	Threshold	77.5	65.4	78.6
	FA	15.0	6.0	3.3
200	Threshold	57.5	57.2	72.0
	FA	17.3	11.9	2.5
400	Threshold	59.2	52.9	68.8
	FA	10.5	3.3	9.5
800	Threshold	63.1	26.1	57.3
	FA	13.3	6.7	5.7
1600	Threshold	56.9	42.8	55.3
	FA	3.3	11.6	11.1
3200	Threshold	48.1	30.2	52.7
	FA	8.8	4.1	3.2
6400	Threshold	31.4	19.2	43.5
	FA	5.4	2.8	4.1

Comparisons of aerial and underwater pressure thresholds corrected for reference level but not for impedance differences are shown in Fig. 3(A). Aerial pressure thresholds for the sea lion average 14 dB lower than underwater thresholds from the same animal, with the 1600-Hz threshold being nearly identical in both media. Pressure thresholds for the harbor seal are much more similar, with relatively minor differences (averaging 8 dB), except for the 800-Hz thresholds in which aerial sensitivity is nearly 30 dB greater than underwater sensitivity. For the elephant seal, aerial thresholds are greater than underwater thresholds for all frequencies tested. The differences in this case average 15 dB. Figure 3(B) shows the same comparisons but in this case sound-pressure levels have been converted to sound intensity levels, to correct for the impedance difference between air and water. Intensity thresholds for all the subjects are higher in air than in water, with average differences being 21, 27, and 51 dB for the sea lion, harbor seal, and elephant seal, respectively.

III. DISCUSSION

A. Near-field sensitivity

The low-frequency audiograms obtained from these subjects show a typical mammalian form (Fay, 1988). However,

TABLE II. Underwater sound detection thresholds (in dB *re*: 1 μPa) and false alarms (in percentage of catch trials) for Rocky and Rio (*Zalophus californianus*), Sprouts (*Phoca vitulina*), and Burnyce (*Mirounga angustirostris*).

Frequency (Hz)		Rocky (*Zalophus*)	Rio (*Zalophus*)	Sprouts (*Phoca*)	Burnyce (*Mirounga*)
75	Threshold	120.6	111.9	101.9	98.3
	FA	13.3	3.9	2.3	1.1
100	Threshold	119.4	116.3	95.9	89.9
	FA	6.6	10.0	5.3	2.6
200	Threshold	103.7	100.1	83.8	72.8
	FA	4.0	12.0	7.9	3.9
400	Threshold	100.0	88.9	83.9	74.9
	FA	11.1	4.7	8.8	4.1
800	Threshold	105.6	84.2	79.8	73.5
	FA	3.3	2.9	10.1	3.6
1600	Threshold	78.7	69.3	67.1	73.4
	FA	6.5	8.0	3.3	2.2
3200	Threshold				73.3
	FA				3.4
6300	Threshold				59.0
	FA				2.7
6400	Threshold	79.8	57.1	62.8	
	FA	3.3	10.2	6.0	

the sea lion's 75-Hz thresholds deviate from the expected monotonic low-frequency up-slope. The younger sea lion's threshold at 75 Hz is 3 dB lower than her threshold at 100 Hz while that of the older is no higher than her threshold at 100 Hz. This is particularly noteworthy considering that the two phocid audiograms show a fairly steep rise below 100 Hz. It is likely that the low thresholds at 75 Hz for the two sea lions resulted from a sensory modality shift (sound detection to vibration detection). Turl (1993) found a similar effect in dolphins at low frequencies which he attributed to sensation of particle motion Close to the projector, especially with long test wavelengths, such an effect is likely because of the high amplitude of particle movements associated with the near field (Siler, 1969). Close to the sound source, hydrodynamic flow dominates the sound field, and effects other than pressure (particle motion) can be responsible for sensation of a test stimulus. Beyond the near field, the effects of particle motion are less noticeable, as the ratio of pressure to particle velocity approaches that of a plane wave (Siler, 1969; Rogers and Cox, 1988). If the subject was able to sense particle motion in the near field, then it may report a signal, even if the measured sound pressure of the stimulus was at or below ambient noise levels. Turl (1993) found that after an initial response plateau corresponding to thresholds obtained for

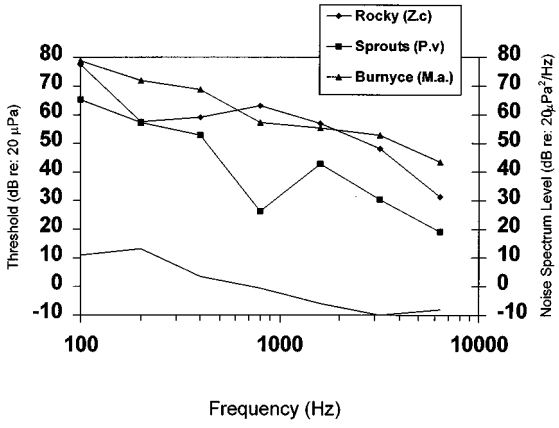


FIG. 1. Aerial sound detection thresholds (100–6400 Hz) and typical noise spectrum levels recorded beneath the headphones (at the level of the external meatal opening) for *Zalophus, Phoca,* and *Mirounga*.

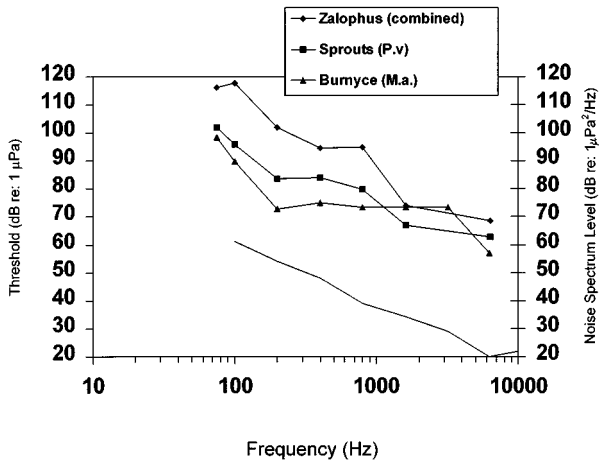


FIG. 2. Underwater sound detection thresholds (75–6400 Hz) and typical noise spectrum levels recorded at the chin station for *Zalophus* (average of two animals), *Phoca,* and *Mirounga*.

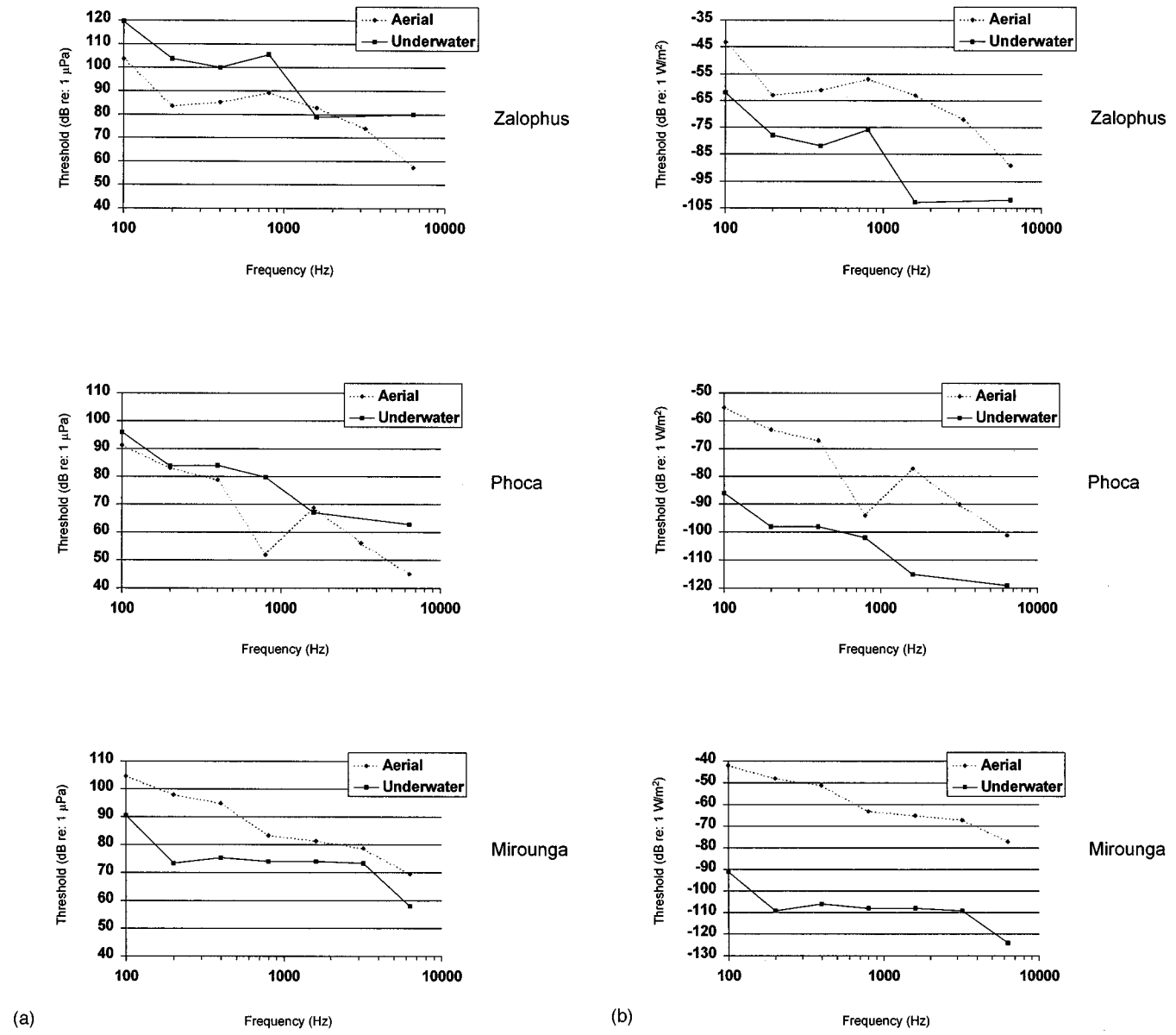


FIG. 3. Comparisons of aerial and underwater sound-pressure detection thresholds (A) and sound intensity thresholds (B) for *Zalophus*, *Phoca*, and *Mirounga*. Aerial thresholds have been corrected with respect to reference level (converted from 20 to 1 μPa). Underwater thresholds were estimated using the pressure/intensity relationship for a plane progressive wave in the far field.

Tursiops by Johnson (1967), his dolphin began responding to signal levels at sound-pressure levels below that of the ambient noise.

The sea lions in this study likely responded in the same way. The wavelength of a 75-Hz signal produced underwater is 20 m, and the approximate near-field–far-field boundary (Siler, 1969) is $\lambda/2\pi$ or 3.18 m. Although the separation of projector and chin station in these experiments was on the order of 5 m, the distinction between near and far field in the reverberant enclosure is not as clear cut, and significant deviations from expected magnitudes of particle motion were likely present outside the predicted near field (Fahy, 1989). It is thus likely that the drop in threshold at 75 Hz is due to sensation of particle motion, possibly by the sea lions' vibrissae, an effect that should be more evident with increasing wavelength. This conclusion is supported by the fact that Rocky, the older of the two sea lions, had considerable difficulty with this frequency, yet responded more often than not at stimulus levels lower than expected based on her 100-Hz threshold. The modality shift (hearing to feeling) probably caused considerable confusion for this subject.

Of further interest is the fact that neither of the phocids responded in a similar way to the 75-Hz signal. There are two possible explanations for this lack of response. The first is that neither seal was sensitive to particle motion; therefore both responded solely to auditory stimulation. This is unlikely given the bulk of anatomical and behavioral work on seal vibrissae (Renouf, 1979, 1980; Hyvaarinen, 1989) which suggests that the vibrissae are sensitive to minute near-field displacements. The second explanation is that particle motion as a cue was overshadowed by auditory cues. Whatever the case, particle motion associated with the acoustic near field appeared to control the behavior of both California sea lions and therefore might be used as a cue to

detect the presence of swimming prey at close range (Renouf, 1980).

B. Air–underwater comparisons

In order for an auditory threshold to be biologically meaningful, it must take into consideration the particular stimulus detected by the animal in nature. Thus there is some concern about whether to use sound pressure or intensity to compare aerial versus underwater hearing sensitivity. For instance, Lombard *et al.* (1981) concluded that the appropriate measure of comparison between aerial and underwater hearing in the bullfrog (*Rana catesbeiana*) is intensity. The rationale for this conclusion was that at the time, sufficient evidence that the frog ear operates in a similar fashion in both media did not exist. Because Lombard *et al.* did not think they could justify a pressure mode of underwater sound reception they compared aerial and underwater thresholds in terms of intensity. However, in a subsequent paper (Hetherington and Lombard, 1982) it was shown that under water, the ears of several frog species responded to pressure rather than particle motion. If one can reasonably assume similarity of aerial and underwater functioning of the ear, then comparisons should be made on the basis of the relevant parameter, in this case, acoustic pressure. Intensity comparisons based solely on accepted practice do little to shed light on the physiological nature of the detection process, and indeed, imply that the auditory system of amphibious animals functions differently underwater than in air.

There are various reasons to suspect that the sound pressure is conducted through the pinniped middle ear under water. Møhl and Ronald (1975) found that the optimal site for sound entry under water was over the external meatus, suggesting that it acts as an acoustic waveguide (see also Ramprashad, 1975). Additionally, experiments by Møhl and Ronald (1975) suggest that the effective interaural distance corresponds to the span between the outer ear openings rather than the separation of the inner ears. These investigators also reported that the auditory reflex could be induced underwater by loud sounds (with a corresponding 10-dB increase in threshold), again suggesting that the middle ear is functional in water. Though these data do not eliminate the possibility that bone conduction of acoustic particle motion occurs, they do indicate a preferential sound pathway similar to the one used in air. Ramprashad (1975) suggested that the sound pathway may be the walls or lumen of the external meatus, and that the inner ear might be stimulated via pressure fluctuations in the middle ear (either via the ossicles or directly through the middle ear cavity). In this respect it is interesting that there is a spatial separation of the round and oval windows in the phocids (Repenning, 1972). This separation may be an adaptation for receiving sound directly through the middle ear, functioning to maintain a high acoustic pressure difference between the oval and round windows. This pressure difference is the primary stimulus for auditory sensation in conventional hearing (Voss *et al.*, 1996).

It is possible that under water, pressure waves stimulate the pinniped inner ear in one or more of five primary ways: (1) via direct ossicular action, given that the ossicles function normally in a middle ear space of increased air density, or surrounded by cavernous tissue; (2) direct window stimulation through the middle ear cavity, with little loss due to the good impedance match between the cavernous tissue and water along with the separation of the oval and round windows; (3) via the middle ear cavity air space which should vibrate under pressure fluctuations, transmitting particle displacements to the inner ear (Lipatov, 1992, inferred a similar function for the air-filled external meatus); (4) via the skull, through translational bone conduction, in which pressure fluctuations cause the skull and tympanic bones to vibrate around the ossicles; or (5) direct compression of the cochlear capsules via bone conduction. With respect to (2) above, the relatively low ratio between the tympanic membrane and the oval window in pinnipeds is thought to be an adaptation to protect the inner ear from the generally greater sound-*pressure* levels underwater (Repenning, 1972). Thus the spatial separation of the oval window and round window may serve two functions, (1) to maintain the pressure difference between the two windows when sound waves impinge upon the middle ear; and (2) to shield the round window from the corpus cavernosum at depth, in order to maintain its compliance.

There has been considerable interest in the route of sound conduction to the pinniped inner ear since the first air/water comparisons were made (Møhl, 1968b). The acoustic impedance difference between air and water is typically used to predict a 30-dB loss in sensitivity (with respect to intensity) underwater for an air adapted ear and vice versa. Such a view of the ear as an air/water interface appears to be an oversimplification. Because the pinniped head ought to be acoustically transparent or nearly so under water, it is often stated that sound energy enters the inner ears via the skull and head tissues. However, Møhl and Ronald (1975) experimentally rejected the idea of an acoustically transparent head under water. Additionally, pinniped underwater hearing sensitivity cannot be explained on the basis of vibration of the skull by particle motion in the acoustic far field, especially at high frequencies. The amplitude of acoustic particle motion at a particular pressure threshold should not be sufficient to cause sensation, even if were transmitted without attenuation to the inner ear. For instance, at a distance of 5 m from a sound source, the particle displacement corresponding to a pressure level of 60 dB *re*: 1 μPa at 6400 Hz (elephant seal threshold) is on the order of 10^{-5} nm (Harris and van Bergeijk, 1962). Given that at threshold, hair-bundle deflection is on the order of 0.3 nm (Hudspeth, 1989), it is extremely unlikely that particle motions associated with underwater sound at any significant distance from the source (i.e., far field) are responsible for sensation by the mammalian ear. Rather, acoustic pressure is likely to be the predominant stimulus. It is known that fish with pressure transducers (swim bladders) have an acoustic advantage in frequency range as well as absolute sensitivity over fish without such transducers (Popper and Fay, 1993). By exploiting acoustic pressure, these fish are able to detect sound sources at much greater ranges than predicted on the basis of sensitivity to particle motion alone. It is extremely likely that pinnipeds as well as cetaceans are sensitive to acoustic pressure under water, because patterns of auditory sensitivity in these ma-

rine mammals reflect an expanded frequency range and a high degree of sensitivity. These characteristics are not seen in organisms sensitive strictly to acoustic particle motion. Clearly, experiments in which pressure and particle motion can be spatially separated are needed to confirm this suggestion. Until such experiments are conducted we can only speculate on sound conduction pathways, relative sensitivity to pressure versus particle motion, and the amphibious adaptations of the middle ear.

In addition to the theoretical matters, there are practical concerns when determining whether to compare auditory thresholds using pressure or intensity. Because of the acoustic properties of small test tanks, intensity calculations based on pressure measurements are likely to be *underestimates* (Parvulescu, 1964; Hetherington and Lombard, 1982). Ridgway *et al.* (1997) recently obtained sound detection thresholds in the open ocean. The beluga whale (*Delphinapterus leucas*) pressure thresholds in this study were similar to those obtained by Awbrey *et al.* (1988) and Johnson *et al.* (1989) in a test tank and a shallow-water pen, respectively. However, the intensity values calculated from pressure in the latter two studies (tanks or shallow-water situations) are probably inaccurate. Intensity values calculated from pressure in the open ocean study should more closely conform to the $p^2/\rho c$ relationship. Thus *similar* pressure thresholds among the various belugas tested are highly likely to correspond to *different* intensity thresholds which are based solely on the configuration of the test tank and not on the receiving characteristics of the animals. In fact, intensity thresholds may be specific to various experimental procedures (e.g., presence of reflective boundaries), whereas pressure thresholds are relatively independent of such factors, provided that the sound fields are carefully mapped. Accurate estimates of intensity are impossible to obtain from earlier marine mammal studies, without making unwarranted assumptions about the acoustic characteristics of the testing configuration.

Similar tests at depth with seals and sea lions will be necessary to draw firm conclusions regarding sound reception by the pinniped ear. As in the beluga, comparable hearing thresholds in tanks, shallow open water, and at depth would be indicative of pressure detection.

Finally, aerial/underwater hearing data obtained from an essentially terrestrial mammal, the human, have been compared in terms of pressure rather than intensity in 11 out of 12 studies reviewed by Kirkland *et al.* (1989). Based on a thorough consideration of the evidence as well as standard practice, we believe that pressure is the relevant parameter to use when examining underwater hearing in marine mammals.

C. Low-frequency sound and anthropogenic noise

The low-frequency thresholds obtained from these subjects suggest that the California sea lion is relatively insensitive to the frequencies associated with most types of anthropogenic sound in the ocean. At frequencies around 100 Hz the sea lion's threshold appears to be much higher than typical sources of man-made noise at moderate distances from the source.

The harbor seal is about 20 dB more sensitive to signals at 100 Hz than the California sea lion. It is thus more likely to hear low-frequency sounds from man-made sources such as ships and oil-drilling platforms. The effects of such sounds, if heard, are unknown, but potentially deleterious. For instance, harbor seal males produce low-frequency underwater sounds during the breeding season (Hanggi and Schusterman, 1994). It is possible that even if no behavioral reaction to anthropogenic noise is evident, masking of intraspecific signals may occur.

The elephant seal is the most sensitive to underwater low-frequency sound. It is thus most likely to hear low-frequency noise, but, as in other pinnipeds, the effects of such sounds are unknown. A potential factor to consider when assessing the possible effects of sound on elephant seals is that these animals are routine deep divers (LeBoeuf *et al.*, 1988). Dives to or below the deep sound channel may expose these animals to higher sound levels than would be predicted based on simple propagation models. In addition, there is some evidence that elephant seals (in contrast to sea lions and harbor seals) do not readily habituate to certain types of sound (Schusterman and Kastak, 1996), but may actually become sensitized not only to disturbing noises, but to environmental features associated with the noises.

Firm conclusions cannot be drawn regarding the effects of noise on animals, even when thresholds are known. It is possible, however, to calculate ranges of detectability for different types of sound sources. For instance, the sound used by the Acoustic Thermometry of Ocean Climate project (ATOC) has a source pressure level of 195 dB *re*: 1 μPa and a center frequency of 75 Hz. Ranges of detection estimated by assuming a simplified propagation model of spherical spreading ($20 \log R$) to a distance of 1000 m followed by $15 \log R$ (or ''lossy'' cylindrical) spreading, and the 75 Hz thresholds listed in Table II are approximately 9–34 km for a California sea lion, 160 km for a harbor seal, and 279 km for an elephant seal. Based on ocean noise measurements cited in Richardson *et al.* (1995), detection of the signal by these animals could be limited to shorter ranges at sea states of about four or higher, or in the presence of shipping or other noise (assuming critical ratios at 75 Hz to be over 20 dB). Although this simplified example does not take into account real propagation effects, it offers an illustration of how sensitivity differences might reflect differences in susceptibility to the effects of anthropogenic noise.

The presumed functioning of the middle ear as an underwater pressure transducer has important implications regarding the predictions related to both tolerable levels of sound and the physical acoustic parameters that should be used when comparing underwater sounds (such as ATOC) with their airborne counterparts. For example, as evidenced by the recent ATOC controversy, the corrections for intensity and reference levels between airborne and underwater sound has led physical acousticians and scientists dealing with the public to simply ''subtract 60 dB'' from underwater sound levels to convert to a scale purportedly comparable to the in-air scale more familiar to the public. Such a translation not only neglects the characteristics of underwater auditory systems (such as the pinniped ear), but is misleading because

TABLE III. Average differences between aerial and underwater sound detection thresholds for three phocid species, two otariid species, and one odontocete cetacean (bottlenose dolphin–*Tursiops truncatus*). Thresholds are compared with respect to pressure (column 4), corrected for reference level, and intensity (column 5), calculated from pressure measurements using $I=p^2/\rho c$. Positive values indicate greater sensitivity (lower thresholds) underwater; negative values indicate greater sensitivity in air.

Family	Species	*N*	Avg. pressure difference (dB)	Avg. intensity difference (dB)	References
Phocidae	*Phoca vitulina*	3	−10.1	25.6	Mohl, 1968a; Terhune, 1989, 1991; this study
Phocidae	*Phoca groenlandica*	1	−8.3	27.4	Terhune and Ronald, 1971, 1972
Phocidae	*Mirounga angustirostris*	1	15.9	51.6	This study
Otariidae	*Zalophus californianus*	3	−23.6	12.1	Schusterman *et al.*, 1972; Schusterman, 1974; this study
Otariidae	*Callorhinus ursinus*	3	−27.0	8.7	Babushina *et al.*, 1991; Moore and Schusterman, 1987
Delphinidae	*Tursiops truncatus*	1	23.5	59.2	Babushina, 1979

it artificially reduces (by 60 dB) the presumed amplitude of such stimuli. While these conversions are defensible from a strictly physical standpoint, they fail to address the relevant biological issues involved in the detection of underwater sound by marine animals. Without an understanding of the hearing abilities of these animals (i.e., complete audiometric assessment), perhaps the best approach is to estimate sound exposure levels based on a ''best guess'' as to the functioning of the particular auditory systems involved. We believe that from both a physiological and ecological viewpoint, the appropriate approach for comparing aerial and underwater hearing in pinnipeds should take into consideration sound-pressure levels rather than, or in addition to, sound intensity levels. Figure 3(A) and (B) illustrates the differences between these two comparisons, as well as between the conclusions drawn independently from each measure (e.g., using the intensity comparison, the sea lion ear appears ''water-adapted'' while using the pressure comparison it appears ''air-adapted'').

D. Comparative aspects

Previous intensity comparisons of other otariids were used as evidence that the otariid ear is ''water-adapted'' (Moore and Schusterman, 1987; Babushina *et al.*, 1991). Paradoxically, a study of otariid ear morphology reveals what are considered to be only minor modifications for sound reception underwater (King, 1983). When aerial and underwater thresholds are compared in terms of pressure, however, this lack of aquatic adaptation is clear. For the sea lion tested in both media, the aerial thresholds obtained in this study average 14 dB better than the underwater thresholds. From an adaptive standpoint, this implies that efficient aerial functioning is necessary for this species, and it has sacrificed some degree of aquatic adaptation in order to retain aerial sensitivity.

Compared to the otariids, *Phoca* has a more extreme sensitivity bias underwater when viewed in terms of sound intensity. These differences, on the order of 15–30 dB (Møhl, 1968b; Terhune, 1989, 1991) suggest an ear highly adapted to hearing underwater. When viewed in terms of pressure, however, the similarity of aerial and underwater thresholds obtained for the *Phoca* in this study is striking. The only great difference in sensitivity occurs at 800 Hz, and is due to the presence of a notch at this frequency in the aerial audiogram. The notch is suggestive of a resonance in the external auditory meatus. The similarity between pressure thresholds in the two media strongly suggests that, as Møhl (1968b) and others have hypothesized, the harbor seal has maximized both aerial and underwater hearing sensitivity. Absolute aerial thresholds are higher than those of most terrestrial carnivores, and underwater thresholds are higher than those of strictly aquatic mammals (e.g., cetaceans), when comparing within the range of best sensitivity.

The pressure comparisons for *Mirounga* strengthen the conclusion that the elephant seal is water-adapted. In each case, the aerial threshold is greater than the corresponding underwater threshold. From a morphological standpoint, the elephant seal ear does not appear well adapted to detect aerial acoustic signals. The external meatus is long, narrow, and closed, despite the presence of musculature attached to its cartilaginous supporting structures. It is highly unlikely that a continuous air space exists between the environment and the tympanic membrane, a necessity for efficient aerial hearing. In addition, the tympanic membrane-oval window ratio is on the order of 10:1 (Repenning, 1972), probably conferring protection from sound pressure underwater, but extremely inefficient for pressure amplification in air.

Using results from the present study and those of earlier investigations, we compared the aerial and underwater hearing sensitivity of certain phocids and otariids. These are summarized in Table III (which includes data from five phocid seals, five otariid seals, and one odontocete cetacean). These comparisons are quite revealing in demonstrating that differences in sensitivity are less variable within families than between families. The elephant seal data from Table III are significantly different at the 0.05 level (Tukey–Kramer HSD) from that of each of the other pinnipeds, and the relative differences between aerial and underwater hearing in this subject corroborate predictions based on gross morphology and ecology. Interestingly, the elephant seal data are not

TABLE IV. Ecological factors likely to have influenced low-frequency aerial and underwater hearing sensitivity in the California sea lion (*Zalophus californianus*), the harbor seal (*Phoca vitulina*), and the northern elephant seal (*Mirounga angustirostris*). Pluses indicate the relative degree to which each factor is present in each species (e.g., *Zalophus* is extremely vocal in air).

	Zalophus	*Phoca*	*Mirounga*
Aerial sound production	+++ moderate–high freq.	+ low freq.	+++ low freq. (adults)
Underwater sound production	+ broadband	+++ low freq./broadband	?
Time spent at sea	+	++	+++
Deep dives	shallow	? (most shallow; some evidence of deep dives)	+++

significantly different from that of the bottlenose dolphin, an odontocete presumably well adapted to hear underwater and poorly adapted to hear in air. The similarities imply a convergence in medium-specific auditory sensitivity between the two species; however, much more data must be collected from each species both in air and underwater in order to strengthen the conclusions based on these comparisons.

Table III also shows the same data when comparisons are made with respect to intensity rather than pressure. While the trends among family groupings are similar, the degree of aquatic adaptation appears dramatically different using such comparisons. All pinniped ears, regardless of phylogenetic relationship or life history appear water adapted when intensity is used to compare aerial and underwater hearing. While this may be somewhat informative from a physical standpoint, it fails to reveal what we believe are the appropriate relationships between life history, ecology, and amphibious hearing that emerge when comparisons are made with respect to pressure.

The elephant seal appears to be an aquatic specialist. Auditory sensitivity for this species is greater underwater regardless of whether pressure or intensity is used as a measure of comparison. These extreme modifications likely reflect the elephant seal's ability to dive regularly to depths of 500–700 m (LeBoeuf *et al.*, 1989). The external meatus, already closed in air, essentially disappears upon submergence to even shallow depths. The air normally occupying the middle ear space should yield to corpus cavernosum as pressure increases from 1 atm at the surface to over 100 atm at maximum dive depths.

The nearly complete closure of the external meatus decreases the aerial hearing sensitivity of *Mirounga* relative to other pinnipeds. However, aerial vocalizations produced by elephant seals are exceptionally loud, and are highly correlated with visual and seismic cues (Shipley *et al.*, 1992). Thus although the elephant seal has relatively poor aerial sensitivity, it has compensated to allow effective aerial communication.

Table IV shows the ecological factors that are likely involved in hearing, specifically at low frequencies. The tendencies illustrated by this table suggest that *Zalophus* should have the best aerial hearing sensitivity, followed by *Phoca*, then *Mirounga*. If underwater hearing is related to deep diving and the overall amount of time spent at sea, then *Mirounga* ought to be the most sensitive to underwater sound, followed by *Phoca*, then *Zalophus*. Overall, both *Mirounga* and *Phoca* should have good low-frequency hearing, followed by *Zalophus*.

Thus, from an ecological standpoint, the results of this study come as no surprise; *Zalophus* relies on aerial communication to a greater extent than the other species, in that it must recognize subtle signal characteristics involved in such aerial vocalizations as pup attraction calls (Gisiner and Schusterman, 1991; Schusterman *et al.*, 1992). In addition, most sea lion vocal signals have greatest intensities at 1000 Hz or higher (Peterson and Bartholomew, 1969; Schusterman, 1978), so low-frequency hearing might be expected to be relatively poor.

The harbor seal produces aerial vocalizations, but it is unclear to what extent subtle differences in signal content are detected and utilized. Harbor seal pups call simultaneously in air and underwater, thus good amphibious hearing is indicated. Apart from the pup calls, harbor seal vocalizations are either broadband clicking or creaking sounds, or low-frequency, guttural roars (Hanggi and Schusterman, 1994). Harbor seals, therefore, are expected to have relatively good low-frequency hearing and relatively good underwater hearing.

The northern elephant seal produces loud low-frequency vocalizations in air (Bartholomew and Collias, 1962). Although Poulter (1968) claims to have recorded underwater vocalizations from the northern elephant seal, evidence for this occurrence is sparse (Fletcher *et al.*, 1996). At this point it is difficult to determine whether the underwater sensitivity of the elephant seal is related to adaptive pressures to hear well under water, or to pressure-compensate at depth, or both.

IV. SUMMARY

As a group, the pinnipeds have exploited the aquatic environment but have remained tied to the shore. Retention

of the sensitivity of one sensory modality in air places constraints upon the ability of this sense to function well underwater. The phylogenetic constraints of the fissiped ear remain present to some extent in all members of this group. The hearing mechanism has been modified by the selective pressures brought about by the life histories of each species. This is most evident within the phocids, and can be seen in a comparison of the aerial and underwater hearing abilities of two species with drastically different ecologies, the harbor seal and northern elephant seal. Additional work on the sound reception pathways, hearing at depth, absolute sensitivities, frequency discrimination, sound localization, and masking in all species is needed before any more firm conclusions are drawn from behavioral data.

ACKNOWLEDGMENTS

Financial support for this study was provided by a grant to R.J.S. from the Office of Naval Research. The authors would like to thank David Levenson, Colleen Reichmuth, and Brandon Southall for assistance in experimental design, animal training, and conducting experimental sessions. C. Reichmuth additionally provided thoughtful criticisms of earlier versions of this manuscript. This paper was greatly improved by the comments of W. John Richardson and two anonymous reviewers. Thanks to the student volunteers at Long Marine Laboratory for providing logistical support and assisting with animal care and experiments. Thanks also to Gene Spurlock and Guy Oliver, who were instrumental in the acoustic mapping of the test tank.

Awbrey, F. T., Thomas, J. A., and Kastelein, R. A. (**1988**). ''Low-frequency underwater hearing sensitivity in belugas, *Delphinapterus leucas*,'' J. Acoust. Soc. Am. **84**, 2273–2275.

Babushina, Ye. S. (**1979**). ''Localization by the dolphin of the source of tonal and pulse signals in water and in air,'' Vestr. Leningr. Univ. **3**, 119–121.

Babushina, Ye. S., Zaslavskii, G. L., and Yurkevich, L. I. (**1991**). ''Air and underwater hearing characteristics of the northern fur seal: Audiograms, frequency and differential thresholds,'' Biophysics **36**, 909–913.

Bartholomew, G. A., and Collias, N. E. (**1962**). ''The role of vocalization in the social behavior of the northern elephant seal,'' Animal Beh. **10**, 7–14.

Carey, W. M. (**1995**). ''Standard definitions for sound levels in the ocean,'' IEEE J. Ocean Eng. **20**, 109–113.

Castellini, M. A., Milsom, W. K., Berger, R. J., Costa, D. P., Jones, D. R., Castellini, J. M., Rea, L. D., Bharma, S., and Harris, M. (**1994**). ''Patterns of respiration and heart rate during wakefulness and sleep in elephant seal pups,'' Am. J. Physiol. **266**, R863–R869.

Cornsweet, T. N. (**1962**). ''The staircase method in psychophysics,'' Am. J. Psychol. **75**, 485–491.

Dixon, W. J., and Mood, A. M. (**1948**). ''A method for obtaining and analyzing sensitivity data,'' J. Am. Stat. Assoc. **43**, 109–126.

Evans, W. E., and Haugen, R. M. (**1963**). ''An experimental study of the echolocation ability of a California sea lion, *Zalophus californianus* (Lesson),'' Bull. Southern California Acad. Sci. **62**, 165–175.

Fahy, F. J. (**1989**). *Sound Intensity* (Elsevier, London).

Fay, R. R. (**1988**). *Hearing in Vertebrates: A Psychophysics Databook* (Hill-Fay, Winnetka).

Finney, D. J. (**1971**). *Probit Analysis* (Cambridge U.P., Cambridge, England), 3rd ed.

Fletcher, S., LeBoeuf, B. J., Costa, D. P., Tyack, P. L., and Blackwell, S. B. (**1996**). ''Onboard acoustic recording from diving northern elephant seals,'' J. Acoust. Soc. Am. **100**, 2531–2539.

Gisiner, R., and Schusterman, R. J. (**1991**). ''California sea lion pups play an active role in reunions with their mothers,'' Animal Beh. **41**, 364–366.

Hanggi, E. B., and Schusterman, R. J. (**1994**). ''Underwater acoustic displays and individual variation in male harbour seals, *Phoca vitulina*,'' Animal Beh. **48**, 1275–1283.

Harris, G. G., and van Bergeijk, W. A. (**1962**). ''Evidence that the lateral-line organ responds to near-field displacements of sound sources in water,'' J. Acoust. Soc. Am. **34**, 1831–1841.

Hetherington, T. E., and Lombard, R. E. (**1982**). ''Biophysics of underwater hearing in anuran amphibians,'' J. Exp. Biol. **98**, 49–66.

Hollien, H., and Brandt, J. F. (**1969**). ''Effect of air bubbles in the external auditory meatus on underwater hearing thresholds,'' J. Acoust. Soc. Am. **46**, 384–387.

Hudspeth, A. J. (**1989**). ''How the ear's works work,'' Nature (London) **341**, 397–404.

Hyvaarinen, H. (**1989**). ''Diving in darkness: whiskers as sense organs of the ringed seal (*Phoca hispida saimensis*),'' J. Zool. London **218**, 663–678.

Johnson, C. S. (**1967**). ''Sound detection thresholds in marine mammals,'' in *Marine Bioacoustics*, edited by W. N. Tavolga (Pergamon, Oxford), Vol. 2, pp. 247–260.

Johnson, C. S., McManus, M. W., and Skaar, D. (**1989**). ''Masked tonal hearing thresholds in the beluga whale,'' J. Acoust. Soc. Am. **85**, 2651–2654.

Kalmijn, A. J. (**1989**). ''Functional evolution of lateral line and inner ear sensory systems,'' in *The Mechanosensory Lateral Line: Neurobiology and Evolution*, edited by S. Coombs, P. Goerner, and H. Muenz (Springer-Verlag, New York), pp. 187–216.

King, J. E. (**1983**). *Seals of the World* (British Museum of Natural History, London), 2nd ed.

Kirkland, P. C., Pence, E. A., Dobie, R. A., and Yantis, P. A. (**1989**). ''Underwater noise and the conservation of divers' hearing: A review,'' Technical report APL-UW TR8930, University of Washington.

LeBoeuf, B. J., Costa, D. P., Huntley, A. C., and Feldkamp, S. D. (**1988**). ''Continuous, deep diving in female northern elephant seals, *Mirounga angustirostris*,'' Can. J. Zool. **66**, 446–458.

LeBoeuf, B. J., Naito, Y., Huntley, A. C., and Asaga, T. (**1989**). ''Prolonged, continuous, deep diving by northern elephant seals,'' Can. J. Zool. **67**, 2514–2519.

Lipatov, N. V. (**1992**). ''Underwater hearing in seals: the role of the outer ear,'' *in Marine Mammal Sensory Systems*, edited by J. A. Thomas, R. A. Kastelein, and A. Ya. Supin (Plenum, New York), pp. 249–256.

Lombard, R. E., Fay, R. R., and Werner, Y. L. (**1981**). ''Underwater hearing in the *frog, Rana catesbeiana*,'' J. Exp. Biol. **91**, 57–71.

Lombard, R. E., and Hetherington, T. E. (**1993**). ''Structural Basis of Hearing and Sound Transmission,'' in *The Skull, Vol. 3. Functional And Evolutionary Mechanisms*, edited by J. Hanken and B. K. Hall (University of Chicago Press, Chicago), pp. 241–302.

Møhl, B. (**1967**). ''Seal ears,'' Science **157**, 99.

Møhl, B. (**1968a**). ''Auditory sensitivity of the common seal in air and water,'' J. Aud. Res. **8**, 27–38.

Møhl, B. (**1968b**). ''Hearing in seals,'' in *The Behavior and Physiology of Pinnipeds*, edited by R. J. Harrison, R. C. Hubbard, R. S. Peterson, C. E. Rice, and R. J. Schusterman (Appleton-Century-Crofts, New York), pp. 172–195.

Møhl, B., and Ronald, K. (**1975**). ''The peripheral auditory system of the harp seal, *Pagophilus groenlandicus* (Erxleben 1777),'' Rapp. P.-v. Reun. Cons. Int. Explor. Mer. **169**, 516–523.

Moore, P. W. B., and Schusterman, R. J. (**1987**). ''Audiometric assessment of northern fur seals *Callorhinus ursinus*,'' Mar. Mammal Sci. **3**, 31–53.

Odend'hal, S., and Poulter, T. C. (**1966**). ''Pressure regulation in the middle ear cavity of sea lions: A possible mechanism,'' Science **153**, 768–769.

Parvulescu, A. (**1964**). ''Problems of propagation and processing,'' in *Marine Bio-acoustics*, edited by W. N. Tavolga (Pergamon, New York), pp. 87–100.

Peterson, R. S., and Bartholomew, G. A. (**1969**). ''Airborne vocal communication in the California sea lion,'' Animal Beh. **17**, 17–24.

Popper, A. N., and Fay, R. R. (**1973**). ''Sound detection and processing by teleost fishes: A critical review,'' J. Acoust. Soc. Am. **53**, 1515–1529.

Popper, A. N., and Fay, R. R. (**1993**). ''Sound detection and processing by fish: critical review and major research questions,'' Brain Beh. Evolution **41**, 14–38.

Poulter, T. C. (**1963**). ''Sonar signals of the sea lion,'' Science **139**, 753–755.

Poulter, T. C. (**1968**). ''Underwater vocalization and behavior of pinnipeds,'' in *The Behavior and Physiology of Pinnipeds*, edited by R. J.

Harrison, R. C. Hubbard, R. S. Peterson, C. E. Rice, and R. J. Schusterman (Appleton-Century-Crofts, New York), pp. 69–84.
Ramprashad, F. (**1975**). ''Aquatic adaptations in the ear of the harp seal *Pagophilus groenlandicus* (Erxleben, 1777),'' Rapp. P.-v. Reun. Cons. Int. Explor. Mer. **169**, 102–111.
Ramprashad, F., Corey, S., and Ronald, K. (**1972**). ''Anatomy of the seal's ear (*Pagophilus groenlandicus*) (Erxleben, 1777),'' in *Functional Anatomy of Marine Mammals*, edited by R. J. Harrison (Academic, London), Vol. 1, pp. 264–305.
Renouf, D. (**1979**). ''Preliminary measurements of the sensitivity of the vibrissae of harbour seals (*Phoca vitulina*) to low frequency vibrations,'' J. Zool., London **188**, 443–450.
Renouf, D. (**1980**). ''Fishing in captive harbor seals (*Phoca vitulina concolor*): a possible role for the vibrissae,'' Netherlands J. Zool. **30**, 504–509.
Renouf, D., and Davis, M. B. (**1982**). ''Evidence that seals use echolocation,'' Nature (London) **300**, 635–637.
Repenning, C. A. (**1972**). ''Underwater hearing in seals: functional morphology,'' *in Functional Anatomy of Marine Mammals*, edited by R. J. Harrison (Academic, London), Vol. 1, pp. 307–331.
Richardson, W. J., Greene, Jr., C. R., Malme, C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, San Diego).
Ridgway, S., Carder, D., Smith, R., Kamolnick, T., and Elsberry, W. (**1997**). ''First audiogram for marine mammals in the open ocean and at depth: Hearing and whistling by two white whales down to 30 atmospheres,'' J. Acoust. Soc. Am. **101**, 3136(A).
Rogers, P. H., and Cox, M. (**1988**). ''Underwater sound as a biological stimulus,'' in *Sensory Biology of Aquatic Animals*, edited by J. Atema, R. R. Fay, A. N. Popper, and W. N. Tavolga (Springer-Verlag, New York), pp. 131–149.
Schusterman, R. J. (**1967**). ''Perception and determinants of underwater vocalization in the California sea lion,'' in *Les Systèmes Sonars Animaux, Biologie et Bionique*, edited by R. G. Busnel (Jouy-en-Josas, France, Laboratoire de Physiologie Acoustique), pp. 535–617.
Schusterman, R. J. (**1974**). ''Auditory sensitivity of a California sea lion to airborne sound,'' J. Acoust. Soc. Am. **56**, 1248–1251.
Schusterman, R. J. (**1978**). ''Vocal communication in pinnipeds,'' in *Behavior of Captive Wild Animals*, edited by H. Markowitz and V. J. Stevens (Nelson-Hall, Chicago), pp. 247–308.
Schusterman, R. J. (**1981**). ''Behavioral capabilities of seals and sea lions: a review of their hearing, visual, learning and diving skills,'' Psychol. Record **31**, 125–143.
Schusterman, R. J., Balliet, R. F., and Nixon, J. (**1972**). ''Underwater audiogram of the California sea lion by the conditioned vocalization technique,'' J. Exp. Anal. Beh. **17**, 339–350.
Schusterman, R. J., Hanggi, E. B., and Gisiner, R. (**1992**). ''Acoustic signaling in mother-pup reunions, interspecies bonding, and affiliation by kinship in California sea lions (*Zalophus californianus*),'' in *Marine Mammal Sensory Systems*, edited by J. A. Thomas, R. A. Kastelein, and A. Ya. Supin (Plenum, New York), pp. 533–551.
Schusterman, R. J., and Kastak, D. (**1996**). Pinniped acoustics: habituation and sensitization to anthropogenic signals,'' 33rd Annual Meeting of the Animal Behavior Society, Flagstaff, Az.
Shipley, C., Stewart, B. S., and Bass, J. (**1992**). ''Seismic communication in northern elephant seals,'' in *Marine Mammal Sensory Systems*, edited by J. A. Thomas, R. A. Kastelein, and A. Ya. Supin (Plenum, New York), pp. 553–562.
Siler, W. (**1969**). ''Near- and farfields in a marine environment,'' J. Acoust. Soc. Am. **46**, 483–484.
Stebbins, W. C. (**1970**). ''Principles of animal psychophysics,'' in *Animal Psychophysics: The Design and Conduct of Sensory Experiments*, edited by W. C. Stebbins (Appleton-Century-Crofts, New York), pp. 1–19.
Terhune, J. M. (**1989**). ''Can seals alter the acoustical impedance of the outer and middle ears?'' in *Proceedings of the Annual Meeting of the Canadian Acoustical Association*, edited by A. J. Cohen, Halifax, Nova Scotia.
Terhune, J. M. (**1991**). ''Masked and unmasked pure tone thresholds of a harbor seal listening in air,'' Can. J. Zool. **69**, 2059–2066.
Terhune, J. M., and Ronald, K. (**1971**). ''The harp seal, *Pagophilus groenlandicus* (Erxleben, 1777). X. The air audiogram,'' Can. J. Zool. **49**, 285–290.
Terhune, J. M., and Ronald, K. (**1972**). ''The harp seal, *Pagophilus groenlandicus* (Erxleben, 1777). III. The underwater audiogram,'' Can. J. Zool. **50**, 565–569.
Terhune, J. M., and Ronald, K. (**1973**). ''Some hooded seal (*Cystophora cristata*) sounds in March,'' Can. J. Zool. **51**, 319–321.
Terhune, J. M., and Ronald, K. (**1975**). ''Underwater hearing sensitivity of two ring seals (*Pusa hispida*),'' Can. J. Zool. **53**, 227–231.
Turl, C. W. (**1993**). ''Low-frequency sound detection by a bottlenose dolphin,'' J. Acoust. Soc. Am. **94**, 3006–3008.
Turnbull, S. D., and Terhune, J. M. (**1990**). ''White noise and pure tone masking of pure tone thresholds of a harbor seal listening in air and underwater,'' Can. J. Zool. **68**, 2090–2097.
Urick, R. J. (**1986**). *Ambient Noise in the Sea* (Peninsula, Los Altos).
Urick, R. J. (**1983**). *Principles of Underwater Sound* (McGraw-Hill, New York), 3rd ed.
Voss, S. E., Rosowski, J. J., and Peake, W. T. (**1996**). ''Is the pressure difference between the oval and round window the effective acoustic stimulus for the cochlea?,'' J. Acoust. Soc. Am. **100**, 1602–1616.
Wainwright, W. N. (**1958**). ''Comparison of hearing thresholds in air and underwater,'' J. Acoust. Soc. Am. **30**, 1025–1029.
Wartzok, D., Schusterman, R. J., and Gailey-Phipps, J. (**1984**). ''Seal echolocation?,'' Nature (London) **308**, 753.
Watkins, W. A., and Wartzok, D. (**1985**). ''Sensory biophysics of marine mammals,'' Mar. Mammal Sci. **1**, 219–260.
Yochem (personal communication).

LETTERS TO THE EDITOR

Early decay times in the Christchurch and Wellington concert halls, New Zealand

M. Barron
Department of Architecture and Civil Engineering, University of Bath, Bath BA2 7AY, England

(Received 24 June 1997; revised 18 November 1997; accepted 30 December 1997)

The concert halls in Christchurch Town Hall and the Michael Fowler Centre, Wellington, New Zealand, have large suspended reflecting surfaces which provide early lateral reflections to listeners. One objective outcome of this is that measured early decay time (EDT) values are shorter than the reverberation time. EDT values for these two halls published in the book, *Auditorium Acoustics and Architectural Design* (Spon/Chapman & Hall, London, 1993) are now known to be significantly in error, due to a software fault in the signal analysis program used. Corrected diagrams including the EDT are presented as well as a brief discussion of sound decay in these halls. © *1998 Acoustical Society of America.* [S0001-4966(98)03504-8]

PACS numbers: 43.55.Fw, 43.55.Gx [JDQ]

INTRODUCTION

While engaged in an acoustic survey of British auditoria, the author was generously invited by Professor A. H. Marshall to visit New Zealand in September 1983 at the time of the opening of the Michael Fowler Centre in Wellington. During the stay we took the opportunity to make objective measurements in both the new Wellington concert hall and the well-known Christchurch Town Hall concert hall. Professor Marshall kindly gave permission for these measurement results to be published in Appendix C of Ref. 1.

For the British survey and the New Zealand measurements, the measurement procedure was to record impulse responses on analog tape and to analyze them by computer at the Institute of Sound and Vibration Research in Southampton, using a program written by the author. Inaccurate early decay time (EDT) values arose due to a software error in this processing program. The results most affected by the error were those of the two New Zealand halls. This letter provides corrected diagrams as a replacement of those in Ref. 1 (first impression).

The recorded tapes have been reanalyzed on a PC-based system using newly written software. This new software has been validated by participation in the admirable round-robin exercise organized by Dr. Bradley.[2] The recalculated EDT values have been reported in a paper about interpretation of EDT's in concert halls;[3] this paper also listed the average errors for British halls between the old and new analyses, duplicated here in Table I. In British halls the largest error occurred in the Royal Albert Hall, London, with a difference of 0.44 s at mid-frequencies (500–2000 Hz). This is a hall with many peculiarities; for all other British halls the errors are within ± 0.10 s. Thus in British halls the errors at mid-frequencies are generally small, so that use of the old EDT values in Refs. 4 and 5 is valid. (In Ref. 5 the principal correlation with EDT is with perceived reverberance; using the new EDT data the regression coefficient is increased slightly from 0.53 to 0.59, when data from the Barbican Concert Hall, London, is omitted in each case.) In the New Zealand halls, the discrepancy at mid-frequencies between old and new is significant; the reasons for this will be described below.

The major errors in British halls occur at 125 Hz, but whether this is subjectively important is a matter of debate. For listeners the main concern regarding the bass is perceived timbre; a timbre with a rich bass is considered to possess ''warmth.'' At least two subjective studies have found correlations between the slope of the early decay time characteristic with frequency and perceived timbre.[6,7] An alternative view[8,9] is that the temporal sensitivity of the ear to low-frequency sound may be poor and that warmth is determined by the relative level of the bass sound. A recent subjective study by Bradley and colleagues[10] supports this relative level hypothesis.

Returning to objective matters, Johan Nielsen and Miklin Halstead, currently at the Acoustics Research Centre of the University of Auckland, New Zealand, have recently remeasured the EDT at the Michael Fowler Centre and obtained results in good agreement with those presented here (with a maximum discrepancy at an octave frequency of 0.09 s). While EDT figures in the first impression of my own

TABLE I. Mean corrections to EDT values for 17 British concert spaces in Ref. 1.

Octave center frequency (Hz)	125	250	500	1000	2000
EDT error (=new−old) (s)	−0.33	−0.16	+0.01	+0.05	+0.05

book[1] are in error, those supplied by me for Beranek's more recent book[11] are correct. EDT values against my name for British halls and Christchurch Town Hall in Appendix 4 of Ref. 11 are the new corrected figures.

I. THE PROGRAM FOR CALCULATING EARLY DECAY TIMES

The software program used to derive early decay times analyzed octave filtered impulse responses and followed the traditional technique. The filtered pressure responses are squared (p^2) and integrated according to the Schroeder integral. Schroeder[12] proposed that by integrating the squared impulse response, $p^2(t)$, in reverse time one derives the function identical to the average of an infinite number of noise decays, $\langle n^2(t)\rangle$:

$$\langle (n^2 t)\rangle = \int_t^\infty p^2(\tau)d\tau = \int_0^\infty p^2(\tau)d\tau - \int_0^t p^2(\tau)d\tau.$$

For computer analysis, the second form of the equation is generally more convenient. The first term involves the integrated energy from zero time to infinity; the upper limit of integration ''infinity'' requires careful consideration and is discussed in detail by Vorländer and Bietz[13] for the real situation with background noise. The EDT is derived from the slope of the first 10 dB of the decay by performing a linear regression.[14] For the EDT the choice of ''infinity'' as half the reverberation time is normally acceptable. Because the Schroeder integral is performed in reverse time, the integrated decay is a *monotonic decreasing* function.

The slope of the first 10 dB can either be derived from all the points on the decay or a sample of them. The size of the sample should be large enough for it not to influence the result; 100 points is normally sufficient. (While for slow computers sampling might save computation time, with current fast PCs sampling no longer seems appropriate.) The software program at Southampton used sampling and had been checked on pure exponential decays, for which it gave no error. It later became apparent that errors occurred for nonexponential decays. The error was found at the end of the processing part of the program at the point where the slope of the decay curve is calculated. The slope of the decay was determined from a sample, size **n**, of the total points, **m**, in the decay. But in the equation used to calculate the slope, the original number of points, **m**, was used in error instead of **n**.

The effects of this error have been checked on exponential and nonexponential decays with superimposed random monotonic decreasing fluctuations. The effect of the error in **n** depends on the shape of the decay. For a linear decay (EDT=reverberation time, RT) the effect of the error is minimal. For a sagging decay (EDT<RT) the error causes the EDT to be too short, whereas for a ballooning decay (EDT>RT) the error causes the EDT to be too long. In other words, the error exaggerates differences between the EDT and the RT. In the case of the New Zealand halls, the error led to excessively small EDT/RT ratios.

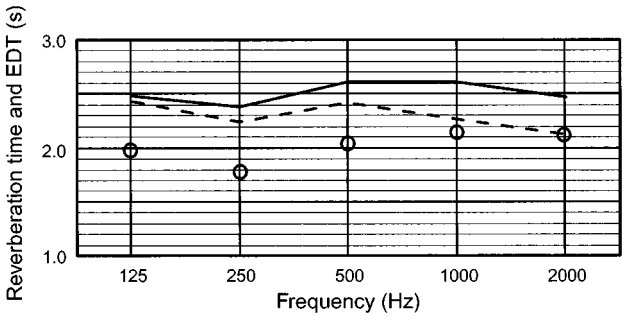


FIG. 1. Measured unoccupied (—) and occupied (---) reverberation times in Christchurch Town Hall plus mean measured early decay times (○) in the unoccupied condition.

II. EARLY DECAY TIMES IN THE TWO NEW ZEALAND LATERAL DIRECTED REFLECTION HALLS

Figures 1 and 2 show the measured reverberation times and now correct EDT's in the two New Zealand halls. The format is identical to that used in Ref. 1. Both the unoccupied and occupied reverberation times are given. Since the EDT was measured in the unoccupied condition the unoccupied results are plotted; values measured under balcony overhangs are omitted in calculating the mean EDT.[3] To interpret these figures one compares the unoccupied RT's and EDT's. If there is a discrepancy between EDT and RT, then this is also likely to occur in the occupied case.

In Ref. 3 it is suggested that the EDT/RT ratio at mid-frequencies is a valuable parameter to consider and that it constitutes a measure of the ''directedness'' of a concert hall. In a hall with diffusing wall surfaces, such as the Vienna Musikvereinssaal, the EDT/RT ratio is unity. Values of the ratio are more often less than unity and this is particularly the case in halls with surfaces which direct early reflections onto the audience. The New Zealand halls are interesting because they direct lateral reflections down onto the audience. This creates an intimate sound but it has the secondary effect of reducing the sense of reverberation, as measured by the EDT. The Christchurch and Wellington halls have EDT/RT

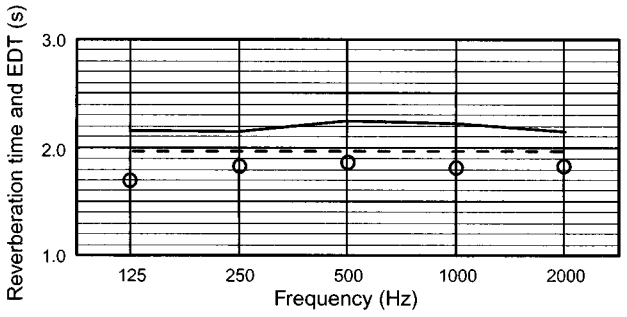


FIG. 2. Measured unoccupied (—) and occupied (---) reverberation times in the Michael Fowler Centre, Wellington, plus mean measured early decay times (○) in the unoccupied condition.

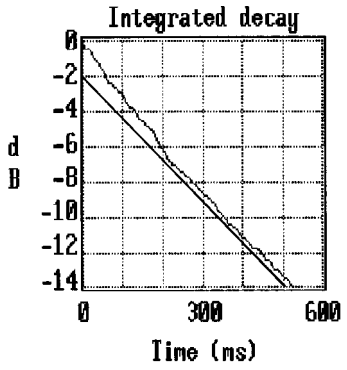


FIG. 3. Integrated decay for a Gallery position in Christchurch Town Hall at the 1-kHz octave. A straight line parallel to the terminal decay has been added for clarity.

ratios of 0.82 and 0.83 at mid-frequencies (as opposed to 0.71 for both halls quoted in Appendix C of Ref. 1).

The nature of the decay in these halls is of interest. For seats not overhung, the decays tend to have double slopes rather than a continuously decreasing slope (sometimes referred to as a sagging decay curve).[3] Figure 3 shows a typical example from the Gallery of Christchurch Town Hall. In this case the change of slope occurs around 210 ms after the direct sound. The range of delays at which there is a change of slope is from 120 to 300 ms in this hall. In the Wellington hall the corresponding range of delays is 30–165 ms. One can reasonably assume that the suspended reflectors in these halls to a certain extent divide the acoustic space in two, so that the steeper decay rate is associated with sound reflected within the volume below the top of the reflectors and at the time when the slope changes sound begins to arrive at the listener from the space above the reflectors as well.

These two New Zealand halls have been categorized as ''lateral directed reflection sequence halls.'' A characteristic of such halls is thus to have an EDT shorter than the RT. With a shorter EDT, satisfactory reverberance can be obtained in directed reflection halls by having a longer RT than normal. As already mentioned, the EDT values have been measured in unoccupied halls. One can estimate the mean occupied EDT by multiplying the measured occupied RT by the unoccupied EDT/RT ratio. This correction process gives values for the probable mean occupied EDT (at 500/1000 Hz) of 1.92 s in Christchurch Town Hall and 1.75 s in the Wellington hall.

Thus with enhanced lateral reflections in these designs, one can achieve both intimacy, source broadening (due to early lateral reflections), and reverberance. This is an exciting combination for the listener, though the acoustic character of these halls is different from that of classical rectangular halls, for instance.

[1] M. Barron, *Auditorium Acoustics and Architectural Design* (Spon/Chapman & Hall, London, 1993).
[2] J. S. Bradley, *An International Comparison of Room Acoustics Measurement Systems* (Institute for Research in Construction Internal Report IRC-IR-714, National Research Council Canada, Ottawa, 1996).
[3] M. Barron, ''Interpretation of early decay times in concert auditoria,'' Acustica **81**, 320–331 (1995).
[4] M. Barron and L-J. Lee, ''Energy relations in concert auditoriums, I,'' J. Acoust. Soc. Am. **84**, 618–628 (1988).
[5] M. Barron, ''Subjective study of British symphony concert halls,'' Acustica **66**, 1–14 (1988).
[6] P. Lehmann, ''Uber die Ermittlung raumakustischer Kriterien und deren Zusammenhang mit subjektiven Beurteilungen der Hörsamkeit,'' dissertation, TU Berlin, 1976.
[7] O. Warusfel, E. Kahle, and J-P. Jullien, ''Relationships between objective measurements and perceptual interpretation: The need for considering spatial emission of sound sources,'' J. Acoust. Soc. Am. **93**, 2281–2282 (1993).
[8] T. J. Schultz and B. G. Watters, ''Propagation of sound across audience seating,'' J. Acoust. Soc. Am. **36**, 885–896 (1964).
[9] M. Barron, ''Bass sound in concert auditoria,'' J. Acoust. Soc. Am. **97**, 1088–1098 (1995).
[10] J. S. Bradley, G. A. Soulodre, and S. Norcross, ''Factors influencing the perception of bass,'' J. Acoust. Soc. Am. **101**, 3135 (1997).
[11] L. L. Beranek, *Concert and Opera Halls: How They Sound* (Acoustical Society of America, Woodbury, NY, 1996).
[12] M. R. Schroeder, ''New method of measuring reverberation time,'' J. Acoust. Soc. Am. **37**, 409–412 (1965).
[13] M. Vorländer and H. Bietz, ''Comparison of methods for measuring reverberation time,'' Acustica **80**, 205–215 (1994).
[14] ISO 3382 (1997) ''Acoustics—Measurement of the reverberation time of rooms with reference to other acoustical parameters.''

Efficient numerical simulation of stochastic internal-wave-induced sound-speed perturbation fields

John A. Colosi
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Michael G. Brown
Rosenstiel School of Marine and Atmospheric Science, University of Miami, 4600 Rickenbacker Causeway, Miami, Florida 33149

(Received 24 November 1996; accepted for publication 24 November 1997)

An efficient method is presented to numerically simulate stochastic internal-wave-induced sound-speed perturbation fields in deep ocean environments. The sound-speed perturbation field is represented as an internal-wave eigenfunction expansion in which WKB amplitude scaling and stretching of the depth coordinate are exploited. Individual realizations of the sound-speed perturbation field are constructed by evaluating a multidimensional fast Fourier transform of a complex-valued function whose modulus has a known simple form and whose phase is random. Approximations made are shown to be consistent with approximations built into the Garrett–Munk internal-wave spectrum, which is the starting point of this analysis. Both time-varying internal-wave fields in three space dimensions and frozen fields in a vertical plane are considered. © *1998 Acoustical Society of America.* [S0001-4966(98)02703-9]

PACS numbers: 43.30.Pc, 43.30.Ft [JHM]

INTRODUCTION

Internal-wave-induced sound-speed fluctuations are the dominant source of high-frequency variability of acoustic wave fields in the ocean.[1–3] Numerical simulations of acoustic wave fields in the presence of internal-wave-induced sound-speed perturbations have been shown to be extremely useful in the interpretation of measured wave fields[4–6] and for testing theoretical predictions.[7–11] A critical shortcoming of the numerical models that have been used previously to generate random realizations of internal-wave-induced sound-speed perturbation fields is their complexity; these models are as complicated and computationally intensive as the sound propagation models that are used to model the transmission of sound through the fields generated. The purpose of this paper is to describe a simple and efficient method to construct statistically realistic random realizations of internal-wave-induced sound-speed perturbation fields. We restrict our attention to deep ocean environments and assume that the statistics of the internal-wave field are described by the empirical Garrett–Munk (GM) spectrum.[12,13]

In addition to their complexity, previously used GM-based models of internal-wave-induced sound-speed perturbation fields have made compromises that violate the GM model. For example, Flatté and Tappert[2] only modeled internal waves which propagated in the direction of acoustic propagation and Colosi *et al.*[4] and Colosi and Flatté[10] generated several 16-km sections of statistically independent internal-wave fields which were patched together in range. Other approximations have been made.[6,7,11] This paper presents an efficient numerical technique for generating GM internal-wave-induced sound-speed perturbations without making the compromises of previous work.

Our new technique to construct sound-speed perturbation fields is based on an internal-wave eigenfunction expansion. The WKB amplitude and depth scaling are shown to lead to an extremely simple analytical description of the internal-wave vertical modes. Our use of WKB scaling is shown to be consistent with approximations already built into the GM spectrum. Thus, sound-speed perturbation fields produced with our simple technique have statistics which are consistent with predictions based on the GM spectrum. We consider both time-varying internal-wave fields in three space dimensions and frozen fields in a vertical plane. In both cases, the bulk of the computation required to construct a sound-speed perturbation field is the evaluation of a multidimensional fast Fourier transform.

I. INTERNAL-WAVE-INDUCED SOUND-SPEED FLUCTUATIONS

In this section we describe the mathematical formulation of our new method to construct sound-speed perturbation fields. Much of the material presented below is not new; our main contribution is the synthesis of this material into our new method. We include this material so that our presentation is self-contained. Our presentation of this background material is similar to the presentation in Ref. 11.

Internal-wave-induced sound-speed perturbations δc are proportional to internal-wave-induced vertical displacements ζ of a fluid parcel,[1,3]

$$\delta c = \left(\frac{\partial c}{\partial z}\right)_{\text{pot}} \zeta \approx c\left(\frac{\mu}{g}\right) N^2 \zeta. \tag{1}$$

Here $(\partial c/\partial z)_{\text{pot}}$ is the potential sound-speed gradient, $\mu \approx 24.5$ is a dimensionless constant, $g = 9.8\ \text{ms}^{-2}$ is the gravitational acceleration, c is sound speed, and N is buoyancy frequency. In practice, c in this equation can be treated as a constant, N can be treated as a slowly varying function of depth z [and perhaps also lateral position (x,y)], while δc and ζ are rapidly varying functions of x, y, z and t. Equation (1) provides a simple link between the statistics of $\zeta(x,y,z,t)$ and $\delta c(x,y,z,t)$.

 0001-4966/98/103(4)/2232/4/$10.00

Statistics of $\zeta(x,y,z,t)$ are described by the empirical GM internal-wave spectrum,[12,13]

$$F_\zeta(j,\omega)=\frac{2B^2E}{\pi M}\,\frac{N_0}{N}\,\frac{f}{\omega^3}\,(\omega^2-f^2)^{1/2}\,\frac{1}{(j^2+j_*^2)}\,. \tag{2}$$

Here ω is the internal-wave angular frequency, j is the mode number, $j_*=3$ and $E=6.3\times10^{-5}$ are empirical dimensionless constants, f is the inertial frequency [$f=2\Omega$ sin (latitude) where $\Omega=2\pi/1$ day is the angular velocity of the earth], $B\approx1$ km is the thermocline depth scale, N_0 is the surface extrapolated buoyancy frequency, and $M=\Sigma_{j=1}^{\infty}(j^2+j_*^2)^{-1}\approx\frac{1}{2}\,j_*^{-2}(\pi j_*-1)$ is a normalization constant. It is often convenient to assume $N(z)=N_0\,\exp(z/B)$. More generally, we may define $N_0B=\int_{-h}^{0}N(z)dz$, where the ocean surface and bottom lie at $z=0$ and $z=-h$, respectively. The GM spectrum is normalized so that $\langle\zeta^2\rangle=\int_f^N d\omega\;\Sigma_{j=1}^{\infty}F_\zeta(j,\omega)=\frac{1}{2}\,B^2E\;(N_0/N)$, where the angular brackets denote average. It follows from the normalization that the product $(1/2)B^2E$ can be interpreted as ζ_0^2, the mean-square displacement at the depth at which $N=N_0$. The vertically integrated internal-wave potential energy density is $(\rho/2)\int_{-h}^{0}dz\;N^2\langle\zeta^2\rangle=(\rho/4)\;B^3N_0^2E$, where ρ is the density of sea water.

For some purposes it is convenient to replace the dependence of F_ζ on j or ω by dependence on horizontal wave number $k=(k_x^2+k_y^2)^{1/2}$ or vertical wave number m. These transformations require use of the dispersion relation:

$$\omega^2=N^2\,\frac{k^2}{k^2+m^2}+f^2\,\frac{m^2}{k^2+m^2}\approx\left(\frac{N_0B}{j\pi}\right)^2k^2+f^2\,. \tag{3}$$

Both forms of the dispersion relation follow from a WKB analysis. The j-dependent form is a large j asymptotic result. The functions $F_\zeta(j,k_x,k_y)$ and $F_\zeta(j,k_x)$ are required below. Noting that $F_\zeta\;dj\;d\omega=F_\zeta\;dj\;dk(\partial\omega/\partial k)$ and using the approximate form of the dispersion relation [Eq. (3)] gives

$$F_\zeta(j,k)=\frac{2B^2E}{\pi M}\,\frac{N_0}{N}\,\frac{1}{(j^2+j_*^2)}\,\frac{k_jk^2}{(k^2+k_j^2)^2}, \tag{4}$$

where $k_j=(f/N_0)(\pi j/B)$. When a change of variables is made, we normalize F_ζ so that the sum and/or integral of F_ζ over the domain of its arguments leaves $\langle\zeta^2\rangle$ unchanged. Because $2\pi\int_0^\infty dk\;kF(k)=\int_{-\infty}^{\infty}dk_x\int_{-\infty}^{\infty}dk_y\;F(\sqrt{k_x^2+k_y^2})$, we may write

$$F_\zeta(j,k_x,k_y)=\frac{B^2E}{\pi^2M}\,\frac{N_0}{N}\,\frac{1}{(j^2+j_*^2)}\,\frac{k_j\sqrt{k_x^2+k_y^2}}{(k_x^2+k_y^2+k_j^2)^2}. \tag{5}$$

The reduction of $F_\zeta(j,k_x,k_y)$ to $F_\zeta(j,k_x)$ can be done analytically:

$$F_\zeta(j,k_x)=\int_{-\infty}^{\infty}dk_y\;F_\zeta(j,k_x,k_y)=\frac{I(j,k_x)}{j^2+j_*^2}\,\frac{B^2E}{\pi^2M}\,\frac{N_0}{N}\,, \tag{6}$$

where

$$I(j,k_x)=\frac{k_j}{k_x^2+k_j^2}+\frac{1}{2}\,\frac{k_x^2}{(k_x^2+k_j^2)^{3/2}}\,\ln\left(\frac{\sqrt{k_x^2+k_j^2}+k_j}{\sqrt{k_x^2+k_j^2}-k_j}\right). \tag{7}$$

This result was originally derived by Brill and Dozier.[14]

It is important to note that the GM spectrum is closely linked to WKB analysis. We have already used the WKB result (3) to derive $F_\zeta(j,k_x,k_y)$ and $F_\zeta(j,k_x)$. This usage is consistent with the manner in which the empirical GM spectral weights were chosen; to find the spectral weights which gave the best fit to a wide variety of internal-wave measurements (dropped, towed, fixed point, and combinations thereof). To the extent that these equations introduce a bias in the spectrum for small j and k, this bias has been compensated for by the choice of spectral weights. We emphasize this point because we make use of WKB scaling below; the approximations we make are entirely consistent with approximations already built into the GM spectrum.

For the purpose of generating random realizations of $\zeta(x,y,z,t)$ [and hence $\delta c(x,y,z,t)$] it is most useful to express ζ as a sum over internal-wave vertical modes in which the weights in the expansion are functions of j, k_x, and k_y. Let G_ζ denote F_ζ after renormalizing and omitting the depth dependence of the latter. The new normalization constant is chosen so that the vertically integrated internal-wave potential energy density is unchanged. These considerations give $G_\zeta=2BN_0NF_\zeta$, so

$$G_\zeta(j,k_x,k_y)=\frac{2B^3N_0^2E}{\pi^2M}\,\frac{1}{(j^2+j_*^2)}\,\frac{k_j\sqrt{k_x^2+k_y^2}}{(k_x^2+k_y^2+k_j^2)^2}. \tag{8}$$

The desired modal expansion of ζ is then

$$\zeta(x,y,z,t)=\mathrm{Re}\left[\int\int dk_x\;dk_y\sum_j\;g(j,k_x,k_y)\times W_{jk}(z)e^{i(k_xx+k_yy-\omega_{jk}t)}\right], \tag{9}$$

where each $g(j,k_x,k_y)$ is a complex Gaussian random variable with zero mean and variance $\langle g(j,k_x,k_y)g^*(j',k_x',k_y')\rangle=G_\zeta(j,k_x,k_y)\,\delta_{jj'}\,\delta(k_x-k_x')\,\delta(k_y-k_y')$. Construction of the modes $W_{jk}(z)$ will be discussed below. These are orthonormal, $\int_{-h}^{0}dz(N^2(z)-f^2)W_{jk}(z)W_{j'k}(z)=\delta_{jj'}$, from which it follows that $\int dzN^2(z)W_{jk}^2(z)\approx1$. The mean-square displacement is $\langle\zeta^2\rangle=\frac{1}{2}\int\int dk_x\;dk_y\;\Sigma_jG_\zeta(j,k_x,k_y)W_{jk}^2(z)$, and the vertically integrated internal-wave potential energy density is $(\rho/2)\int dz\;N^2(z)\langle\zeta^2\rangle\approx(\rho/4)\int\int dk_x\;dk_y\;\Sigma_jG_\zeta(j,k_x,k_y)=(\rho/4)\;B^3N_0^2E$, consistent with the result found earlier.

To model sound propagation in a frozen (not evolving in time) vertical slice (range and depth) of the ocean, the double integral over k_x and k_y in (9) can be reduced to a single integral. This is partially justified by the horizontal isotropy of the internal-wave field; we may, without loss of generality, take the propagation plane to coincide with the $y=0$ plane. To treat this problem we require

$$G_\zeta(j,k_x)=\int_{-\infty}^{\infty}dk_y\;G_\zeta(j,k_x,k_y)=\frac{I(j,k_x)}{j^2+j_*^2}\,\frac{2B^3N_0^2E}{\pi^2M}\,, \tag{10}$$

where $I(j,k_x)$ is defined above [Eq. (7)]. The frozen (we set $t=0$ for convenience) vertical slice representation of ζ is then

$$\zeta(x,z)=\mathrm{Re}\left[\int dk_x\sum_j\;g(j,k_x)W_{jk_x}(z)e^{ik_xx}\right], \tag{11}$$

where each $g(j,k_x)$ is a complex Gaussian random variable with zero mean and variance $\langle g(j,k_x)g^*(j',k_x')\rangle = G_\zeta(j,k_x)\delta_{jj'}\delta(k_x-k_x')$. Note that k_y is set equal to zero when the modes $W_{jk_x}(z)$ are constructed. The mean-square displacement is now $\langle\zeta^2\rangle = \frac{1}{2}\int dk_x\ \Sigma_j G_\zeta(j,k_x)W_{jk_x}^2(z)$, and the vertically integrated internal-wave potential energy density is $(\rho/2)\int dz\ N^2(z)\langle\zeta^2\rangle \approx (\rho/4)\int dk_x\ \Sigma_j G_\zeta(j,k_x) = (\rho/4)\ B^3N_0^2E$, consistent with results found previously.

We turn our attention now to the difficult part of evaluating Eqs. (9) and (11). This is the problem of constructing the internal-wave modes, $W_{jk}(z)$, which satisfy

$$\frac{d^2W}{dz^2}+k^2\ \frac{N^2(z)-\omega^2}{\omega^2-f^2}\ W(z)=0 \tag{12}$$

subject to the boundary conditions $W(0)=W(-h)=0$, and the related problem of finding ω_{jk}. {Note, however, that an assumed form of ω_{jk} [Eq. (3)] has already been used.} Consistent with the observation that WKB scaling is already built into (9) and (11), we now invoke WKB scaling of both the modal amplitudes and the depth coordinate by introducing

$$V(z)=(N(z)/N_0)^{1/2}W(z) \tag{13}$$

and

$$\xi(z)=\frac{1}{N_0B}\int_{-h}^{z}N(z')dz'. \tag{14}$$

Similar transformations are routinely used in the analysis of field observations of internal waves.[15–17] It follows from the definition of Eqs. (13) and (14), and the observation that $N>0$, that ξ increases monotonically from 0 at $z=-h$ to 1 at $z=0$. Equation (12) written in terms of $V(z)$ is

$$\frac{3}{4N^2}\left(\frac{dN}{dz}\right)^2V(z)-\frac{1}{2N}\frac{d^2N}{dz^2}\ V(z)-\frac{1}{N}\frac{dN}{dz}\frac{dV}{dz}+\frac{d^2V}{dz^2}$$
$$+k^2\ \frac{N^2(z)-\omega^2}{\omega^2-f^2}\ V(z)=0. \tag{15}$$

It will be shown below that the first two terms in this equation can be neglected. Transforming the remaining terms from z to ξ and noting that $N^2-\omega^2\approx N^2$, consistent with GM approximations, gives

$$\frac{d^2V}{d\xi^2}+k^2B^2\ \frac{N_0^2}{\omega^2-f^2}\ V(\xi)=0 \tag{16}$$

subject to the boundary conditions $V(\xi=0)=V(\xi=1)=0$. Solutions are

$$V_j(\xi)=(1/N_0)(2/B)^{1/2}\ \sin(j\pi\xi) \tag{17}$$

with

$$kBN_0/\sqrt{\omega^2-f^2}=j\pi. \tag{18}$$

Here j is a positive integer and the modes are normalized so that the modal orthogonality condition with $N^2-f^2\approx N^2$ is satisfied. Note that Eq. (18) is identical to Eq. (3) which was used to derive $G_\zeta(j,k)$.

To assess the validity of neglecting the first two terms in Eq. (15), it is insightful to assume $N(z)=N_0\ \exp(z/B)$. Substituting this N into (15) and making use of Eqs. (17) and (18) leads to the conclusion that the first two terms in (15) are negligible when $(2j\pi)^2\gg(N_0/N)^2$. For more general buoyancy frequency profiles, our WKB treatment will be valid wherever the local vertical wavelength of mode j is small compared to the local vertical length scale of N. This condition may be violated for small j. Recall, however, our earlier comments about the GM spectral weights; to the extent that the use of large j asymptotic results bias the spectrum, this bias has been compensated for by the choice of the spectral weights. Again we emphasize that our approximations are consistent with approximations already built into the GM spectrum. Note also that in the deep ocean (where N/N_0 is small) errors in ζ associated with violation of the condition $(2j\pi)^2\gg(N_0/N)^2$ should give negligible errors in δc because δc is proportional to $N^2\zeta$.

The preceding discussion provides a complete mathematical description of our new technique to generate random realizations of ζ [or, using Eq. (1), δc]. We now summarize, rewriting the relevant equations in a form which is convenient for numerical purposes. For the time-dependent field in three space dimensions

$$\zeta(x,y,z,t)=\mathrm{Re}\Bigg[\frac{2B}{\pi}\left(\frac{E}{M}\right)^{1/2}\left(\frac{N_0}{N(z)}\right)^{1/2}(\Delta k_x)^{1/2}(\Delta k_y)^{1/2}$$
$$\times\sum_j\ \frac{k_j^{1/2}\sin(j\pi\xi(z))}{(j^2+j_*^2)^{1/2}}\sum_{k_x}\sum_{k_y}\ \frac{(k_x^2+k_y^2)^{1/4}}{k_x^2+k_y^2+k_j^2}$$
$$\times e^{i\phi(j,k_x,k_y)}e^{i(k_xx+k_yy-\omega_{jk}t)}\Bigg]. \tag{19}$$

Here ω_{jk} is defined by Eq. (18), Δk_x and Δk_y are the step sizes used in the integrals over k_x and k_y, and $\phi(j,k_x,k_y)$ is a delta-correlated random variable with a uniform distribution between 0 and 2π. For the frozen vertical slice problem,

$$\zeta(x,z)=\mathrm{Re}\Bigg[\frac{2B}{\pi}\left(\frac{E}{M}\right)^{1/2}\left(\frac{N_0}{N(z)}\right)^{1/2}(\Delta k_x)^{1/2}$$
$$\times\sum_j\ \frac{\sin(j\pi\xi(z))}{(j^2+j_*^2)^{1/2}}\sum_{k_x}\ (I(j,k_x))^{1/2}e^{i\phi(j,k_x)}e^{ik_xx}\Bigg], \tag{20}$$

where $I(j,k_x)$ is defined by Eq. (7). It should be noted that (19) and (20) are consistent with the GM spectral densities $F_\zeta(j,k_x,k_y)$ [Eq. (5)] and $F_\zeta(j,k_x)$ [Eq. (6)] presented above.

To numerically demonstrate the use of our new description of ζ we have computed $\delta c(x,z)$ using Eqs. (19) and (20) for an exponentially stratified ocean, $N(z)=N_0\ \exp(z/B)$ with $N_0=2\pi/10$ min and $B=1$ km, whose depth is 5 km. [For this simple case $z=B\ \ln(\xi+e^{-5})$.] Simulations, produced by evaluating Eqs. (1) and (19), are shown in Figs. 1 and 2. Figure 1 shows plots of $(N/N_0)^{-3/2}\delta c$ versus range at selected depths, at a fixed time. Figure 2 shows plots of $(N/N_0)^{-3/2}\delta c$ versus time at selected fixed locations. For these simulations we used $512\times512(k_x,k_y)$ values with $2\pi/100\ \mathrm{km}\leqslant|k_x|$, $|k_y|\leqslant 2\pi/1\ \mathrm{km}$, and $j_{\max}=50$. Our numerical estimates of the following statistical properties of δc and $\partial\delta c/\partial z$ have been found to be consistent with predictions[1,3,11] based on the GM spectrum: (i) $(\delta c)_{\mathrm{rms}}$ scales

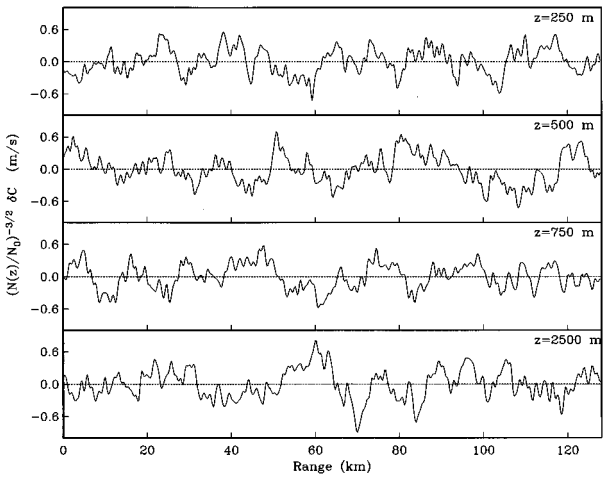


FIG. 1. Plots of $(N/N_0)^{-3/2}\delta c$ versus range at selected depths, at a fixed time, computed using Eqs. (1) and (31).

in depth like $N^{3/2}$; (ii) $(\partial\delta c/\partial z)_{\rm rms}$ scales in depth like $N^{5/2}$; (iii) the horizontal and vertical correlation lengths of δc are approximately $(\pi/8)\times(B/j_*)(N_0/f)/\ln[(N/f)-1/2]$ and $(B/(\pi j_*-1))\times(N_0/N)$, respectively; and (iv) the horizontal and vertical correlation lengths of $\partial\delta c/\partial z$ are approximately $\pi/(k_x)_{\max}$ and $(B/j_{\max})(N_0/N)$, respectively.

II. SUMMARY

In this paper we have described an extremely simple and efficient technique to simulate internal-wave-induced sound-speed perturbation fields. Our analysis is based on the empirical GM internal-wave spectrum and thus should not be expected to apply at very high latitudes, very low latitudes, or in coastal regions. Use of WKB scaling in our analysis is entirely consistent with approximations already built into the GM spectrum. In addition to simplicity and efficiency, the new method has several advantages over previously used techniques: (i) it is true to the GM spectrum (the technique used by Colosi *et al.*[4] and Colosi and Flatté[10] distorts the horizontal wave-number spectrum predicted by GM); (ii) both the surface and bottom boundary conditions are satisfied exactly (the technique used by Brown and Viechnicki[11] assumes an infinitely deep ocean); (iii) no restrictive assumptions about $N(z)$ (e.g., monotonicity, which is assumed by Brown and Viechnicki[11]) are required; (iv) lateral variability of N and/or the ocean depth can easily be treated (Colosi *et al.*,[4] Dozier and Tappert,[7,8] Colosi and Flatté,[10] and Brown and Viechnicki[11] do not address this problem but Wolfson *et al.*[6] does); (v) three-dimensional δc fields can be generated; and (vi) time-dependent δc fields can be generated.

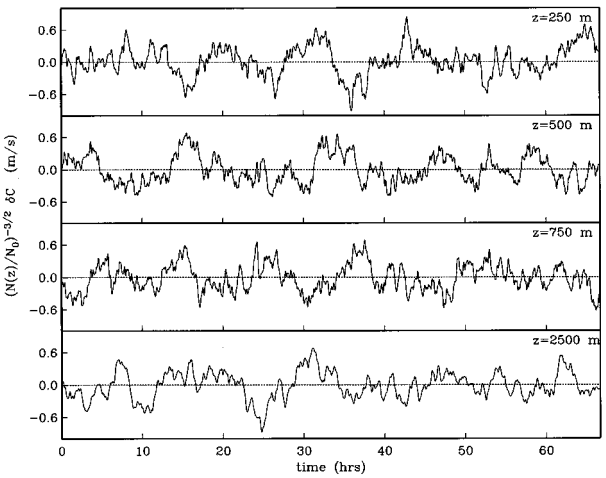


FIG. 2. Plots of $(N/N_0)^{-3/2}\delta c$ versus time at selected fixed locations, computed using Eqs. (1) and (31). The increasing smoothness of the time series plotted as depth increases (especially visible at 2500 m) is consistent with the GM prediction that the correlation time for δc at a fixed location is approximately $(\pi/(2\,fN))^{1/2}$.

ACKNOWLEDGMENTS

We thank W. Munk, F. Tappert, and M. Wolfson for the benefit of our discussions on internal-wave-induced sound speed fluctuations. This work was funded by the Office of Naval Research Grant Nos. N00014-97-1-0046 and N00014-97-1-0068. This is Woods Hole Oceanographic Institution contribution number 9374.

[1] W. H. Munk and F. Zachariasen, ''Sound propagation through a fluctuating stratified ocean: Theory and observation,'' J. Acoust. Soc. Am. **59**, 818–838 (1976).

[2] S. M. Flatté and F. D. Tappert, ''Calculation of the effects of internal waves on oceanic sound transmission,'' J. Acoust. Soc. Am. **58**, 1151–1159 (1975).

[3] S. Flatté, R. Dashen, W. Munk, K. Watson, and F. Zachariasen, *Sound Transmission through a Fluctuating Ocean* (Cambridge U.P., Cambridge, 1979).

[4] J. A. Colosi, S. M. Flatté, and C. Bracher, ''Internal-wave effects on 1000-km oceanic acoustic pulse propagation: Simulation and comparison to experiment,'' J. Acoust. Soc. Am. **96**, 452–468 (1994).

[5] P. F. Worcester *et al.*, ''A test of basin-scale acoustic thermometry using a large-aperture vertical array at 3250-km range in the eastern North Pacific Ocean,'' submitted to J. Acoust. Soc. Am. (1997).

[6] M. A. Wolfson and J. Spiesberger, ''Full wave simulations of the forward scattering of sound in a structured ocean: A comparison with observations,'' submitted to J. Acoust. Soc. Am. (1997).

[7] 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).

[8] L. B. Dozier and F. D. Tappert, ''Statistics of normal mode amplitudes in a random ocean. II. Computations,'' J. Acoust. Soc. Am. **64**, 533–547 (1978).

[9] S. M. Flatté and G. Rovner, ''Path-integral expressions for fluctuations in acoustic transmission in the ocean waveguide,'' Proceedings of the 1997 Aha Hulikoa Winter Workshop (1997).

[10] J. A. Colosi and S. M. Flatté, ''Mode coupling by internal waves for multi-megameter acoustic propagation in the ocean,'' J. Acoust. Soc. Am. **100**, 3607–3620 (1996).

[11] M. G. Brown and J. Viechnicki, ''Stochastic ray theory for long-range sound propagation in deep ocean environments,'' J. Acoust. Soc. Am. (to be published).

[12] C. Garrett and W. Munk, ''Space-time scales of ocean internal waves,'' Geophys. Fluid Dyn. **2**, 225–264 (1972).

[13] W. Munk, ''Internal waves and small-scale processes,'' in *The Evolution of Physical Oceanography*, edited by C. Wunsch and B. Warren (MIT, Cambridge, 1981), pp. 264–291.

[14] M. H. Brill and L. B. Dozier, ''Exact transition from two- to one-dimensional internal wave power spectra,'' J. Acoust. Soc. Am. **77**, 1248–1249 (1985).

[15] J. T. Bell, ''Processing vertical internal-wave spectra,'' J. Phys. Oceanogr. **4**, 669–670 (1974).

[16] R. Pinkel, ''Doppler sonar observations of internal waves: The wave-number-frequency spectrum,'' J. Phys. Oceanogr. **14**, 1249–1270 (1984).

[17] T. F. Duda and C. S. Cox, ''Vertical wavenumber spectra of velocity and shear at small internal-wave scales,'' J. Geophys. Res. **94**, 939–950 (1989).

Erratum: "Approximate meridional leaky ray amplitudes for tilted cylinders: End-backscattering enhancements and comparisons with exact theory for infinite solid cylinders" [J. Acoust. Soc. Am. 102, 358–369 (1997)]

Philip L. Marston
Department of Physics, Washington State University, Pullman, Washington 99164-2814

(Received 7 January 1998; accepted for publication 17 January 1998)

[S0001-4966(98)04404-X]

PACS numbers: 43.40.Rj, 43.20.Fn, 43.30.Vh, 43.10.Vx [CBB]

There are two errors of transcription which should be noted. An overall factor of 2 was omitted from the right side of Eq. (12). In the second part of Eq. (15), the factor $(2/\pi)^{1/2}$ should be replaced by $(2/\pi^{1/2})$ so that erfc(w) is given by the standard definition. These errors are noteworthy since the integral under consideration has applications to scattering problems beyond the specific application given. All of the numerical results given used the corrected equations. Finally, the numerical comparison mentioned briefly in Sec. VI for the scattering by a tilted aluminum cylinder has been graphed in Ref. 1, below.

[1] P. L. Marston, ''Approximations for leaky wave amplitudes in acoustical imaging: applications to high frequency sonar,'' in *Proceedings of the 23rd International Symposium on Acoustical Imaging*, edited by S. Lees and L. A. Ferrari (Plenum, New York, 1997), pp. 369–374.

 0001-4966/98/103(4)/2236/1/$10.00