CODEN: JASMAN

ISSN: 0001-4966

The Journal of the Acoustical Society of America

Vol. 112, No. 2 August 2002

ACOUSTICAL NEWS—USA 347

USA Meetings Calendar 351

ACOUSTICAL NEWS—INTERNATIONAL 357

International Meetings Calendar 357

BOOK REVIEWS 361

REVIEWS OF ACOUSTICAL PATENTS 363

LETTERS TO THE EDITOR

Three-dimensional propagation effects near the mid-Atlantic Bight shelf break (L) — Kevin B. Smith, Chris W. Miller, Anthony F. D'Agostino, Brian Sperry, James H. Miller, Gopu R. Potty — 373

GENERAL LINEAR ACOUSTICS [20]

Transition from transient response to steady state for a layered medium — Shaw-Wen Liu, Chien-Ching Ma — 377

Linear time domain model of the acoustic potential field — Peter J. Lesniewski — 385

Free molecular sound propagation — F. Sharipov, W. Marques, Jr., G. M. Kremer — 395

Scattering from a ribbed finite cylindrical shell with internal axisymmetric oscillators — Michel Tran-Van-Nhieu — 402

Analysis of the time-reversal operator for scatterers of finite size — D. H. Chambers — 411

Planar near-field acoustical holography in a moving medium — Richard J. Ruhala, David C. Swanson — 420

Resonance analysis of a 2D alluvial valley subjected to seismic waves — Juin-Fu Chai, Tsung-Jen Teng, Chau-Shioung Yeh, Wen-Shinn Shyu — 430

NONLINEAR ACOUSTICS [25]

Acoustic streaming related to minor loss phenomenon in differentially heated elements of thermoacoustic devices — Mikhail Mironov, Vitalyi Gusev, Yves Auregan, Pierrick Lotton, Michel Bruneau, Pavel Piatakov — 441

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 112, NO. 2, AUGUST 2002

CONTENTS—*Continued from preceding page*

AEROACOUSTICS, ATMOSPHERIC SOUND [28]

A linearized Eulerian sound propagation model for studies of complex meteorological effects — Reinhard Blumrich, Dietrich Heimann — 446

Breaking wind waves as a source of ambient noise — Pavlo Tkalich, Eng Soon Chan — 456

UNDERWATER SOUND [30]

Wavefronts and waveforms in deep-water sound propagation — Chris T. Tindle — 464

A two-way parabolic equation that accounts for multiple scattering — Joseph F. Lingevitch, Michael D. Collins, Michael J. Mills, Richard B. Evans — 476

Surf-generated noise signatures: A comparison of plunging and spilling breakers — Steven L. Means, Richard M. Heitmeyer — 481

Multibeam volume acoustic backscatter imagery and reverberation measurements in the northeastern Gulf of Mexico — Timothy C. Gallaudet, Christian P. de Moustier — 489

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

Thermodynamic efficiency of thermoacoustic mixture separation — D. A. Geller, G. W. Swift — 504

TRANSDUCTION [38]

Study on the prestressed sandwich piezoelectric ceramic ultrasonic transducer of torsional-flexural composite vibrational mode — Lin Shuyu — 511

STRUCTURAL ACOUSTICS AND VIBRATION [40]

Applications of the direct Trefftz boundary element method to the free-vibration problem of a membrane — Jiang Ren Chang, Ru Feng Liu, Weichung Yeih, Shyh Rong Kuo — 518

Leaky helical flexural wave backscattering contributions from tilted cylindrical shells in water: Observations and modeling — Florian J. Blonigen, Philip L. Marston — 528

NOISE: ITS EFFECTS AND CONTROL [50]

The propagation of sound in narrow street canyons — K. K. Iu, K. M. Li — 537

A multiple regression model for urban traffic noise in Hong Kong — W. M. To, Rodney C. W. Ip, Gabriel C. K. Lam, Chris T. H. Yau — 551

Distribution of risk factors for hearing loss: Implications for evaluating risk of occupational noise-induced hearing loss — Mary M. Prince — 557

ARCHITECTURAL ACOUSTICS [55]

Rating, ranking, and understanding acoustical quality in university classrooms — Murray Hodgson — 568

PHYSIOLOGICAL ACOUSTICS [64]

A model cochlear partition involving longitudinal elasticity — Taha S. A. Jaffer, Hans Kunov, Willy Wong — 576

Inner hair cell loss and steady-state potentials from the inferior colliculus and auditory cortex of the chinchilla — Sally Arnold, Robert Burkard — 590

Acoustic intensity, impedance and reflection coefficient in the human ear canal — B. L. Farmer-Fedor, R. D. Rabbitt — 600

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 112, NO. 2, AUGUST 2002

CONTENTS—*Continued from preceding page*

PSYCHOLOGICAL ACOUSTICS [66]

Temporal pitch mechanisms in acoustic and electric hearing — Robert P. Carlyon, Astrid van Wieringen, Christopher J. Long, John M. Deeks, Jan Wouters — 621

Comodulation masking release in consonant recognition — Bom Jun Kwon — 634

The mid-level hump at 2 kHz — Lance Nizami, Jason F. Reimer, Walt Jesteadt — 642

Buildup and breakdown of echo suppression for stimuli presented over headphones—the effects of interaural time and level differences — Katrin Krumbholz, Andrea Nobbe — 654

The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal — Douglas S. Brungart, Brian D. Simpson — 664

SPEECH PRODUCTION [70]

Broadband sound generation by confined turbulent jets — Zhaoyan Zhang, Luc Mongeau, Steven H. Frankel — 677

Modulation frequency and modulation level owing to vocal microtremor — Jean Schoentgen — 690

Normalized amplitude quotient for parametrization of the glottal flow — Paavo Alku, Tom Bäckström, Erkki Vilkman — 701

SPEECH PERCEPTION [71]

Learning to perceive speech: How fricative perception changes, and how it stays the same — Susan Nittrouer — 711

Word recognition in competing babble and the effects of age, temporal processing, and absolute sensitivity — Karen B. Snell, Frances M. Mapes, Elizabeth D. Hickman, D. Robert Frisina — 720

BIOACOUSTICS [80]

Captive dolphins, *Tursiops truncatus*, develop signature whistles that match acoustic features of human-made model sounds — Jennifer L. Miksis, Peter L. Tyack, John R. Buck — 728

The relative detectability for mice of gaps having different ramp durations at their onset and offset boundaries — James R. Ison, Jason Castro, Paul Allen, Tracy M. Virag, Joseph P. Walton — 740

Auditory temporal resolution in birds: Discrimination of harmonic complexes — Robert J. Dooling, Marjorie R. Leek, Otto Gleich, Micheal L. Dent — 748

ERRATA

Erratum: "Geoacoustic inversion for fine-grained sediments" [J. Acoust. Soc. Am. 111, 1560–1564 (2002)] — Charles W. Holland — 760

Erratum: "Kramers-Kronig relations applied to finite bandwidth data from suspensions of encapsulated microbubbles" [J. Acoust. Soc. Am. 108, 2091–2106 (2000)] — Joel Mobley, Kendall R. Waters, Michael S. Hughes, Christopher S. Hall, Jon N. Marsh, Gary H. Brandenburger, James G. Miller — 760

CUMULATIVE AUTHOR INDEX — 762

ACOUSTICAL NEWS—USA

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

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

First Pan-American/Iberian Meeting on Acoustics

144th Meeting of the Acoustical Society of America
3rd Iberoamerican Congress on Acoustics
9th Mexican Congress on Acoustics

The First Pan-American/Iberian Meeting on Acoustics will be held Monday through Friday, 2–6 December 2002, at the Fiesta Americana Grand Coral Beach Hotel in Cancun, Mexico. Committee meetings and some other events will be held at the Hyatt Regency Hotel, which is across the street from the Fiesta Americana. A block of sleeping rooms has been reserved at each of these hotels.

Technical Program

Contributed papers are welcome in all branches of acoustics. The technical program will consist of lecture and poster sessions. Technical sessions will be scheduled Monday through Friday, 2–6 December 2002.

The special sessions described below will be organized by the Technical Program Committee.

Special Sessions

Acoustical Oceanography (AO)
David Weston memorial session (joint with Underwater Acoustics)
David Weston's seminal contributions in underwater explosions, shallow water propagation, ambient noise, sediment acoustics, propagation of sound in wedges, acoustic invariants, array performance, and attenuation of sound by marine fish
Using ambient sound in the ocean (joint with Animal Bioacoustics)
Using ambient sound for biological or environmental monitoring

Animal Bioacoustics (AB)
Amphibian acoustics/Session in honor of Robert Capranica
Sound production, reception, and communication by amphibians
Bioacoustics as a conservation tool
Animal acoustics with special emphasis on conservation issues. Population estimates, inferences on social structure of a population, geographical variation, methods for censusing, or any other relevant work on acoustic communication applicable to animal conservation
Coral reef bioacoustics
Sound production and reception by coral reef animals
Electrophysiological investigations of animals (joint with Psychological and Physiological Acoustics)

Architectural Acoustics (AA)
Acoustical design at hotels (joint with Noise)
Issues that need to be addressed in the design of hotels and resort properties relating to noise, sound isolation and room acoustics
Acoustic comfort in architecture (joint with Noise)
Criteria, concepts, case-histories, etc. associated with the acoustical characteristics of public spaces such as offices, hotels, and commercial spaces (restaurants, shopping centers, etc.)
Ancient acoustics (joint with Musical Acoustics and Physical Acoustics)
Archeological acoustics—ancient buildings and structures and ancient musical instruments Design of offices to enhance speech privacy and work performance (joint with Noise)
Indoor environmental studies continue to indicate that speech privacy is the #1 distraction in the office workplace
Duct acoustics and vibration (joint with Noise and Structural Acoustics and Vibration)
Standards, specifications, methods of measurement and test, and practical applications of passive noise control techniques to HVAC systems
Music buildings in Mexico (joint with Noise)
State of the art in performance hall design for music spaces
Recording in Mexico
State of the art recording in Mexico
Scattering topics in room acoustics
Importance of including scattering in both real room measurements and model room calculations
Speech intelligibility in buildings and metrics for the prediction and evaluation of intelligibility (joint with Speech Communication)
Which metrics should be used and under what circumstances to evaluate/predict intelligibility or are new metrics needed

Biomedical Ultrasound/Bioresponse to Vibration (BB)
Acoustic microscopy (joint with Signal Processing)
Methods of imaging at very high ultrasonic frequencies
Lithotripsy
Applications and mechanisms related to shockwave destruction of biological stones
Ultrasound and its applications in medicine
Broad overview of topics and techniques in biomedical ultrasound
Ultrasound-mediated drug delivery and gene transfection
Mechanisms and techniques for delivering large molecules through the membranes of living cells using ultrasound

Education in Acoustics (ED)
Development of acoustics programs in Latin America
Discussion of acoustics programs in universities
Hands-on demonstrations in acoustics for high-school students
Demonstrations to introduce high school students to the excitement of acoustics
''Take 5's''—Sharing ideas for teaching acoustics
Bring short presentations and demonstrations (no abstracts necessary)
Undergraduate research poster session
Undergraduate students present their research in poster format
Virtual labs, workshops, and multimedia in acoustics education
Use of computers in various ways for education in acoustics

Engineering Acoustics (EA)
Hearing aid technology (joint with Signal Processing in Acoustics)
Technology and design issues in modern hearing aids
Metrology standards and calibration on acoustics, ultrasound, and vibrations
Issues in standards and calibration methods for acoustic and vibration systems
Session in honor of Per Bruel
Session honoring the professional career of Per Bruel

Musical Acoustics (MU)
Analysis, synthesis, perception, and classification of musical sounds
Guitar acoustics
Acoustical studies of the guitar and related instruments
Hands-on explorations in acoustics
Hands-on activities in acoustics, open to all meeting attenders, but designated primarily for accompanying persons, both adults and children
Musical instruments of the South American dance tradition
Acoustics and use of musical instruments (such as the cuica, beremban, drums, and strings) used in South American dance music (samba, carocirn, tango, etc.)
Wind instrument measurement techniques
Apparatus and techniques for measuring frequency domain or time domain responses in wind instruments

 0001-4966/2002/112(2)/347/10/$19.00

Noise (NS)
Aircraft interior noise (joint with Structural Acoustics and Vibration)
Methods of controlling noise in an aircraft cabin
Environmental noise planning and legislation
Approaches to working at the local and national levels to create quiet communities
Hearing protection
Recent advances in hearing protection methods and assessment
Noise in urban communities (joint with Engineering Acoustics)
Measuring, predicting, and controlling noise in cities
Predicting noise in indoor industrial spaces (joint with Architectural Acoustics)
Phenomena, methodology, and software for predicting noise in cluttered industrial spaces
Soundscapes and community noise (joint with Architectural Acoustics)
New and novel methods to describe, rate, and assess community noise
Pan-American noise standards
Successful and new noise standards and regulations in the Americas

Physical Acoustics (PA)
Bubbles, drops, and foams
Acoustics of bubbles, drops, and foams
Global infrasound monitoring
Infrasound sensors, arrays of sensors, grouping of arrays, and wind noise suppression as well as analysis and interpretation of data
Laser ultrasonics
Optical techniques for the generation and detection of ultrasound
The coda and other stochastic seismic signals (joint with Signal Processing in Acoustics)
Physics and signal processing of microseismic random elastic waves near and on the surface of the Earth, with implications for imaging in the lithosphere

Signal Processing in Acoustics (SP)
Acoustic automatic target recognition (joint with Underwater Acoustics)
Automatic detection and recognition of targets sensed by airborne, underwater acoustic, and ultrasonic systems
Feature extraction and models for speech (joint with Speech Communication)
Recent models and techniques for the analysis, processing, and recognition of speech
Internet protocol voice quality
Seismic signal processing: Detection, estimation, and inversion methods (joint with Physical Acoustics)
Processing of seismic signals including detection, estimation, and subsurface imaging for characterizion and identification of geophysical processes and features

Speech Communication (SC)
Perception and production of hearing impaired
Speech perception and production in hearing-impaired speakers, with examples from both English- and Spanish-speaking populations
Spanish and English in contact
Issues in speech communication raised by the conflict between Spanish and English speakers
Voice processing and phonetic algorithms
Algorithms to process and display voice and speech parameters for use in speech production or perception training for persons with speech or hearing disorders

Structural Acoustics and Vibration (SA)
Analysis, measurements, and control of structural intensity
Energy flow methods in vibroacoustic analysis and control
Aimed at discussing recent advances in analytical and experimental energy flow analysis and control methods for vibroacoustic applications
Reconstruction of acoustic radiation from vibrating structures
Forum to exchange state-of-the-art technologies and development in acoustical holography as an effective noise and vibration diagnostic tool
Vibration of floors of buildings
Criteria, prediction, and measurement; effects on structures, people, and equipment

Underwater Acoustics (UW)
Concurrent mapping and navigation
Use of acoustic sensors to address the problem of concurrent mapping and navigation for autonomous land and undersea vehicles
GeoClutter and boundary characterizations
Recent advances in the analysis of shallow water reverberation including the Geological Clutter and Boundary Characterization experiments of Spring 2001
Littoral environmental variability and its acoustic effects (joint with Acoustical Oceanography)
Environmental variability in shallow water and its effects on acoustic propagation, including results from the recently completed ASIAEX (Asian Seas International Experiment) field effort in the South and East China Seas
Underwater acoustic measurement laboratories
New global perspectives of automatic control and management (joint with Engineering Acoustics)
Explore the potential of the World Wide Web to enable sharing of facilities and laboratory capabilities among colleagues around the world, and how to solve problems which might arise

Exhibit

The instrument and equipment exposition, which will be conveniently located near the registration area and meeting rooms, will open at the Fiesta American Grand Coral Beach Hotel on Monday, 2 December, and will close on Wednesday afternoon, 4 December. The exposition will include computer-based instrumentation, scientific books, sound level meters, sound intensity systems, signal processing systems, devices for noise control and acoustical materials, active noise control systems, and other exhibits on acoustics. Contact the Exhibit Manager, Robert Finnegan, Advertising and Exhibits Division, American Institute of Physics, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, Tel: 516-576-2433; Fax: 516-576-2481; E-mail: rfinneg@aip.org.

History Lectures and Hot Topics Session

A ''Hot Topics'' session sponsored by the ASA Tutorials Committee is scheduled. The ASA Committee on Archives and History will jointly sponsor lectures on the history of acoustics with the ASA Technical Committees on Animal Bioacoustics and Biomedical Ultrasound/Bioresponse to Vibration.

Special Tutorials

Two special tutorial sessions have been scheduled and are described below. They are open to all meeting participants and no separate registration fee is charged to attend. Notes will not be provided. Tutorials will be presented in English; however, questions will be taken in English, Spanish, and Portuguese.

Register in advance using the registration form in the call for papers or the online registration form at ⟨http://asa.aip.org⟩. You may also sign up to attend at the meeting subject to space availability.

Monday, 2 December, 7:00 p.m.–9:00 p.m., ''**Architectural Acoustics**'' by Neil Shaw (Menlo Scientific Acoustics, Inc., Topanga, CA), Rick Talaske (The Talaske Group, Inc., Oak Park, IL), and Sylvio Bistafa (Dept. of Mechanical Engineering, Polytechnic School, University of Sao Paulo, Brazil). This tutorial is intended to provide an overview of current knowledge and practice in architectural acoustics. Topics covered will include basic concepts and history, acoustics of small rooms (small rooms for speech such as classrooms and meeting rooms, music studios, small critical listening spaces such as home theatres) and the acoustics of large rooms (larger assembly halls, auditoria, and performance halls).

Friday, 6 December, 7:00 p.m.–9:00 p.m., ''**Industrial Noise Control**'' by Samir Gerges (Federal University of Santa Catarina, Florianopolis, Brazil) and Brandon Tinianov (Johns Manville, Littleton, CO). This tutorial is intended to provide a practical overview of applied acoustics and industrial noise control, including the fundamentals of applied acoustics, industrial noise control, and case histories. The following topics will be covered:

sound radiation, sound transmission, outdoor sound propagation and indoor sound fields, sound absorption materials, machinery noise and its control (fans, motors, compressors, air jets, pneumatic tools), and the design of sound enclosures.

Special Short Courses

On Sunday, 1 December, two special short courses will be organized and presented by Bruel & Kjaer USA. One course will be held in the morning and the other in the afternoon. A separate registration fee of $10 is charged to attend. Each course is limited to 40 registrants. Lunch, compliments of B&K, will be provided to course participants. Register in advance using the registration form in the call for papers or the online registration form at ⟨http://asa.aip.org⟩. You may also sign up to attend at the meeting subject to space availability. Topics include:

Sound Quality Fundamentals: Basic introduction to sound quality; recording and playback; introduction to sound quality parameters; overview of SQ metrics.

Noise Source Identification Techniques: Introduction to sound intensity; sound intensity metrics; sound intensity equipment; advanced noise source identification techniques.

Paper Collection on CD-ROM

No proceedings will be published for the joint meeting and no paper copying service will be available. However, a CD-ROM containing papers voluntarily submitted by authors will be produced by the Mexican Institute of Acoustics. All authors are cordially invited to submit their material for inclusion on the CD. Guidelines for the preparation and submission of papers will be distributed with acceptance notices. The CD ROM can be purchased at the meeting.

Student Challenge Problem

The Technical Committee on Signal Processing in Acoustics of the Acoustical Society of America invites all students to participate in the Student Challenge Problem competition. The competition is open to both undergraduate and graduate students. The prize is $500 USD. This problem presents a real-world signal processing challenge encountered in the field of acoustics. It is intended to be easily understood by students with a basic understanding of signal processing. A unique solution to the Student Challenge Problem is not expected. Students are encouraged to explore various approaches. Submissions will be judged on clarity, creativity, rigor, and insight.

The Student Challenge Problem is posted online at ⟨http://www.sao.nrc.ca/ims/asa_sp/Challenge_f01_v4.htm⟩.

Submissions are being accepted by e-mail and should be made in MS Word, WordPerfect, PDF, or ASCII with TeX formula conventions. The deadline date for receipt of submissions is 30 September 2002. Submissions are being accepted by e-mail only and should be sent to Dr. Charles Gaumond at ⟨gaumond@nrl.navy.mil⟩ with a copy to Dr. Jose Diaz at ⟨jadiaz@thor.uc.edu.ve⟩.

Your submission should include your name, university attended, professor's name and e-mail address, and your mailing address and your e-mail address both now and after December 2002. Winners will be selected by a panel appointed by the Chair of the Technical Committee on Signal Processing in Acoustics. The panel will meet during the Cancun meeting to choose the winner and the winner will be notified by e-mail. The winning answer will be posted on the Technical Committee on Signal Processing in Acoustics website.

Meeting Program

An advance meeting program summary will be published in the September issue of *The Journal of the Acoustical Society of America* (JASA) and a complete meeting program will be mailed as Part 2 of the November issue. Abstracts will be available on the ASA Home Page (http://asa.aip.org) in September.

Student Transportation Subsidy Program

ASA needs the assistance of students who will be attending the meeting to serve as room monitors and a/v assistants. Students who serve in these positions will be provided travel grants to support their travel to the meeting. Each grant will be $150 per half day served and no student will be asked to serve more than 3 half days.

Duties will include attendance at the full session from 15 minutes prior to the start and 15 minutes following the end of the session to deliver the pointer and timer to the session chair before the meeting and to collect them at the end of the session, raise and lower the lights, assist speakers with loading their slides in the slide projector, assist speakers with connecting their laptops to the projector switch box, contacting the a/v staff for assistance if equipment fails, and other tasks that might be required by the session chair. Note that students serving must be able to speak fluently in English.

Every effort will be made to schedule students to serve in sessions corresponding to their technical areas of interest but we cannot guarantee scheduling in specific sessions. Do not offer to serve on the same half-day that you are presenting a paper.

Acceptance notices will be sent to authors in August. If you are presenting a paper, please wait until you have received your acceptance notice before submitting your application for a subsidy. If you are interested in serving as a room monitor/av assistant, please submit the following information by mail, fax or e-mail (e-mail preferred) not later than 1 October to Jolene Ehl [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, Fax: 516-576-2377, E-mail: asa@aip.org]:

(1) Your name, the university/college you are attending, degree program (i.e., MS, Ph.D.), mailing address, and telephone number
(2) Your current e-mail address and your e-mail address after 1 October (if different)
(3) Planned arrival date and time (estimated arrival time at hotel)
(4) Departure date and time
(5) Half-days that you will be available during the days you will be in attendance. Sessions are usually scheduled from 8:00 a.m. to 12:00 noon and 1:00 p.m. to 5:00 p.m. You must be available 15 minutes before and 15 minutes after the sessions in which you will serve.
(6) Technical areas of interest corresponding to the 13 Technical Committees of the Acoustical Society. Please indicate 1 for first choice and 2 for second choice: Acoustical Oceanography, Animal Bioacoustics, Architectural Acoustics, Biomedical Ultrasound/Bioresponse to Vibration, Engineering Acoustics, Musical Acoustics, Noise, Physical Acoustics, Psychological and Physiological Acoustics, Signal Processing in Acoustics, Speech Communication, Structural Acoustics and Vibration, Underwater Acoustics
(7) Whether or not you are presenting a paper and the day and time of your session.

We will contact you by 15 October to let you know whether or not you have been selected to participate in the travel grant program.

Opening Ceremonies

Opening ceremonies will be held on Monday, 2 December, from 9:30 a.m. to 11:30 a.m. This event will be followed by a reception in the exhibit area where coffee and other refreshments will be served.

Young Investigator Travel Grant

The Committee on Women in Acoustics is sponsoring a Young Investigator Travel Grant to help with travel costs associated with presenting a paper at the Cancun meeting. This award is designed for young professionals who have completed the doctorate in the past five years (not currently enrolled as a student) and who plan to present a paper at the Cancun meeting. Each award will be of the order of $200 USD. It is anticipated that the Committee will grant a maximum of five awards. Applicants should submit a request for support, a copy of the abstract they have submitted for the meeting, and a current resume/vita to (e-mail preferred): Ellen Livingston, Office of Naval Research, Code 3210A, 800 No. Quincy St., Arlington, VA 22217; E-amil: livinge@onr.navy.mil. Deadline for receipt of applications is 25 October 2002.

Plenary Session, Awards Ceremony and Social Events

Complimentary buffet socials with cash bar will be held early on Tuesday and Thursday evenings (3 and 5 December). The Plenary session will be held on Wednesday afternoon at the Fiesta Americana Grand Coral Beach Hotel where awards will be presented and recognition of ASA Fellows will be announced.

Women in Acoustics Luncheon

The Women in Acoustics luncheon will be held on Tuesday, 3 December. Those who wish to attend this luncheon must register using the form in the Call for Papers. The fee is $15 (Students $5) for preregistration by 4 November and $20 (Students $5) thereafter, including on-site registration at the meeting.

Travel Tips

Entry Requirements: U.S. and Canadian citizens require a valid U.S. passport (preferred), or an original birth certificate with a raised seal and a state/federal issued photo id (driver's license). U.S. passports should be applied for well in advance of the meeting to avoid high processing fees. Children traveling alone, with one parent, or with someone else must have written and notarized consent to travel from both parents.

Citizens of countries other than the U.S. and Canada should contact their local Mexican Embassy or Consulate to determine entry requirements for Mexico.

You will receive two forms to fill out before you can enter Mexico. These are normally distributed by the airlines either at the ticket counter or while in transit. The **first form** is your tourist card and you'll need one per person. Be sure to fill out both the top and bottom parts of the form. Upon arrival at the airport you will first go through Immigration, which will inspect your proof of citizenship and validate your tourist card. Save the stamped (blue) copy which you will need to exit the country (make a note of the number in case you lose it). Your tourist card will be endorsed for a 30-day stay. The **second form** is for Hacienda (the Customs and Taxation arm of the government). You will need one of these forms per family.

After passing through Immigration, claim your bags in the baggage claim area and then proceed to the exit. Hand your tourist card to the Customs Agent and push the button on the ''traffic light'' which is set up to randomly turn green for ''Go'' or red for ''Stop and Inspect.'' If you get a red light, they will ask to see the contents of your bags.

During busy periods, particularly on weekends, there can be lengthy delays at Immigration upon arrival in Cancun.

Language: English and Spanish are spoken at all hotels and restaurants on the Cancun peninsula.

Electricity: (Same as U.S.) 110 volts, 60 Hz

Health Tips: The water standards on the Cancun peninsula, where the hotel is located, are as high as those found in the United States. However, for visitors who leave the Cancun peninsula, it is advisable to drink only distilled or bottled water. When buying fresh fruits and vegetables, be sure to wash them with distilled or bottled water before eating; peel them if possible.

Money, Exchange and Credit: While most businesses will accept foreign currency, especially US dollars, foreign currency can be exchanged at one of many casas de cambio (exchange houses) or at banks and hotels. Mexican banking hours are normally Monday–Friday from 9:00 a.m. to 5:00 p.m. However, for money exchange, the hours are normally from 9:00 a.m. to 1:00 p.m. Casas de cambio are open later than most banks and they generally offer quicker service. Credit cards are widely accepted, with Visa, MasterCard, and American Express being the most popular. Automatic Teller Machines (ATM's) are plentiful, including at hotels. Note that the dollar sign ($) is often used to denote pesos.

U.S. State Department Website: www.travel.state.gov

Air Transportation

The Cancun Airport is served by Aeromexico, Air France, American, British Airways, Continental, Delta, Iberia, KLM, Lacsa, Lufthansa, Mexicana, TWA, and US Airways. The airport designation is CUN.

Ground Transportation/Airport–Hotel Transfer

A rental car is not necessary; there are plenty of restaurants and shops within easy walking distance of the hotels. Attendees who do wish to rent cars will find car rental counters at the airport and hotels.

Advance reservations for ground transportation between the airport and the hotel are highly recommended. It is about a 25 minute trip from the airport to the hotel.

Reservations for transportation may be made on-line at the Thomas More Travel web site listed below. You will need to provide your arrival and departure flight information. Remember that the meeting hotels are the Fiesta Americana Grand Coral Beach and the Hyatt Regency Cancun (there are two Fiesta Americana and two Hyatt hotels in Cancun!). Be sure to enter ''ACOUSTICAL SOCIETY'' in the comments section.

http://www.cancun.com/tours/thomasmoretravel/transport.htm

After clearing Customs at the Cancun airport and after exiting the International arrivals building, look for Thomas More travel and the ASA banner; you will be directed to a waiting vehicle. If you have not made advance reservations, you have two choices. After clearing Customs and before exiting the building, look for the taxi shuttle called Chac-Mol for a shared ride from the airport to the hotel for approximately $10.00 USD per person. Alternatively, a taxi to the hotel will cost approximately $20 USD. Taxis are not metered so it is advisable to first negotiate the fare. Remember that the meeting hotels are the Fiesta Americana Grand Coral Beach and the Hyatt Regency Cancun.

Room Sharing

ASA will compile a list of those who wish to share a hotel room and its cost. To be listed, send your name, telephone number, e-mail address, gender, and smoker or nonsmoker by 21 October to the Acoustical Society of America, preferably by e-mail: asa@aip.org or by postal mail to Attn.: Room Sharing, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502. The responsibility for completing any arrangements for room sharing rests solely with the participating individuals.

Weather

Cancun's average winter temperature is 80 °F.

Hotel Accommodations

Blocks of sleeping rooms have been reserved at the Fiesta Americana Grand Coral Beach and the Hyatt Regency Cancun, which is across the street from the Fiesta Americana. All technical sessions will be held at the Fiesta Americana and most administrative meetings will be held at the Hyatt Regency. Please make your reservations directly with the desired hotel. The reservation cut-off date for the special discounted ASA rates is 30 October at the Fiesta Americana and 15 October at the Hyatt Regency; after this date the special ASA rate will not be available. See p. 13 in the call for papers for information on making hotel reservations.

Hotel Reservation Information

Guest rooms at discounted rates have been reserved for meeting participants at Fiesta Americana Grand Coral Beach and the Hyatt Regency Cancun. All technical sessions will be held at the Fiesta Americana. Early reservations are strongly recommended. Note that the special ASA meeting rates are not guaranteed after 1 November at the Fiesta Americana and 15 October at the Hyatt Regency. You must reference the Acoustical Society of America when making your reservation to obtain special ASA meeting rates.

Reservations will only be accepted by fax using the forms that appear on pp. 14 and 15 in the call for papers. Reservations cannot be made by phone. The special ASA discounted rate cannot be obtained by calling the 800 number for Hyatt hotels nor through online reservation services.

Fiesta Americana Grand Coral Beach

⟨http://fiestamexico.com/hotels/fiesta_americana/frameset_fa_hotels2.asp?hotel=5&lang=EN⟩

Reservation cut-off date: 30 October 2002

Please make your reservations directly with the hotel (reference ASA):
Fiesta Americana Grand Coral Beach
Blvd. Kukulkan Km. 9.5 Lote 6
Zona Hotelera
C.P. 77500
Cancún, Quintana Roo, Mexico
Tel: 011-52-9988-81-32-00
Fax: 011-52-9988-81-32-73 or 63
$130.00 Single/Double
plus 12% tax
Recommended tips: Bellman: $5.00 USD per person round-trip
Maid: $3.00 USD per night/per room

Hyatt Regency Cancun Mexico

⟨cancun.hyatt.com/cancu/index.html⟩

Reservation cut-off date: 15 October 2002

Hyatt Regency Cancun Mexico
Blvd Kukulcan Km 8.5
P.O. Box 1201
Hotel Zone
Cancun, Q Roo 77500 Mexico
Tel: 011-52-9988-83 1234/883 0966
Fax: 011-52 9988-83 1349/883 1438
E-mail: hyattreg@hyattregencycancun.com.mx
$120.00 Single/Double or $150 Regency Club Single/Double
$30.00 for third person in room
Mandatory tips: Bellman: $4.50 USD per person round-trip
Maid: $2.50 USD per night/per room
plus 12% tax

Assistive Listening Devices

Anyone planning to attend the meeting who will require the use of an assistive listening device is requested to advise the Society in advance of the meeting: Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, asa@aip.org.

Accompanying Persons Program

Accompanying persons are welcome at the meeting. A hospitality room, specifically designated for accompanying persons, will be open each morning. Information will be available about activities in Cancun and the surrounding area.

Spouses and accompanying persons may register in advance using the registration form in the call for papers. The accompanying persons registration fee includes the opening ceremony, buffet socials on Tuesday and Thursday evenings, and refreshments in the accompanying persons hospitality suite during the meeting.

The Accompanying Persons program will include the following activities.

A presentation describing the area and the culture to acquaint accompanying persons with the Mayan culture and the tours and activities in Cancun and the surrounding area.

A trip to Playa del Carmen (Carmen Beach). This charming little town was once a sleepy fishing village. Take a walk down famous Fifth Avenue where you will find a wide variety of shops to browse and buy regional handicrafts, silver, embroidered clothing, leather goods, and pottery.

A trip to Isla Contoy (Contoy Island), a natural reserve with sea life and a large variety of birds. This tiny island was declared a National wildlife reserve in 1961 and is renowned for its flora and fauna, particularly for its birds. Bring your cameras.

A trip to Tulum and Xel-ha. Tulum, to the south of Cancun, is the only major Mexican ruins to overlook the Caribbean Sea and is the only Mayan port city ever discovered. Tulum is a gleaming-white walled city that invites visitors to roam its small temples and buildings. Xel-ha is the world's largest natural aquarium where you can see colorful fish and other fresh water animals in their own habitat.

Full details about the above tours, including registration, will be available at the meeting.

Post-Meeting Excursion

There will be a post meeting tour to Chichen Itza on Saturday and Sunday, 7 and 8 December. This huge complex of magnificent Mayan ruins is one of the biggest in Mexico. This site dates back to 600 AD and the Mayan civilization. It features the majestic pyramid of Kukulkan. Special acoustical effects will be presented by ASA members at the pyramid of Kukulkan during the tour. A visit to Valladolid and Balancanche Caves is also planned as part of the tour. Full details are available on the ASA website at ⟨http://asa.aip.org⟩.

Preregistration and Registration

The registration desk at the meeting will open on Monday, 2 December, at the Fiesta Americana Grand Coral Beach Hotel. To register use the form on p. 18 in the call for papers or register online at ⟨http://asa.aip.org⟩. **If your registration is not received at the ASA headquarters by 11 November, you must register on-site.**

Registration fees are as follows:

Category	Preregistration by 4 November	Registration after 4 November
ASA/FIA/IMA Members	$270	$320
ASA/FIA/IMA Members One-Day	$135	$160
Nonmembers	$320	$370
Nonmembers One-Day	$160	$185
Nonmember Invited Speakers	$270	$320
(Note: The fee is waived for these speakers if they attend the meeting on the day of their presentation only.)		
Students (with current ID cards)	Fee waived	Fee waived
Emeritus members of ASA	$35	$45
(Emeritus status preapproved by ASA)		
Accompanying Persons	$35	$45
(Spouses and other registrants who will not participate in the technical sessions)		

Nonmembers who simultaneously apply for Associate Membership in the Acoustical Society of America will be given a $50 discount off their dues payment for the first year (2003) of membership. (Full price for dues: $100). Invited speakers who are members of the Acoustical Society of America are expected to pay the registration fee, but **nonmember invited speakers** who participate in the meeting for one day only may register without charge. Nonmember invited speakers who wish to participate in the meeting for more than one day will be charged the member registration fee, which will include a one-year membership in the ASA upon completion of an application form.

NOTE: A $25 PROCESSING FEE WILL BE CHARGED TO THOSE WHO CANCEL THEIR REGISTRATION AFTER 4 NOVEMBER.

Online Registration

Online registration is available at ⟨http://asa.aip.org⟩

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.

2002

12–14 Aug.	Recreational Noise Symposium, Key Largo, FL [WWW: http://www.internoise2002.org].
19–21 Aug.	INTER-NOISE 2002, Dearborn, MI [INTER-NOISE

02 Secretariat, The Ohio State University, Department of Mechanical Engineering, 206 West 18th Ave., Columbus, OH 43210-1107, USA or e-mail: hp@internoise2002.org].

21–25 Aug. International Hearing Aid Research Conference (IHCON), Lake Tahoe, CA [House Ear Inst., 2100 W. 3rd St., Los Angeles, CA 90057; Fax: 213-413-0950; WWW: http://www.hei.org].

16–20 Sept. 7th International Conference on Spoken Lanugage Processing (ICSLP 2002), Interspeech 2002, Denver, CO [http://www.icslp2002.org/home.html].

2–6 Dec. First Pan-American/Iberian Meeting on Acoustics (Joint Meeting: 144th Meeting of the Acoustical Society of America, 3rd Iberoamerican Congress on Acoustics, and 9th Mexican Congress on Acoustics), Cancun, Mexico [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org/cancun/cancun.html].

2003

28 April–2 May 145th Meeting of the Acoustical Society of America, Nashville, TN [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

5–8 May SAE Noise & Vibration Conference & Exhibition, Traverse City, MI [P. Kreh, SAE International, 755 W. Big Beaver Rd., Suite 1600, Troy, MI 48084; Fax: 724-776-1830; WWW: http://www.sae.org].

10–14 Nov. 146th Meeting of the Acoustical Society of America, Austin, TX [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

2004

24–28 May 75th Anniversary Meeting (147th Meeting) of the Acoustical Society of America, New York, NY [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

22–26 Nov. 148th Meeting of the Acoustical Society of America, San Diego, CA [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

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

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

Some indexes are out of print as noted below.

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

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

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

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

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

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

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

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

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

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

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

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

Revision List

New Associates

Abon, Israel A., Bestway Studio, G.P.O. Box 17525 Dugbe Ibadan, Ibadan, Oyo State 30667, Nigeria

Adams, Stephanie A., David L. Adams Associates, Inc., 1701 Boulder St., Denver, CO 80211

Agnew, Nancy E., 229 Moria Place, Aston, PA 19014

Alexander, Mona-Lisa, Signet Scientific, 3401 Aerojet Ave., El Monte, CA 91731-2882

Alfieri, Richard P., Respironics, 1001 Murry Ridge Ln., Murrysville, PA 15668

Avsic, Tom, Jungfernstieg 36, Kiel 24116, Germany

Bates, Bradley O., Undersea Warfare, Johns Hopkins Univ., Applied Physics Lab., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099

Bright, Andrew P., Electroacoustics, Nokia Research Center, Itomerenkatu 11-13, Helsinki 00180, Finland

Browne, Sean D., 2409 Jetty Ln., Columbia, PA 17512

Brownlee, Lyall R., 215 Stewart St., Peterborough, ON K9J 3M6, Canada

Bruce, Ian C., Dept. of Electrical and Computer Eng., McMaster Univ., 1280 Main St., West, Hamilton, ON L8S 4K1, Canada

Buen, Anders, Brekke & Strans Akustikk, Hovfaret 17, Oslo N-0275, Norway

Chen, Henry, 770 Josina Ave., Palo Alto, CA 94306

Chen, Yuming A., 1702 Oak Creek Dr., #304, Palo Alto, CA 94304

Chennoukh, Samir, CONEXANT Systems, Cellular Systems Design Center, 2 Av. Pierre Piffault, Le Mans 72100, France

Christensen, Knud B., Skovvej 2, Ryomgard DK-8550, Denmark

Crandall, Frank A., High Power Ultrasonics, Zevex, Inc., 4314 Zevex Park Ln., Salt Lake City, UT 84123

Dala-Krishna, Praveen, 260 Thunder Cir., Bensalem, PA 19020

Delgutte, Bertran, Eaton-Peabody Lab., Massachusetts Eye and Ear Infirmary, 243 Charles St., Boston, MA 02114

Depireux, Didier A., Anatomy and Neurobiology, Univ. of Maryland at Baltimore, 685 West Baltimore St., HSF222, Baltimore, MD 21201

DeRuiter, Mark, Audiology and Speech Lang. Pathology, Wayne State Univ., 581 Manoogian Hall, Detroit, MI 48202

Diaz-Sanchidrian, Cesar, Cerezo Del Paular 7, Pozuelo de Alarcon, Madrid 28224, Spain

Donaldson, Michael E., Box 61264, Fairbanks, AK 99706

Dozsa, Leslie, 2101 SW Sunset Blvd., B-204, Renton, WA 98055

Drinkwater, Bruce W., Mechanical Engineering, Univ. of Bristol, University Walk, Bristol, Avon B58 1TR, United Kingdom

Dunkin, Bradley M., ITLL, Univ. of Colorado, 522 UCB, Boulder, CO 80309

Dyba, Roman A., 15119 Thatcher Dr., Austin, TX 78717

Elchik, Michael E., Akustica, Inc., 2403 Sidney St., Pittsburgh, PA 15203

Elegbe, Mercy O., 9396 Richmond, #522, Houston, TX 77063
Ezcurra, Amaya, Fisica, Univ. Publica de Navarra, Campus Arrosadia, Pamplona, Navarra 31006, Spain
Fick, Douglas D., BSA Design, 9365 Counselors Row, Suite 300, Indianapolis, IN 46240
Flanders, John M., Alexander Graham Bell Associates, 3714 Volta Place, NW, Washington, DC 20007
Freeberg-Renwick, Timothy J., 607 West Broadway, Fairfield, IA 52556
Gerashehenko, Natalia V., 126-2H Montgomery St., Highland Park, NJ 08904
Ghiringhell, Gian Luca, Dip. Ingegneria Aerospaziale, Politechnico di Milano, via La Masa 34, Milano MI 20156, Italy
Gillespie, Douglas M., International Fund for Animal Welfare, The Old Chapel, Fairview Dr., Bristol, Avon BS6 6PW, United Kingdom
Green, Kent C., Compliance-Acoustics, IBM, 5600 Cottle Rd., San Jose, CA 95193
Guarino, Joe C., Mechanical Engineering, Boise State Univ., 1910 University Dr., Boise, ID 83725
Hah, Zaegyoo, Applied Coating, Inc., 465 Paul Rd., Rochester, NY 14624
Han, Jeong-Im, Dept. of English, Konkuk Univ., Hwayang-dong 1, Gwangain-gu, Seoul 143-701, Korea
Handlin, Syd H., Consultron, P.O. Box 984, Moorhead, MN 56561
Handzel, Amir A., Inst. for Systems Research, Univ. of Maryland, A. V. Williams Bldg., College Park, MD 20742
Hanoglu, Kemal B., Orient Research Consulting Engineers, Bebek Yolu Sokak 15/6, Istanbul, Etiler 80630, Turkey
Haverstick, Gavin A., Auralex Acoustics, Inc., 8851 Hague Rd., Indianapolis, IN 46256
Herrmann, Alix K., Chemin du Couchant, Lausanne VD CH-1007, Switzerland
Hideki, Iwasawa, 16-410 Terada-Kaitogo, Joyo, Kyoto 610-0121, Japan
Hikichi, Takafumi, Media Information Lab., NTT Communication Science Labs., 3-1 Morinosato, Wakamiya, Atsugi-shi, Kanagawa 243-0198, Japan
Huang, Changzheng, 722 Stanford Ct., Irvine, CA 92612
Kim, Doh-Suk, Lucent Technologies, Rm. 14C-373, 67 Whippany Rd., Whippany, NJ 07981
Kim, Ki-Bok, Nondestructive Measurement Group, Yuseong Doroyong dong 1, Daejeon, Chungnam 305-600, Korea
Kim, Wontak, Automotive Systems Div., Bose Corporation, The Mountain, Framingham, MA 01701-9168
Kling, Terry A., 200 Buckhorn Rd., Port Matilda, PA 16870
Kunio, Jason T., 610 Spy Glass Ct., North Aurora, IL 60542
La Cour, Brian R., Applied Research Laboratories, Univ. of Texas at Austin, 10000 Burnet Rd., Austin, TX 78758
Lacerda, Francisco P., Tallasvagen 338, Vallentuna SE-18651, Sweden
Lixue, Wu, National Reserach Council, Inst. for National Measurement Standards, M-36, 1200 Montreal Rd., Ottawa, ON K1A 0R6, Canada
Lloyd, Watts, Applied Neurosystems Corp., 4906 El Camino Real, Los Altos, CA 94022
Lu, Huancai, Apt. 111, 1616 Ouellette Ave., Windsor, ON N8X 1L2, Canada
Machleder, Anton M., 222 Bastian Rd., Rochester, NY 14623
Marn, Borut, Milana Resetara 4, Zagreb 10090, Croatia
Martini, Joerg, Salzstrasse 13, Neuss NRW 41460, Germany
Matsuhiro, Norio, Furuno Electric Co., Ltd., Research Office, 9-52 Ashihara-cho, Nishinomiya-City, Hyougo 662-8580, Japan
McCracken, Larry, LeapFrog, 130D Knowles Dr., Los Gatos, CA 95032
Meireles, Antonio Maria G. M., MediaChip, Project Development, R. St. Barbara, 27-45, V. N. Gaia Porto 4400-289, Portugal
Merlen, Alain M. E., Mathematique Mecanique, USTL, Cite Scientifique Bat. M3, Villeneuve d'Ascq F59655, France
Milburn, Darrell E., 179 East Lothrop St., Beverly, MA 01915
Miraglia, Vincent A., Analog Design Group, 547 Ridgewood Rd., Washington, NJ 07676-4930
Moore, Sue E., NOAA/Fisheries, National Marine Mammal Lab., 7600 Sand Point Way, NE, Seattle, WA 98115
Moore, Thomas R., Dept. of Physics, Rollings College, 1000 Holt Ave., Winter Park, FL 32789
Nagarajan, Srikantan S., Bioengineering, Univ. of Utah, 205 2030 East BPRB506D, Salt Lake City, UT 84112-9458
Navarri, Massimo, via Gramsci 12, San Donato MiLn.se MI 20097, Italy
Neumann, John J., Carnegie Mellon Univ., ECE Dept., 1209 Hamerschlag Hall, 5000 Forbes Ave., Pittsburgh, PA 15213
Nils-Ake, Nilsson, Acoustic Control AB, Tumstocksvaegen 1, Taeby SE-18766, Sweden
O'Neil, Robert K., 19129 Bennington Dr., Strongsville, OH 44136
Olofsson, Niten B., FOI, Swedish Defense Research Agency, Systemteknik, Inst. for Marina Sensor System, Enkopingsvagen 126, Stockholm 17290, Sweden
Ono, Nobutaka, Dept. of Information Physics and Computing, Graduate School of Information Sci. and Tech., Univ. of Tokyo, 7-3-1 Hongo, Bunkyoku, Tokyo 113-0033, Japan
Orfield, Steve J., Orfield Laboratories, Inc., 2709 East 25th St., Minneapolis, MN 55406
Pahuja, Tarun, A29 New Friends Colony, New Delhi 110065, India
Parra, Claudio E., Apdo 47692, Caracas DF 1041-A, Venezuela
Poesio, Pietro, Isaak Hornbeekstraat 26, Delft 2613 HH, The Netherlands
Pritchard, William A., Wilson, Ihrig & Associates, 5776 Broadway, Oakland, CA 94618
Purnell, Thomas C., Dept. of Linguistics, Univ. of Wisconsin, Madison, 1168 Van Hise Hall, 1220 Linden Dr., Madison, WI 53706
Ravn, Gert, Technical-Audiological Lab., DELTA, Sdr. Boulevard 29, Bldg. 3, 3rd Floor, Odensec DK-5000, Denmark
Rhee, Huinam, School of Mechanical and Automotive Eng., Sunchon National Univ., 315 Maegok-Dong, Sunchon, Junnam 540-742, Korea
Ritchie, James A., Speech Group, Ume Voice, Inc., 73 Digital Dr., Novato, CA 94949
Rodriguez, Rose T., Applied Physics Lab., Johns Hopkins Univ., National Security Technology Dept., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099
Ross, Jacob W., Altermatt Associates, Inc., 522 SW 5th Ave., Suite 1200, Portland, OR 97204
Ruhlmann, Lilimar A., 2758 Congress Ct., Waldorf, MD 20603
Ruzzene, Massimo, Mechanical Engineering, The Catholic Univ. of America, 620 Michigan Ave., Washington, DC 20064
Sanchez-Guzman, Alejandro J., Sanchez Guzman Ingenieria Electronica S. A. de C. V., Ave. Lazaro Cardenas #4000, Local 20-C Col. Las Brisas, Monterrey, Neuvo Leon 64780, Mexico
Santos, Ricardo, Periodical Consulting, Av. Nicolau Mercadante #48, Jacarei, S. Paulo 12327-680, Brazil
Saunders, Melvin L., Pelton Marsh Kinsella, 1420 West Mockingbird Ln., Suite 400, Dallas, TX 75247
Schmidt, Peter L., 813 Callahan Place, Franklin, TN 37069
Scott, Jerome E., 317 Locust Thorn Ct., Millersville, MD 21108
Sellschopp, Jurgen, FWG, Klausdorfer Weg 2-24, Kiel S-H, 24148, Germany
Seung, Kim H., Konkuk Univ., CPO Box 6775, Seoul 100-667, South Korea
Shatalov, Michael Y., CSIR, Manufacturing and Materials Technology, P.O. Box 395, Pretoria 0001, South Africa
Sowizal, John, G&H Acoustics, LLC, 1419 Forest Dr., Suite 205, Annapolis, MD 21403
Spahn, Kevin K., Severns, Reid & Associates, Inc., 821 South Neil St., Champaign, IL 61820
Sprague, John F., 143 Myrtle Ave., Allendale, NJ 07401-1516
Streit, Roy L., Naval Undersea Warfare Center, 1176 Howell St., Newport, RI 02841
Svec, Jan, W. J. Gould Voice Center, Denver Center for the Performing Arts, 1245 Champa St., Denver, CO 80206
Swallow, John C., John Swallow Associates, Ltd., 366 Revus Ave., Unit 23, Mississauga, ON L5G 4S5, Canada
Tenghamm, Rune S. L., Petroleum Geo-Services, 16010 Barker's Point Ln., Houston, TX 77079
Terzi, Duilio, Rua Monte Alegre 470 apto. 22, Sao Paulo 05014-000, Brazil
Uber, John E., Environmental Engineering, Siemens Westinghouse Power Corporation, 4400 Alafaya Trail MC Q2-388, Orlando, FL 32826-2399
van Halteren, Aart Z., Research and Development, Microtronic Nederland BV, P.O. Box 61200, Amsterdam 1005 HE, The Netherlands
Varady, Mark J., OmniSonics Medical Technologies, 66 Concord St., Wilmington, MA 01887
Veirs, Val, 585 Smugglers Cove, Friday Harbor, WA 98250
Vilming, Stale, Heiasvingen 9C, Oslo N-1177, Norway
Vykhodtseva, Natalia L., Radiology Dept., Brigham and Women's Hospital, Harvard Medical School, 221 Longwood Ave., LMRC G013, Boston, MA 02115

Wellsted, David M., Psychology Dept., Univ. of York, York Y010 5DD, United Kingdom
Whalen, Joseph E., 27504 Berkshire Hills Place, Valencia, CA 91354
Wiley, Sara E., Shen Milson & Wilke, 3300 North Fairfax Dr., #302, Arlington, VA 22201
Will, Eric M., 6 Pinkham Ave., Mont Vernon, NH 03057
Wismer, Margaret G., Electrical Engineering, Dana Engineering, Bucknell Univ., Lewisburg, PA 17837
Wolf, Steven, Arsons Brinckerhoff, 505 South Main St., Orange, CA 92868
Wood, Michael N., Home Theater Magazine, 6400 Independence Ave., Woodland Hills, CA 91367
Yap, Adrian, DSO National Laboratories, Signal Processing Lab., 20 Science Park Dr., Singapore 118230
Yudichak, Thomas W., Applied Research Lab., Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713
Zamzes, Joseph P., Long Communication Group, Tyco Electronics Corp., 961 Burke St., Winston-Salem, NC 27101
Zhang, Xioaming, Mayo Clinic, Basic Ultrasond Research Lab., Dept. of Physiology and Biophysics, 200 First St., SW, Rochester, MN 55905
Zwerling, Eric M., Envrionmental Sciences, Rutgers Noise Technical Assist. Ctr., 14 College Farm Rd., New Brunswick, NJ 08901

New Students

Aasland, Wendi A., 606-4760 Cote-des-Neiges, Montreal, PQ H3V 1G3, Canada
Abdul Razak, Aishah, Faculty of Information Technology, Multimedia Univ., Jalan Multimedia, Cyberjaya, Selangor 63100, Malaysia
Acharyya, Ranjan, 100 East Univ. Boulevard, Apt. G20, Melbourne, FL 32901
Aguilar, Rolando N., Technical Univ. of Delft, ITS, Mekelweg 4, Room 13.100, Delft 2628, The Netherlands
Akana, Keola, 5246 Agnes Ave., #105, Valley Village, CA 91607
Armstrong, Jill C., 279 West 119th St., #4, New York, NY 10026
Babich, Jeffrey E., 2300 24th Rd., South, Apt. 547, Arlington, VA 22206
Barker, Heather M., 1350 King College, 130, Bristol, TN 37620
Bescup, John C., 113 Porter Hall, 60 South College Dr., Platteville, WI 53818
Bharitkar, Sunil G., 2461 Coolidge Ave., #1, Los Angeles, CA 90064
Bindell, Ariela R., 75 East Main St., Rockaway, NJ 07866
Bolat, Bulent, Electronic and Telecommunications Engineering, Yildiz Technical Univ., YTU Elk. ve Hab. Mh. Bl. Barnaros Blv. Besiktas, Istanbul 80750, Turkey
Boley, Jonathan, 1306 North Lincoln Ave., #202, Urbana, IL 61801
Bose, Arpita, Speech Language Pathology, Univ. of Toronto, 6 Queen's Park Crescent West, Toronto, ON M5S 3H2, Canada
Bunta, Ferenc, 840 East Joan D Arc, Phoenix, AZ 85022
Burcsak, Stephen M., Applied Research Lab., Univ. of Texas, 10000 Burnet Rd., Austin, TX 78758
Cairns, Richard L., 410 East Mineral St., Platteville, WI 53818
Carney, Melinda J., 6701 North Greenview, #2A, Chicago, IL 60626
Carroll, Samuel V., 200 Don Ave., Rumford, RI 02916
Cartwright, Mark B., Northwestern Univ., 1865 Sherman Ave., Room 239, Evanston, IL 60201
Cederroth, Christopher R., 5 chemin de la Boisserette, Geneva CH-1208, Switzerland
Chan, Kar-Ki, Computer Science and Engineering, The Chinese Univ. of Hong Kong, HSH Engineering Bldg., Shatin, Hong Kong
Chapin, Sean R., 3805 Houma Blvd., Apt. A206, Metairie, LA 70006
Coleman, Mark N., 2 Delafield Woods, St. James, NY 11780
Coren, Amy E., Psychology Dept., Univ. of Texas at Austin, Graduate Student Office, Austin, TX 78705
Dashtseren, Erdenebat, Graduate School of Science and Engineering, Shizuoka Univ., System Engineering, Johoku 3-5-1, Hamamatsu, Shizuoka 4328011, Japan
Deshpande, Parijat D., Volrat Thams Gatan 20-3412, Gothenburg 41260, Sweden
Doellinger, Michael, Dept. of Phoniatrics & Pediatric Audiology, Univ. of Erlangen-Nuremberg, Bohlenplatz 21, Erlangen, Bayern 91054, Germany
Englund, Anneli M., Sunlea 15 Hartlands Rd., Cork, Glasheen, Ireland
Epstein, Melissa A., Linguistcs, UCLA, 3125 Campbell Hall, Los Angeles, CA 90095-1543
Faber, Benjamin M., 167 West 2060 North, #194, Provo, UT 84604
Fisher, Rob W., 4930 Timothy Way, Fair Oaks, CA 95628
Furr, Elizabeth A., National Center for Physical Acoustics, Univ. of Mississippi, Coliseum Dr., University, MS 38677
Galea, Vincent J., P.O. Box 460, Sunbury VIC 3429, Australia
Gazagnaire, Julia, 545 Trace Cir., #212, Deerfield Beach, FL 33441
Gee, Kent L., 736 Wymount Terrace, Provo, UT 84604
Gerig, Anthony L., Medical Physics—153 MSC, Univ. of Wisconsin—Madison, 1300 University Ave., Madison, WI 53706
German, Brian J., 325677 Georgia Tech Station, Atlanta, GA 30332
Gopalan, Harish, Aerospace and Mechanical Engineering, 110 Cummington St., Boston, MA 02215
Hamaoui, Kamil, 4051-A Miramar St., La Jolla, CA 92037
Hoffman, Joel C., Fisheries Science, Virginia Inst. of Marine Science, P.O. Box 1346, Greate Rd., Gloucester Point, VA 23062
Holden, Craig, B&M Longworth, Ltd., Sett End Rd. North, Blackburn, Lancashire BB1 2QG, United Kingdom
Hollingsworth, Bill A., 109 College Station Rd., J104, Athens, GA 30605
Holt, Donann C., Educational Leadership, Univ. of Georgia, Rivers Crossings, Room 310, Athens, GA 30602
Hubbard, Roger J., 46 Victoria Terrace, Stafford, Staffordshire, Staffs ST163HB, England
Kaf, Wafaa A., 24 Welsford St., Apt. 2, Pittsburgh, PA 15213
Karner, Christopher J., 1321 West Farwell, Apt. 2, Chicago, IL 60626
Kawenski, Ted, 70 LeClair St., #1, Winooski, VT 05404
Kelly, John K., Speech and Hearing Science, Ohio State Univ., 110 Pressey Hall, 1070 Carmack Rd., Columbus, OH 43210
Kenderian, Shant, Center for Nondestructive Evaluation, Johns Hopkins Univ., 810 Wyan Park Dr., Baltimore, MD 21211
Langlois, Christian, Mechanical Engineering, Univ. de Sherbrooke, 2500 Blvd. de L'Universite, Sherbrooke, PQ J1K 2R1, Canada
Larue, James P., 3920 Evangeline Ave., Chalmette, LA 70043
Lee, Bowon, Electrical and Computer Eng., Beckman Inst., MC 251, 405 North Mathews Ave., Urbana, IL 61801
Lee, Sunwoong, 1010 Massachusetts Ave., #1, Cambridge, MA 02138
Levi, Susannah V., Linguistics Dept., Univ. of Washington, Box 354340, Seattle, WA 98195
Lin, Yi-Jiun K., Communication Sciences, Univ. of Connecticut, 850 Bolton Rd., U-85, Storrs, CT 06269-1085
Locey, Lance L., Brigham Young Univ., Physics and Astronomy, N283 ESC, Provo, UT 84602
Luk'yanchuk, Mikhail B., N. N. Andreev Acosutics Inst., Russian Academy of Sciences, Svernika St. 4, Moscow 117036, Russia
Matveev, Konsantin I., Caltech, 301-46, Pasadena, CA 91125
McCabe, Terrence, 45 Excelsior Ave., Troy, NY 12180
McDermott, Scott D., 1431 St. Claude Ave., New Orleans, WA 70116
McLennan, Sean, Linguistics/Cognitive Science, Indiana Univ., 322 Memorial Hall, 1021 East Third St., Bloomington, IN 47405-7005
Morrison, Andrew C. H., Physics Dept., Northern Illinois Univ., 202 Faraday West, DeKalb, IL 60115
Narayan, Chandan R., 439 3rd St., #7, Ann Arbor, MI 48103
Newman, Matt M., 733 $\frac{1}{2}$ Sannoner Ave., Florence, AL 35630
Nguyen, Thu T. A., School of English, Media Studies & Art History, Univ. of Queensland, QLD 4072, Australia
Osgood, Jonathan H., 501 East College ''A,'' Apt. 2017, Wheaton, IL 60187
Outing, Donald A., 4-B Mordecai Dr., Watervliet, NY 12189-4008
Page, David A., 4352 Soldiers Home Rd., West Lafayette, IN 47906
Pawluk, Ann Marie, 221-17 69th Ave., Bayside, NY 11364
Pegors, Katherine, 550 Windsor Halls, Lafayette, IN 47906
Pilchta, Bartlomiej, 1313D University Village, East Lansing, MI 48823
Poling, Jeremy R., Room 308, RWJ, P.O. Box 353, Milwaukee, WI 53201-0353
Poole, Travis L., 150 Webster Ave., #2, Cambridge, MA 02141
Poon, Becky B., 305 Memorial Dr., 614B, Cambridge, MA 02139
Qin, Michael K., Harvard–MIT Div. of Health Sci. and Tech., 77 Massachusetts Ave., 36-769, Cambridge, MA 02139
Reilly-Raska, Laurel K., Math Sciences, Rensseleaer Polytechnic Inst., 110 15th St., Amos Eaton Room 301, Troy, NY 12180
Robinson, Mark C., P.O. Box 704, Mount Washington, KY 40047
Roedelbronn, Michael F., Physics Dept., Univ. of Illinois, 1110 West Green St., Urbana, IL 61801
Scott, David E., 6033 North Paulina, 1A, Chicago, IL 60660

Semenova, Tatiana V., 620 West Forest Ave., Apt. 27, Detroit, MI 48201-1152
Sinha, Shiva R., Neuroscience and Cognitive Science Program, Psychology, Univ. of Maryland, College Park, MD 20742
Smith, Zachary M., Massachusetts Eye and Ear Infirmary, Peabody Lab., 243 Charles St., Boston, MA 02114
Snook, Bradley M., 1635 Noriega St., San Francisco, CA 94122
Suh, Sanghoon, School of Mechanical Eng., Purdue Univ., 1077 Ray W. Herrick Labs., West Lafayette, IN 47907
Swick, Andrew H., Wiess College, Rice Univ., 6340 Main St., Houston, TX 77005
Takamaru, Keiichi, 2-3-5-31 Higashisapporo, Shitoishi-ku, Sapporo, Hoklaido 0030002, Japan
Tam, Mei-Wa T., Audiology, Univ. of South Florida, 4202 East Fowler Ave., Tampa, FL 33620
Thomson, Scott L., Mechanical Engineering, Purdue Univ., 76 HERL, West Lafayette, IN 47907
Trenton, John, 118 East Kingwood Dr., Apt. D7, Murfreesboro, TN 37130
Turk, Oytun, Kurtulus Cad. Baysingur Sk. Ferah, Apt. No. 245-247, D:10, Istanbul 80240, Turkey
Ulrich, T. J., 1555 North Sierra St., #133, Reno, NV 89503
Van Cauwelaert, Javier, 71 Chester St., Apt. 4, Allston, MA 02134
van Etten, Chris P., P.O. Box 209, Tuart Hill, WA 6939, Australia
van Alphen, Petra M., MPI for Psycholinguistics, P.O. Box 310, Nijmegen, Gelderland 6500 AH, The Netherlands
Vazquez-Perez, Gabriel J., Venezuelan Navy, Inst. of Sound and Vibration Research, Univ. of Southampton, Southampton SO17 2NU, United Kingdom
Wang, Fei, Engineering Science and Mechanics, Pennsylvania State Univ., 212 EEC Bldg., University Park, PA 16802
Watson, Jennifer A., 100 Massmill Dr., #510, Lowell, MA 01852
Watson, Kevin D., Edge Hill College of Higher Education, English Language and Literature, St. Helens Rd., Ormskirk, Lancashire L39 4QP, United Kingdom
Wezensky, Eryn M., 3800 Maiden St., Waterford, MI 48329
Wiese, Cassandra H., 57 West Montana, Glendale Heights, IL 60139
Williamson, Rene G., Dept. of Mechanical Engineering, Yale Univ., 9 Hillhouse Ave., ML219, New Haven, CT 06520
Won, Jungun, 1400 Martin St., #2098, State College, PA 16803
Woolley, Jonathan A., 300 East St., Golden, CO 80403
Yates, Gregory D., 7 Berkeley Ln., Topsham, ME 04086
Yen, KunLung, 1F, 96 Dong Ming St., Hsinchu, 300 Taiwan, Republic of China
Yoon, Sung Y., 2136 5th Ave., Basement, Troy, NY 12180
Yoon, Yang-soo, Indiana Univ., 107 South Indiana Ave., Bloomington, IN 47405

New Electronic Associates

Anderson, Linda M., 28 Olive Ave., San Anselmo, CA 94960
Arbel, Dan, Dan Arbel Risk Eng., 27 Albert Schweizer, Haifa IL 34995, Israel
Bass, Andrew H., Neurobiology and Behavior, Cornell Univ., Mudd Hall-Tower Rd., Ithaca, NY 14853
Bettinzana, Silvano, Casella Postale, Calino BS 25040, Italy
Blank, William T., Microsoft, One Microsoft Way, Redmond, WA 98052
Bonori, Maurizio, Dipt. di Fisica, Univ. Degli Studi di Roma ''La Sapienza,'' Ple A. Moro 2, Roma 00185, Italy
Bradley, Hare A., OmniSonics Medical Technologies, 66 Concord St., Wilmington, MA 01887
Cetinkaya, Cetink, Clarkson Univ., CAMP 241 Box 5725, Potsdam, NY 13699-5725
Herlin, Bo, Garbage Collector AB, Hogvallen 106, Undersaker, Jamtland 830-10, Sweden
Herman, Vladimir, CTRL Systems, Inc., Research Lab., 1004 Littlestown Pike, Suite H, Westminster, MD 21157
Karpov, Sergey, Parametric Technology, 140 Kendrick St., Needham, MA 02494-2714
Keane, Jim, Sound Alert Technology, Tarus Park, Europa Blvd., Warrington, Cheshire WA5 7YT, United Kingdom
Kocak, Ismail, Istabul Surgery Hospital, Dept. of Otolaryngology, Ferah Sok, No. 18 Nisantasi, Istanbul 80200, Turkey
Loggins, Chester D., Sonatech, Inc., 879 Ward Dr., Santa Barbara, CA 93111
Mohammed, Ichchou, ECL, MSGMGC, 36 Ave. Guy de Collongue, Ecully 69130, France
Monroe, James W., W. L. Gore Associates, Inc., 1901 Barksdale Rd., Newark, DE 19713
Riedel, Richard F., Riedel Audio Services, 443 Potter Boulevard, Brightwaters, NY 11718
Rutkiewicz, Mark S., St. Croix Medical Inc., 5301 East River Rd., Minneapolis, MN 55421
Succi, Geroge P., Sentech, 38 Montvale Ave., Stoneham, MA 02180
Ulrich, Kathleen M., Sonus, USA, 201 East Ogden Ave., #126, Hinsdale, IL 60521
Wootters, Adrienne H., Mass. College of Liberal Arts, 375 Church St., North Adams, MA 01247
Wouters, Johan, OGI School of Sci. and Eng., Oregon Health & Sciences Univ., Electrical and Computer Eng., 20000 NW Walker Rd., Beaverton, OR 97006

New Corresponding Electronic Associates

Goncalves, Anastacio P., Av. Oceanica 52 Ondina, Salvador, Bahia 170-010, Brazil
Higuti, Ricardo T., Electrical Engineering, Unesp-FEIS, Avenida Brasil, 364, Ilja Solteira SP 15385-000, Brazil
Ivanov, Alexei V., 14-4 Shafarnyanskaya Str., Minsk 220125, Belarus

Associates Elected Members

D. C. Bruck, B. S. Cazzolato, F. M. Guillot, S. Kenji, D. Khismatullin, M. Kumaresan, S. K. Lehman, T. W. Murray, K. T. Nakayama, H. Olsen, I. Park, T. J. Poterek, C. H. Reyes, C. J. Robinson, T. M. Sheridan, U. P. Svensson, J. T. Tibbitts, C. Wang, L. M. Wang, W. Xu, K. Yamazaki, W. Zhu

Students to Associates

T. J. Adam, J. Auriemmo, K. M. Becker, W. J. H. Choi, K. Chung, S. L. Curtin, J. M. Dodson, M. V. Golden, S. D. Hansen, J. F. Heake, G. M. Hughes, J. A. Jones, R. Koehler, T. M. Mueller, R. Shrivastan, N. E. Singer, S. E. Solberg, A. V. G. Sousa, A. E. Stefaniw, S. P. Y. Stenfelt, A. D. Sweeney, C. P. Walsh, K. R. Waters, O. Weisser

Associate to Student

K. M. Rosen

Associate to Electronic Associate

H. M. Rodriguez-Davila

Member to Corresponding Electronic Associate

I. L. Oboznenko

Reinstated

P. A. Chinnery, G. P. Gibbs, D. A. McCurdy, M. D. Slurzberg, Y. M. Szymko-Bennett—*Members*
E. Tzoukermann—*Associate*

Resigned

F. F. Romanow—*Fellow*
G. A. Alers, W. T. Buhl, D. S. Emmerich, A. G. H. Gabrielsson, L. R. LeBlanc, B. B-P. Lim, J. C. Page, C. G. Rudisill, W. F. Tjaden, T. Wang—*Members*
P. Coulter, O. Gasparini, C. D. Henze, H. D. Luke, D. L. Sorkin, J. E. Stangel, M. Ziv—*Associates*
F. Duchassin, R. Hill, T. Hollingset, N. Scholtz, G. C. Smith, G. Vacca—*Students*

Deceased

R. H. Bolt, D. L. Johnson, W. A. Rosenblith, J. E. Wesler—*Fellows*
L. S. Jones, B. Lundquist, R. Milroy, D. J. Ramsdale, E. M. Relkin—*Members*

Dropped

Bill G. Watters—*Fellow*

Naomi Aizawa, Tuncay Akal, Federico Albano Leoni, John G. Alekna, Ziad K. Alhamdani, Charles W. Allen, Eric Neil Angevine, David Robert Bacon, Theresa A. Baus, Alice O. Berkowitz, Hugo F. Bezdek, Ning Bi, Barry C. Birdwell, J. Alton Burks, Curtis I. Caldwell, Kathleen C. M. Campbell, John H. Casseday, Edward D. Chaika, Bishwajit Chakraborty, Vladimir Chaloupka, Olga Chernets, Young-Chung Cho, Ronald A. Cole, Brian J. Conolly, William L. Dalton, Robert A. Day, Steven P. Dear, Alexander A. Doinikov, Jan Edwards, N. Scott Emery, John J. Galvin, Martin A. Gold, Igor I. Gorban, Karl F. Graff, Robert T. Green, James R. Gundlach, Steven M. Haas, Robert J. Hallermeier, David John Hill, Edward Hojan, Sung-Rung Huang, John H. Huckans, Kenzo Ishizaka, Christopher T. Kello, Jeffrey J. Kelly, Heung Sik Kim, Cornelis W. Koutstaal, Aleksandr G. Kovalyuk, Clifton L. Kussmaul, Akhlesh Lakhtakia, Richard J. Lataitis, Kin Tat Lee, Peter Y. Lee, John J. Leonard, Johan C. W. A. Liljencrants, Thomas M. Logan, Richard J. Lucas, Timothy S. Lucas, Helmut A. Mangold, John F. Miller, Kurt N. Milligan, Mark J. Moeller, Eric C. Mousset, Bakulesh B. Patel, Ernest A. Peterson, Jorge A. Porti, Jean Ripoche, Jean C. Risset, William E. Ryan, Armen Sarvazyan, Marina V. Shitikova, Malcolm Slaney, Catherine Stamoulis, Foster B. Stulen, Tomohiko Taniguchi, Maureen B. Tracy, Richard Evan Doeren, Jerry R. Veen, H. Joseph Venne, Easwaran Viswanathan, HisashiWakita, Bernd Weinberg, Gregory L. Wojcik, Yangguang Xu, Qiang Xue, Tsang-Chiang Yang, Graeme K. Yates, Zhen Ye—*Members*

Douglas A. Abraham, Barbara E. Acker, Jeremy A. Agnew, Grant Allan, Orson L. Anderson, Erik W. Andrews, Norma S. Antopnanzas-Barroso, Thomas J. Asaki, Henrik T. Ax, Mohammed Bahobail, Justin E. Baird, Martin J. Ball, James D. Bass, Kirstin E. Beach, Alison Behrman, Carlos I. Beltran, Erlend Bjorndal, Mark E. Blasky, Eric Boler, Michael F. Boucher, Edward D. Breitenbach, Patrick C. Brown, William S. Brown, Daniel V. Brown, Jeff M. Brunstrom, John A. Busenitz, Ross H. Casey, Michael D. Cerna, Chris M. Chanko, Cheng-Hsiung Chen, Rathinavelu Chengalvarayan, Sunil Chojar, Kim Hoi Chou, John T. Christine, Lou J. Clark, Andre M. Cooper, Cameron J. Craigie, Stephen J. Crawford-Bourke, Brian J. Davis, Joseph L. Day, Marco T. DeBiasi, Rory A. DePaolis, Fahrettin L. Degertekin, Phalguni S. Deliwala, Robert L. Dion, Thomas G. Dolan, Donald D. Dorfman, Tadeusz M. Drzewiecki, Alberto D. Duenas, Scott D. Duncan, Hathaichanok Duriyabunleng, Ira B. Ekhaus, Theodore R. Farrell, N. Douglas Fast, Vladimir E. Fedorov, Colm M. Flannery, John W. Flood, Mark A. Friesel, Leo J. Gallenstein, Robert B. Gampert, William G. Gardner, Arnold L. Garlick, Mercedes Gonzalez Suarez, Michael Gottfried, Walter B. Green, David G. Hackman, Jeffrey L. Hanson, P. Greig Happoldt, Yoko Hasegawa, Zulkifli B. Hashim, Rachel M. Hemphill, Kwon Young Heon, Chow K. Hin, James W. Hiney, Margarita S. Hnilo, John E. Hogden, Brian W. Holmes, Gary W. Holmes, Nigel J. Holt, Kambiz Homayounfar, Christopher A. Homola, Douglas N. Honorof, Zi Qiang Hu, Susan L Hura, Michael E. Huster, Dejun Jiang, Paul L. Joitke, Andrew Jones, Yoram Kadman, Shoji Kajita, Kenneth L. Kantor, Suzanne Keilson, Robert P. Kendig, Michael Kim, Kerry A. Klingborg, Matthew T. Korporaal, Guenter J. Krauss, James A. Kvikstad, Chris Kyriakakis, Stephen G. Lacker, Steven A. Lane, James E. Lankford, Claus Larsen, Sungbok Lee, Torsten Lenk, Mireille F. Levy, Anders Lindstrom, Matthew K. Ling, Mark R. Loewen, John S. Logan, David Lord, Fang-Ling Lu, Alex S. Lung, Ernest L. Madsen, Luciano Maffei, Sergey N. Makarov, Nirmal K. Mandal, Douglas S. Mandic, Robert D. Marciniak, Mehmet Marsan, Francesco Marulo, Peter Marvit, Didier Mauuary, J. Devin McAuley, Diane L. Meador, Lee Mendoza, Mary E. Meskan, Douglas R. Messner, Sela M. Meyouhas, Christophe Micheyl, Thomas K. Mills, Bethany A. Milner, John E. Moe, Jaime Moreno, Dennis R. Morgan, Richard E. Morrow, Annie M. Moulin, Yongke Mu, Ronald A. Mucci, Murray J. Munro, Edouard G. Nesvijski, Douglas G. Niessen, Susan J. Norton, Michael J. Nutting, Michael W O'Brien, Hiroyuki Ogawa, Won T. Oh, Louis P. Oligee, Robert F. Orlikoff, Chandramouli Padmanabhan, Dinesh K. Pai, Nils Paz, Norman I. Perle, Hoang Pham, Andy D. Phelps, Kenneth L. Pilgreen, Hans G. Piroth, Michel Piterman, Todd A. Pitts, Curtis W. Ponton, David C. Potter, Jeff Power, Wayne E. Prather, Mark R. Prausnitz, Daniel V. Rabinkin, Bohdan Raczynski, Krishna M. Rajana, Kuchimanchi V. S. Rama Rao, Guillermo Ramirez, Ashwin P. Rao, Thibault Rigaudias, Joseph M. Riley, Eric G. Rodgers, Damion Romero, Galina Rovner, Howard W. Sabrin, David E. Sanders, Miguel A. Sattler, Robert E. Schlegel, Sima Segev-Gilad, Vimal V. Shah, Ilya Shoykhet, Vernon P. Simmons, Jonathan Z. Simon, Kathleen A. Siren, Paula M. T. Smeele, Linda L. Smith, Eric Smith, Vikas S. Sonwalkar, John F. Spadaro, Bradley M. Starobin, John M. Storyk, Shawn Strange, James A. Stratman, Alvin W. Y. Su, Bruno Suner, Seetharaman Swarnamani, Theodore C. Tanner, Rob H. Telling, David D. Thiel, Roger M. Thies, Vaughn M. Thomas, Biao Tian, William Tobocman, Christopher E. Truman, Richard D. Tschiegg, Timo K. Tuori, Douglas E. Vetter, John T. Viechnicki, Ir Wahyudi, Karen L. Wallace, Chao-Nan Wang, Lora G. Weiss, Robert L. Whitehead, Allen B. Wright, Timothy F. Wright, Denny Yim, Jeongbin Yim, Miida Yoshiro, Mark Zagaeski, Shengke Zeng Wei Zhou, Moche Ziv—*Associates*

Nazan Akman, Shaker H. Al-Alawi, Meagan K. Andrews, Antonios Apostolou, Dhany Arifianto, Robert L. Asmussen, Heriberto Avelino, Azrael Azrael, Michael R. Backfish, Charlie E. Baynes, Tobias P. Berndt, Denise Bloom, Candace D. Bourland, Denise M. Bowman, Cliff S. Burgess, Jason S. Chan, Michael T. Chan, Xiaohui Chen, Kuang-Tao Chiao, Taehong Cho, Joseph A. Cichock, Cory L. Clarke, Daniel A. Cook, Joseph J. Corrigan, Omar Cruz, Mariapaola D'Imperio, Michael J. Daley, Peter W. De Bonte, Nicole M. Devitt, Leanne E. Dodge, Fang Dong, Michael V. Drexel, Andrew L. Dunlap, Sorin V. Dusan, Patrick L. Edson, Erhan Ermisoglu, Annika M. Fain, Brandon C. Faulds, Patrick McDowell Fitzsimmons, Katherine J. Frisch, Alvaro R. Fuentes Cabrera, Joseph Barry Gaalaas, Jose A. Gallegos, Valdez L. Gant, Roderick C. Gauss, Vincente E. Giuseppe, Alexander M. Goberman, Ari Z. Goldberg, Andrew Gorton, Dimitar P. Gueorguiev, Tom W. Gunderson, Dave A. Harris, David A. Heesch, Janelle L. Helser, Merrill G. Hibbs, Jennifer C. Hill, Steven R. Hill, Changmin Hu, George Hwang, Choi Jaeseung, Ben Alexander Johnson, Matthew K. Johnson, Megan M. Jones, Kristin Kaschner, Steven M. Kaye, Michael F. Keane, Kate Kelly, Daniel B. Kilfoyle, Sang Kun Kim, Bumjun Kim, Mi Ryoung Kim, Kayo Kimotsuki, Michael D. Krein, B. Suresh Krishna, Jeremy K. Laabs, Michelle Langlais, Pierre Ph. Ledevedec, Young Jin Lee, Young-Sup Lee, Hai Lin, Wensen Liu, Steven B. Luka, Gretchen A. Maia, Kelvin C. Maliti, Yutaka Mamada, Sarah E. Marsh, Alex G. Militello, Christine R. Montgomery, Benjamin D. Murphy, Kyle T. Nakamoto, Aravind K. Namasivayam, Flavio Nardi, Paul H. Nelson, Nicole G. Olson, Hyun Paek, Blair O. Parker, Blas G. Payri, Tony R. Pegg, Trudy L. Philip, Ioana N. Pieleanu, Robert J. Podesva, Kelly L. Poort, Tyrone M. Porter, Mario M. Rollo, Douglas Rollow, Sarah M. Rotter, Peter L. Russo, Robert J. Sanchez, M. Loreto Sanchez, Sven E. Sawin, Diana L. Schenck, John C. Schmitt, Elijah A. Schutz, Danielle M. Seablom, Liat Seiger, Pusadee Seresangtakul, Lisa C. Shoaf, Paulus J. W. Siwabessy, Christian M. Skinner, Kelly A. Smith, Dana M. Smith, Thomas H. Stainsby, Takao Suzuki, Wolfgang R. Svensson, Joyce L. Tachner, Rachod Thongprasirt, Prisilia Tirtabudi, Katie E. Toop, Christos Tsakostas, Georgios Tserdanelis, Alex F. Turkovic, Albert F. Turri, Peng Wang, Sameh S. Wanis, Andrew Whitten, Thomas C. Willett, Brandon S. Woll, Zhongyu Yan, Jun Yao, River H. Yuan, Jiahong Yuan, Wei Zhang—*Students*

Fellows	997
Members	2714
Associates	2627
Students	961
	7299

ACOUSTICAL NEWS—INTERNATIONAL

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

New Home Pages

Some National or International Acoustical Societies or Associations have recently changed their Web addresses or have started new Home Pages. At this writing these new URL are:

Czech Acoustical Society	www.czakustika.cz
European Acoustics Association	www.euracoustics.org
French Acoustical Society	www.sfa.asso.fr
International Commission for Acoustics	www.icacommission.org
Latvian Acoustical Society	www.akustika.lv
New Zealand Acoustical Society	www.acoustics.org.nz
Norwegian Acoustical Society	www.akustisk-selskap.com
Slovak Acoustical Association	www.tuzvo.sk/skas
Swiss Acoustical Society	www.sga-ssa.ch

International Meetings Calendar

Below are announcements of meetings and conferences to be held abroad. Entries preceded by an * are new or updated listings.

August 2002

19–23 **16th International Symposium on Nonlinear Acoustics (ISNA16)**, Moscow, Russia. (Fax: +7 095 126 8411; Web: acs366b.phys.msu.su/isna16/)

26–28 **2nd Biot Conference on Poromechanics**, Grenoble, France. (Web: geo.hmg.inpg.fr/biot2001)

26–28 **Joint Baltic-Nordic Acoustical Meeting 2002**, Lyngby, Denmark. (Fax: +45 45 88 05 77; Web: www.dat.dtu.dk/~b-nam)

September 2002

10–12 **32nd International Acoustical Conference—EAA Symposium**, Banská Štiavnica, Slovakia. (Fax: +421 45 532 1811; Web: alpha.tuzvo.sk/skas/acoustics)

11 **Dutch and Belgian Acoustical Societies Joint Meeting**, Utrecht, The Netherlands. (e-mail: info@nag-acoustics.nl)

11–13 **10th International Meeting on Low Frequency Noise and Vibration**, York, UK. (Fax: +44 1277 223 453; Web: www.lowfrequency2002.org.uk)

16–18 **International Conference on Noise and Vibration Engineering**, Leuven, Belgium. (Fax: +32 1632 2987; Web: www.isma.isaac.be)

16–21 **Forum Acusticum 2002 (Joint EAA-SEA-ASJ Meeting)**, Sevilla, Spain. (Fax: +34 91 411 7651; Web: www.cica.es/aliens/forum2002)

26–28 **Autumn Meeting of the Acoustical Society of Japan**, Akita, Japan. (Fax: +81 3 5256 1022; Web: wwwsoc.nii.ac.jp/asj/index-e.html)

30–1 ***Hellenic National Conference ACOUSTICS 2002**, Patras, Greece. (D. Skariatos, Department of Mechanics and Aeronautical Engineering, University of Patras, 26500 Patras, Greece; e-mail: skariat@mech.upatras.gr)

October 2002

9–11 **Acoustics Week in Canada**, Charlottetown, PE, Canada. (Fax: +1 902 628 4359; Web: caa-aca.ca/PEI-2002.html)

24–25 ***International Symposium on Past, Present and Future of the Transmission Index [cosponsored by ASA]**, Soesterberg, The Netherlands. (S. J. van Wijngaarden, Kampweg 5, 3769 DE Soesterberg, The Netherlands; Fax: +31 346 35 39 77; Web: www.sti-symposium.org)

26–28 **6th National Congress of the Turkish Acoustical Society**, Kars, Turkey. (Fax: +90 212 261 0549; Web: www.takder.org/kongre-2002/kongre-2002.html)

November 2002

7–8 ***Fall Meeting of the Swiss Acoustical Society**, Wallis, Switzerland. (SGA-SSA, Akustik, SUVA, P.O. Box 4358, 6002 Luzern, Switzerland; Fax: +41 419 6213; Web: www.sga-ssa.ch)

13–15 **Australian Acoustical Society Conference 2002**, Adelaide, Australia. (Fax: +61 8 303 4367; Web: www.mecheng.adelaide.edu.au/aasconf)

15–17 **Reproduced Sound 18: Perception, Reception, Deception**, Stratford-upon-Avon, UK. (Fax: +44 1727 850553; Web: www.ioa.org.uk)

21–22 **New Zealand Acoustical Society 16th Biennial Conference**, Auckland, New Zealand. (e-mail: graham@marshallday.co.nz)

30–6 **Joint Meeting: 9th Mexican Congress on Acoustics, 144th Meeting of the Acoustical Society of America, and 3rd Iberoamerican Congress on Acoustics**, Cancún, Mexico. (Web: asa.aip.org/cancun.html)

December 2002

9–13 **International Symposium on Musical Acoustics (ISMA Mexico City)**, Mexico City, Mexico. (Fax: +52 55 5601 3210; Web: www.unam.mx/enmusica/ismamexico.html)

19–21 ***3rd WSEAS International Conference on Acoustics, Music, Speech and Language Processing**, Tenerife, Canary Islands, Spain. (Web: www.wseas.org/conferences/2002/tenerife/icamsl/)

March 2003

17–20 **German Acoustical Society Meeting (DAGA2003)**, Aachen, Germany. (Fax: +49 441 798 3698; E-mail: dega@aku-physik.uni-oldenburg.de)

18–20 **Spring Meeting of the Acoustical Society of Japan**, Tokyo, Japan. (Fax: +81 3 5256 1022; Web: wwwsoc.nii.ac.jp/asj/index-e.html)

April 2003

7–9 **WESPAC8**, Melbourne, Australia. (Web: www.wespac8.com)

June 2003

8–13 **XVIII International Evoked Response Audiometry Study Group Symposium**, Puerto de la Cruz, Tenerife, Spain. (Web: www.ierasg-2003.org)

16–18 **Acoustics 2003—Modeling & Experimental Measurements**, Cadiz, Spain. (Fax: +44 238 029 2853; Web: www.wessex.ac.uk/conference/2003/acoustics/index.html)

29–3 **8th Conference on Noise as a Public Health Problem**, Amsterdam–Rotterdam, The Netherlands. (Fax: +31 24 360 1159; E-mail: office.nw@prompt.nl)

 0001-4966/2002/112(2)/357/2/$19.00

July 2003

7–10 **10th International Congress on Sound and Vibration**, Stockholm, Sweden. (Fax: +46 88 661 9125; Web: www.congex.com/icsv10)

14–16 **8th International Conference on Recent Advances in Structural Dynamics**, Southampton, UK. (Web: www.isvr.soton.ac.uk/sd2003)

August 2003

6–9 **Stockholm Music Acoustics Conference 2003 (SMAC03)**, Stockholm, Sweden. (Web: www.speech.kth.se/music/smac03)

25–27 **Inter-Noise 2003**, Jeju Island, Korea. (Fax: +82 42 869 8220; Web: www.icjeju.co.kr)

September 2003

1–4 **Eurospeech 2003**, Geneva, Switzerland. (Web: www.symporg.ch/eurospeech2003)

7–10 **World Congress on Ultrasonics**, Paris, France. (Fax: +33 1 46 33 56 73; Web: www.sfa.asso.fr/wcu2003)

October 2003

15–17 **34th Spanish Congress on Acoustics**, Bilbao, Spain. (Fax: +34 91 411 7651; Web: www.ia.csic.es/sea/index.html)

March 2004

22–26 **Joint Congress of the French and German Acoustical Societies (SFA-DEGA)**, Strasbourg, France. (Fax: +49 441 798 3698; e-mail: sfa4@wanadoo.fr)

April 2004

5–9 **18th International Congress on Acoustics (ICA2004)**, Kyoto, Japan. (Web: www.ica2004.or.jp)

September 2004

13–17 **4th Iberoamerican Congress on Acoustics, 4th Iberian Congress on Acoustics, 35th Spanish Congress on Acoustics**, Guimarães, Portugal. (Fax: +351 21 844 3028; e-mail: dsilva@lnec.pt)

BOOK REVIEWS

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

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

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

Real-time Adaptive Concepts in Acoustics: Blind Signal Separation and Multichannel Echo Cancellation

Daniël W. E. Schobben

Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001. xii + 160 pp. Price: $87.00 (hardcover) ISBN: 0792371097.

Adaptive signal processing techniques have many applications in acoustics including echo cancellation, noise control, array processing, filtering, and acoustic communication. Many books are available on the general subject of adaptive signal processing. The book by Schobben focuses on real-time adaptive algorithms for blind signal separation and multichannel echo cancellation. Blind signal separation (BSS) is an active and productive research topic on which a large number of research papers have been published over the last decade. Multichannel echo cancellation is a technology-driven engineering topic that has wide commercial applications to multichannel audio communication and speech recognition systems. The novel feature of this book, as delineated in its title, is emphasis on real-time, adaptive concepts and two specific but important applications: blind signal separation and multichannel echo cancellation.

The book consists of ten chapters and six appendices. Included with the book is a compact disk that can be played on any compact disk (CD) player to play the audio tracks of the experiments that are described in the book. The CD also contains MATLAB implementation of the algorithms that have been described in the text. The audio data can be used by the reader for experimentation. More audio tracks are available on the author's home page. The book is divided into three parts: an introductory part, entitled ''Real-time adaptive concepts in acoustics: an introduction,'' and two parts that deal with specific algorithms. The second part is entitled ''Acoustic echo cancellation'' and the third part is called ''Blind signal separation.'' Chapter 1 is a survey of BSS and echo cancellation and other techniques of related adaptive signal processing techniques. This chapter briefly reviews multichannel signal processing concepts such as echo cancellation, BSS, deverberation, and noise reduction. It also has a brief discussion of room acoustics, properties of microphones, and speech. Obviously, these important topics are not treated in depth and are only presented to provide some motivation and background for adaptive signal processing concepts. The end of the first chapter contains a definition and list of notation that is subsequently used. This feature greatly facilitates the more detailed study of algorithms that are presented in following chapters.

Chapter 2, entitled ''Array processing techniques,'' gives an overview of array processing techniques with particular emphasis on acoustic applications. An overview of beamforming includes a brief discussion of delay-sum beamforming, filter-sum beamforming, constrained beamforming, and side-lobe canceler. Related topics that are briefly discussed in this chapter are blind deconvolution, detection of multiple signals, inverse room acoustics, system modeling, signal separation, and noise cancellation. Obviously a brief chapter can provide only a cursory overview of these wide-ranging topics. For a more thorough study, the reader should turn to specialized texts on array processing, beamforming, and adaptive signal processing.

The next chapter, entitled ''Efficient filtering using FFTS,'' is an introduction to the efficient implementation of large adaptive filters in the frequency domain. This chapter provides a good introduction to frequency domain filtering for the readers who are not familiar with the overlap-save techniques in frequency domain filtering. Again, for more thorough treatment of FFT base adaptive filters, the reader is referred to specific texts such as S. Haykin, *Adaptive Filter Theory* (Prentice Hall, Englewood Cliffs, NJ, 1996). This chapter does provide new insight into low-latency frequency domain filters. Low latency is important in active noise cancellation because the filtered noise (anti-noise) must be time aligned with physical noise that is to be cancelled. Chapter 3 concludes the introductory part of the book.

The first chapter in the second part of the book is Chap. 4, entitled ''An efficient adaptive filter implementation.'' This chapter provides an introduction to adaptive filters, reviews time domain adaptive filters, and presents a detailed derivation of the block frequency domain adaptive filter. The emphasis here is on efficient computation of FFTs and filter updating. This chapter points out that an important requirement in applications of adaptive filters for acoustic noise cancellation is that adaptive filters and algorithms must have small delays. It is also pointed out that computational efficiency is important for acoustic applications.

Chapter 5, entitled ''Efficient multichannel RLS,'' derives, in detail, a computationally efficient adaptive multichannel algorithm which is based on the classical multichannel recursive least squares (RLS) algorithm. The derivation of the new algorithm is a generalization of the block frequency domain adaptive filter. The resulting algorithm is a computationally efficient approximation to the well-known RLS algorithms. Its main application is multichannel acoustic echo cancellation. A brief overview of this application is given. An adaptive filter is implemented using block processing in frequency domain. The multichannel RLS is a generalization of the single channel case that was presented in the previous chapter. The resulting algorithm is called *multichannel block frequency domain adaptive filter* (MC-BFDAF) and has been implemented in MATLAB. The chapter concludes with a presentation of experimental results. The MC-BFDAF algorithm is used to cancel correlated signals that are projected by two loudspeakers and picked up by a single microphone in a room. The objective is to suppress the loudspeaker signals at the microphone. Experimental results show that the MC-BFDAF algorithm suppresses the microphone output power by 32 dB after 10 s. If a single-channel BFDAF is used separately on each channel using the same experimental setup as before, then only 17 dB suppression is obtained. This clearly shows that a multichannel algorithm is required for effective echo suppression.

Part III of the book is devoted to blind signal separation. Chapter 6 is entitled ''Blind signal separation, an overview,'' and it reviews existing methods of blind signal separation (BSS). BSS is a method that extracts desired signals from a linear mixture of multiple signals. The word ''blind'' is used to describe this method because details of the mixing system are not known in advance, only the general structure of the mixing system is postulated. The mixing system is either assumed to be an instantaneous mixture, modeled by a mixing matrix, or a convolutive mixture, modeled by a matrix of filters. Signal separation is achieved by use of a nonlinear steepest decent algorithm to adjust parameters of an unmixing system. Algorithms are nonlinear because they are based on optimization of either higher-order statistics, probability density functions (Kullback–Liebler divergence), or information theoretic criteria (information maximization, entropy or mutual information). Unlike optimization of second-order statistics (means square error or other squared function of data), gradients of these criterion functions are nonlinear functions. The chapter concludes with an argument that using second-order statistics is sufficient to achieve separation of acoustic signals, and therefore simpler linear algorithms can be used. Chapter 7 pre-

 0001-4966/2002/112(2)/361/2/$19.00

sents a new blind signal separation algorithm, which is based on second-order statistics. Experiments with recorded data show that the algorithm has an acceptable performance when used for separation of speech and music. The algorithm can be implemented in real time on a personal computer.

Chapter 8 deals with joint blind signal separation and echo cancellation problems that arise in teleconferencing and voice-controlled machinery. In teleconferencing there are loudspeakers in the same room with a local speaker at a microphone. The contributions of these loudspeakers must be cancelled from the microphone signal and then the remaining signals are separated to recover individual local speakers. The author presents a low computational complexity algorithm to accomplish this task. Performance of the algorithm is verified by data that has been recorded in a realistic acoustic environment.

The focus of Chap. 9 is on issues relating to the evaluation of blind source separation algorithms such as recording setups, performance measures, generation of synthetic data, and data mixing. The last chapter summarizes the main contributions of the book, discusses a number of open issues, and presents suggestions for future research. This reviewer feels that future research should emphasize evaluation of algorithms in real-time dynamic situations that include time-varying multipaths and source motion. In addition to ten chapters, the book also has six appendices entitled ''A—Efficient Beam forming,'' ''C—Computational Complexity of CoBliSS,'' ''B—On the Indeterminacies of Convolutive Blind Signal Separation Based on Decorrelation,'' ''D—Efficient Computation of Windowed FFT's,'' ''E—Contents of the Compact Disc,'' and ''F—Glossary.'' Included with the book is a compact disc that contains simulation results that have been described in the text. It also contains MATLAB implementation of new algorithms along with audio data for experimentation.

This book is clearly written and well organized. The author has spent considerable effort to present material in a clear and concise manner. This book can be recommended for readers who are interested in a quick overview of basic concepts and acoustic application of blind signal separation and echo cancellation, and for the readers who are familiar with these applications, but who want to understand and use new algorithms that have been presented in the book. The algorithms are based on the author's original work and constitute a valuable addition to blind signal separation and multichannel echo cancellation technology. Readers who are not experts in digital signal processing, in adaptive signal processing, or in array processing should acquire the necessary background to fully benefit from this book. This book is definitely too concise to be a first text on adaptive signal processing, but it is a good supplementary text for knowledgeable readers who are interested in specific acoustic applications.

LEON H. SIBUL
Applied Research Laboratory
The Pennsylvania State University
University Park, Pennsylvania 16802

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ALIREZA DIBAZAR, *Department of BioMed Engineering, University of Southern California, Los Angeles, California 90089*
HASSAN NAMARVAR, *Department of BioMed Engineering, University of Southern California, Los Angeles, California 90089*
DAVID PREVES, *Micro-Tech Hearing Instruments, 3500 Holly Lane No., Suite 10, Plymouth, Minnesota 55447*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
CARL J. ROSENBERG, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
WILLIAM THOMPSON, JR., *Pennsylvania State University, University Park, Pennsylvania 16802*

6,341,661

43.30.Yj BOW DOME SONAR

Ernest Theodore Bick *et al.*, assignors to L3 Communications Corporation
29 January 2002 (Class 181/110); filed 17 May 2000

This patent describes a particular arrangement of both transmitting and receiving transducer elements that produces a compact, modular, bow-mounted, sonar assembly. The transmitting elements are preferably flextensional transducers. The cited advantages of this design are good performance at a low cost and with a reduction in size and weight compared to other bow dome sonars.—WT

6,346,126

43.35.Ty ACOUSTIC-ENERGY-ASSISTED REMOVAL OF SOIL FROM FABRIC IN A GASEOUS ENVIRONMENT

Sidney C. Chao and Nelson W. Sorbo, assignors to Raytheon Company
12 February 2002 (Class 8/149.2); filed 2 December 1999

This involves a sort of ''dry-cleaning'' method whereby a soiled piece of fabric is cleaned by mechanical agitation while the fabric is simultaneously subjected to acoustic energy in a gaseous environment. Mechanical action may be provided either by gas jet action and/or by tumbling. The acoustic energy, which may lie in the audible or ultrasonic range, vibrates the fibers of the fabric to enhance the cleaning action. The source of the acoustic energy is a driver driven by a function generator.—DRR

6,339,719

43.35.Yb APPARATUS AND METHOD FOR DETECTING SOUNDS FROM A HUMAN BODY AND DISPLAYING THE DETECTED SOUND INFORMATION

Myun Woo Lee *et al.*, assignors to Hitouch, Incorporated
15 January 2002 (Class 600/511); filed in the Republic of Korea 2 February 1998

The apparatus consists of a stethoscope for collecting sounds, a piezoelectric device for converting the collected sounds into an electric signal, a speaker to convert the electric signal into audible sound, an A/D converter for digitizing the electric signal, and a memory for storing the digitized sound data. In addition, a microprocessor is included to generate the switch control signal, analyze the extracted principal parameters statistically, store the analyzed statistical figures in the memory, and read out the statistical features from the memory. A LCD panel provides the signal displays.—DRR

6,340,348

43.35.Yb CONTRAST AGENT IMAGING WITH DESTRUCTION PULSES IN DIAGNOSTIC MEDICAL ULTRASOUND

Sriram Krishnan *et al.*, assignors to Acuson Corporation
22 January 2002 (Class 600/447); filed 2 July 1999

The subject device provides diagnostic medical ultrasound imaging. In the preferred embodiment, high pulse repetition frequency (HPRF) destruction pulses are fired at a rate higher than necessary for receiving returning echoes. The first or second pulse provides imaging but subsequent pulses can destroy the contrast agent, thereby providing further physiological information to be compared with the contrast-agent-influenced images obtained initially. Pulse parameters can also be changed between the pluralities of contrast-agent-destroying pulses. A timer can trigger destruction frames that consist of a plurality of destruction pulses or some fixed part of a physiological signal, such as an ECG signal.—DRR

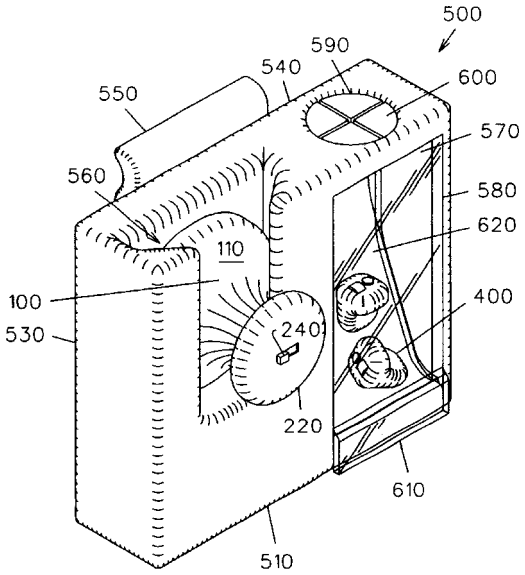
6,340,350

43.35.Yb TRANSMITTER/RECEIVER STETHOSCOPE AND HOLDER THEREFOR

Juanita P. Simms, Rantoul, Kansas
22 January 2002 (Class 600/528); filed 6 January 2000

An electronic stethoscope and holder consists of a chestpiece, an earpiece, and a case for holding these parts. The chestpiece constitutes a sound sensor for detecting auscultatory sounds and converting them into an

 0001-4966/2002/112(2)/363/9/$19.00



electrical signal. The chestpiece also contains a radio wave transmitter for the electrical signal. The earpiece includes a receiver for the transmitted electrical signal and a means of converting the signal into audible form.—DRR

6,343,512

43.35.Yb ULTRASOUND PROBE INCLUDING A HYDROPHILIC COUPLANT

Simon Bourne *et al.*, assignors to Cranfield University
5 February 2002 (Class 73/644); filed in the United Kingdom 15 September 1998

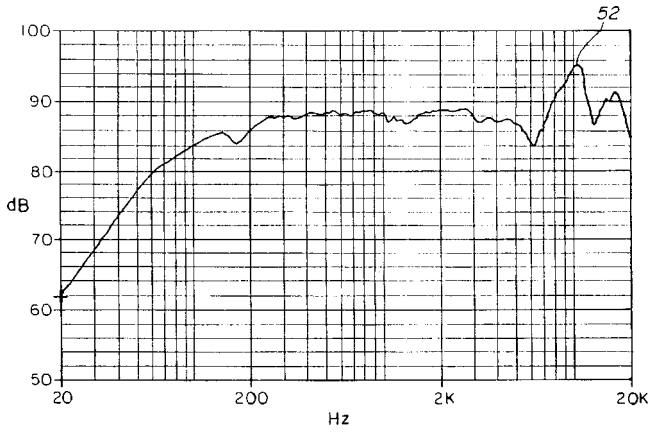
Couplants are materials that overcome the mismatch at the air interface between transducers (both generative and receiving) and the specimen. The ultrasound probe described in the patent includes the usual components for generating ultrasound, means of transmitting ultrasound from a remote source, and a cross-linked hydrophilic material as an integral couplant. The hydrophilic material is capable of transmitting a frequency in the range of 5 to 20 MHz and exhibits attenuation of transmission less than 1.5, and preferably less than 1 dB/mm at 5 MHz. The hydrophilic material can be integrated into the probe or supplied independently so as to provide a choice of impedances to allow matching with the test structure.—DRR

6,327,372

43.38.Ja CERAMIC METAL MATRIX DIAPHRAGM FOR LOUDSPEAKERS

Allan O. Devantier and An D. Nguyen, assignors to Harman International Industries, Incorporated
4 December 2001 (Class 381/426); filed 5 January 1999

At one time or another, loudspeaker cones and domes have been made from metal, shellac-impregnated fabric, balsa wood, polystyrene foam, injection-molded plastic, carbon fiber, aerogel, and felted paper. A longer, complete list would be augmented by various coatings and layered configurations. This patent describes a ceramic/metal/ceramic sandwich that is easily fabricated by anodizing both sides of an aluminum cone or dome. Stated



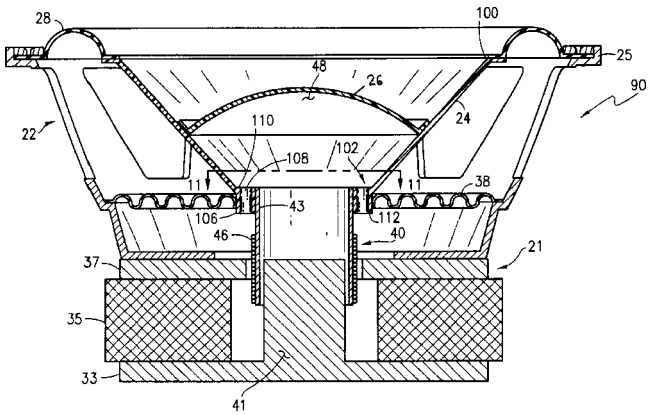
benefits include consistent performance over a wide range of temperature and humidity, as well as desirable acoustical properties. On-axis response of a 100-mm ceramic metal matrix loudspeaker shows its first major cone resonance **52** at about 10 kHz. For a comparable aluminum cone speaker the resonance occurs at about 7 kHz.—GLA

6,330,340

43.38.Ja LOUDSPEAKER WITH A DIAPHRAGM HAVING INTEGRAL VENT BORES

Lucio Proni, assignor to JL Audio, Incorporated
11 December 2001 (Class 381/397); filed 30 August 2000

The invention is an improved version of earlier United States patents 5,734,734 and 6,118,884. The object is to provide more efficient cooling, and correspondingly greater power ratings, for loudspeakers with relatively



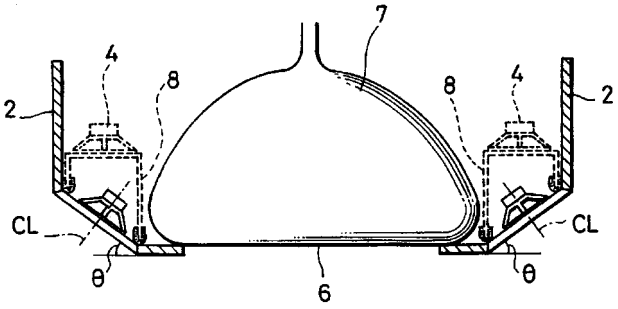
small voice coil diameters. Air trapped under center dome **26** is vented through cooling openings **108** into the undesignated cavity behind centering spider **38**, which is presumably porous.—GLA

6,335,974

43.38.Ja SPEAKER SYSTEM FOR TELEVISION RECEIVER WITH SOUND DUCTS AND PERFORATED PANELS

Isao Kunimoto, assignor to Sony Corporation
1 January 2002 (Class 381/306); filed in Japan 19 March 1997

Numerous earlier patents address the problem of providing good stereo TV sound while keeping cabinet width to a minimum. In this Sony design, tweeters are mounted on angled panels. Below them, woofers **4** sit on the



backs of waveguides **8** which are sealed against, but mechanically isolated from, cabinet **2**. Sound produced by each woofer travels forward through the waveguide and emerges from a simple perforated grill. The latter is intended to change the direction of the sound wave by diffraction. A quick review of Huygens' principle will show that such is not the case.—GLA

6,344,940

43.38.Md METHOD AND APPARATUS FOR ADDING 1/f FLUCTUATION TO AN AUDIO SIGNAL

Yoichiro Sako and Heitaro Nakajima, assignors to Sony Corporation
5 February 2002 (Class 360/32); filed in Japan 29 November 1996

The purpose of the device is to enhance digital signal reproduction by adding the 1/f fluctuation characteristic in the temporal and amplitude directions of digitized audio and video data. The digital signal recording apparatus includes a means for generating a 1/f fluctuation, a means for converting an analog signal into a digital signal while controlling a generalized 1/f fluctuation in the A/D conversion characteristic, and a system for recording the digital signal.—DRR

6,337,908

43.38.Si TELEPHONE HANDSET, CORDLESS TELEPHONE OR MOBILE TELEPHONE FOR IMPROVING ACOUSTIC PROBLEMS IN A FREQUENCY RANGE

Morten Kjeldsen Andersen, assignor to Kirk Acoustic A/S
8 January 2002 (Class 379/433.02); filed in Denmark 28 November 1996

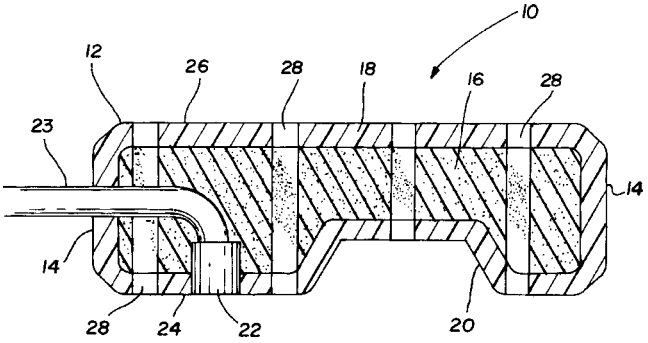
Miniaturized mobile telephones are easy to carry but too small to operate efficiently as handsets. Low frequencies can be severely attenuated because of air leaks between the transducer and the user's ear. To counteract this problem the invention provides an internal cavity and several additional sound apertures (which may be shaped to form the manufacturer's logo).—GLA

6,337,915

43.38.Si EARPHONE

Michael Lewis, Mt. Clear, Victoria 3350, Australia
8 January 2002 (Class 381/380); filed 7 April 2000

The invention is intended to be a comfortable, insert-type earphone for use with personal stereo systems. A spongy, thimble-shaped inner core **16** is encased in a plastic skin **14**. The assembly houses speaker **22**, which appears



to be no larger than a grain of rice. "A plurality of bores through the housing are adapted for equalizing air pressure between the atmosphere and the ear of the user."—GLA

6,330,486

43.38.Vk ACOUSTIC PERSPECTIVE IN A VIRTUAL THREE-DIMENSIONAL ENVIRONMENT

Thomas J. Padula, assignor to Silicon Graphics, Incorporated
11 December 2001 (Class 700/94); filed 16 July 1997

In a virtual reality environment, the apparent position of a given sound source must match the visual impression of the scene. In effect, camera pan, tilt, and zoom must be linked to corresponding pan, tilt, and zoom of the three-dimensional sound environment. The patent describes an efficient computer algorithm for achieving this goal, even when sound sources lie outside the field of vision. "By identifying the observer's field of view and the observer's perceived azimuths, the present invention ensures that the rendering engine renders audio signals emitted by audio sources as if the sources were at the locations seen by the observer."—GLA

6,332,026

43.38.Vk BASS MANAGEMENT SYSTEM FOR HOME THEATER EQUIPMENT

Juha Kuusama *et al.*, assignors to Flextronics Design, Finland Oy
18 December 2001 (Class 381/17); filed in Finland 6 August 1996

Most surround-sound formats include a separate subwoofer channel. In a home theater system, the same subwoofer may be called upon to reproduce low-frequency information from other channels as well. Allowing for various combinations of loudspeakers and loudspeaker locations requires a complicated assortment of adjustable delays, filters, and summing amplifiers. Trying to perform all of these functions in the digital domain is discouraging because of the wide range of possible signal levels in the various channels. The inventors propose a system in which filters and delays are digital but summing is performed in the analog domain. "By doing this the flexibility of a digital system is maintained and the dynamic range of DA-converters can be fully exploited."—GLA

6,345,688

43.55.Ev METHOD AND APPARATUS FOR ABSORBING SOUND

Gerald R. Veen and Marvin R. Mealman, assignors to Johnson Controls Technology Company
12 February 2002 (Class 181/290); filed 23 November 1999

In this process, the heat and pressure applied to a glass fiber pad cause the fibers to create perforations in a film facing. In this manner the flow

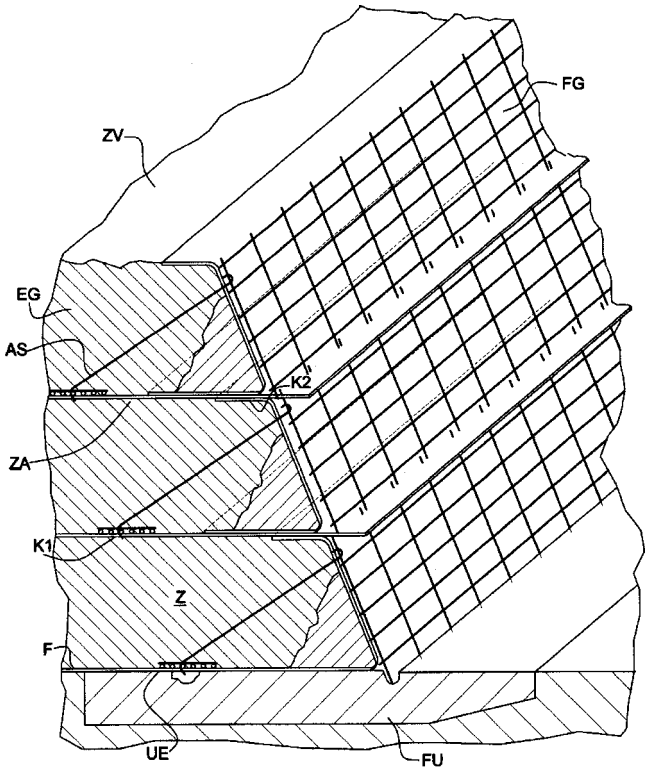
resistance of the film—and hence the absorptive or attenuation properties of the pad—can be adjusted for desired acoustical properties. The pad could be used as a trim panel in an automobile.—CJR

6,341,922

43.55.Ti CELLULAR CONSTRUCTION, IN PARTICULAR SUPPORTING OR SOUND INSULATING CONSTRUCTION CAPABLE OF BEING GREENED, AND PROCESS FOR PRODUCING THE SAME

Felix Paul Jaecklin, assignor to Felix P. Jaecklin
29 January 2002 (Class 405/280); filed 17 June 1999

The invention incorporates a simple and inexpensive way to distribute tension bracing across the face of a stackable series of modules. Together



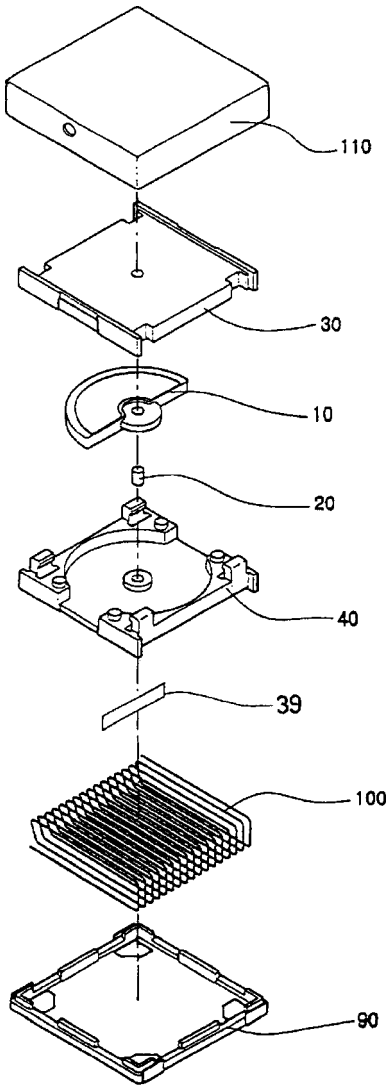
these form a layered cellular structure that can be filled with greenery, earth, or bulk material to create a sound barrier wall.—CJR

6,342,742

43.58.Wc VIBRATOR

Jong-Kyu Kim, assignor to Won-II Communics Company, Limited
29 January 2002 (Class 310/81); filed in the Republic of Korea 4 January 2000

A signalling vibrator is disclosed which includes a rotary member having a center plate and an arc-shaped, unipolar magnet positioned eccentrically on the outer surface of the center plate. The configuration of this



vibrator is intended to minimize its volume and to reduce power requirements for use in cell phones and mobile equipment.—DRR

6,342,776

43.58.Wc ACOUSTIC CHARGER FOR ELECTRONIC DEVICES

Elizabeth B. Taylor *et al.*, assignors to Motorola, Incorporated
29 January 2002 (Class 320/137); filed 17 January 2001

The device described in this patent charges the battery of an electronic device by converting acoustic energy into electric current. In one embodiment, the protective layer of a liquid crystal display screen includes a magnetic material. The housing of the liquid crystal display includes a corresponding coil of wire. The protective layer has freedom of motion allowed by the liquid crystal panel, so when acoustic energy is incident upon that layer, it moves, causing the magnetic material to generate a changing magnetic field in the coil. In another embodiment, the protective layer is connected to piezoelectric transducers, actuated when acoustic energy is incident upon the protective layer. In a third embodiment, the electronic device includes an input for receiving concentrated energy from various sources including loudspeakers and waveguides.—DRR

6,341,166

43.60.Gk AUTOMATIC CORRECTION OF POWER SPECTRAL BALANCE IN AUDIO SOURCE MATERIAL

Peter L. Basel, assignor to LSI Logic Corporation
22 January 2002 (Class 381/103); filed 12 March 1997

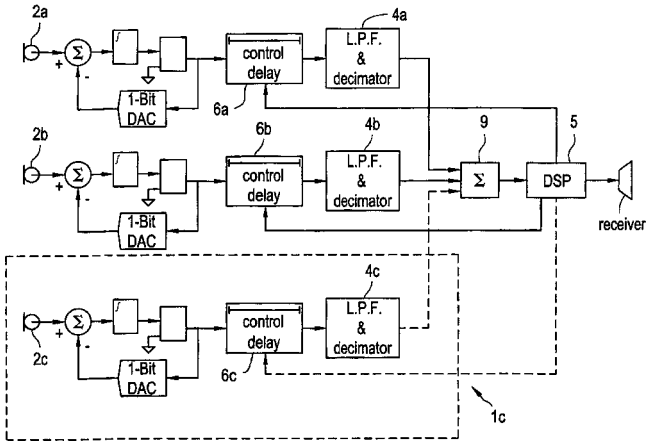
The patented object is an audio system that automatically corrects for variations in spectral balance in audio source material. The system consists of a power spectrum analyzer and a source correction equalizer unit coupled to a spectral balance correction data memory. A reference spectral balance may be initially established by playing a reference source with the desired sound attributes. The power spectrum analyzer establishes the spectral balance of the reference material, storing the results in a nonvolatile memory. When another source is played, an equalizer-setting computing unit compares the spectral balance of the reference and that of the source and then calculates the equalizer correction setting, which is then stored in the memory.—DRR

6,339,647

43.66.Ts HEARING AID WITH BEAM FORMING PROPERTIES

Henning Hougaard Andersen and Carl Ludvigsen, assignors to Topholm & Westermann APS
15 January 2002 (Class 381/312); filed 26 February 2001

Variation of the polar directivity pattern is achieved with at least two microphones attached to separate analog-to-digital converters and at least one controllable very short delay. A very fast sampling rate is used with a



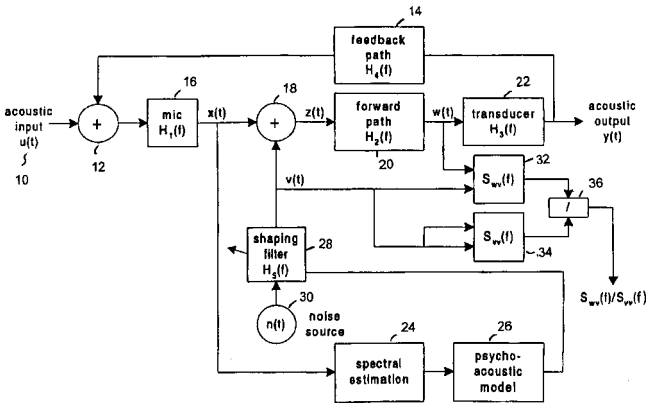
1-bit sigma-delta A-to-D converter to achieve fine control of the beamforming characteristic via a digital signal processor making fine adjustments to the delay time.—DAP

6,347,148

43.66.Ts METHOD AND APPARATUS FOR FEEDBACK REDUCTION IN ACOUSTIC SYSTEMS, PARTICULARLY IN HEARING AIDS

Robert Brennan and Anthony Todd Schneider, assignors to Dspfactory Limited
12 February 2002 (Class 381/318); filed 16 April 1998

A control signal is formed by making a spectral estimate of a hearing aid input signal and feeding that estimate to an auditory system model of the hearing aid wearer. A noise probe signal is used to continuously characterize the acoustic feedback path. The noise signal is spectrally shaped by the



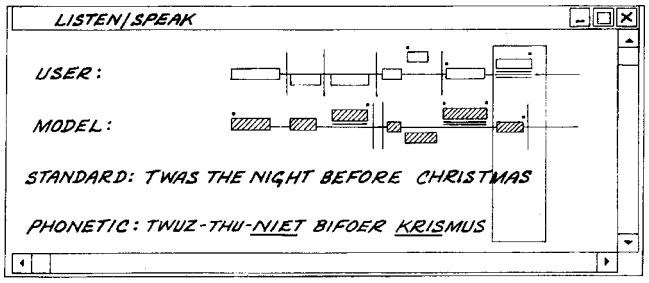
control signal, making it inaudible to the hearing aid wearer. The shaped noise is added to the input signal. The resultant combined signal is processed and amplified and is then analyzed to detect the presence of acoustical feedback. The forward path of the system is modified to attenuate the signal at the frequencies at which feedback is detected.—DAP

6,336,089

43.70.Kv INTERACTIVE DIGITAL PHONETIC CAPTIONING PROGRAM

Michael Everding, Venice, California
1 January 2002 (Class 704/1); filed 20 August 1998

This language-learning workstation assists the user by presenting a comparative display of a model and the user's pronunciation, presented in a form which highlights the pitch, amplitude, and duration information. The



displays may be separated, as shown in the figure, or overlapped for more detailed analysis. Facial and x-ray-style section views of the head are also available.—DLR

6,341,958

43.71.Hw METHOD AND SYSTEM FOR ACQUIRING A FOREIGN LANGUAGE

Arkady G. Zilberman, Powell, Ohio
29 January 2002 (Class 434/157); filed 7 February 2001

One of the typical impediments for an adult in learning a foreign language is the natural tendency to translate the foreign material into the speaker's native language, rather than learning to think directly in the target langauge. Using a series of reading and listening drills, this device trains the user to experience the materials directly and rapidly, not giving an opportunity to think explicitly about the content or to mentally rehearse in a way which might encourage mental translation.—DLR

6,334,777

43.71.Ky METHOD FOR ADAPTIVELY TRAINING HUMANS TO DISCRIMINATE BETWEEN FREQUENCY SWEEPS COMMON IN SPOKEN LANGUAGE

William M. Jenkins *et al.*, assignors to Scientific Learning Corporation
1 January 2002 (Class 434/185); filed 24 June 2000

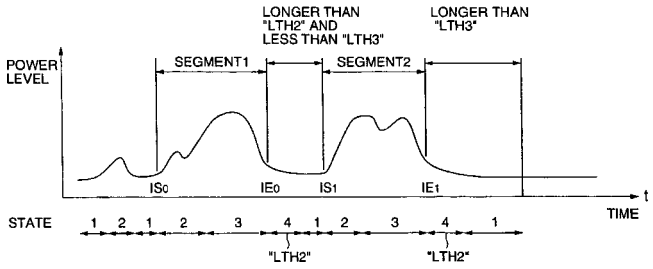
The flow continues of patents produced by the present assignee which cover devices to help the recovery from speech perception disorders. The general pattern addressed here, as in previous patents, is the perception of frequency sweeps, such as occur in formant transitions. Here, the listener must decide whether a sweep moves up or down in frequency. Upon successes, the sweep rate is gradually increased.—DLR

6,317,711

43.72.Ar SPEECH SEGMENT DETECTION AND WORD RECOGNITION

Tetsuya Muroi, assignor to Ricoh Company, Limited
13 November 2001 (Class 704/253); filed in Japan 25 February 1999

This system for segmenting a speech signal measures frame power levels (often referred to as energy levels) and uses a decision tree analysis



procedure to assign the speech end points. One curious feature is the use of power minima, rather than maxima, to set normalization levels.—DLR

6,335,962

43.72.Ar APPARATUS AND METHOD FOR GROUPING AND PRIORITIZING VOICE MESSAGES FOR CONVENIENT PLAYBACK

Syed S. Ali *et al.*, assignors to Lucent Technologies, Incorporated
1 January 2002 (Class 379/88.11); filed 27 March 1998

This device allows you to listen to your recorded phone messages in a sequence other than the way they were recorded. A processor classifies incoming messages according to speaker, language, or topic, and also allows various users of the system, when calling in to retrieve messages, to specify the presentation sequence. In this case, any of several users of a system may also be identified by their voice patterns.—DLR

6,330,023

43.72.Ew VIDEO SIGNAL PROCESSING SYSTEMS AND METHODS UTILIZING AUTOMATED SPEECH ANALYSIS

Tsuhan Chen, assignor to American Telephone and Telegraph Corporation
11 December 2001 (Class 348/14.13); filed 18 March 1994

Certain video processing may involve the deletion of occasional frames in order to reduce the bitrate or to speed up the presentation. In these cases, it is desirable to adjust the rate of speech on the sound track to match the video. However, video frames with useful articulation information may have been deleted. This system uses the available information and attempts to smooth across any deleted regions in order to match video data with the speech signal.—DLR

6,335,973

43.72.Ew SYSTEM AND METHOD FOR IMPROVING CLARITY OF AUDIO SYSTEMS

Eliot M. Case, assignor to Qwest Communications International Incorporated
1 January 2002 (Class 381/61); filed 23 October 1997

This patent covers a system and method to enhance the quality, clarity, and intelligibility of hearing aids, telephone systems, data compressed low bit rate audio, and other audio systems by adding even harmonic distortion to selected frequencies of the audio signals.—HHN

6,338,038

43.72.Ew VARIABLE SPEED AUDIO PLAYBACK IN SPEECH RECOGNITION PROOFREADER

Gary Robert Hanson, assignor to International Business Machines Corporation
8 January 2002 (Class 704/500); filed 2 September 1998

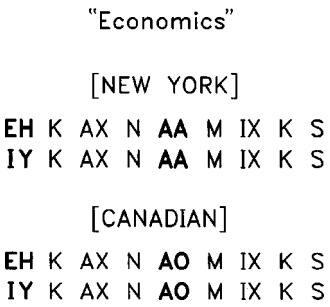
This system allows speech to be played back slowly, basically, a word at a time, for the purpose of audio proofchecking of the recorded material. In setting up a playback, the user specifies a desired delay to be inserted between words. Upon playback, each word is spoken at the normal rate with the programmed delay occurring only between words.—DLR

6,343,270

43.72.Ja METHOD FOR INCREASING DIALECT PRECISION AND USABILITY IN SPEECH RECOGNITION AND TEXT-TO-SPEECH SYSTEMS

Lalit R. Bahl and Paul S. Cohen, assignors to International Business Machines Corporation
29 January 2002 (Class 704/257); filed 9 December 1998

This phonetic transcription system, primarily intended for speech synthesizers, uses a method of encoding the phonetic structure in a way that allows the reconstruction, by rules, of any of a variety of dialect patterns in



the phonetic representation. Coded baseforms of words and phrases are included in a system lexicon. Sets of rules are provided which allow the generation of a specific dialect in the phonetic stream for use by the synthesizer.—DLR

6,347,298

43.72.Ja COMPUTER APPARATUS FOR TEXT-TO-SPEECH SYNTHESIZER DICTIONARY REDUCTION

Anthony J. Vitale *et al.*, assignors to Compaq Computer Corporation
12 February 2002 (Class 704/260); filed 26 February 2001

This patent describes a fairly obvious procedure for eliminating those words from a phonetic dictionary which can be precisely reproduced using a given set of rules. The idea was fully worked out in the 1970s or earlier. The rules would allow reconstruction of a phonetic string, either directly or by morphemic decomposition of the spelled word. Any words thus reconstructable are then deleted from the dictionary.—DLR

6,317,712

43.72.Ne METHOD OF PHONETIC MODELING USING ACOUSTIC DECISION TREE

Yu-Hung Kao and Kazuhiro Kondo, assignors to Texas Instruments Incorporated
13 November 2001 (Class 704/256); filed 21 January 1999

This speech recognizer constructs phonetic models by using a decision tree to analyze triphone sequences. The phonetic segments of each triphone are structured according to traditional phonetic features. These features are then used to guide the decision tree to form clusters as the training data is processed.—DLR

6,321,196

43.72.Ne PHONETIC SPELLING FOR SPEECH RECOGNITION

Carlos Antonio Franceschi, assignor to International Business Machines Corporation
20 November 2001 (Class 704/243); filed 2 July 1999

For improved handling of spelled input, this speech recognition system has a spelling input mode distinct from the normal word recognition mode. When training the system for spelling mode, the user may choose any word or may list a set of word choices for use as the cue to each letter.—DLR

6,334,102

43.72.Ne METHOD OF ADDING VOCABULARY TO A SPEECH RECOGNITION SYSTEM

James R. Lewis and Kerry A. Ortega, assignors to International Business Machines Corporation
25 December 2001 (Class 704/255); filed 13 September 1999

This patent describes a method for use in an automatic speech recognition system for adding a new vocabulary item by making use of an existing language model. The speech recognition system automatically creates language model statistics for any word not in the current vocabulary by locating another, similar word already in the vocabulary. The frequency characteristics of the similar word are then copied to provide language model statistics for the new word.—HHN

6,337,899

43.72.Ne SPEAKER VERIFICATION FOR AUTHORIZING UPDATES TO USER SUBSCRIPTION SERVICE RECEIVED BY INTERNET SERVICE PROVIDER (ISP) USING AN INTELLIGENT PERIPHERAL (IP) IN AN ADVANCED INTELLIGENT NETWORK (AIN)

Tommy Alcendor *et al.*, assignors to International Business Machines Corporation
8 January 2002 (Class 379/88.17); filed 31 March 1998

Certain computer networks which are enhanced by the addition of technically advanced human interface capabilities are described in Bellcore telephony standards as automated intelligent networks (AINs). The AIN described in this patent uses voice recognition to maintain customer access via the Internet to various subscription services offered by an Internet service provider. The voice interface protocols are described in some detail.—DLR

6,345,254

43.72.Ne METHOD AND APPARATUS FOR IMPROVING SPEECH COMMAND RECOGNITION ACCURACY USING EVENT-BASED CONSTRAINTS

James R. Lewis *et al.*, assignors to International Business Machines Corporation
5 February 2002 (Class 704/275); filed 29 May 1999

This patent provides a system for improving the recognition accuracy of speech recognition systems for spoken software-control commands. Since commands are typically one or two words, syntactic language models are generally ineffective. Instead, the context in which a command is spoken can be used to improve performance. This system gathers information on current, ongoing events and uses that information as a command context. The system includes an efficient arrangement for monitoring the current states of the voice-controlled programs currently running in the computer.—AAD

6,347,296

43.72.Ne CORRECTING SPEECH RECOGNITION WITHOUT FIRST PRESENTING ALTERNATIVES

Steven J. Friedland, assignor to International Business Machines Corporation
12 February 2002 (Class 704/231); filed 23 June 1999

This patent discusses some of the problems and issues involved in correcting text documents using a computer editor with voice interaction for

command and text inputs. The emphasis is on managing blocks of text and, in particular, how alternate replacement text portions can be quickly and conveniently selected and tried on a temporary basis.—DLR

6,347,297

43.72.Ne MATRIX QUANTIZATION WITH VECTOR QUANTIZATION ERROR COMPENSATION AND NEURAL NETWORK POSTPROCESSING FOR ROBUST SPEECH RECOGNITION

Safdar M. Asghar and Lin Cong, assignors to Legerity, Incorporated
12 February 2002 (Class 704/243); filed 5 October 1998

This patent introduces a hybrid speech recognition system which combines matrix quantization (MQ), vector quantization (VQ), hidden Markov models (HMMs), and neural network technology. In order to achieve robust speech recognition, the spectral envelope of the speech signal is first quantized using MQ and VQ and then HMM-based classifiers are applied. The stochastic output data of the HMMs serve as the input to the neural network. According to this patent, the combination of two-stage classification and MQ provides robust recognition and enhances accuracy.—AAD

6,347,299

43.72.Ne SYSTEM FOR NAVIGATION AND EDITING OF ELECTRONIC RECORDS THROUGH SPEECH AND AUDIO

Thomas G. Holzman *et al.*, assignors to NCR Corporation
12 February 2002 (Class 704/270); filed 29 July 1998

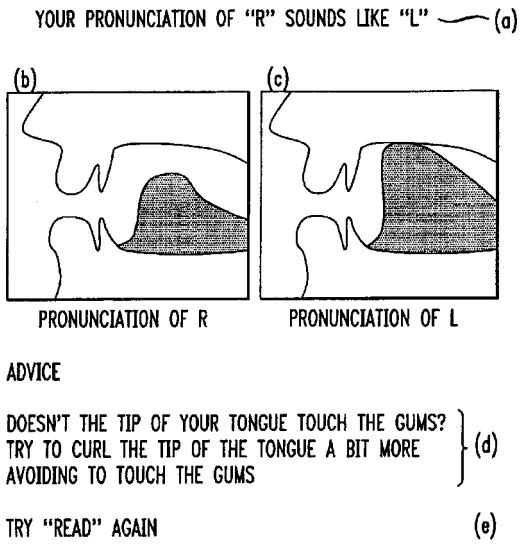
This system for voice interactive database management includes several ideas intended to make the system ''user friendly,'' by which is meant that the user actively selects the region of interest and describes the intended data contents for insertion or editing, rather than having the system proceed through a fixed sequence of actions. Audio feedback of the current database status may include tones and beeps, as well as speech. Key words and phrases, both input and output, are, in general, easily reprogrammable from one setting to another.—DLR

6,347,300

43.72.Ne SPEECH CORRECTION APPARATUS AND METHOD

Ayako Minematsu, assignor to International Business Machines Corporation
12 February 2002 (Class 704/270); filed in Japan 17 November 1997

This cross-language speech recognizer provides special capabilities for dealing with foreign language accents. In a normal recognition mode, the system tests for certain highly probable errors which typically result from native-language interference for a particular speaker. The example given is the difficulty Japanese speakers frequently have when trying to learn the



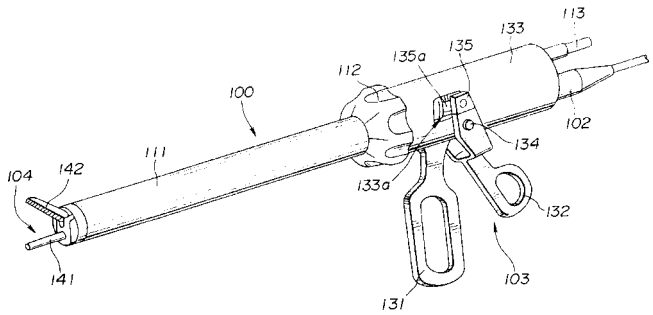
English l-r distinction. The authors assert that, in a training mode, the system would be able to provide detailed feedback to the user by providing a measure of the pronunciation accuracy, including an ''x-ray'' view of the tongue. It is not clear whether this would be done based on a detailed analysis of the user's pronunciation or whether the system would merely respond to a false phonemic classification of the input phonetic quality. There seems to be no provision for a more detailed presentation of the actual phonetic difficulties, such as the lateral articulation issue raised by this particular example.—DLR

6,340,352

43.80.Sh ULTRASOUND TREATMENT SYSTEM

Mitsumasa Okada *et al.*, assignors to Olympus Optical Company, Limited
22 January 2002 (Class 601/2); filed in Japan 6 April 1995

The patent describes an ultrasound treatment system for coagulating or resecting the region of a living tissue to be treated. The system includes transducers for generating ultrasonic signals. These are housed within a handpiece that also contains a probe connected to the transducers and serves as a vibration conveyor to a distal member that constitutes a stationary



section of a treatment unit used to treat living tissue. A protective sheath covers the probe and a movable holding unit for clamping the living tissue. A means of manipulation is also provided for clamping or freeing the tissue and to rotate the treatment unit with the axial direction of the ultrasonic transducers as a center.—DRR

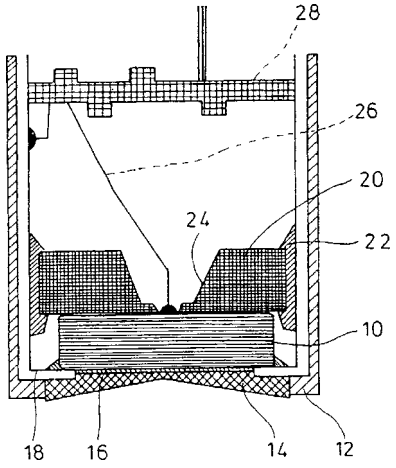
6,344,024

43.80.Sh MULTIFREQUENCY ULTRASOUND PROBE

Rainer Brucher *et al.*, assignors to DWL Elektronische Systeme GmbH

5 February 2002 (Class 600/459); filed in Germany 12 May 1997

The device is an ultrasonic probe intended for use in medical diagnosis via ultrasonography. The probe includes an ultrasound sensor configured to generate an electrically triggered, ultrasound send signal and to receive a signal reflected from an observed medium. The ultrasound sensor is held in a supporting device and is fitted with coupling and transmission devices to relay the transmission signal into the observation medium. The transmission signal is a multifrequency signal consisting of at least two different frequencies. The probe also contains a means for evaluating the reflected signal in relation to the separate signals.—DRR



Three-dimensional propagation effects near the mid-Atlantic Bight shelf break (L)

Kevin B. Smith, Chris W. Miller, and Anthony F. D'Agostino
Naval Postgraduate School, Monterey, California 93943

Brian Sperry
Woods Hole Oceanographic Institute, Woods Hole, Massachusetts 02543

James H. Miller and Gopu R. Potty
University of Rhode Island, Narragansett, Rhode Island 02882

(Received 14 February 2002; accepted for publication 7 May 2002)

Significant three-dimensional (3-D) environmental variability exists in the vicinity of the shelf break along the mid-Atlantic Bight. This study examines the influence of azimuthal coupling due to this variability on acoustic propagation in this region. Numerical studies employing a 3-D ray code, a hybrid ray-mode code, and a 3-D parabolic equation model are used to study the significance of azimuthal coupling on various propagation paths. These paths include up-slope, slant-slope, and cross-slope propagation. The numerical analysis suggests that, for the propagation ranges less than 60 km examined, the influence of azimuthal coupling is negligible compared to the inherent uncertainty in the environment itself. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1490559]

PACS numbers: 43.30.Dr, 43.30.Re, 43.30.Zk [SAC-B]

I. INTRODUCTION

Continental slopes represent some of the steepest large-scale bathymetric features in the world's oceans, averaging 2–3 degrees and extending for several miles out from the continental shelf. The boundary between these regions and the shallow, littoral zones on the shelf is the shelf break, typically an area of significant oceanographic variability. Such three-dimensional environmental variability exists in the vicinity of the shelf break along the mid-Atlantic Bight and is characterized by relatively strong bathymetric slopes, cold shelf water interacting with warm Gulf Stream waters, and diurnal generation of internal solitary waves onto the shelf. The influence of such a highly variable environment on acoustic propagation is of great interest as is the impact on inversion techniques such as tomography and localization, and sonar system performance estimations.

In the summer of 1996, the Office of Naval Research sponsored an acoustic field study, called the Shelf Break Primer, off the coast of New Jersey near the mid-Atlantic Bight. Details of the experimental design and data collection are provided by Lynch *et al.* (1997).

This letter provides an overview of several different numerical propagation studies performed for this region. Specifically, each study examined the predicted influence of azimuthal coupling due to the three-dimensional character of the environment on acoustic propagation. A 3-D ray code was used to estimate horizontal refraction of propagation paths and differences in arrival times when contrasted with simple 2-D ray calculations. A hybrid ray-mode code was also used to investigate the influence of the refracted paths on mode amplitude and phase. The modes were assumed to be adiabatic in this treatment. Finally, a fully coupled 3-D parabolic equation model was used to investigate travel time differences and to determine the dominant mechanism for 3-D effects.

II. ENVIRONMENTAL/EXPERIMENTAL DESCRIPTION

The primary assets used in the experiment consisted of three tomography arrays with sources moored near the bottom. All of these sources were deployed on the slope about 30 km seaward of the shelf break at depths of 300–350 m. On the shelf were deployed two, 16-element vertical line arrays (VLAs). The depth of the lowest array element was 3 m off the seafloor with 3.5-m element separation for a total aperture of 52.5 m. The configuration of the sources and VLAs relative to the local bathymetry based on standard USGS contours is displayed in Fig. 1.

In addition to the acoustics, several assets were continuously devoted to environmental assessment. These included a string of thermistors located near the southwest corner of the experiment and a nearly continuous towing of a high-resolution SeaSoar (Gawarkiewicz *et al.*, 2001). This device was able to measure temperature, salinity, and various other water properties as a function of depth. During the deployment, four cross-shelf transects, roughly 45 km in length, were conducted daily. Each transect was separated by 10 km in the along-shelf direction to resolve the three-dimensional structure of the shelf break region. An example sound speed profile generated from such data is displayed in Fig. 2. Note the cold ''tongue'' of shelf water extending out over the shelf break. Gradients as large as 2 s^{-1} can be seen at the boundaries of this protrusion.

III. NUMERICAL RESULTS

Several different source/receiver tracks were investigated in the analysis, including two directly upslope onto the shelf (SW2NW, SE2NE) and two crossing diagonally across the shelf (SW2NE, SE2NW).

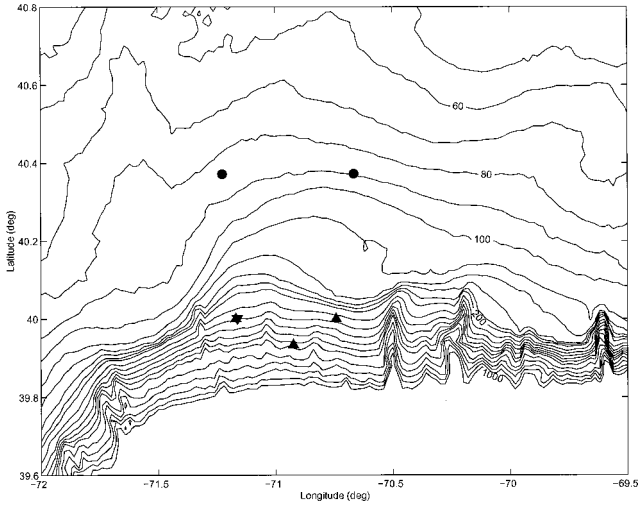


FIG. 1. Bathymetry and location of deployed acoustic elements for Primer experiment. The circles indicate the locations of the vertical line arrays (VLAs), the upright triangles indicate the locations of the 400-Hz tomography sources, and the upside-down triangle indicates the location of the 240-Hz tomography source (co-located with the southwest 400-Hz source).

The majority of the ray modeling work was done prior to the actual experiment in an attempt to provide some *a priori* knowledge of the expected influence of 3-D propagation (D'Agostino, 1996). Therefore, the environmental characterization, specifically the sound speed profiles, was based on empirical knowledge of shelf break regions. The bottom was treated as a simple reflector. To account for the critical angle effects, no rays which interacted with the bottom at any point along the path with grazing angles greater than ~25 degrees were kept. This led to ray vertical launch angles between −7 and +7 degrees with a chosen resolution of 0.01 degree.

The 3-D ray code used was a modified version of HARPO adapted to compute the necessary properties of eigenrays (see D'Agostino, 1996). The eigenray pressure magnitudes, phases, and travel times, including the effects of the boundary reflections, ray turning, and raytube spreading, were then computed.

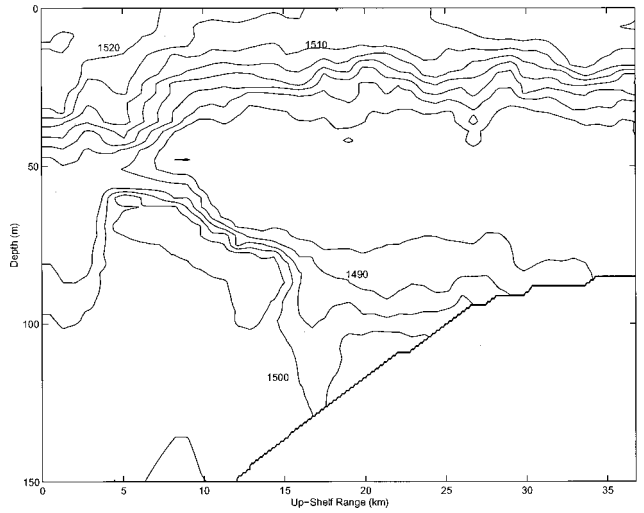


FIG. 2. Two-dimensional slice of sound speed profiles as computed from temperature and salinity measurements made by the SeaSoar.

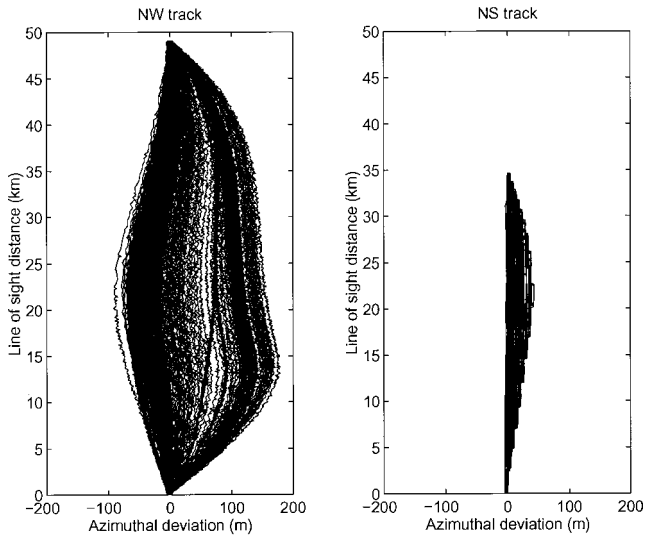


FIG. 3. Plan view of eigenray paths from the southeast source position to the northwest (left) and northeast (right) VLA positions.

Figure 3 is a plan view of the eigenray geometries showing the azimuthal deviations from the line-of-sight as a function of range for all the 3-D eigenrays for both the SE2NE and SE2NW cases. The SE2NE track bisects the ocean front and bathymetry at approximately right angles and thus the azimuthal deviations are minimal, less than 50 m over ~30 km. The SE2NW track bisects the ocean front and bathymetry along a diagonal and is more strongly affected by 3-D effects from both the ocean frontal structure and the bathymetry. Upon inspection, it was found that the horizontal refraction of the steep eigenrays is controlled by the bathymetry while the small-angle eigenrays were more affected by the frontal structure. Still, the maximum horizontal deviation is less than 150 m over ~45 km.

The ray-mode analysis was conducted after the experiment. Because bottom interactions were significant and affect both the ray-mode and PE calculations, it was important to provide a good characterization of the sea-floor acoustic properties. Cores have been taken at nearby sites on the shelf (Hathaway *et al.*, 1976). Additional analysis of the ambient noise field and explosive SUS charges measured on the northeast array provide similar values for sediment sound speed (Potty *et al.*, 2000).

Horizontal mode refraction calculations were made of the SE2NE and SW2NE paths. Note the bathymetry for the direct upslope tracks are identical to that used in the previous ray results. The SW2NE track has a similar inverted bathymetric influence as the previous SE2NW track. For each track, two different environments were considered. Both are based on the objectively analyzed SeaSoar data product, which is a 3-D block of temperature and salinity over the entire SeaSoar domain (Sperry, 1999). The first, a cross-shelf reference slice, was created by averaging all of the SeaSoar data in the along-shelf direction, as well as over all seven days of SeaSoar data collection. This gave a 2-D cross-shelf slice that ended up very closely resembling climatology and the historical database used in the ray work. This single slice was then replicated in the along-shelf direction to create the reference environment. The reference case then has relatively

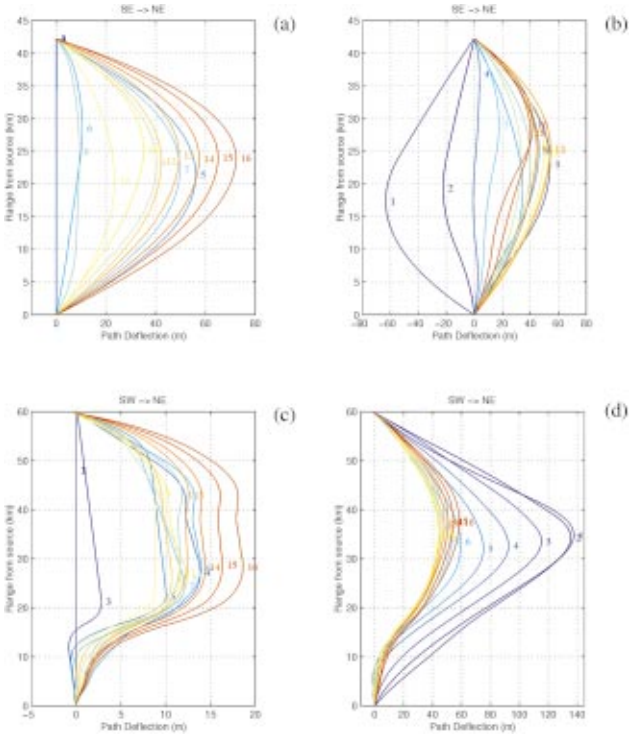


FIG. 4. Modal horizontal deflection paths: (a) SE2NE track for the reference profiles; (b) SE2NE track for the Day 7/31 environment; (c) SW2NE track for the reference profiles; and (d) SW2NE track for the Day 7/31 environment.

smooth variations in the cross-shelf direction and no variability in the along-shelf direction. The second environment, simulating Day 7/31 of the experiment, was taken directly from the objectively analyzed SeaSoar data containing the full 3-D variability of the oceanography.

Figure 4 shows the resulting mode paths for the two environments and two source/receiver paths. For the reference propagation case [Figs. 4(a) and (c)], the lowest two to three modes make the trip with no refraction at all. There is no bottom interaction for these modes, and the cross-slope variability in phase speeds is slight. The more bottom-interacting higher modes are all refracted to varying degrees toward the right, in accordance with local bathymetric gradients, consistent with the ray paths in Fig. 3. Although the deflections are as great as 70 m from the direct path, the total change in path length is no more than one meter in the worst case. Note also that this is similar in magnitude and effect to the SE2NW ray results.

To compute the influence of 3-D propagation using a full-wave model, the 3-D parabolic equation (PE) algorithm of Smith (1999) was used. Four different tracks were studied with this method: SW2NW, SE2NW, SW2NE, and a fourth, imaginary track from the southeast source directly west for 60 km, hereafter referred to as the SE2W track, in an attempt to generate a stronger 3-D effect.

For travel time predictions, multiple frequencies were computed, specifically 32 frequencies over a 32-Hz bandwidth centered at 400 Hz. This bandwidth and frequency sampling were determined as minimal for travel time predictions in this region. Figure 5 displays the data from the reference environment along the SE2NW track after plane-

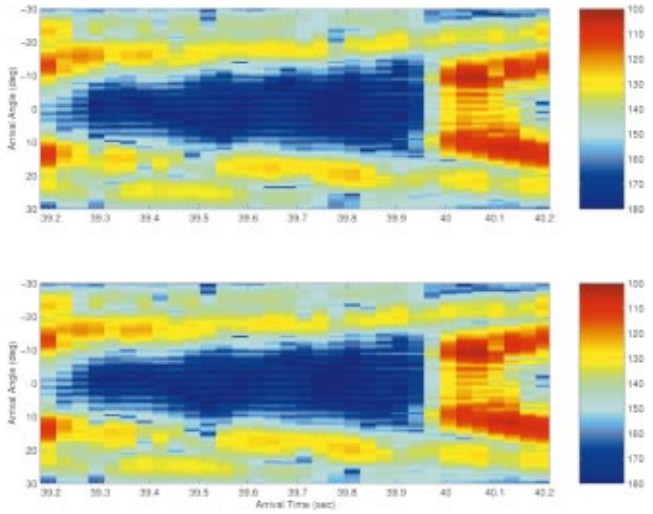


FIG. 5. A comparison of the arrival time structure versus arrival angle between the 2-D (upper plot) and 3-D (lower plot) PE models for the SE2NW track with the reference 3-D interpolated sound speed field.

wave beamforming in the vertical. Note the numerical wraparound effect due to insufficient frequency sampling, so that the first arrivals and loudest part of the signal appear at ~40 s. In this case, no significant differences exist with variations on the order of ± 5 dB in level and $< \pm 25$ ms in travel times (less than the resolution of the calculation here). This is consistent with similar values found in the ray analysis.

IV. SUMMARY

Both the 3-D ray and ray-mode solutions predicted negligible 3-D influences, due either to bathymetry or sound speed, in the directly upslope direction, SE2NE. Path deflections were typically <60 m with travel time variations $<\frac{1}{2}$ ms. This is consistent with the lack of noticeable 3-D influences on the SW2NW results obtained from the PE model.

Both the 3-D ray code and the ray-mode calculations suggest horizontally deflected paths that may be separated from the direct line-of-sight path by as much as 100 m or more for diagonal cross-slope propagation. However, in each case this accounts for less than a 2-m change in total path length. Such path length variations contribute only as much as ~1 ms difference in travel time over a ~40-km path. Such a difference would be on the order of a total sound speed perturbation of ~0.1 m/s or, similarly, a total temperature variation of 0.02 °C. Thus, the environmental resolution required to clearly identify such phenomena is not currently achievable.

Both the ray and ray-mode analyses indicated that the strongest 3-D effect of the environment was due to bathymetry gradients near the shelf break. Consequently, the highest modes (and highest propagation angles) were deflected the most in a manner consistent with a wedgelike boundary. However, all the numerical results also indicate that the lowest modes contain the most travel time sensitivity and variability, even though their corresponding paths are more azimuthally stable. Considerably more variability and 3-D influences were observed in the ray-mode analysis when the more complex 3-D sound speed field was used instead of the

symmetrical reference environment, as may be expected. Still, a general statement may be made that travel time fluctuations in this region are highest for the lower modes and dominated by variations in the sound speed field along the track rather than azimuthal path variations along the track itself.

All of the travel time results, using a 3-D ray, ray-mode, or 3-D PE model, produced similar results. Essentially, travel time differences were not distinguishable on the resolution scale available. The only exception found was the hypothetical SE2W path which considered propagation approximately parallel to the shelf for 60 km. In this case, travel time differences of hundreds of ms were observed. Because such large travel time differences would require total sound speed perturbations of ~5 m/s (or 1 °C), the 3-D influences along this path may be distinguishable from environmental uncertainty.

From all of the numerical analysis conducted, there appears to be no clear signature of azimuthal coupling in the acoustic propagation. Horizontal arrival angles do provide some evidence of azimuthal refraction. However, for these environments and ranges, unrealistically long horizontal arrays would be needed to resolve such features. Furthermore, the maximum distances of propagation did not exceed any mode horizontal shadow zones and no other 3-D propagation phenomenon produces anything distinctive which differentiates it from a variable 2-D propagation path.

D'Agostino, A. (**1996**). ''Simulation of 3-D acoustic arrival structure in the Middle Atlantic Bight and its effect on the inverse tomography solution,'' Master's thesis, Naval Postgraduate School, June 1996.

Gawarkiewicz, G., Bahr, F., Beardsley, R., and Brink, K. (**2001**). ''Interaction of a slope eddy with the shelfbreak front in the Middle Atlantic Bight,'' J. Phys. Oceanogr. **31**, 2783–2796.

Hathaway, J. C., Schlee, J. S., and Poag, C. W. (**1976**). ''Preliminary Summary of the 1976 Atlantic Margin Coring Project of the U.S. Geological Survey,'' Open File Report, No. 76-844.

Lynch, J. F., Gawarkiewicz, G. G., Chiu, C.-S., Pickart, R., Miller, J. H., Smith, K. B., Robinson, A. R., Brink, K. H., Beardsley, R., Sperry, B., and Potty, G. (**1997**). ''Shelfbreak PRIMER—An integrated acoustic and oceanographic field study in the Middle Atlantic Bight,'' in Proceedings of International Conference on Shallow Water Acoustics, Beijing, China, 21–25 April, pp. 205–212.

Potty, G. R., Miller, J. H., Lynch, J. F., and Smith, K. B. (**2000**). ''Tomographic mapping of sediments in shallow water,'' J. Acoust. Soc. Am. **108**, 973–986.

Smith, K. B. (**1999**). ''A three-dimensional propagation algorithm using finite azimuthal aperture,'' J. Acoust. Soc. Am. **106**, 3231–3239.

Sperry, B. (**1999**). ''Analysis of Acoustic Propagation in the Region of the New England Continental Shelfbreak,'' Ph.D. dissertation, Massachusetts Institute of Technology/Woods Hole Oceanographic Institution, June 1999.

Transition from transient response to steady state for a layered medium

Shaw-Wen Liu
Department of Electrical Engineering, Chin Min College, 110 Shyue Fu Road, Tou-Fen, Taiwan 351, Republic of China

Chien-Ching Ma[a)]
Department of Mechanical Engineering, National Taiwan University, Taipei, Taiwan 10617, Republic of China

(Received 19 June 2001; revised 15 December 2001; accepted 22 May 2002)

The transition from transient response to steady-state for a layered medium subjected to antiplane loadings is studied. The steady-state formula for a layered medium is derived and the solutions for a layered half-space are then expressed explicitly in the form of wave number integrals. The transient response solutions for a layered half-space are obtained by the convolution of time harmonic loading function with transient response formula derived analytically from an effective matrix method. Two layered half-spaces with different ratios of wave velocities in the layer and half-space are considered and investigated by means of extensive numerical results to show their quite different transition behavior. The numerical results indicate that transient responses will approach steady state after certain characteristic times when the transient effects die away. The transition phenomena and characteristic times are investigated in detail through the responses from near field to far field as well as from low frequency to high frequency. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1494806]

PACS numbers: 43.20.Bi, 43.20.El, 43.20.Rz, 43.40.At, 43.40.Kd [ANN]

I. INTRODUCTION

This work is a continuation of the study of dynamic responses of a layered medium subjected to antiplane loadings given by Ma *et al.* (2001) in which the attention was focused on the investigation of the transition from transient response to static value in a layered half-space. Not only the characteristic times of the transition could be defined, but also some unusual phenomena were found during the transition period. In this study, a further investigation is performed to understand the transition behavior from transient response to steady state. This study is equivalent to the acoustic problem of sound wave propagating in layered fluid.

The elastic wave propagation in a layered medium has been the interest of considerable studies in recent years, mainly because of its significance in various technological and geophysical problems. Transient responses or steady-state solutions of layered media subjected to time-dependent loadings have been treated by so many authors that it is impossible to give an exhaustive list here. In principle, the steady-state analysis of wave propagation is made equivalent mathematically to the transient-pulse problem by means of the Fourier integral theorem. However, the study related to the transition from transient response to steady state is very limited so far. The steady state can be obtained after some characteristic time when the transient effect dies away.

A tremendous amount of work was devoted in the past to the problem of transient waves excited by an impulsive force, as summarized by Achenbach (1980) and Miklowitz (1978). During the past, several solution methods for this problem have been published. Most of the transient solutions are derived in terms of Laplace or Fourier transform. The evaluation of the inverse transformation integrals that are usually very complicated needs a great deal of computational effort to achieve the desired accuracy. Therefore, various methods or quadrature techniques have been developed to obtain either exact or approximate solutions. It is not too surprising that all these methods are intimately connected, even though they are different in spirit.

The method of expansion in eigenfunctions (Weaver and Pao, 1982; Santosa and Pao, 1989), known also as the method of expansion in normal modes, is usually used to study a body bounded in all directions. This expansion method is particularly suited for obtaining transient responses for large source–receiver separation and large time.

The method of obtaining complete synthetic seismograms, based on the generalized ray theory, was implemented for the structure of one homogeneous layer over a homogeneous half-space for an SH-torque pulse by Pekeris *et al.* (1963). Based on the generalized ray theory and Cagniard's method (Cagniard, 1939), Pao *et al.* (1979) investigated the transient waves generated by various loadings in an infinite plate. The generalized ray expansion for a structure containing any number of layers welded on top of a half-space subjected to an *SH* pulse was presented by Abramovici (1984), along with a rigorous mathematical proof based on the corresponding initial-boundary value problem. The theory of generalized rays is an exact method. An important characteristic of the generalized ray theory is its strong connection with the physical picture of rays propagating through the considered medium. However, it becomes extremely cumbersome for larger receiver–source separation and longer time

a)Electronic mail: cma@w3.me.ntu.edu.tw

because a very large number of generalized ray integrals must be evaluated numerically.

A modified propagator matrix method was used by Franssens (1983) to calculate elastodynamic Green's functions in layered media. The major advantages of this approach are the elimination of the numerical loss of precision problems associated with the Thomson–Haskell formulation (Thomson, 1950; Haskell, 1953). Kundu and Mal (1985) applied a modified wave number integral approach to the calculation of the motion produced in a multilayered solid by dynamic sources. The well-known numerical difficulties associated with the calculation of the integrands are avoided through the use of delta matrices. Special numerical integration schemes are then used to compute the integrals accurately at smaller distances and higher frequencies. When source and receiver depths are close, it is difficult to compute Green's functions of the layered medium, because the integrands, whose variable of integration is the horizontal wave number, oscillate with only slowly decreasing amplitude. To remedy this problem, Hisada (1994) derived the asymptotic solutions to obtain accurate Green's functions.

Recently, the responses of a layered medium subjected to dynamic antiplane impact loading were investigated by Ma and Huang (1995) using the Laplace transform and Cagniard–de Hoop method. The transient responses for both stresses and displacement in each layer are expressed in a closed form. Based on the similar approach, the closed-form solutions for the case of in-plane impact loadings were also given by Ma and Huang (1996). An effective analytical method for determining transient solutions in a strip was developed by Ma and Lee (1999). The analytical solution in the Laplace-transformed domain is expressed in matrix form. It is then decomposed into infinite wave groups. Each group of reflected waves has the same reflection coefficient. Each multireflected wave can be verified by the theory of generalized ray. The inverse transform is performed by the Cagniard method. Both numerical calculations and experimental results were presented. Later, an effective matrix method expanding the matrix solution into a power series of the phase-related reflection and transmission matrix which characterizes the multiple reflections and transmissions of all waves in every layer was developed by Lee and Ma (2000) for a multilayered medium subjected to in-plane loadings. Then, numerical calculations and experimental measurements for a layered half-space were given by Ma and Lee (2000).

In this study, the steady-state solution is solved by a method (Xu and Mal, 1987) based on the classical wave number integral representation and the delta matrix formulation of the elastodynamic field. The transient solutions are obtained by the convolution of the time harmonic loading function with transient solution formula given by Ma *et al.* (2001). Two layered half-spaces with different wave velocity ratios between layer and half-space are considered to show their quite different transition phenomena. The numerical results of displacements and stresses from near field to far field are presented to show the transition from transient response to steady state. The effects caused by different frequencies are also considered.

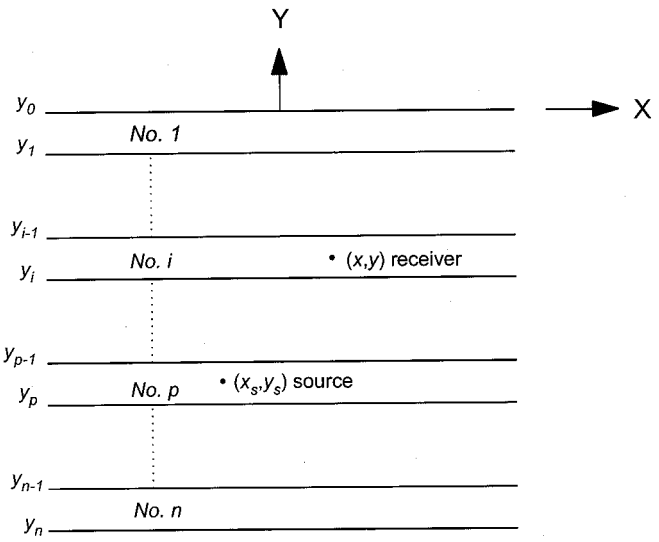


FIG. 1. Configuration and coordinate system of an *n*-layered medium.

II. FORMULATION OF STEADY STATE

Consider a multilayered medium consisting of n layers as shown in Fig. 1. Each layer is assumed to be isotropic, homogeneous, and elastic, and can therefore be characterized by its density ρ and shear wave velocity c_T for the antiplane problem. A time harmonic loading with magnitude σ_0 is applied at the point (x_s, y_s) in the pth layer. Let the loading, displacement (w), and stresses $(\sigma_{xz}, \sigma_{yz})$ have time harmonic behavior of the form $e^{-i\omega t}$, where ω is the circular frequency. For convenience, the time-dependence factor $e^{-i\omega t}$ will be dropped in the following equations. The equation of motion in the layer without body forces is given by

$$\frac{\partial^2 w}{\partial x^2}+\frac{\partial^2 w}{\partial y^2}+k_T^2 w=0, \tag{1}$$

where $k_T=\omega/c_T$ is the shear wave number. Similarly, the equation of motion in the source layer can be written as

$$\frac{\partial^2 w}{\partial x^2}+\frac{\partial^2 w}{\partial y^2}+k_T^2 w=-\frac{\sigma_0}{\mu}\delta(x-x_s)\,\delta(y-y_s), \tag{2}$$

where μ and δ denote the shear modulus and Dirac delta function, respectively. The shear stresses are given by the constitutive relations

$$\sigma_{xz}=\mu\frac{\partial w}{\partial x}, \quad \sigma_{yz}=\mu\frac{\partial w}{\partial y}. \tag{3}$$

This problem is solved by the classical wave number integral representation and delta matrix formulation of the elastodynamic field. Let $\hat{w}(y;k)$ denote the Fourier transform of $w(x,y)$ so that

$$w(x,y)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\hat{w}(y;k)e^{ikx}\,dk, \tag{4}$$

where k stands for the wave number along the x axis. Substitution of Eq. (4) in the differential equations (1) and (2) yields the following first-order, ordinary differential equations:

$$\frac{d^2\hat{w}}{dy^2}+\gamma_T^2\hat{w}=0, \tag{5}$$

$$\frac{d^2\hat{w}}{dy^2}+\gamma_T^2\hat{w}=-\frac{\sigma_0}{\mu}e^{-ikx_s}\delta(y-y_s), \tag{6}$$

where $\gamma_T=\sqrt{k_T^2-k^2}$. The solution of Eq. (5) can be easily obtained and the corresponding solution of Eq. (1) can be expressed in terms of the wave number integral

$$w(x,y)=\int_{-\infty}^{+\infty}[w_-(k)e^{-i\gamma_T(y-y_r)} + w_+(k)e^{i\gamma_T(y-y_r)}]e^{ik(x-x_r)}dk, \tag{7}$$

where w_- and w_+ are unknown functions of k, (x,y) are the coordinates of the receiver, and (x_r,y_r) are the coordinates of a reference point. Both w_- and w_+ will be determined by the boundary and continuity conditions. It should be noted that k in Eq. (7) is real. In a perfectly elastic material, for $k<k_T$, the terms $e^{i\gamma_T y}$ represent upgoing waves in the layer, whereas the terms $e^{-i\gamma_T y}$ represent downgoing waves. For $k>k_T$, the waves become inhomogeneous with exponential increase or decrease with y.

For the solution in the source layer, the following additional term is needed to yield discontinuity in the stress σ_{yz} and continuity in the displacement w across the plane $y=y_s$,

$$w_s=\int_{-\infty}^{+\infty}a_p(k)e^{i\gamma_{T_p}|y-y_s|+ik(x-x_s)}dk, \tag{8}$$

where $a_p(k)=i\sigma_0/(4\pi\mu_p\gamma_{T_p})$. This term represents the wave field produced by the force in an infinite homogeneous medium whose properties are identical to those of the source layer.

Let $\{C\}$ denote the coefficient vector $[w_-,w_+]^T$ and $\{B(y)\}$ the displacement-stress vector $[W,\Sigma]^T$ in each layer such that

$$[w,\sigma_{yz}]^T=\int_{-\infty}^{+\infty}[W,\Sigma]^T e^{ik(x-x_s)}\,dk. \tag{9}$$

From Eqs. (3), (7), and (8), it can be shown that $\{B(y)\}$ in the ith layer may be represented by means of the matrix product

$$\{B(y)\}=[Q_i][E(y-y_{i-1})]\{C_i\}+\{B(y)^{\pm}\}, \tag{10}$$

where

$$[Q_i]=\begin{bmatrix}1 & 1\\ -i\mu_i\gamma_{T_i} & i\mu_i\gamma_{T_i}\end{bmatrix},$$

$$[E(y-y_{i-1})]=\begin{bmatrix}e^{-i\gamma_{T_i}(y-y_{i-1})} & 0\\ 0 & e^{i\gamma_{T_i}(y-y_{i-1})}\end{bmatrix},$$

$$\{B(y)^{\pm}\}=\begin{Bmatrix}1\\ \pm i\mu_i\gamma_{T_i}\end{Bmatrix}a_p(k)e^{i\gamma_{T_i}|y-y_s|},\quad i=p$$

$$=\{0\},\quad i\neq p.$$

Then, by successively applying the interface continuity conditions, the displacement-stress vector at the ith layer can be related to those at the top and bottom surfaces. This is the well-known Thomson–Haskell problem (Thomson, 1950; Haskell, 1953). For the multilayered plate, upon enforcing the traction-free boundary conditions at the top surface and the appropriate boundary conditions at the bottom surface, the formal solution for $\{B(y)\}$ can be obtained through linear operation of matrices $[Q]$, $[E]$ of each layer, and $\{B^{\pm}\}$ of the source layer. A similar procedure can be used for a multilayered half-space where the radiation condition is used in the bottom half-space. Note that the stress component, σ_{xz}, is not included in $\{B(y)\}$. However, it can be derived directly from Eqs. (3) and (9). Thus, the wave number integral representation of the complete displacement-stress vector is obtained. The steady-state solutions can then be easily obtained from the displacement-stress vector multiplied by the factor $e^{-i\omega t}$.

The magnitude of the elements in the matrix products in the kernels of Eq. (9) should decay as $e^{-k|y-y_s|}$ when the wave number k becomes large. However, the direct evaluation of the matrix product becomes erroneous for large k due to a loss of precision problem. The root cause of this phenomenon is the improper cancellation of the leading terms in the product matrices when they are numerically calculated for large k. It should be mentioned that the values of the integrals for large k are often required for the evaluations of the wave number integrals. Here, the delta matrix technique is successively employed to reformulate the problem in avoiding the loss of precision. The details of the delta matrix reformulation are omitted here, and the description of the delta matrix technique can be found in a number of papers (e.g., Kundu and Mal, 1985). Then, the numerical integration is achieved through an effective quadrature scheme (Xu and Mal, 1987) in which the kernels of the wave number integrals are represented by means of polynomials in finite and semi-infinite integrals, and the resulting oscillatory integrals are evaluated analytically. The accuracy of the results is controlled by using an adaptive procedure whereby the number of panels is increased successively until the desired accuracy is reached, with no previous function evaluations wasted.

III. STEADY-STATE SOLUTIONS FOR A LAYERED HALF-SPACE

In order to understand in detail the transition phenomena from transient response to steady state, the numerical examples of a layered half-space are considered. Therefore, the explicit expressions of the steady-state solutions for a layered half-space are given in this section. The thickness of the layer is h. A time harmonic force of magnitude σ_0 is applied at the point (x_s,y_s) within the layer. The complete steady-state solutions can be obtained from Eqs. (3), (9), and (10) by considering appropriate boundary and continuity conditions. By applying the boundary condition of traction free at the top surface, radiation condition in the half-space, and continuity conditions of displacement and stress at the interface, the expressions of the steady-state solutions in the layer are given by

$$w^{(1)}(x,y,\omega)=\int_{-\infty}^{+\infty}(w_{-}^{(1)}e^{-i\gamma_{T_1}y}+w_{+}^{(1)}e^{i\gamma_{T_1}y}+a_1e^{i\gamma_{T_1}|y-y_s|})e^{ik(x-x_s)}\,dk, \tag{11}$$

$$\sigma_{yz}^{(1)}(x,y,\omega)=i\mu_1\int_{-\infty}^{+\infty}\gamma_{T_1}(-w_{-}^{(1)}e^{-i\gamma_{T_1}y}+w_{+}^{(1)}e^{i\gamma_{T_1}y}\pm a_1e^{i\gamma_{T_1}|y-y_s|})e^{ik(x-x_s)}dk, \tag{12}$$

where $+:y>y_s$; $-:y<y_s$, and

$$\sigma_{xz}^{(1)}(x,y,\omega)=i\mu_1\int_{-\infty}^{+\infty}k(w_{-}^{(1)}e^{-i\gamma_{T_1}y}+w_{+}^{(1)}e^{i\gamma_{T_1}y}+a_1e^{i\gamma_{T_1}|y-y_s|})e^{ik(x-x_s)}dk. \tag{13}$$

The solutions in the half-space are expressed as follows:

$$w^{(2)}(x,y,\omega)=\int_{-\infty}^{+\infty}w_{-}^{(2)}e^{-i\gamma_{T_2}y}e^{ik(x-x_s)}dk, \tag{14}$$

$$\sigma_{yz}^{(2)}(x,y,\omega)=-i\mu_2\int_{-\infty}^{+\infty}\gamma_{T_2}w_{-}^{(2)}e^{-i\gamma_{T_2}y}e^{ik(x-x_s)}dk, \tag{15}$$

$$\sigma_{xz}^{(2)}(x,y,\omega)=i\mu_2\int_{-\infty}^{+\infty}kw_{-}^{(2)}e^{-i\gamma_{T_2}y}e^{ik(x-x_s)}dk. \tag{16}$$

The expressions of a_1, $w_{+}^{(1)}$, $w_{-}^{(1)}$, and $w_{-}^{(2)}$ in Eqs. (11)–(16) are given by

$$a_1=\frac{i\sigma_0}{4\pi\mu_1\gamma_{T_1}},\quad w_{+}^{(1)}=\frac{a_1s^2p(r^2+1)}{r(q-ps^2)},$$

$$w_{-}^{(1)}=\frac{a_1(q+pr^2s^2)}{r(q-ps^2)},\quad w_{-}^{(2)}=\frac{is(r^2+1)}{2\pi tr(q-ps^2)},$$

where

$$p=\mu_1\gamma_{T_1}-\mu_2\gamma_{T_2},\quad q=\mu_1\gamma_{T_1}+\mu_2\gamma_{T_2},$$

$$r=e^{i\gamma_{T_1}y_s},\quad s=e^{i\gamma_{T_1}h},\quad t=e^{i\gamma_{T_2}h}.$$

Note that the terms with coefficient a_1 in Eqs. (11)–(13) are the contributions from the source in the layer.

Because the expressions of γ_{T_1}, γ_{T_2}, a_1, $w_{+}^{(1)}$, $w_{-}^{(1)}$, and $w_{-}^{(2)}$ are even functions of k, the kernels of the wave number integrals in Eqs. (11)–(16) are either symmetric or antisymmetric functions of k. Therefore, the integrals of Eqs. (11)–(16) can be simplified as follows:

$$w^{(1)}=2\int_{0}^{+\infty}(w_{-}^{(1)}e^{-i\gamma_{T_1}y}+w_{+}^{(1)}e^{i\gamma_{T_1}y}+a_1e^{i\gamma_{T_1}|y-y_s|})\times\cos[k(x-x_s)]dk, \tag{17}$$

$$\sigma_{yz}^{(1)}=2i\mu_1\int_{0}^{+\infty}\gamma_{T_1}(-w_{-}^{(1)}e^{-i\gamma_{T_1}y}+w_{+}^{(1)}e^{i\gamma_{T_1}y}\pm a_1e^{i\gamma_{T_1}|y-y_s|})\cos[k(x-x_s)]dk, \tag{18}$$

$$\sigma_{xz}^{(1)}=-2\mu_1\int_{0}^{+\infty}k(w_{-}^{(1)}e^{-i\gamma_{T_1}y}+w_{+}^{(1)}e^{i\gamma_{T_1}y}+a_1e^{i\gamma_{T_1}|y-y_s|})\sin[k(x-x_s)]dk, \tag{19}$$

$$w^{(2)}=2\int_{0}^{+\infty}w_{-}^{(2)}e^{-i\gamma_{T_2}y}\cos[k(x-x_s)]dk, \tag{20}$$

$$\sigma_{yz}^{(2)}=-2i\mu_2\int_{0}^{+\infty}\gamma_{T_2}w_{-}^{(2)}e^{-i\gamma_{T_2}y}\cos[k(x-x_s)]dk, \tag{21}$$

$$\sigma_{xz}^{(2)}=-2\mu_2\int_{0}^{+\infty}kw_{-}^{(2)}e^{-i\gamma_{T_2}y}\sin[k(x-x_s)]dk. \tag{22}$$

The simplified integrals can be evaluated more efficiently.

IV. TRANSIENT SOLUTIONS FOR A LAYERED HALF-SPACE

In this section, the transient solutions of a layered half-space subjected to a time harmonic loading $f(t)=\sigma_0\sin(\omega t)H(t)$ applied at $(x,y)=(0,-h_s)$ in the layer will be given. It can be expected that these transient solutions will approach the steady-state solutions as time increases. The transient solutions here are obtained by using the similar formulations and techniques presented by Ma *et al.* (2001). Thus, the details of the formulations are omitted here. Only the solutions related to a layered half-space are given. The main difference between Ma *et al.* (2001) and this paper is caused by the loading functions. The loading function used in this study is $f(t)=\sigma_0\sin(\omega t)H(t)$, while $f(t)=\sigma_0H(t)$ is used in Ma *et al.* (2001). Therefore, the transient solutions are derived directly from the results presented in Ma *et al.* (2001). For any given loading function $f(t)$ applied at a point in the layer, the general transient solutions of the displacement and stresses can be obtained by the following convolution techniques:

$$w(x,y,t)=f(t)*w^{(1)}(x,y,t), \tag{23}$$

$$\sigma_{yz}(x,y,t)=L^{-1}[\hat{f}(p)\cdot p]*\sigma_{yz}^{(1)}(x,y,t), \tag{24}$$

$$\sigma_{xz}(x,y,t)=L^{-1}[\hat{f}(p)\cdot p]*\sigma_{xz}^{(1)}(x,y,t), \tag{25}$$

where $\hat{f}(p)$ is the Laplace transform of $f(t)$ with transform parameter p, L^{-1} denotes the inverse Laplace transform, and $w^{(1)}(x,y,t)$, $\sigma_{yz}^{(1)}(x,y,t)$, and $\sigma_{xz}^{(1)}(x,y,t)$ can be obtained from Eqs. (43a), (43b), and (43c) in Ma *et al.* (2001), respectively.

For $f(t)=\sigma_0\sin(\omega t)H(t)$ considered here, the transient solutions in the layer can be obtained from the convolutions as follows:

$$w(x,y,t)=\int_{0}^{t}sin[\omega(t-\tau)]w^{(1)}(x,y,\tau)d\tau, \tag{26}$$

$$\sigma_{yz}(x,y,t)=\omega\int_{0}^{t}\cos[\omega(t-\tau)]\sigma_{yz}^{(1)}(x,y,\tau)d\tau, \tag{27}$$

$$\sigma_{xz}(x,y,t)=\omega\int_{0}^{t}\cos[\omega(t-\tau)]\sigma_{xz}^{(1)}(x,y,\tau)d\tau. \tag{28}$$

The Gaussian quadrature is then used to evaluate the integrals. For a high-frequency, long-time, and far-field response, however, direct computations of the convolutions presented in Eqs. (26), (27), and (28) could be very time consuming. Here, the convolutions can be rearranged by taking $\sin(\omega t)$ and $\cos(\omega t)$ out of the integrals because the time t, which is the upper bound of the integrals, is not function of τ. Thus Eqs. (26)–(28) can be rewritten as

$$w=\sin(\omega t)\int_0^t \cos(\omega\tau)w^{(1)}(\tau)d\tau-\cos(\omega t)\times\int_0^t \sin(\omega\tau)w^{(1)}(\tau)d\tau, \tag{29}$$

$$\sigma_{yz}=\omega\,\cos(\omega t)\int_0^t \cos(\omega\tau)\sigma_{yz}^{(1)}(\tau)d\tau+\omega\,\sin(\omega t)\times\int_0^t \sin(\omega\tau)\sigma_{yz}^{(1)}(\tau)d\tau, \tag{30}$$

$$\sigma_{xz}=\omega\,\cos(\omega t)\int_0^t \cos(\omega\tau)\sigma_{xz}^{(1)}(\tau)d\tau+\omega\,\sin(\omega t)\times\int_0^t \sin(\omega\tau)\sigma_{xz}^{(1)}(\tau)d\tau. \tag{31}$$

Now the entire time histories of transient responses from $\tau=0$ to $\tau=t$ are obtained by computing the integration once only. It is not necessary to compute the integration starting from $\tau=0$ for each moment t, because the previous integration value can be used for the next moment t. This saves a lot of computational time particularly for the evaluation of transient responses at higher frequencies. In addition, the values of integrals in Eqs. (29)–(31) will oscillate faster for higher frequency. Therefore, an adaptive quadrature procedure depending on the frequency is adopted to achieve the desired numerical accuracy.

V. NUMERICAL RESULTS AND DISCUSSIONS

The transition phenomena from transient response to steady state for a layered half-space are investigated in detail in this section by means of extensive numerical results. However, only the representative ones are shown in this paper.

The layered half-space is composed of a layer with thickness equal to h and a half-space. All of the wave velocities shown in the numerical results are normalized to the velocity in the layer. The loading is applied at the point $(x,y)=(0,-h/2)$ within the layer. Three receivers located at $(5h,-h/2)$, $(50h,-h/2)$, and $(200h,-h/2)$ are used to study near-field, intermediate-field, and far-field responses, respectively. Numerical results are presented for the displacement (w) and stress (σ_{yz}). The stress (σ_{xz}) is not presented because of the similar characteristics to σ_{yz}. In order to study the effects caused by different frequencies, the low ($\omega=1$), intermediate ($\omega=6$), and high ($\omega=12$) frequencies are taken into account in this paper. Note that the wavelength at $\omega=6$ is of the order of the thickness (h) of the layer. In the following numerical results, two layered half-spaces with different wave velocity ratios ($c_{T(2)}/c_{T(1)}=0.771$ and 1.297) are considered to show the completely different transition behavior. The amplitude spectra of stresses located at near field, intermediate field, and far field are shown in Figs. 2 and 3 for the layered half-spaces with

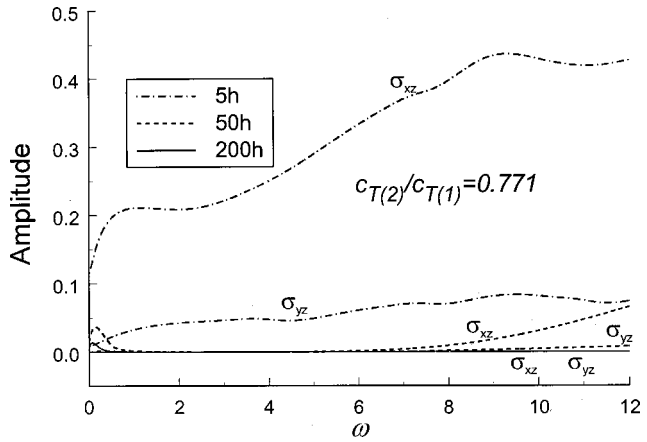


FIG. 2. Amplitude spectra of stresses in a layered half-space with faster wave velocity in the layer.

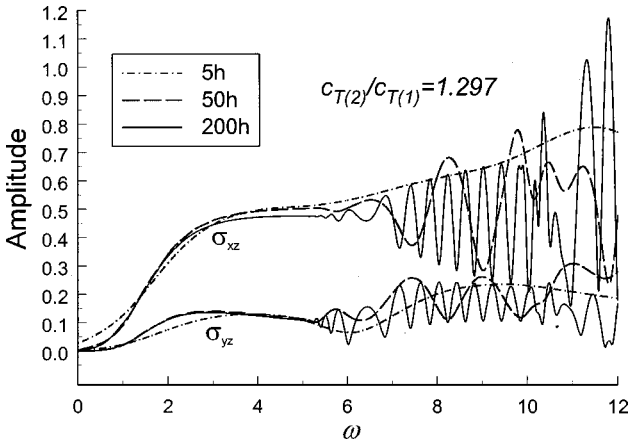


FIG. 3. Amplitude spectra of stresses in a layered half-space with slower wave velocity in the layer.

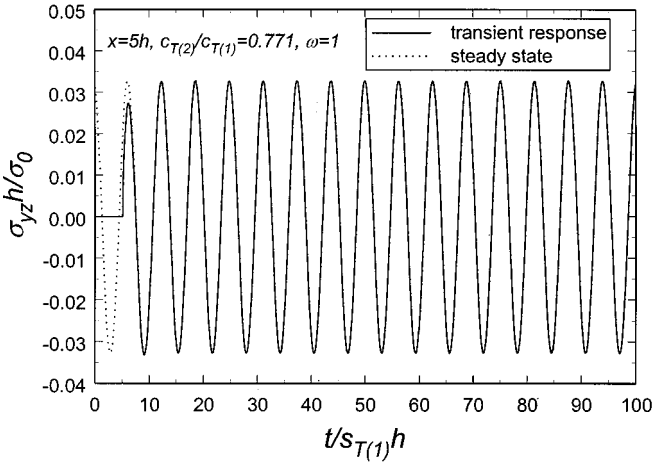


FIG. 4. Transient response and steady state of the stress (σ_{yz}) at $(x,y)=(5h,-h/2)$ in the layer with faster wave velocity, circular frequency $\omega=1$.

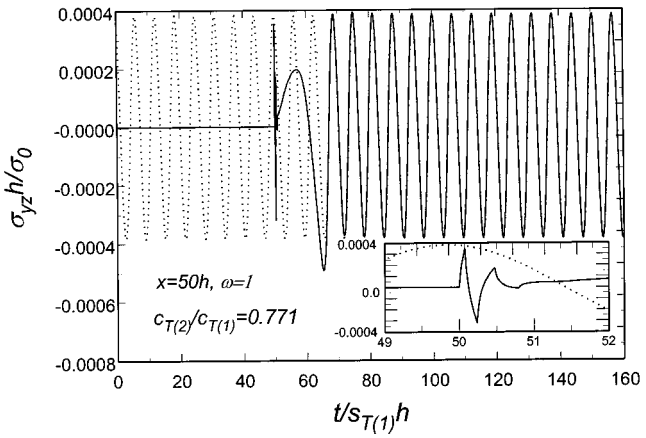


FIG. 5. Transient response and steady state of the stress (σ_{yz}) at (x, y) $=(50h, -h/2)$ in the layer with faster wave velocity, circular frequency $\omega=1$.

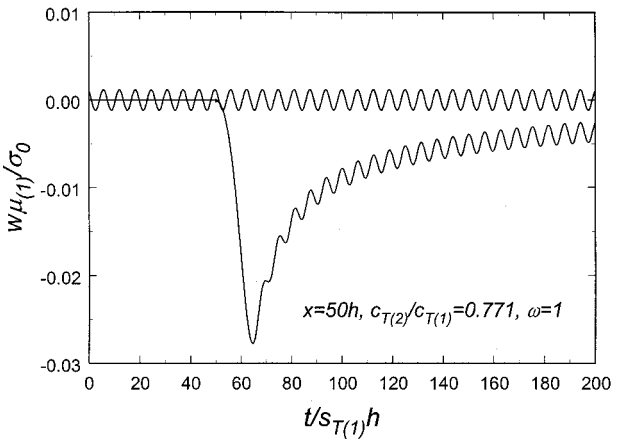


FIG. 7. Transient response and steady state of the displacement at (x, y) $=(50h, -h/2)$ in the layer with faster wave velocity, circular frequency $\omega=1$.

wave velocity ratios $c_{T(2)}/c_{T(1)}=0.771$ and 1.297, respectively. It is obvious from Fig. 2 that the amplitudes of stresses decrease as the distances between source and receiver increase for the case with faster wave velocity in the layer. However, this phenomenon is not observed in the case with slower wave velocity in the layer, as shown in Fig. 3. Furthermore, the amplitudes begin to oscillate at frequency $\omega\approx 5.5$. Note that the wavelength at this frequency is close to the thickness (h) of the layer. The oscillations are more serious for far-field responses.

A. Layer with faster wave velocity

The numerical results for the layered half-space with wave velocity ratio $c_{T(2)}/c_{T(1)}=0.771$ are presented and discussed in this subsection. The stresses at low frequency ($\omega=1$) are shown in Figs. 4, 5, and 6 for near-field, intermediate-field, and far-field responses, respectively. The corresponding displacement for the intermediate field is shown in Fig. 7. The transient responses at both intermediate field and far field as shown in Figs. 5 and 6 are quite different from that of near field as shown in Fig. 4. It is found from Figs. 4–7 that the amplitudes for both displacement and stress responses decrease as the distance between source and receiver increases. The transient response of the displacement shown in Fig. 7 reaches a maximum value at around $t/s_{T(1)}h=xc_{T(1)}/c_{T(2)}h$ (i.e., $t=x/c_{T(2)}$), and then takes a very long time to approach to steady state. It is observed from the numerical results that it takes a much longer time for far-field displacement to transfer from transient response to steady state. However, the transient responses of stresses shown in Figs. 4, 5, and 6, go to steady state right after $t=x/c_{T(2)}$. Note that the transient stresses at both intermediate field and far field as shown in the small windows in Figs. 5 and 6 vary tremendously at the very beginning, when a large number of reflected waves arrive in a very short time right after the initial wave. A common feature is observed in the transient responses at low frequency for both displacements and stresses. Although a time-harmonic function is applied in the layer, the transient responses vary smoothly most of the time before $t=x/c_{T(2)}$ without oscillations. The interesting phenomenon found in Fig. 10 presented in Ma *et al.* (2001) is also observed in Fig. 6 here. During the tran-

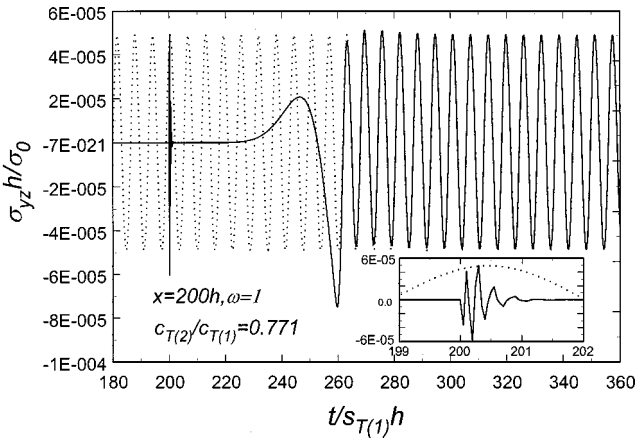


FIG. 6. Transient response and steady state of the stress (σ_{yz}) at (x, y) $=(200h, -h/2)$ in the layer with faster wave velocity, circular frequency $\omega=1$.

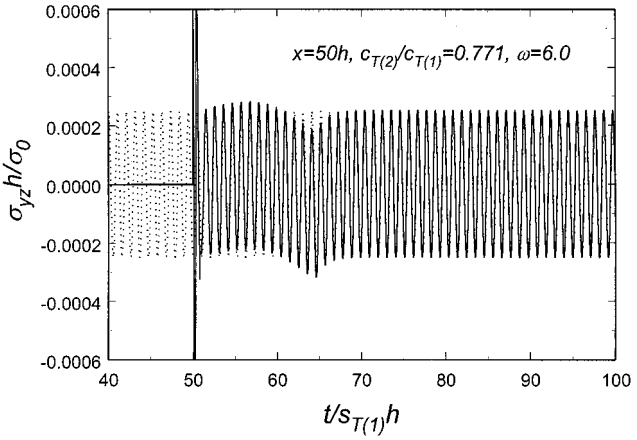


FIG. 8. Transient response and steady state of the stress (σ_{yz}) at (x, y) $=(50h, -h/2)$ in the layer with faster wave velocity, circular frequency $\omega=6$.

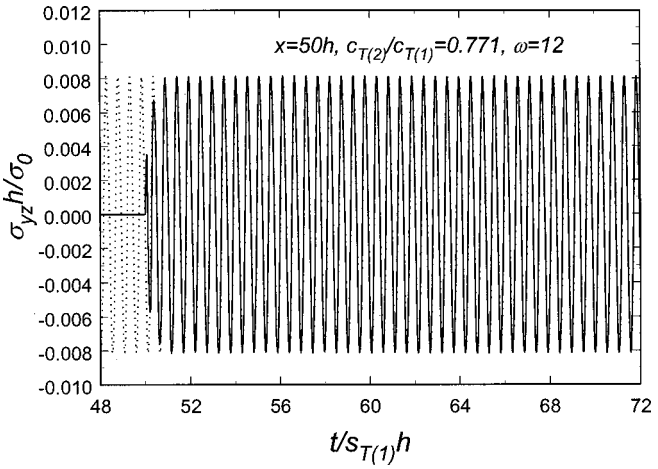


FIG. 9. Transient response and steady state of the stress (σ_{yz}) at (x, y) $=(50h, -h/2)$ in the layer with faster wave velocity, circular frequency $\omega=12$.

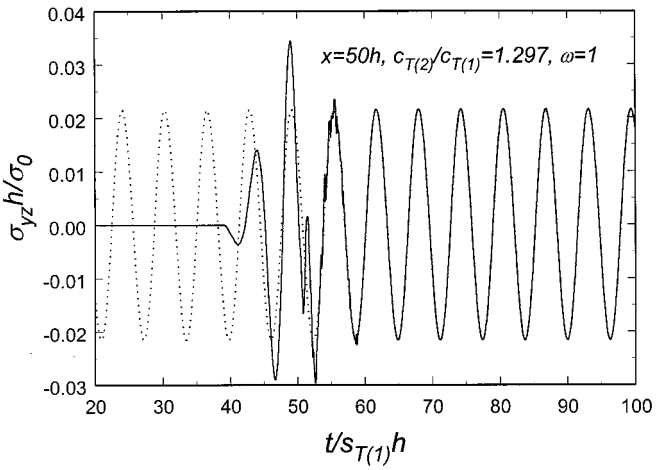


FIG. 11. Transient response and steady state of the stress (σ_{yz}) at (x, y) $=(50h, -h/2)$ in the layer with slower wave velocity, circular frequency $\omega=1$.

sition from transient response to steady state, there is a long period ($t/s_{T(1)}h \approx 202-230$) where the responses are very small and no variation is observed.

The effects caused by different frequencies are considered next. Here, only the intermediate-field stress responses are presented for illustration. The responses at intermediate frequency ($\omega=6$)and high frequency ($\omega=12$) are shown in Figs. 8 and 9, respectively. Compared with transient stress responses at low frequency ($\omega=1$) as shown in Fig. 5, the responses at higher frequencies ($\omega=6$ and 12) as shown in Figs. 8 and 9 can transfer to steady state much faster. The transient displacement responses also have a similar tendency. The maximum values observed at low-frequency responses are not so evident at high-frequency responses. It is noticed that the transient stress as shown in Fig. 9, in which the transient effect is not so obvious, becomes steady state much earlier before $t=x/c_{T(2)}$. From the numerical results studied above, it can be concluded that during the transition stage the behavior of low-frequency responses is strongly related to the Heaviside step function $H(t)$ as studied in Ma *et al.* (2001), whereas high-frequency responses are dominated by the harmonic function $\sin(\omega t)$.

B. Layer with slower wave velocity

The numerical results for a layered half-space with slower wave velocity in the layer are considered here. Figures 10, 11, and 12 show the stress responses at low frequency ($\omega=1$) for near field, intermediate field, and far field, respectively. Due to the existence of head waves, the transition phenomena from transient response to steady state are quite different from the last case with faster wave velocity in the layer. The arrival and end times of head waves can be obtained from the formula given by Ma *et al.* (2001). These times and arrival times of body waves play important roles to the transition phenomena.

Unlike the previous case, the amplitudes of stress responses as shown in Figs. 10–12 apparently do not decrease from near field to far field because more energy can be kept inside the layer. Displacement responses, which are not shown here, have similar behavior. From the numerical results of the intermediate-field ($x=50h$) responses at higher

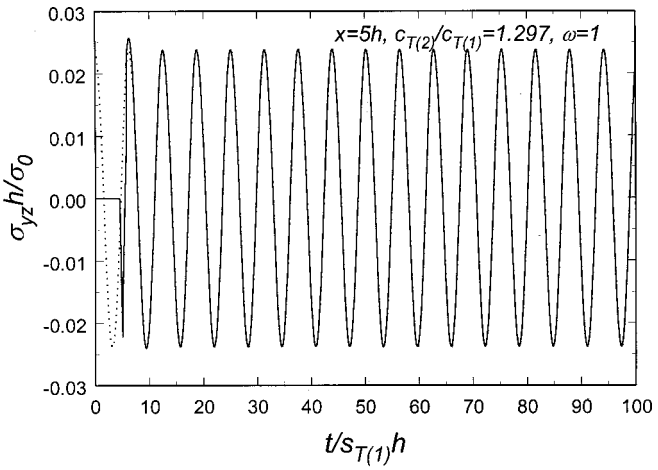


FIG. 10. Transient response and steady state of the stress (σ_{yz}) at (x, y) $=(5h, -h/2)$ in the layer with slower wave velocity, circular frequency $\omega=1$.

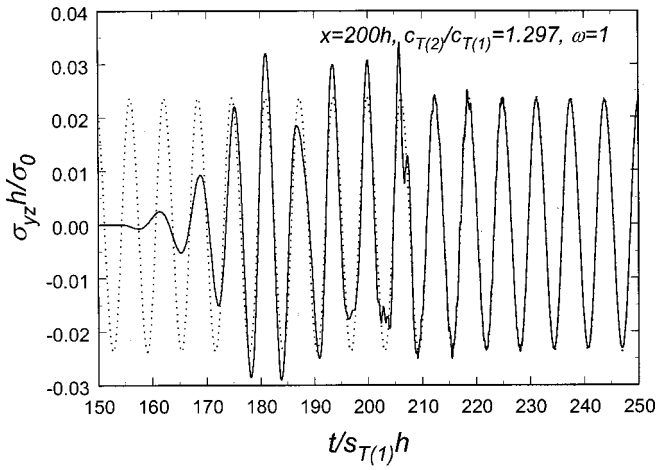


FIG. 12. Transient response and steady state of the stress (σ_{yz}) at (x, y) $=(200h, -h/2)$ in the layer with slower wave velocity, circular frequency $\omega=1$.

frequencies (not shown here), it is found that no matter low- or high-frequency responses, the characteristics of Heaviside step function are not observed in transient displacements. Compared with body waves, the contribution of head waves is smaller for higher frequency responses. From the numerical results obtained in this study, the transient responses can transfer to steady state between the arrival of first body wave ($t=x/c_{T(1)}$) and the end of last head wave ($t=xc_{T(2)}/c_{T(1)}^2$).

VI. CONCLUSIONS

The transition from transient response to steady state for a layered half-space subjected to an antiplane harmonic loading has been investigated in detail by extensive numerical results. The transient responses are obtained by the convolution of time harmonic loading function with transient solutions derived analytically from an effective matrix method given by Ma *et al.* (2001). The steady-state problem is solved by the wave number integral representation. The numerical integration is achieved through an effective quadrature scheme. For the layered half-space with faster wave velocity in the layer, the transition behavior of low-frequency responses is strongly related to the Heaviside step function while high frequency responses are dominated by the harmonic function. Usually high-frequency responses approach steady state faster than low-frequency responses, which can go to steady state after $t=x/c_{T(2)}$. For the layered half-space with slower wave velocity in the layer, the transient responses transfer to steady state between the arrival of first body wave ($t=x/c_{T(1)}$) and the end of first head wave ($t=xc_{T(2)}/c_{T(1)}^2$) no matter low- or high-frequency responses.

ACKNOWLEDGMENT

The research support of the National Science Council, Republic of China, through Grant NSC 89-2212-E002-018 at National Taiwan University is gratefully acknowledged.

Abramovici, F. (**1984**). ''The exact solution to the problem of an SH pulse in a layered elastic half-space,'' Bull. Seismol. Soc. Am. **74**, 377–393.

Achenbach, J. D. (**1980**). *Wave Propagation in Elastic Solids* (North-Holland, Amsterdam).

Cagniard, L. (**1939**). *Reflexion et Refraction des Ondes Seismiques Progressives* (Cauthiers-Villars, Paris); Translated into English and revised by E. A. Flinn and C. H. Dux, 1962. *Reflection and Refraction of Progressive Seismic Waves* (McGraw Hill, New York).

Franssens, G. R. (**1983**). ''Calculation of the elasto-dynamic Green's function in layered media by means of a modified propagator matrix method,'' Geophys. J. R. Astron. Soc. **75**, 669–691.

Haskell, N. (**1953**). ''The dispersion of surface waves on multilayered media,'' Bull. Seismol. Soc. Am. **43**, 17–34.

Hisada, Y. (**1994**). ''An efficient method for computing Green's functions for a layered half-space with sources and receivers at close depths,'' Bull. Seismol. Soc. Am. **84**, 1456–1472.

Kundu, T., and Mal, A. K. (**1985**). ''Elastic waves in a multilayered solid due to a dislocation source,'' Wave Motion **7**, 459–471.

Lee, G. S., and Ma, C. C. (**2000**). ''Transient elastic waves propagating in a multi-layered medium subjected to in-plane dynamic loadings. I. Theory,'' Proc. R. Soc. London, Ser. A **456**, 1355–1374.

Ma, C. C., and Huang, K. C. (**1995**). ''Wave propagation in layered elastic media for antiplane deformation,'' Int. J. Solids Struct. **32**, 665–678.

Ma, C. C., and Huang, K. C. (**1996**). ''Analytical transient analysis of layered composite medium subjected to dynamic inplane impact loadings,'' Int. J. Solids Struct. **33**, 4223–4238.

Ma, C. C., and Lee, G. S. (**1999**). ''Theoretical analysis, numerical calculation, and experimental measurement of transient elastic waves in strips subjected to dynamic loadings,'' Int. J. Solids Struct. **36**, 3541–3564.

Ma, C. C., and Lee, G. S. (**2000**). ''Transient elastic waves propagating in a multi-layered medium subjected to in-plane dynamic loadings. II. Numerical calculation and experimental measurement,'' Proc. R. Soc. London, Ser. A **456**, 1375–1396.

Ma, C. C., Liu, S. W., and Lee, G. S. (**2001**). ''Dynamic responses of a layered medium subjected to antiplane loadings,'' Int. J. Solids Structures **38**, 9295–9312.

Miklowitz, J. (**1978**). *The Theory of Elastic Waves and Wave Guides* (North-Holland, Amsterdam).

Pao, Y. H., Gajewski, R. R., and Ceranoglu, A. N. (**1979**). ''Acoustic emission and transient waves in an elastic plate,'' J. Acoust. Soc. Am. **65**, 96–105.

Pekeris, C. L., Alterman, Z., and Abramovici, F. (**1963**). ''Propagation of an SH-torque pulse in a layered solid,'' Bull. Seismol. Soc. Am. **53**, 39–57.

Santosa, F., and Pao, Y. H. (**1989**). ''Transient axially asymmetric response of an elastic plate,'' Wave Motion **11**, 271–295.

Thomson, W. T. (**1950**). ''Transmission of elastic waves through a stratified solid medium,'' J. Appl. Phys. **21**, 89–93.

Weaver, R. L., and Pao, Y. H. (**1982**). ''Axisymmetric elastic waves excited by a point source in a plate,'' J. Appl. Mech. **49**, 821–836.

Xu, P.-C., and Mal, A. K. (**1987**). ''Calculation of the inplane Green's functions for a layered viscoelastic solid,'' Bull. Seismol. Soc. Am. **77**, 1823–1837.

Linear time domain model of the acoustic potential field

Peter J. Lesniewski[a)]
School of Electrical and Information Engineering, University of South Australia, Mawson Lakes, South Australia 5095, Australia

(Received 27 June 2001; revised 25 March 2002; accepted 29 April 2002)

A new time domain formulation of the acoustic wave is developed to avoid approximating assumptions of the linearized scalar wave equation that limit its validity to low Mach particle velocity modeling or to a smooth potential field in a stationary medium. The proposed model offers precision of the moving frame while retaining the form of the widely used linearized scalar wave equation although with respect to modified coordinates. It is applicable to field calculations involving transient waves with unlimited particle velocity, propagating in inhomogenous fluids or in those with time varying density. The model is based on the exact flux continuity equation and the equation of motion, both using the moving reference frame. The resulting closed-form free space scalar wave equation employing total derivatives is converted back to the partial differential form by using modified independent variables. The modified variables are related to the common coordinates of space and time following integral expressions involving transient particle velocity representing wave radiated by each point of a stationary source. Consequently, transient field produced by complex surface velocity sources can be calculated following existing surface integrals of the radiation theory although using modified coordinates. The use of the proposed model is presented in a numerical simulation of a transient velocity source vibrating at selected magnitudes, leading to the determination of the propagating pressure and velocity wave at any point. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1488941]

PACS numbers: 43.20.Bi, 43.20.Px [DEC]

I. INTRODUCTION

The general dynamic formulation underlying linear modeling of the acoustic field in fluids can be grouped into three broad categories:

(i) models based on the fully linearized scalar wave equation,[1,2]
(ii) models using the Navier–Stokes equations for fluids, partly linearized, aimed at investigating the effects of the macroscopic motion of the medium,[3]
(iii) models based on the energy field approach using the Lagrangian or Hamiltonian formulations.[4]

The first category, extensively used in its harmonic version, is used in the time domain by the analytical spatial impulse response method[5] dedicated to unsteady wave applications such as imaging and transducer field modeling. In order to reach its anticipated accuracy it has been modified by subdividing integrals into simpler and physically relevant blocks[6] structured into a multidimensional convolution[7] and extended to secondary reflections,[8] but eventually arriving at excessively complex formulations when involving the Kirchhoff's integral to model the radiating surface.[9] The second category is often used as a starting point for the investigation of wave propagation in moving media—which is also approached by a less general method of ray acoustics. The third category is suitable for the investigation of propagation of shock waves and high level or supersonic noise which may also be nonlinear. Since the first category is approximate and the second and the third ones are very complex, and usually limited to harmonic signals or plane waves,[10] engineering applications would benefit from improving the modeling accuracy without increasing the complexity of computations.

Throughout the latter half of 1700s Euler, d'Alembert, and Lagrange have formulated laws governing the acoustic wave phenomena including the equation of state, the equation of motion and the flux continuity equation, leading to the free-space linearized scalar wave equation for fluids (1) presented along with underlying formulations by Rayleigh,[2]

$$\nabla^2\Phi - \frac{1}{c^2}\frac{\partial^2\Phi}{\partial t^2} = 0. \tag{1}$$

The static medium density ρ_0 and the wave velocity c characterize the medium, while the velocity potential Φ is related to the acoustic pressure p and to the particle velocity $\mathbf{v}$ as

$$p = \rho_0 \frac{\partial\Phi}{\partial t}, \tag{2a}$$

$$\mathbf{v} = -\nabla\Phi. \tag{2b}$$

The above forms emerge only after approximations in the constituent formulations, necessary to obtain the hyperbolic equation (1) that engineering mathematics handles efficiently. For a block diagram representing the acoustic governing relations refer to Fig. 1.

The proposed model is developed in two steps. The standard approximations are highlighted in Sec. II while deriving exact relations. The exact motion (Euler's) equation is expressed in the moving (Lagrangian) frame using total derivatives. Euler's equation which is well known in acoustics,[1,2,11]

a)Electronic mail: peter.lesniewski@unisa.edu.au

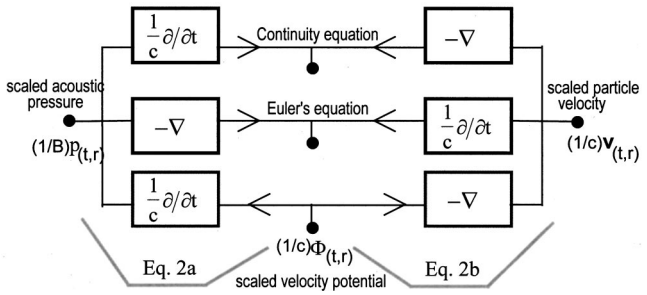


FIG. 1. Block diagram illustrating linearized relations (1), (2a), (2b), and constituent equations using linear operators described in Sec. II. The bulk modulus, $B=\rho_0 c^2$.

is independently derived in this paper employing perspective useful in further discussion. Similarly, the exact continuity equation known in fluid dynamics,[12] but not used in acoustics is derived in the moving frame. In Sec. III these two are combined into a symmetrical set producing the exact form of the wave equation in the moving frame. In Sec. IV, all total derivatives in the above formulation are converted to partial derivatives taken over modified variables. The new variables, which are formally defined through this conversion, are then related to the standard variables using line integrals involving particle velocity of a particular wave path.

II. EXACT CONSTITUENT EQUATIONS IN THE MOVING REFERENCE FRAME

The equation of state, uniquely linking instantaneous pressure P and density ρ of a given fluid, can be expressed using Taylor series at the point of atmospheric equilibrium P_0, ρ_0. This leads to characterizing the medium through the linearized bulk modulus $B=\rho_0(\partial P/\partial\rho)|_{\rho\to\rho_0}$ which along with the acoustic pressure $p=P-P_0$ reduces the equation of state to $p=BS$ with condensation $S=(\rho-\rho_0)/\rho_0$.

The assumption about medium's linearity is justified for small variations of density and pressure, superimposed on larger steady values but nonlinear terms related to the particle velocity vector should not be removed since they represent distortion of the inertial reference frame and appear for a perfectly linear vibration.

Euler's equation (4) considers the infinitesimal volume $dV=dx\,dy\,dz$ where distributed force $\mathbf{F}$ produces the net pressure accelerating the medium. *[This notation is common in acoustics. A more strict but longer derivation (leading to the same results) would define the minute volume as $\delta V=\delta x\,\delta y\,\delta z$ and density as $\rho=\delta m/(\delta x\,\delta y\,\delta z)$ avoiding the arbitrary use of the increment "d" as in derivatives.]* The acceleration using the exact form of the Newton's law is expressed as the total (known as *substantial* in fluid dynamics) increment of velocity, following the infinitesimal path of the particle divided by the elapsed time:

$$d\mathbf{F}=dm\frac{D\mathbf{v}}{Dt}\quad\text{while}\quad d\mathbf{F}=-\nabla p\cdot dV. \tag{3}$$

Eliminating $d\mathbf{F}/dV$, leads to the formulation in the inertial (Eulerian) frame

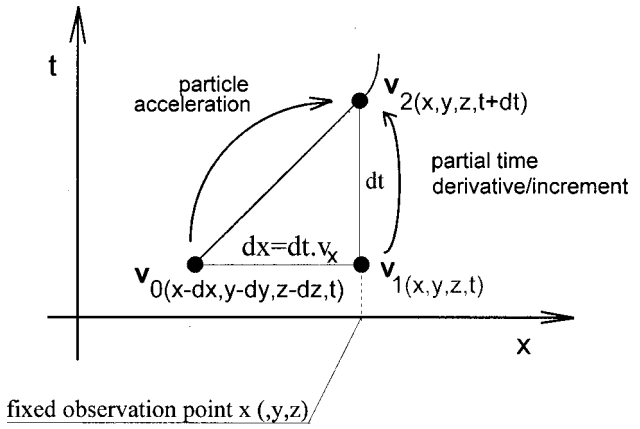


FIG. 2. Variation of a particle velocity vector shown in one plane in the moving frame.

$$-\nabla p=\rho_0\left(\frac{\partial\mathbf{v}}{\partial t}+(\mathbf{v}\cdot\nabla)\mathbf{v}\right),$$

where

$$(\mathbf{v}\cdot\nabla)\mathbf{v}=v_x\frac{\partial\mathbf{v}}{\partial x}+v_y\frac{\partial\mathbf{v}}{\partial y}+v_z\frac{\partial\mathbf{v}}{\partial z}. \tag{4}$$

For a plane harmonic wave, the normally disregarded term $(\mathbf{v}\cdot\nabla)\mathbf{v}$ related to the spatial variation of velocity in Eq. (4) is assumed negligible only when the particle velocity is small compared with the wave velocity.[11] This requirement ($|\mathbf{v}|\ll c$) may not be satisfied in the case of highly mobile (e.g., turbulent) medium, shock waves[13] or near field ultrasonic signals and causes difficulty with the definition of impulsive elementary waves, such as Green's or test functions. Similarly, rapid spatial and temporal variation of the higher order transient spherical waves[14,15] and of the class of nondiffracting waves[16] considered for ultrasonic imaging should benefit from removing this constraint.

Unfortunately, the resulting strain tensor in fluids (4) appears nonlinear to inertial frame observers. It leads directly to the general Navier–Stokes equation, modeling flow of viscous fluids including turbulence and is reduced to simpler forms according to the properties of the fluid.[3] Modeling the particle and medium velocity field is often approached by changing the adopted frame of reference. The inertial-moving frame translation is possible by computationally intensive methods, e.g., by relating coordinate systems using series expansion[13] or through Jacobian operators.[3] Exact translation between the Lagrangian (moving) frame and the inertial frame, possible for elementary waves[12] has not yet been proposed at the stage of the formulation of the general wave equation. Operating on fields of motion using the Lagrangian model directly and curvilinear spatial coordinates, as it is done in fluid dynamics, leads to significant complexity and may not be helpful for linear modeling with arbitrary acoustic signals.[17]

An independent derivation of the equation of motion relates to Fig. 2, highlighting the difference between total and partial derivatives with respect to one spatial variable x (in one plane). A moving particle passes the observation point $(x,y,z,t+dt)$ while reaching velocity $\mathbf{v}_2$. The incre-

mented velocity $\mathbf{v}_2$ is the original vector $\mathbf{v}_0$ which over the time dt has accelerated but which has also relocated to (x,y,z). This velocity difference represents the acceleration vector

$$\mathbf{v}_2-\mathbf{v}_0=dt\,D\mathbf{v}/Dt, \tag{5}$$

where the spatial increments relate to the velocities within dV as $dx=v_x\,dt$, $dy=v_y\,dt$, $dz=v_z\,dt$. The vector $\mathbf{v}_1$ is related to $\mathbf{v}_0$ by extrapolating it to the observation point using only its spatial variation

$$\mathbf{v}_1=\mathbf{v}_0+dx\,\partial\mathbf{v}/\partial x+dy\,\partial\mathbf{v}/\partial y+dz\,\partial\mathbf{v}/\partial z. \tag{6}$$

The partial time derivative is defined at the fixed observation point

$$\mathbf{v}_2-\mathbf{v}_1=dt\,\partial\mathbf{v}/\partial t. \tag{7}$$

The acceleration field, obtained from Eq. (5) by dividing it by dt and substituting velocities from Eqs. (6) and (7) leads to

$$\frac{D\mathbf{v}}{Dt}=\frac{\partial\mathbf{v}}{\partial t}+\frac{\partial\mathbf{v}}{\partial x}v_x+\frac{\partial\mathbf{v}}{\partial y}v_y+\frac{\partial\mathbf{v}}{\partial z}v_z. \tag{8}$$

The above total variation of the momentum, valid for some fixed point (x,y,z) assumes that the path $\mathbf{v}_0\rightarrow\mathbf{v}_2$ approaches straight line, limiting the consideration to irrotational fields (i.e., with common direction of the acceleration, force and velocity vectors).

The flux continuity equation (10) relates the imbalance of the medium mass flow within volume dV to its local condensation, i.e., compression manifested as the acoustic pressure. Thus the variation of the flux momentum $\rho\mathbf{v}$ over boundaries dV relates to the rate of the increment of the mass within the volume

$$\frac{dm}{dt}=-dV\left(\frac{\partial(\rho v_x)}{\partial x}+\frac{\partial(\rho v_y)}{\partial y}+\frac{\partial(\rho v_z)}{\partial z}\right)$$

$$\text{equivalent to}\ \frac{\partial\rho}{\partial t}=-\nabla\cdot(\rho\mathbf{v}). \tag{9}$$

For small density variation within the minute volume, equivalent to considering dV much smaller than the wavelength of the highest spectral component, Eq. (9) is approximated as

$$\frac{\partial S}{\partial t}=-\nabla\cdot\mathbf{v}\quad\text{equivalent to}\ \frac{\partial p}{\partial t}=-B\nabla\cdot\mathbf{v}. \tag{10}$$

The above relation integrated in time states that the buildup of the acoustic pressure in the volume dV is caused by the imbalance in the displacement (divergence) of the medium over that volume. The approximation (10) assuming spatial invariance of the density, leads to a certain inconsistency since the density is still regarded time variant in other expressions. The above formulation, typical for technical acoustics is inaccurately related to the inertial frame, in spite of allowing the flux balance within dV, i.e., the net velocity to be nonzero. Therefore, the frame of reference aligned with the average velocity moves rather than remains at one fixed point. An exact version of Eq. (9) using total derivative of density, obtained through mathematical manipulation, is

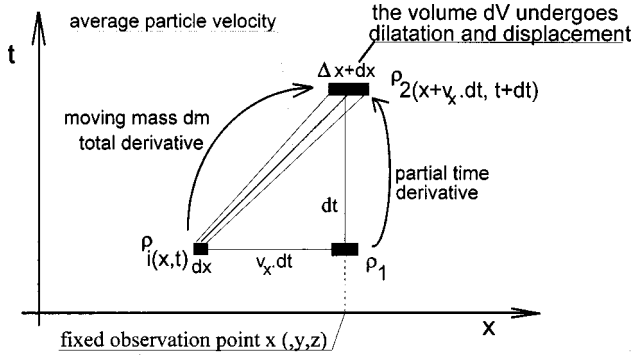


FIG. 3. Divergence of velocity presented in one plane in the moving frame. Note: rectangles represent symbolically the varying volume.

known but it is further simplified narrowing the description to incompressible media in fluid dynamics[3,12] (liquids) where $\nabla\cdot\mathbf{v}=0$.

The standard expression (9), representing the varying density of the mass dm with diverging velocity ignores the drift of the position of the frame, from the point x to $x+dt\,v_x$. Refer to Fig. 3 identifying the correct increment to ρ_2 from ρ_i (rather than ρ_1) to be determined as the limit

$$D\rho=\rho_2-\rho_i,\quad\text{where}\ \ \rho_2=\frac{dm}{(dx+\Delta x)(dy+\Delta y)(dz+\Delta z)}$$

$$\text{and}\quad\rho_i=\frac{dm}{dx\,dy\,dz}, \tag{11}$$

therefore

$$\lim_{\Delta\ll d}(\rho_2-\rho_i)=\frac{dm}{dx\,dy\,dz}\left(-\frac{\Delta x}{dx}-\frac{\Delta y}{dy}-\frac{\Delta z}{dz}\right), \tag{12}$$

where

$$\frac{\Delta x}{dx}=\frac{\Delta x/dt}{dx}dt=\frac{Dv_x}{dx}dt$$

(and the same for y,z) leading to

$$\frac{D\rho}{Dt}=-\rho_0\nabla\cdot\mathbf{v},\quad\text{where}\ \ \rho_0=\frac{dm}{dx\,dy\,dz}. \tag{13}$$

The symbols dx, dy, dz in Eq. (11) determine the initial infinitesimal volume dV while $\Delta x, \Delta y, \Delta z$ represent also infinitesimal extension of that volume due to divergent velocity. That volume, with its density varying from ρ_i to ρ_2 undergoes both, movement as well as dilatation, changing its boundary to $(dx+\Delta x, dy+\Delta y, dz+\Delta z)$. Thus the particle velocity varies by $D\mathbf{v}=(\Delta x, \Delta y, \Delta z)/dt$. With the above substitutions and provided that $\Delta\ll d$, i.e., that the variation of velocity is relatively small compared to the infinitesimal grid, the expression (11) leads to Eqs. (12) and finally (13). The latter is similar to the one used in the standard derivation of the continuity equation (10) but with the medium density taken out of the divergence because the total derivative is taken along the average velocity vector, central to the divergent vectors. The density $\rho=dm/dx\,dy\,dz$ is now only the local property of the medium and therefore it is replaced in Eq. (13) by ρ_0 without approximation.

In the required inertial frame description, the difference between the total increment in Eq. (13), related to the increment found from the spatial extrapolation of the density field at constant time $\rho_2-\rho_1=(\rho_2-\rho_i)-(\rho_1-\rho_i)$ is

$$\frac{\partial\rho}{\partial t}=\frac{D\rho}{Dt}-\frac{\partial\rho}{\partial x}v_x-\frac{\partial\rho}{\partial y}v_y-\frac{\partial\rho}{\partial z}v_z. \tag{14}$$

With Eq. (9) relating the acoustic pressure to the density variation, Eq. (13) substituted into Eq. (14) produces Eq. (15) that should replace the inaccurate expression (10),

$$\frac{\partial p}{\partial t}=-B\nabla\mathbf{v}-\frac{\partial p}{\partial x}v_x-\frac{\partial p}{\partial y}v_y-\frac{\partial p}{\partial z}v_z. \tag{15}$$

The above formulation is precise for spatial profile of the velocity vector being time invariant within dV, i.e., for irrotational fields with locally rectilinear motion.

The linearized wave equation for p is obtained by differentiating Eq. (10) with respect to time and substituting $\partial\mathbf{v}/\partial t$ found from Eq. (4), which is linearized by neglecting the second term, assuming small condensation (approximating the variable ρ by the constant ρ_0) in Eq. (10). This also leads to expressing the wave velocity as $c=\sqrt{B/\rho_0}$. The same wave equation for the velocity potential Φ is obtained by substituting p and $\mathbf{v}$ defined by Eq. (2) into Eq. (10).

Note that the velocity, density, and the pressure are deformed by the frame translation in the same way. The relations (8) and (14) and their equivalents for the pressure, between the two different frames, should be considered in precision applications like metrology or imaging, when utilizing time derivatives of the variables, since measurands are typically observed under conditions close to the inertial frame. Similarly, the boundary condition nulling the pressure on a chosen freely moving boundary (such as the pressure release baffle condition) would have to be reconsidered. This applies to any boundary that could move following the passage of the wave. In fact, nulling the pressure on any arbitrarily shaped boundary of this type is inaccurate since the fine profile of such a boundary is affected by the wave motion itself (causing a transient surface wave).

III. THE ACOUSTIC WAVE EQUATION IN THE MOVING FRAME

The presented moving frame formulation leads to exact Eqs. (16) and (17) describing the acoustic potential field only with static linearization of the equation of state. Combining the equation of state with (13) leads to the continuity equation in the moving frame

$$\frac{Dp}{Dt}=-B\nabla\cdot\mathbf{v} \tag{16}$$

while the exact equation of motion for a sourceless field (4) can be shown as

$$\rho_0\frac{D\mathbf{v}}{Dt}=-\nabla p. \tag{17}$$

Consideration of the above expressions in the inertial frame would lead to complex algorithms with approximations required to make the numerical modeling for high frequency ultrasonic applications realistic. An example of such an approach is the formulation of the finite difference transient wave propagation[18] obtained from the inertial frame following a nonlinear progressive wave equation with damping proposed earlier.[19] For computational reasons, the inertial frame is finally translated to the moving frame associated with the propagating wavefront (assumed to be normal to the particle velocity vector). Such a moving frame however is not equivalent to the moving frame of the velocity field considered here because the wavefront progresses while the particles vibrate at their own velocities.

The moving frame wave equation, separate for velocity and pressure, is determined by combined Eqs. (16) and (17) while the latter alone determines the acoustic potential. The continuity equation (17) involving total derivative brings total derivative in the relation (18) between the acoustic potential and the pressure. It may be justified in detail by taking divergence of Eq. (17) and showing that, at the microscopic level, equivalence $D(\nabla\cdot\mathbf{v})/Dt=\nabla\cdot(D\mathbf{v}/Dt)$ is applicable to velocity or potential in irrotational fields with continuous velocity. Note that the definition of acoustic potential as velocity gradient is not affected by the exact model

$$\mathbf{v}=-\nabla\Phi \quad \text{and} \quad p=\rho_0\frac{D\Phi}{Dt}. \tag{18}$$

In order to obtain the acoustic wave equation, the divergence of velocity found from Eq. (17) should be substituted into another total derivative in Eq. (16). This produces the alternative form (19) of the wave equation for the acoustic pressure. Similarly, the alternative wave equation for velocity can be obtained by differentiating Eq. (17) and substituting Eq. (16). The exact wave equation (19) for the acoustic potential is obtained by substituting Eq. (18) into Eq. (16),

$$\frac{D^2p}{Dt^2}-c^2\nabla^2p=0$$

and (19)

$$\frac{D^2\Phi}{Dt^2}-c^2\nabla^2\Phi=0, \quad \text{where} \quad c^2=B/\rho_0.$$

A short notation $D^2/Dt^2=D(D/Dt)/Dt$ has been used to highlight this operator's similarity with the standard partial form. Note that the total differentiation is a linear operator which may be shown as $D/Dt=dt(\partial/\partial t+\mathbf{v}\cdot\nabla)/dt$, i.e., relating to the same time increment dt common for both parts.

For comparison with the common linearized model (Fig. 1), the moving frame relations are presented by linear blocks in Fig. 4 while the Eq. (20) defines scaled variables. The scaling brings full symmetry to now dimensionless constituent relations (21) and (22) and is convenient in further consideration,

$$p'=p/B, \quad \mathbf{v}'=\mathbf{v}/c, \quad t'=tc, \quad \Phi'=\Phi/c,$$
$$\text{and} \quad \rho'=\rho/\rho_0=S, \tag{20}$$

$$\frac{Dp'}{Dt'}=-\nabla\cdot\mathbf{v}' \text{ and } \frac{D\mathbf{v}'}{Dt'}=-\nabla p', \tag{21}$$

$$p' = \frac{D\Phi'}{Dt'} \text{ and } \mathbf{v}' = -\nabla\Phi'. \tag{22}$$

IV. LINEAR REPRESENTATION OF THE NEW MODEL

Although the proposed moving frame model is compact, symmetrical, and exact (not limited to small particle velocity) yet it cannot be directly implemented by modifying existing formulations because it uses total derivatives—represented by expressions appearing nonlinear in the common inertial frame. Expressing all total derivatives in the inertial frame, would lead to the mathematical complexity exceeding the one found in the existing approximate time domain formulations[19] limited to beams or to the far field. Modifying the spatial coordinate system is often helpful to simplify the form of a solution but it used to be restricted to sound beams and specific source geometry.[20] The use of exact Lagrangian spatial coordinates of the moving frame alone, has lead to complex finite difference modeling algorithms with problems of unsteady performance even in two dimensions.[21]

In the currently proposed model, intended to be more general, total derivatives are replaced by partial derivatives without involving approximations, following Eq. (23a). This is possible by introducing modified variables for the position (represented by $\mathbf{r}\sim$) aligned with the moving frame, thus D/Dt should not involve variation of the position $\mathbf{r}\sim$. Introducing modified time is then required to maintain the original laws represented by new equations. The concept is outlined in Fig. 5 where some dependent variable $W_{(t',\mathbf{r})}$ normally related to $(t',\mathbf{r})$ is shown as $W_{(u',\mathbf{r}\sim)}$ with respect to $(u',\mathbf{r}\sim)$. The new coordinate u is aligned with a streamline, thus progressing along a streamline corresponds to incrementing u without changing $r\sim$. On the other hand, moving along a path normal to streamlines corresponds to varying $r\sim$ while keeping u constant. The distance along any path u' is always larger than the corresponding time t', although the difference would only be noticeable for large magnitude velocities. For simplicity, Fig. 5 uses a single axis r only to represent vector $\mathbf{r}$. One should imagine that r is assembled from individual segments $|d\mathbf{r}|$, with their spatial orientation important in derivations although not shown. Thus, the new coordinates may be regarded curvilinear in the domain $(t,\mathbf{r})$.

Strictly speaking, du' has the same spatial direction as the velocity vector $\mathbf{v}$, i.e., $d\mathbf{u} = (\mathbf{v}/|\mathbf{v}|)\cdot|d\mathbf{u}|$, because it represents a time path of a particle/streamline of a potential field. It may be interpreted as the local time of a particle/streamline which is the common time modified through its "tilting" in the direction of the velocity vector, i.e., looking differently along each axis x, y, z.

The differential expressions are analyzed using local spatial variables represented by $\mathbf{r}\sim$, locally orthogonal to u' which is related to the standard variables $t,\mathbf{r}$.

The modified time denoted u in Eq. (23b), represents the time t modified by velocity (refer to Fig. 5). This means that field irregularities related to particle velocities or to displacements are removed by this (very fine in common cases) local variation of time

$$\frac{DW_{(t',\mathbf{r})}}{Dt'} = \frac{dW}{dt'}\bigg|_{t',\mathbf{r}\to t'+dt',\mathbf{r}+\mathbf{v}\,dt'} = \frac{\partial W_{(u'(t',\mathbf{r}),\mathbf{r}\sim)}}{\partial u'}\,\frac{du'}{dt'}, \tag{23a}$$

where W is any field property (dependent variable) defined at a point $(t,\mathbf{r})$,

$$du' = \sqrt{|d\mathbf{r}|^2 + dt'^2}. \tag{23b}$$

Note that the new variable u' defined above, as well as all other variables (except of the distance) are used in their scaled form (20) again, denoted by an apostrophe—for clarity of the subsequent expressions.

Converting the total derivative in the constitutive Eqs. (21) to the new coordinates following Eq. (23a) and using some field variables $W1$, $W2$ leads to

$$\left(\frac{\partial W1_{(u',\mathbf{r}\sim)}}{\partial u'}\right)\frac{du'}{dt} = -\nabla W2_{(t',\mathbf{r})} \tag{24}$$

with du' and $d\mathbf{r}\sim$ having the same orientation as the vector $d\mathbf{r}$.

Equation (24) may be applied either to the case where $W1$ is a vector and $W2$ a scalar or the opposite but then certainly $\nabla W2$ would appear as $\nabla\cdot\mathbf{W2}$.

To complete the conversion, the RHS of Eq. (24) has yet to be shown as a function of $(u',\mathbf{r}\sim)$. Considering the local tilting of du' with respect to dt' in Fig. 5 and the fact that $d\mathbf{r}\sim/d\mathbf{r}$ is a scalar, the relation between spatial increments in the two coordinate systems is determined in Eq. (25). The increment along the direction $d\mathbf{r}\sim$, normal to du' assumes

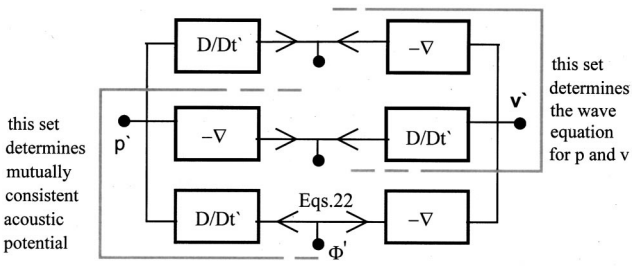


FIG. 4. Block diagram representing scaled moving frame relations (20), (21) and (22).

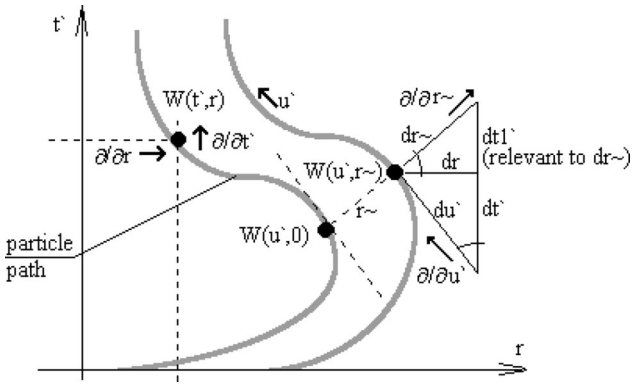


FIG. 5. Relation between the new coordinate system $(u',\mathbf{r}\sim)$ and the original system $(t',\mathbf{r})$. Variable W can be expressed in either system. Note: The two curves represent two adjacent streamlines of an acoustic field and the variable W represents some local property of that field.

that $W2$ is constant over the path du' and time dt', consistent with locally rectilinear although divergent field,

$$\frac{\partial W2_{(t',\mathbf{r})}}{\partial r} = \frac{\partial W2_{(u',\mathbf{r}\sim)}}{\partial r\sim}\frac{dr\sim}{dr},$$

$$\text{where } \frac{dr\sim}{dr} = \frac{du'}{dt'} = \sqrt{1+|\mathbf{v}'|^2}. \tag{25}$$

After eliminating terms du'/dt', Eqs. (24) and (25) lead to Eq. (26) where the unit vector contained in brackets applies only to the case when $W1$ is a vector but $W2$ is a scalar,

$$\frac{\partial W1_{(u',\mathbf{r}\sim)}}{\partial u'} = \frac{-\partial W2_{(u',\mathbf{r}\sim)}}{\partial r\sim}\left(\frac{d\mathbf{r}\sim}{|d\mathbf{r}\sim|}\right). \tag{26}$$

Considering that removing du'/dt' did not change the spatial orientation in Eq. (25), the RHS of Eq. (26) has the same spatial orientation as the original expression (24). Therefore the operator $\boldsymbol{\nabla}$ is used in Eq. (27) for differentiation of $W2$ along $d\mathbf{r}$ or $d\mathbf{r}\sim$.

$$\boldsymbol{\nabla} W2_{(t',\mathbf{r})} = \frac{du'}{dt'}\boldsymbol{\nabla} W2_{(u',\mathbf{r}\sim)}. \tag{27}$$

Thus Eq. (27) substituted into Eq. (24) and related to p' and $\mathbf{v}'$ following Eqs. (21) leads to new linear constituent equations, the continuity equation (28) and the equation of motion (29). It is useful to note that the relation (25) does not involve variation in time, i.e., it relates to the spatial variation only. The translation of the spatial derivatives is the result of a local rotation of the coordinate system, applied to a given point only, e.g., $(u',\mathbf{r}\sim=0)$. Since the time t is only included within u' which has not changed over $d\mathbf{r}\sim$, there is no time increment involved. An alternative explanation is to say that the local input forces represented by the spatial field operator $\boldsymbol{\nabla}$ still have to represent the same physical law irrespective of the local coordinates chosen. Here, only a spatial profile of $W2$ should produce a given effect at the point $(u',\mathbf{r}\sim=0)$, although proportionally scaled down due to the local rotation of each spatial axis

$$\frac{\partial p'_{(u',\mathbf{r}\sim)}}{\partial u'} = -\boldsymbol{\nabla}\cdot\mathbf{v}'_{(u',\mathbf{r}\sim)}, \tag{28}$$

$$\frac{\partial \mathbf{v}'_{(u',\mathbf{r}\sim)}}{\partial u'} = -\boldsymbol{\nabla} p'_{(u',\mathbf{r}\sim)}. \tag{29}$$

The complete set of linear equations equivalent to the ones presented in Fig. 4 but with partial derivatives replacing the total ones as $D/Dt' \to \partial/\partial u'$ and with respect to their modified coordinates $(u'_{(t,\mathbf{r})},\mathbf{r}\sim)$ is shown in Fig. 6. The potential, now defined in the new coordinates, is consistent with Eq. (29).

The closed-form of the new linear wave equation (30) is identical with Eq. (1) but uses variables in the new coordinate system $(u'_{(t,\mathbf{r})},\mathbf{r}\sim)$. This means that traditional solutions based on the linearized wave equation (1) and the associated relations should be considered exact with respect to locally modified coordinates

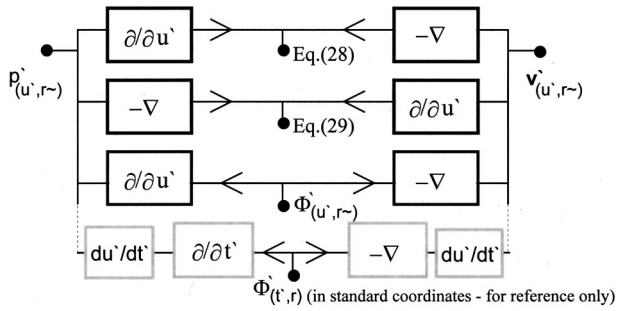


FIG. 6. Block diagram representing scaled relations of the wave propagation model in the new coordinate system $(u',\mathbf{r}\sim)$ produced by modifying $(t,\mathbf{r})$.

$$\boldsymbol{\nabla}^2\Phi_{(uc,\mathbf{r}\sim)} - \frac{1}{c^2}\frac{\partial^2\Phi_{(uc,\mathbf{r}\sim)}}{\partial u^2} = 0. \tag{30}$$

The main difference is in the variation of time t, represented by its counterpart u. Thus solutions obtained using the common, linear form of the wave equation are in fact expressed with respect to time u' modulated by local velocity and with an offset $\mathbf{r}\sim$. An interesting feature of the new coordinates is that the approximate time u' is very close to t' near the local point represented but with the discrepancy accumulating with the distance and with increasing particle velocity.

V. APPLICATION OF THE NEW MODEL

Practical use of the derived model requires only efficient translation of known linear solutions between the standard and modified coordinates. Figure 7 illustrates the time paths of imaginary particles that represent the wave motion of the medium, originating at some arbitrarily vibrating spherical surface positioned at r_0 and propagating along $\mathbf{r}$. The origin of the propagating transient is located on the line representing the wavefront propagating at velocity c relevant to infinitesimal signals. The path $\mathbf{r}\sim$ in the new coordinates is interpreted as a set of adjacent points $(t',\mathbf{r})$ such that $u'_{(t',\mathbf{r})}$

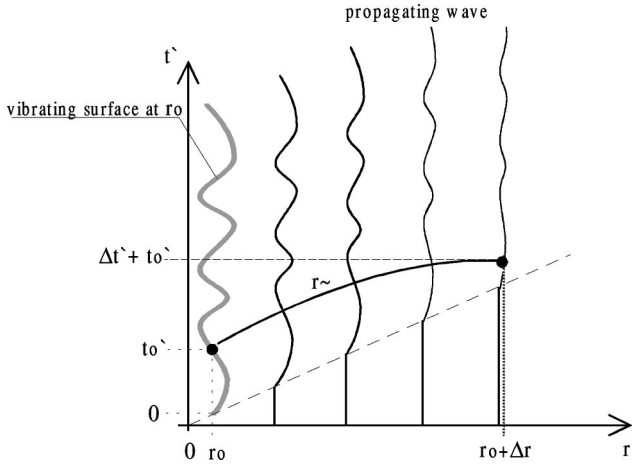


FIG. 7. Relation between the new coordinate system $(u',r\sim)$ and the orthogonal system $(t',\mathbf{r})$ as related to time-paths of particles associated with progressing wavefronts.

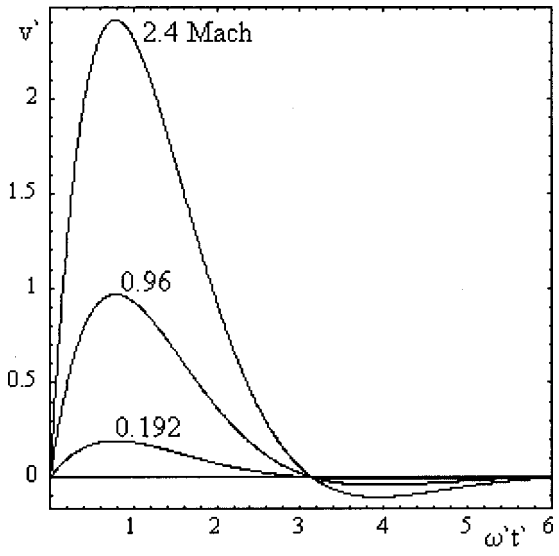


FIG. 8. The source surface velocity (exponentially decaying sinewave at $Q=0.707$) at various peak amplitudes vs $\omega' t'$ in radians ($v'=v/c$, $t'=ct$, $\omega'=\omega/c$) used in simulations verifying the new coordinate system.

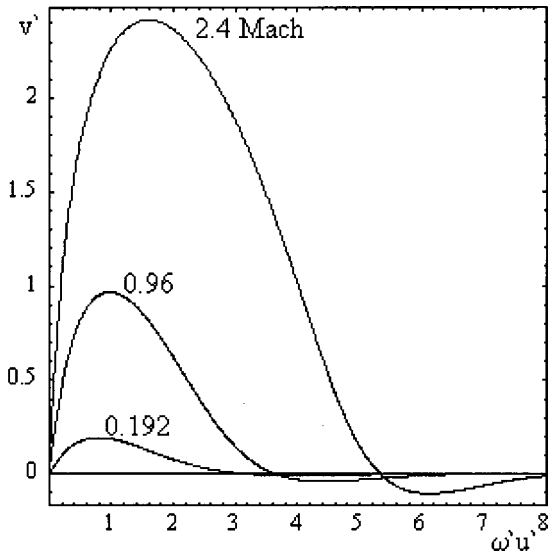


FIG. 9. The velocity transient at three peak levels in the medium at the radiating boundary vs the new time coordinate u' ($v'=v/c$, $u'=cu$, $\omega'=\omega/c$).

$=$const. It depends on local velocity but it approaches a horizontal line for diminishing particle velocities in the far field.

Following Figs. 5 and 7, the translation from the proposed coordinates to the common system $(t',\mathbf{r})$ could involve progressing first along the curve $u'(t',\mathbf{r})$ from the point of origin $u'(0,r_0)$ to $u'(t_0,\mathbf{r}_0+\Delta r_0)$ and then along the curve $\mathbf{r}\sim$ over the increments $\Delta t'$ and $\Delta\mathbf{r}$, formulated by line integrals represented by Eqs. (31) and (32) related to local vibration (the actual path is optional for a potential field). In applications using analytical solutions the integration would more conveniently follow the straight wavefront at the onset of a transient, where $\mathbf{r}\sim=\mathbf{r}$ and only then along the variable u'. The latter would also use Eq. (31) as representing vibrations of an imaginary surface within the field. Note that the integral (32) for the time increment produces a scalar result because $\mathbf{v}$ and $d\mathbf{r}\sim$ have the same spatial orientation. The sign of the scalar result varies with the local velocity, directed either as the vector $d\mathbf{r}\sim$ or opposite to it,

$$\Delta t_0'=\int_{u'}\frac{du'}{\sqrt{1+|\mathbf{v}'_{(u',\mathbf{r}\sim=0)}|^2}}$$

and (31)

$$\Delta\mathbf{r}_0=\int_{u'}\frac{\mathbf{v}\,du'}{\sqrt{1+|\mathbf{v}'_{(u',\mathbf{r}\sim=0)}|^2}},$$

$$\Delta t'=\int_{r\sim}\frac{-\mathbf{v}\,d\mathbf{r}\sim}{\sqrt{1+|\mathbf{v}'_{(u',\mathbf{r}\sim)}|^2}}$$

and (32)

$$\Delta\mathbf{r}=\int_{r\sim}\frac{d\mathbf{r}\sim}{\sqrt{1+|\mathbf{v}'_{(u',\mathbf{r}\sim)}|^2}}.$$

The conversion of the velocity at the vibrating surface from the original to the new coordinates requires only relating the new variable u' to $(t,\mathbf{r})$, following integral (33). The velocity, required in the translation can be obtained from benchmark solutions using retarded time, but interpreted using the new coordinates, following relations in Fig. 6,

$$u'=\int_{t'}\sqrt{1+|\mathbf{v}'_{(t',r)}|^2}\,dt'. \tag{33}$$

The transient waveform, determined by the solution of the wave equation for a particular wave type and propagating along r with radially reduced amplitude replicates the one at the vibrating surface but with some linear distortion and deformation of its profile along $(t,\mathbf{r})$. Time domain solutions of the wave equation with separable variables[14] and the more general ones taking the form of higher order spherical Green's functions[15] can be expressed directly in the proposed coordinates by replacing t' by u' and r by $|\mathbf{r}\sim|$. The free

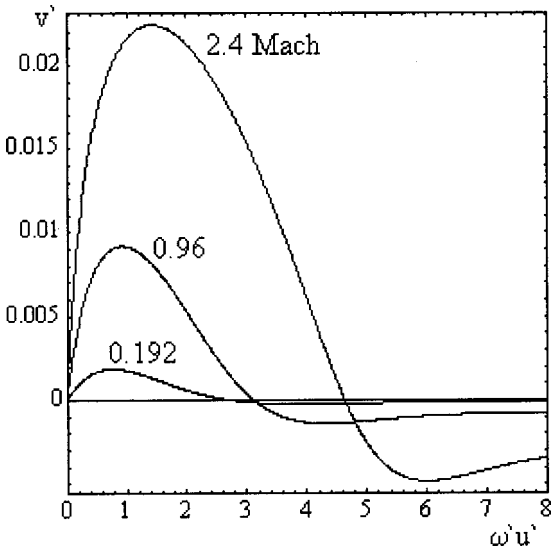


FIG. 10. The particle velocity transients in the medium at the distance $r\sim=100r_0$ with respect to the radiating source, related to the new time coordinate u' ($v'=v/c$, $u'=cu$, $\omega'=\omega/c$, r_0 is the source radius).

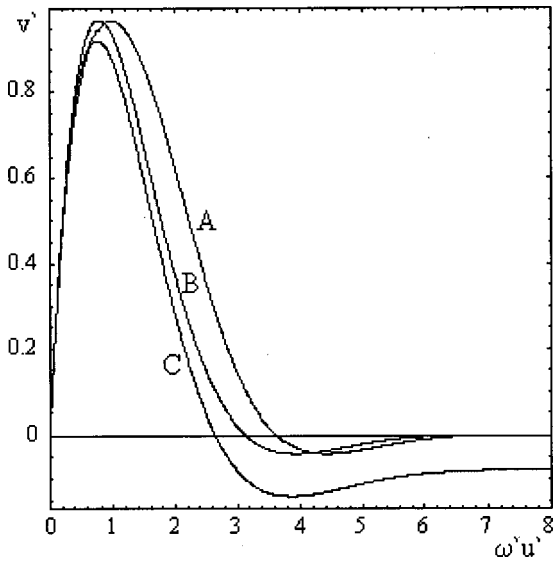


FIG. 11. Comparison of the deformation of the velocity transients for $v_{0\text{peak}}/c=0.96$ ($v'=v/c$, $u'=cu$, $\omega'=\omega/c$, r_0 is the source radius). A, in the medium at the radiating boundary vs $\omega' u'$; B, the original vs $\omega' t'$; C, at the distance $r\sim=100r_0$ vs $\omega' u'$ (scaled up by 100).

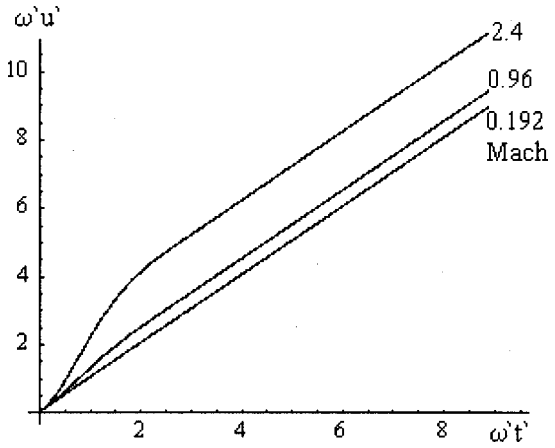


FIG. 13. Elongation of the coordinate u' vs the original time t' increasing with the particle velocity magnitude ($v'=v/c$, $u'=cu$, $t'=ct$, $\omega'=\omega/c$).

space Green's function used in the spatial impulse response formulations[5,8] corresponds to the zero order wave of the form $\Phi'=f(u'-|\mathbf{r}\sim|)/|\mathbf{r}_0+\mathbf{r}\sim|$ where $f(u')$ represents any time function determined by $f(t)$ at the vibrating surface (the boundary at $\mathbf{r}_0$). Higher order waves with other than radial velocity components should be modeled considering varying direction of the velocity vector (similarly along $d\mathbf{r}\sim$ and du') in expressions (31)–(33).

Computational results shown in Figs. 8–15 were obtained by applying the model to the free space Green's function (zero order) radiated by source excited as the same transient waveform but with different surface velocity magnitudes. Results plotted versus variables u' and t' multiplied by the angular frequency $\omega'=\omega/c$ (in radians) are valid for any frequency. The input transient in Fig. 8 is an exponentially decaying sinewave ($Q=0.707$, i.e., slightly less than critically damped) with peak surface velocity levels at the boundary $v_{0\text{peak}}/c=0.192$, 0.96, and 2.4 Mach. The acoustic pressure at the distance $r=100r_0$ reaches 106 dB for the surface peak velocity of 0.96 Mach. The translation of the boundary velocity to the new coordinates $(u',r\sim=0)$ following the integral (33) can be computed numerically or, for selected time functions, analytically. The two stronger transients in Fig. 9, when translated to the new time coordinate u', near the source, are visibly deformed. The same transients at the distance $r\sim=100r_0$ are shown in Fig. 10. Comparison between the original transient of the velocity amplitude 0.96 Mach vs modified time u', on the surface and at a distance $r\sim=100r_0$ (magnified by 100) is shown in Fig. 11.

The surface plots in Figs. 12(a) and (b) compare the propagating transients of the lowest and highest amplitude in the new coordinate system (with the time scale u' adjusted to the wavefront). The relation between the modified time u' and the original t' in Fig. 13 identifies the degree of the local deformation of the field.

The transients in Fig. 14, represent particle velocities translated to the inertial frame of the receiver at various distances from the source as compared with the original excitation velocity on the surface r_0. For the type of transient wave used the velocity and pressure have similar forms (refer to Fig. 15 for a comparison).

Generally, the translation from the moving to the inertial frame can be done through numerically generated matrices of

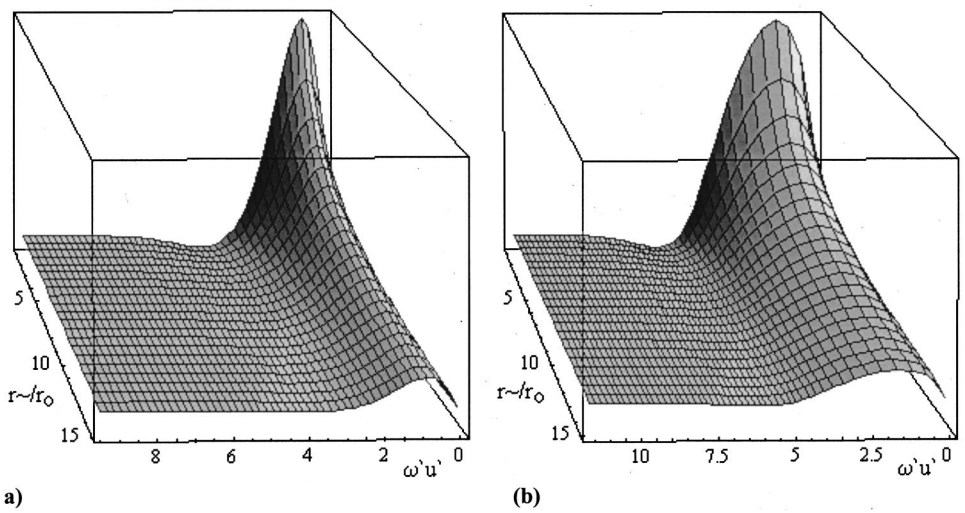


FIG. 12. The minimum (a) and, scaled down maximum strength particle velocity pulse (b) both propagating linearly in the new coordinate system over distance $r\sim/r_0=1-15$.

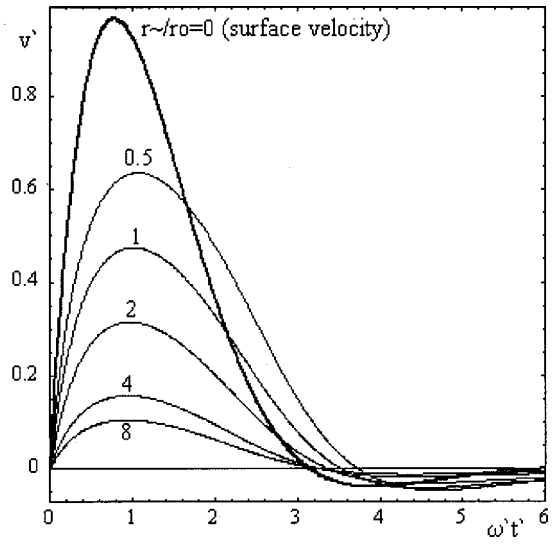


FIG. 14. Comparison of transient waveforms for $v_{0\text{peak}}/c=0.96$ at various distances from the radiating surface ($r\sim/r_0=8$, 4, 2, 1, 0.5 and 0 on the source surface) as indicated along the curves, translated back to the inertial frame ($v'=v/c$, $t'=ct$, $\omega'=\omega/c$).

field variables converted point by point from the domain $(u',\mathbf{r}\sim)$ to $(t',\mathbf{r})$ within some small grid of increments and covering a limited area of interest. Similarly a purely numerical approach assisting finite difference simulation may apply to arbitrary fields or to higher order waves, difficult for symbolic evaluation. For spherical waves the translation is done close to the local position of the propagating wave. The passage $u'=t'$, $\mathbf{r}\sim=\mathbf{r}$ along the ideal wavefront, leads to the surface of interest at $(u',\mathbf{r}\sim)$. The departure from the offset point along u' is calculated relating Eq. (31) to the initial point on the wavefront but removing spatial variations since the observation is in the inertial frame. This requires varying also $\mathbf{r}\sim$ by employing Eq. (32) so that $(\Delta\mathbf{r}_0+\Delta\mathbf{r})=\text{const}$. This determines the correction $d\mathbf{r}\sim$ to the path u' needed to compensate for the position of the surface of interest due to its vibration, equivalent to making $\mathbf{v}'\,du'=-d\mathbf{r}\sim$. This allows for modification of the two integrals for t into a combined expression useful in the required translation. That translation gradually restores the initial deformation of the time scale along the integration path as it can be seen in Fig. 14.

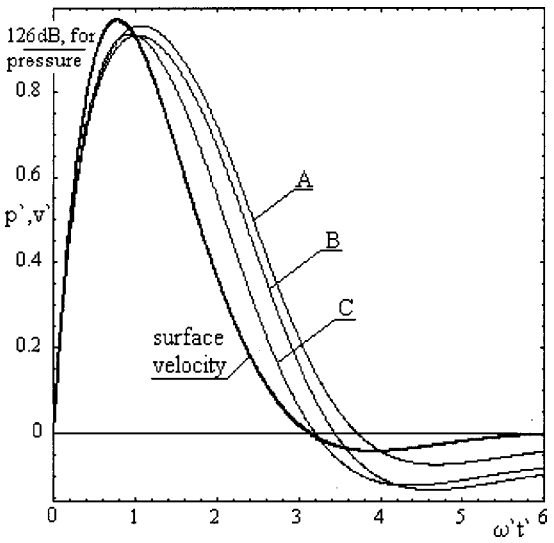


FIG. 15. Source surface velocity for $v_{0\text{peak}}/c=0.96$ compared with scaled up propagating transients at selected distances $r\sim/r_0$ ($v'=v/c$, $p=p/B$, $t'=ct$, $\omega'=\omega/c$). (A) $1.5v'$ at $r\sim/r_0=0.5$, (B) $1.5p'$ at $r\sim/r_0=0.5$, (C) $9p'$ at $r\sim/r_0=8$.

VI. CONCLUSION

The proposed exact formulation of the acoustic wave, based on the moving frame description is suitable for largely symbolic time domain consideration of spherical sources such as those used in the spatial impulse response and Green's function methods. Frequency domain relations for known solutions may be used if the field is determined as composed of such waves. For irregular/arbitrary fields, the model may be implemented numerically using the finite difference or finite element methods, similarly as the more complex schemes using Lagrangian coordinates.[21] Such numerical methods, based on the nonlinear Lagrangian formulation of energy propagation, from 1950s[22] are aimed at modeling detonation and shock waves characterized by high level transport of mass and energy along with the acoustic pressure wave. Therefore those methods are of limited relevance to pulsed acoustic field for which the proposed model is intended. The advantage in adopting the proposed model is in its simplicity resulting from extending the Lagrangian approach to moving frame (modified) coordinates involving both space and time. Another advantage of this model is that it adopts the existing form of the linearized wave equation for which many useful solutions are already available. Therefore, it can extend the established simulation methods to the near field, high frequency and high level signals by the addition of the interface between the moving and the inertial frame (at the source-medium and medium-observer boundaries).

In spite of tedious derivation the proposed formulation is transparent, once the user follows the mathematical formulations, because it does not alter any existing physical interpretation. The moving frame form of the new model (Fig. 4) represents an exact description of the inertial frame but using total derivatives. It is consistent with the exact formulation of the continuity equation, known in fluid mechanics[12] but not utilized in acoustics. That exact formulation has resulted in fully symmetrical expressions for acoustic pressure and particle velocity enabling formulation of the linear form of the model (Fig. 6) in modified coordinates. The apparent nonlinearities of high particle velocity sound waves in the inertial frame disappear in the new coordinates indicating that this distortion is only the result of the vibrational character of the true moving frame of reference. This, for a steady observer in the inertial frame, appears as a variation of variables or operators manifested mathematically through additional products involving particle velocity and leading to the effect of multivalued functions in the description of shock waves.[1] Exact solutions in inertial frame coordinates are determined by the translation between the new and the standard coordinate systems following given line integrals, which may be evaluated symbolically or numerically, de-

pending on field complexity. Since the medium density, denoted ρ_0 in the exact constituent equations, is not a variable of the wave the presented model is applicable to inhomogeneous fluids or to those with time varying density.

[1] A. D. Pierce, *Acoustics, an Introduction to its Physical Principles and Applications* (McGraw-Hill, New York, 1981), Chaps. 1 and 11.
[2] Rayleigh (J. W. Strutt), *Theory of Sound* (Dover, New York, 1945), Vol. 2, Chap. XI.
[3] A. Rutheford, *Vectors, Tensors, and the Basic Equations of Fluid Dynamics* (Prentice-Hall, New York, 1962) Secs. 8 and 3.1.
[4] N. S. Tjøtta and S. Tjøtta, ''Interaction of sound waves, part I basic equations and plane waves,'' J. Acoust. Soc. Am. **82**, 1425–1428 (1987).
[5] P. R. Stepanishen, ''Transient radiation from pistons in an infinite planar baffle,'' J. Acoust. Soc. Am. **49**, 1629–1638 (1971).
[6] H. Lasota, R. Salamon, and B. Delannoy, ''Acoustic diffraction analysis by the impulsive response method: a line impulse response approach,'' J. Acoust. Soc. Am. **76**, 280–289 (1984).
[7] B. Delannoy, H. Lasota, C. Bruneel, R. Torquet, and E. Bridoux, ''The infinite planar baffles problem in acoustic radiation and its experimental verification,'' J. Appl. Phys. **50**, 5189–5195 (1979).
[8] A. Lhémery, ''Impulse response method to predict echo response from targets of complex geometry. Part I: Theory,'' J. Acoust. Soc. Am. **90**, (1991).
[9] Q. Hu and S. F. Wu, ''An explicit integral formulation for transient acoustic radiation,'' J. Acoust. Soc. Am. **104**, 3251–3258 (1998).
[10] D. G. Crighton, A. P. Dowling, J. E. Williams Ffowcs, M. Heckl, and F. G. Leppington, *Modern Methods in Analytical Acoustics* (Springer-Verlag, Berlin, 1992), Chap. 24.
[11] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), Chap. 6.
[12] R. B. Bird, W. E. Stewart, and E. N. Lightfoot, *Transport Phenomena* (Wiley, New York, 1960), Secs. 3.1 and 3.2.
[13] R. T. Beyer, *Nonlinear Acoustics* (The Acoustical Society of America, New York, 1997).
[14] O. M. Buyukdura and S. S. Koc, ''Two alternative expressions for the spherical wave expansion of the time domain scalar free-space Green's function and an application: scattering by a soft sphere,'' J. Acoust. Soc. Am. **101**, 87–91 (1997).
[15] P. Lesniewski and R. Thorn, ''Modeling transient acoustic waves using Green's function,'' Acoust. Lett. **21**, 172–176 (1998).
[16] P. R. Stepanishen, ''Acoustic bullets/transient Bessel beams: near to far field transition via an impulse response approach,'' J. Acoust. Soc. Am. **103**, 1742–1751 (1998).
[17] T. P. Liu, ''Compressible Navier–Stokes equations with zero heat conductivity,'' J. Diff. Eqns. **153**, 225–291 (1999).
[18] G. P. J. Too and J. H. Ginsberg, ''Nonlinear progressive wave equation model for transient and steady-state sound beams,'' J. Acoust. Soc. Am. **91**, 59–68 (1992).
[19] B. E. McDonald and W. A. Kuperman, ''Time domain solution of the parabolic equation including nonlinearity,'' Comput. Math. Appl. **11**, 843–851 (1985).
[20] T. Kamakura, T. Ishiwata, and K. Matsuda, ''A new theoretical approach to the analysis of nonlinear sound beams using the oblate spheroidal coordinate system,'' J. Acoust. Soc. Am. **105**, 3083–3086 (1999).
[21] M. Shashkov and B. Wendroff, ''A composite scheme for gas dynamics in Lagrangian coordinates,'' J. Comput. Phys. **150**, 502–517 (1999).
[22] H. Brode, ''Numerical solutions of spherical blast waves,'' J. Appl. Phys. **26**, 767–775 (1955).

Free molecular sound propagation

F. Sharipov,[a] W. Marques, Jr.,[b] and G. M. Kremer[c]
Departamento de Física, Universidade Federal do Paraná, Caixa Postal 19044, 81531-990 Curitiba, Brazil

(Received 25 September 2001; accepted for publication 10 May 2002)

The sound propagation through a gas in the free-molecular regime is studied on the basis of the linearized collisionless Boltzmann equation. The two principal quantities that characterize the sound propagation, namely the phase and amplitude of the perturbation, are determined by taking into account the influence of the receptor. It is shown that at a small distance between the source and the receptor the presence of the last changes qualitatively the sound characteristics. Two phase velocities are introduced: a differential and an integral, which are different in the free molecular regime. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1490360]

PACS numbers: 43.20.Bi, 51.10.+y [SGK]

I. INTRODUCTION

In the present paper we consider the propagation of plane waves in rarefied gases along the axis x'. The classical theory of sound propagation is based on the assumption that all fields (density, pressure, temperature, velocity, etc.) denoted generically by ψ can be written as a sum of an equilibrium value ψ_0 plus a plane harmonic wave in the form[1]

$$\psi=\psi_0+\overline{\psi}\exp[i(kx'-\omega t')], \tag{1}$$

where $\overline{\psi}$ represents the amplitude of the wave which is considered to be small, ω is the frequency, k is the wave vector, and t' is the time. If we insert the representation (1) into the field equations of the thermodynamics of irreversible processes, we get a system of linear equations for the amplitudes whose solution is called the dispersion relation. This relation has the form $\mathcal{F}(k,\omega)=0$, and by solving it for k taken as a complex quantity, one can obtain the phase velocity v_{ph} and the attenuation coefficient a of the plane harmonic wave

$$k=\frac{\omega}{v_{ph}}+ia. \tag{2}$$

For very small frequencies the phase velocity reduces to the adiabatic sound speed

$$c=\sqrt{\gamma\frac{k_BT_0}{m}}, \tag{3}$$

where k_B is the Boltzmann constant, T_0 is the equilibrium gas temperature, m is the molecular mass, $\gamma=c_p/c_v$ is the ratio of the specific heats c_p and c_v.

Furthermore, the classical theory of sound propagation predicts that both the phase velocity and the attenuation coefficient are independent of the coordinate x'. These results are valid for low sound frequency, i.e., when ω is significantly smaller than the intermolecular collision frequency ν_{coll}. However, when $\omega\sim\nu_{\text{coll}}$ the wavelength $\lambda=2\pi c/\omega$ has the order of the molecular mean free path l and the gas cannot be considered as a continuous medium. In this case the Boltzmann kinetic equation must be applied to describe the sound propagation. The gas rarefaction is usually characterized by the Knudsen number. For the problem in question, it is defined via the wavelength λ as

$$\text{Kn}=\frac{l}{\lambda}. \tag{4}$$

Note that the ratio ω/ν_{coll} has the order the Knudsen number Kn, i.e.,

$$\frac{\omega}{\nu_{\text{coll}}}\sim\text{Kn}.$$

The experiments,[2–4] realized at a low pressure, showed that both sound velocity and attenuation depend on the Knudsen number. To explain such dependence the authors of the paper[5] analyzed the sound propagation on the basis of the Burnett equation, which is valid near the hydrodynamic regime, i.e., when Knudsen number Kn is still small. The works[6–8] provide the results on the sound velocity and attenuation over the whole range of the Knudsen number. These papers confirmed the dependence of the sound propagation on the Knudsen number, but their results do not agree with the experimental data[2–4] near the free molecular regime, i.e., for a large Knudsen number. The trouble is that the solution of the kinetic equation solved in Refs. 6–8 was presented in the form (1) assuming the wave vector k to be independent of the coordinate x', while the experimental data[3,4] showed that both sound velocity and attenuation depend on the coordinate x' in a rarefied gas. It was confirmed numerically in the papers.[9–11] So, the representation (1) is not correct for high values of the Knudsen number.

To take into account the dependence of the wave vector k on x', some attempts, see, e.g., Refs. 12–17, were made to obtain the solution in the free molecular regime, i.e., when $\text{Kn}\to\infty$. But, these papers are not able to describe the free molecular sound propagation when the source–receptor distance is small. In this case there is a discrepancy between the results given in these papers. As a consequence they cannot explain the nonmonotonic dependence of the amplitude on the distance presented in Fig. 3 of the experimental work.[4] Such behavior can be explained by the influence of the receptor on the sound propagation characteristics.

[a]Electronic mail: sharipov@fisica.ufpr.br
[b]Electronic mail: marques@fisica.ufpr.br
[c]Electronic mail: kremer@fisica.ufpr.br

The aim of the present paper is to obtain the free-molecular solution of the sound propagation by considering the receptor influence. It will be shown that the receptor significantly changes the characteristics of the sound propagation in a rarefied gas. This is important to treat and to explain experimental data, since the size of the receptor in many situations cannot be neglected.

II. STATEMENT OF THE PROBLEM

We consider an infinite plate surrounded by a rarefied gas placed at $x'=0$, which oscillates harmonically with a normal velocity to the plate that depends on the time t' according to

$$u'(t')=u_0'\cos(\omega t')=\Re\{u_0'\exp(-i\omega t')\}. \tag{5}$$

The motion of the plate, henceforth called a sound source, is characterized by the angular frequency ω and the amplitude u_0', which is assumed to be small when compared with the most probable molecular velocity v_m of the gas, i.e.,

$$u_0' \ll v_m, \quad v_m=\left(\frac{2k_BT_0}{m}\right)^{1/2}. \tag{6}$$

Since the most probable velocity v_m has the same order as the sound velocity, the condition (6) means that the velocity of the plate is subsonic. Usually, this situation is realized in experiments, see, e.g., Refs. 2–4.

Another plate, which is parallel to the source and fixed at $x'=L'$, records a sound and is called a sound receptor. We assume that the molecular mean free path l is significantly larger than the distance L', i.e., the particles of the gas travel from the source to the receptor and back without undergoing any intermolecular collisions. Under this condition, the collisionless Boltzmann equation

$$\frac{\partial f}{\partial t'}+\mathbf{v}\cdot\frac{\partial f}{\partial \mathbf{r}'}=0, \tag{7}$$

can be applied. Here, $f(t',\mathbf{r}',\mathbf{v})$ is the one-particle distribution function, $\mathbf{r}'$ is the position vector, and $\mathbf{v}$ is the molecular velocity. It should be noted that to neglect the collision integral in the Boltzmann equations, only the condition $l \gg L'$ is enough. The relation of the mean free path to the receptor and source sizes does not matter, because the condition $l \gg L'$ means that the gaseous particles undergo significantly more collisions with the source and receptor surfaces than interparticle collisions.

In the next sections we shall obtain solutions corresponding to the following cases: (i) the receptor is absent; (ii) the receptor size is very small compared with the separation distance L'; and (iii) the receptor size is very large compared with L'. A comparison of the first and second solutions will show us that even if the receptor is small, its interaction with the gas must be taken into account. The third solution will give us the maximum value of the receptor influence to the sound propagation.

It is obvious that in the first and third cases the solution does not depend on the y' and z' coordinates. If the receptor size is finite, then the solution depends on all coordinates. However, the dependence of this solution on y' and z' coordinates is proportional to the area of the receptor surface, so that in the limit of very small receptor, i.e., the second case, this dependence can be neglected. Thus, in all above-mentioned cases the solution depends only on the x' coordinate, i.e., $f(t',\mathbf{r}',\mathbf{v})\equiv f(t',x',\mathbf{v})$.

For further derivations it is convenient to introduce the following dimensionless quantities:

$$c_x=\frac{v_x}{v_m}, \quad t=\omega t', \quad x=\frac{\omega}{v_m}x', \quad L=\frac{\omega}{v_m}L'. \tag{8}$$

Due to the condition (6), the distribution function $f(t',x',\mathbf{v})$ can be linearized as

$$f(t',x',\mathbf{v})=f_0\left[1+h(t,x,c_x)\frac{u_0'}{v_m}\right], \tag{9}$$

where $h(t,x,c_x)$ is the perturbation, f_0 is the equilibrium distribution function

$$f_0=n_0\left(\frac{m}{2\pi k_BT_0}\right)^{3/2}\exp\left(-\frac{mv^2}{2k_BT_0}\right), \tag{10}$$

and n_0 denotes the equilibrium number density of the gas. The linearized collisionless Boltzmann equation that follows from (7) and (9) reads

$$\frac{\partial h}{\partial t}+c_x\frac{\partial h}{\partial x}=0. \tag{11}$$

The perturbation function will be divided into two parts whether the direction of the molecular velocity is positive along to the x axis or negative, i.e.,

$$h(t,x,c_x)=\begin{cases} h^+(t,x,c_x) & \text{for } c_x>0, \\ h^-(t,x,c_x) & \text{for } c_x<0. \end{cases} \tag{12}$$

We assume the diffuse reflection of particles on the plates, i.e., at $x=0$ we have

$$h^+(t,0,c_x)=(\nu_0+2c_x)e^{-it}. \tag{13}$$

The quantity ν_0 is calculated from the condition

$$\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}\exp(-c_x^2)h(t,0,c_x)c_x\mathrm{d}c_x=e^{-it}, \tag{14}$$

which means the plate is impenetrable for the gas, i.e., the bulk velocity of the gas at $x=0$ is equal to the plate velocity. At the receptor surface $x=L$, we have

$$h^-(t,L,c_x)=\nu_L e^{-it}. \tag{15}$$

To calculate the quantity ν_L, the following condition is applied:

$$\int_{-\infty}^{\infty}\exp(-c_x^2)h(t,L,c_x)c_x\mathrm{d}c_x=0, \tag{16}$$

which means that the receptor is also impenetrable and the bulk velocity of the gas is zero at its surface.

The sound receptor records a normal flux of alternating momentum on its surface, which is related to the distribution function as

$$P_{xx}'(t',L')=m\int f(t',L',\mathbf{v})v_x^2\,\mathrm{d}\mathbf{v}-P_0, \tag{17}$$

where $P_0 = n_0 k T_0$ is the equilibrium pressure. In terms of the dimensionless quantities, the normal momentum flux is defined by

$$P_{xx}(t,L) = \frac{P'_{xx}}{2P_0} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} \exp(-c_x^2) h(t,L,c_x) c_x^2 \mathrm{d}c_x, \tag{18}$$

which can be represented as

$$P_{xx}(t,L) = p(L) \exp[i(\varphi(L) - t)]. \tag{19}$$

The quantity $p(L)$ is the amplitude of the momentum normal flux P_{xx}, while $\varphi(L)$ is its phase.

Usually, the attenuation coefficient is calculated via the derivative of the amplitude with respect to the distance L. Following this definition, we introduce the dimensionless attenuation as

$$\alpha = -\sqrt{\frac{\gamma}{2}} \frac{\mathrm{d} \ln p(L)}{\mathrm{d}L}. \tag{20}$$

We shall express the phase velocity v_{ph} through the dimensionless phase parameter β as

$$v_{\mathrm{ph}} = \frac{c}{\beta}, \tag{21}$$

where β is calculated through the phase $\varphi(L)$. Usually, this parameter is calculated as the slope of the phase $\varphi(L)$ with respect to the distance L. The phase velocity calculated through such defined β characterizes the signal transmission velocity between two close points in the gap between the source and receptor. But, if one is interested in the signal transmission velocity from a source to a receptor, one should define the parameter β through the phase difference between the source and the receptor. So, it makes sense to introduce two phase parameters β

$$\beta_d = \sqrt{\frac{\gamma}{2}} \frac{\mathrm{d}\varphi(L)}{\mathrm{d}L}, \tag{22}$$

and

$$\beta_i = \sqrt{\frac{\gamma}{2}} \frac{\varphi(L)}{L}. \tag{23}$$

It is appropriate to call the first β_d a differential phase parameter and the second β_i an integral phase parameter. As a result, we have the two-phase velocity: the differential one calculated by (21) with β_d and the integral one calculated by the same manner with β_i. In the hydrodynamic regime it is possible to show that $\beta_d = \beta_i = 1$, i.e., both differential and integral phase velocities coincide. Below, it will be shown that in the free-molecular regime these two velocities are different.

III. SOLUTION WITHOUT RECEPTOR

Here, we shall obtain the solution when the receptor is absent. In this case the distribution function of the particles that are going to the source is not perturbed, i.e.,

$$h^-(t,0,c_x) = 0. \tag{24}$$

Substituting (13) and (24) into (14) and integrating the resulting equation, we obtain $\nu_0 = \sqrt{\pi}$.

Furthermore, the integration of (11) with respect to x from 0 up to some distance L leads to the expression

$$h^+(t,L,c_x) = h^+\left(t - \frac{L}{c_x}, 0, c_x\right) = (\sqrt{\pi} + 2c_x) \exp\left[-i\left(t - \frac{L}{c_x}\right)\right]. \tag{25}$$

Physically, this equation means that the faster particles reach the receptor earlier than the slow ones. We substitute (24) and (25) into (18) and get the expression for the normal momentum flux

$$P_{xx}(t,L) = \left(I_2 + \frac{2}{\sqrt{\pi}} I_3\right) e^{-it}, \tag{26}$$

where I_n denotes the following integral:

$$I_n = \int_0^\infty c^n \exp\left(-c^2 + \frac{iL}{c}\right) \mathrm{d}c. \tag{27}$$

Using the power series representation given by Eq. (27.5.4) of Ref. 18, we obtain the asymptotic behavior of the amplitude $p(L)$ and the phase $\varphi(L)$ at small distances $L \ll 1$

$$p(L) \to \left(\frac{\sqrt{\pi}}{4} + \frac{1}{\sqrt{\pi}}\right), \tag{28}$$

$$\varphi(L) \to \left(\frac{\sqrt{\pi}}{4} + \frac{1}{\sqrt{\pi}}\right)^{-1} L. \tag{29}$$

Then, the limiting values of the attenuation α and of the phase parameters β_d and β_i can be easily calculated

$$\lim_{L \to 0} \alpha(L) = 0, \tag{30}$$

$$\lim_{L \to 0} \beta_d(L) = \lim_{L \to 0} \beta_i(L) = \sqrt{\frac{\gamma}{2}} \left(\frac{\sqrt{\pi}}{4} + \frac{1}{\sqrt{\pi}}\right)^{-1}. \tag{31}$$

For large distances $L \gg 1$, the normal momentum flux (26) reduces to

$$P_{xx} \to \frac{2}{\sqrt{\pi}} I_3, \tag{32}$$

so that by using the representation (27.5.8) given in Ref. 18, we have

$$p(L) \to \frac{L}{\sqrt{3}} \exp\left[-\frac{3}{2}\left(\frac{L}{2}\right)^{2/3}\right], \tag{33}$$

$$\varphi(L) \to \frac{3^{3/2}}{2}\left(\frac{L}{2}\right)^{2/3}. \tag{34}$$

Hence, the asymptotic representations for the attenuation and phase parameter are

$$\alpha \to \sqrt{\frac{\gamma}{2}}\frac{1}{(4L)^{1/3}}, \tag{35}$$

$$\beta_d \to \sqrt{\frac{\gamma}{2}}\frac{\sqrt{3}}{(4L)^{1/3}}, \quad \beta_i \to \frac{3}{2}\beta_d. \tag{36}$$

We note that the above expressions for α and β_d coincide with those obtained for large separation distances by Maidanik *et al.* [see Eq. (6.6) of Ref. 12] and by Toba [see Eq. (14) of Ref. 15 and Eq. (4.7) of Ref. 16 in the limit $Z\to\infty$].

IV. SOLUTION FOR VERY SMALL RECEPTOR

The solution (26) would be valid in the presence of the receptor too if it could absorb all incident particles. However, the receptor reflects the particles of the gas and as a result it receives an additional quantity of the linear momentum that was not taken into account in all previous papers. Here, we consider that the receptor size is so small that it does not perturb the distribution function of particles which are moving towards the sound source. In this case, the perturbation function in the whole space between the source and receptor is the same as in the previous section except at the point where the receptor is placed. At this point the bulk velocity of the gas is zero and the quantity ν_L is obtained by substituting (15) and (25) into (16). After the integration, we obtain

$$\nu_L = 2\sqrt{\pi} I_1 + 4 I_2. \tag{37}$$

Further, with the help of (15), (18), (25), and (37), we determine the quantity P_{xx} on the receptor surface

$$P_{xx}(t,L) = \left[\frac{\sqrt{\pi}}{2} I_1 + 2 I_2 + \frac{2}{\sqrt{\pi}} I_3\right] e^{-it}. \tag{38}$$

One can see that even if the receptor size is negligible, its interaction with the gas modifies the solution (26).

Using the expansion (27.5.4) of Ref. 18, the asymptotic behaviors of the amplitude $p(L)$ and the phase $\varphi(L)$ at small distances $L \ll 1$ are obtained

$$p(L) \to \left(\frac{3\pi}{4} + \frac{1}{\sqrt{\pi}}\right), \tag{39}$$

$$\varphi(L) \to \left(\frac{\pi+6}{3\sqrt{\pi}+4/\sqrt{\pi}}\right) L. \tag{40}$$

From the above equations it is easy to calculate the limiting values of the quantities α, β_d, and β_i

$$\lim_{L\to 0} \alpha(L) = 0, \tag{41}$$

$$\lim_{L\to 0} \beta_d(L) = \lim_{L\to 0} \beta_i(L) = \sqrt{\frac{\gamma}{2}}\left(\frac{\pi+6}{3\sqrt{\pi}+4/\sqrt{\pi}}\right). \tag{42}$$

Note that this expression for the phase parameters β_d and β_i are different from that obtained in the previous section; see Eq. (31).

One can verify that in the limit $L \gg 1$ the behavior of the normal momentum flux P_{xx} is the same as in the previous case; see (32).

V. SOLUTION FOR VERY LARGE RECEPTOR

In the previous section it has been shown that even a small receptor modifies the expression for the normal momentum flux at the receptor. Hence, the influence of the receptor should be taken into account in order to compare the theoretical results with the experimental data. To evaluate how strong the influence of the receptor is, we consider in this section the other limit with respect to the receptor size, namely, we suppose the receptor as an infinite plate parallel to the source plate.

If we integrate Eq. (11) from L to 0, we obtain the perturbation function h^- at the sound source $x=0$

$$h^-(t,0,c_x) = h^-\left(t + \frac{L}{c_x}, L, c_x\right) = \nu_L \exp\left[-i\left(t + \frac{L}{c_x}\right)\right]. \tag{43}$$

We substitute the function $h^+(t,0,c_x)$ at the sound source (13) together with (43) into the condition (14) and get the following relation between ν_0 and ν_L:

$$\frac{1}{2}\nu_0 - I_1 \nu_L = \frac{\sqrt{\pi}}{2}. \tag{44}$$

In order to determine the perturbation function $h^+(t,L,c_x)$ at the receptor, we integrate Eq. (11) from 0 to L and obtain

$$h^+(t,L,c_x) = h^+\left(t - \frac{L}{c_x}, 0, c_x\right) = (\nu_0 + 2c_x)\exp\left[-i\left(t - \frac{L}{c_x}\right)\right]. \tag{45}$$

Now, we insert (15) and (45) into the condition (16) and obtain one more relationship between ν_0 and ν_L

$$I_1 \nu_0 - \tfrac{1}{2}\nu_L = -2 I_2. \tag{46}$$

From the system of equations (44) and (46) for the quantities ν_0 and ν_L, it follows that

$$\nu_0 = \frac{\sqrt{\pi} + 8 I_1 I_2}{1 - 4 I_1^2}, \tag{47}$$

$$\nu_L = \frac{4 I_2 + 2\sqrt{\pi} I_1}{1 - 4 I_1^2}. \tag{48}$$

The expression for the quantity P_{xx} is obtained through the substitution of (47) and (48) into (45) and (15), respectively. Then, the insertion of the resulting equations into (18) and integration leads to

$$P_{xx}(t,L) = \left[\frac{4 I_2 \sqrt{\pi} + \pi I_1 + 16 I_1 I_2^2}{2\sqrt{\pi}(1 - 4 I_1^2)} + \frac{2}{\sqrt{\pi}} I_3\right] e^{-it}. \tag{49}$$

The asymptotic behavior of the amplitude $p(L)$ and the phase $\varphi(L)$ when $L \ll 1$ is obtained with the help of Eq. (27.5.4) of Ref. 18

$$p(L) \to \frac{1}{2L}, \tag{50}$$

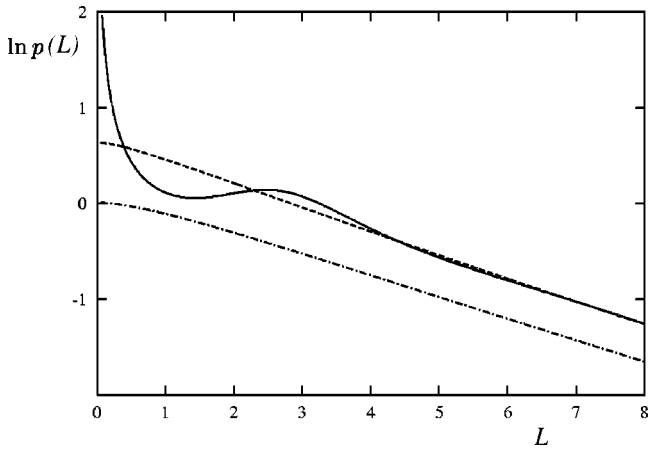


FIG. 1. Amplitude logarithm $\ln p$ vs L: dashed-pointed curve—receptor is absent; dashed curve—very small receptor; solid curve—very large receptor.

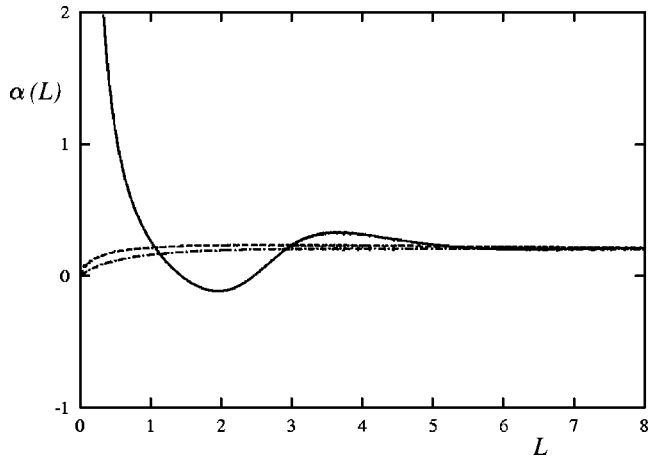


FIG. 3. Attenuation α vs L: dashed-pointed curve—receptor is absent; dashed curve—very small receptor; solid curve—very large receptor.

$$\varphi(L)\to\frac{\pi}{2}+\frac{1}{\sqrt{\pi}}L\ln L. \tag{51}$$

Then, by using the definitions (20), (22), and (23), we obtain

$$\alpha\to\sqrt{\frac{\gamma}{2}}\frac{1}{L}, \tag{52}$$

$$\beta_d\to\sqrt{\frac{\gamma}{2\pi}}\ln L,\quad \beta_i\to\frac{\pi}{2L}. \tag{53}$$

In case of large distances $L\gg 1$, the same asymptotic behavior for P_{xx} as in the two previous situations follows; see (32).

VI. DISCUSSION

The logarithm of the amplitude $\ln p(L)$ and the phase $\varphi(L)$ are plotted in Figs. 1 and 2, respectively, as function of the distance L for the three situations: (i) the solution without receptor is represented by the dashed-pointed curve; (ii) the solution for the very small receptor is shown by the dashed curve; and (iii) the solution for the very large receptor is traced by the solid curve. One can see that the second solution gives a larger amplitude $p(L)$ than the first one, because the particle reflection taken into account in the second solution increases the linear momentum quantity transfered from the gas to the surface. But, the shapes of both plots are the same in the whole range of the parameter L. Both solutions give practically the same phases $\varphi(L)$.

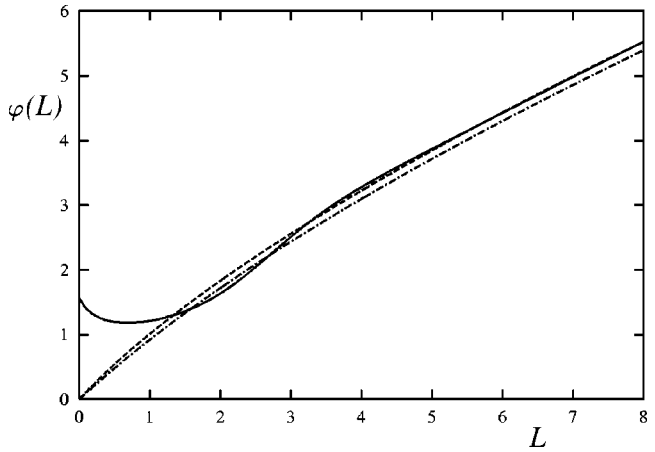


FIG. 2. Phase φ vs L: dashed-pointed curve—receptor is absent; dashed curve—very small receptor; solid curve—very large receptor.

Comparing the second and the third solutions, we see that the difference between them is significant for small values of L. This is explained by the fact that when a large receptor is close to the source, i.e., $L\ll 1$, there are many particles of the gas, which undergo multiple reflections at the source and receptor surfaces during one cycle of the sound source, which is equal to $2\pi/\omega$. But, when the receptor is very small the particles colliding with it undergo only one reflection from the source. So, under the condition $L\ll 1$ the large receptor records significantly more molecular collisions per unity of area than the small one. Comparing Eqs. (39) and (40) with Eqs. (50) and (51), one can infer that the asymptotic behaviors of these two solutions are quite different. Moreover, the amplitude $p(L)$ tends to infinity for the large receptor, because by decreasing the distance L the number of collisions with the large receptor sharply increases. It must be noted that the solid curve in Fig. 1 has a local minimum at $L\approx 1$. Such behavior qualitatively agrees with the experimental curve given in Fig. 3 of Ref. 4. Unfortunately, a quantitative comparison with these experimental results is impossible since the author of this paper did not provide the geometrical characteristics of the source and receptor. To our knowledge, no reliable experimental data providing such geometrical characteristics are available in the open literature.

When the distance is large, $L\gg 1$, the sound source makes several cycles while the particles of the gas cross the gap between the source and the receptor. In this situation both large and small receptors record the same number of molecular collisions per unity of area. From Figs. 1 and 2 one can see that, beginning from $L=5$, the second and third solutions are practically the same. As has been mentioned above, these two solutions have the same asymptotic behaviors when $L\gg 1$.

The attenuation α calculated by (20) assuming the gas to be monoatomic ($\gamma=5/3$) is plotted in Fig. 3. The first and

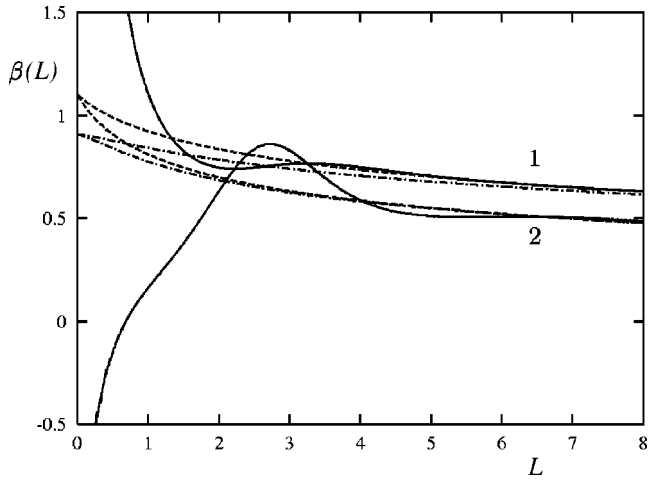


FIG. 4. Phase parameter β vs L: dashed-pointed curve—receptor is absent; dashed curve—very small receptor; solid curve—very large receptor; 1—differential; 2—integral.

second solutions provide α very close to each other. The third solution gives quite a different behavior of the attenuation α: at $L \ll 1$ it tends to infinity, while the first and second solutions give $\alpha = 0$. The other feature of the third solution is that the attenuation α has a negative value near $L = 2$, which seems to be an unphysical situation. However, we note the present solution is constructed as a superposition of incident and reflected particles. As a result of this superposition, the amplitude of perturbation may both increase and decrease when the distance L varies, such that the attenuation may be both positive and negative.

At the large distance $L \gg 1$, all solutions provide the same asymptotic value given by (35). Note that the attenuation does not tend to a constant value, but slowly decreases.

The phase parameters β_d and β_i calculated by (22) and (23), respectively, are plotted in Fig. 4 for the three solutions. The first and second solutions differ slightly for small distances $L \ll 1$. Both solutions provide equality between the differential and integral phase parameters, i.e., $\beta_d = \beta_i$ at $L = 0$, while the third solution gives the quite different behavior of β_d and β_i for small values of the distance L. From (53) one can see that $\beta_d \to -\infty$ and $\beta_i \to -\infty$ when $L \to \infty$. Moreover, the differential phase parameter β_d is negative in the range $0 \leqslant L < 1$. This leads to a negative sound velocity, which is an unphysical behavior. As has been explained above, the differential phase parameter characterizes the sound transmission between two points close to each other. However, if one uses a large receptor to measure the phase in two close points, one measures two different waves, because a small displacement of the large receptor significantly changes the sound characteristics. As a result, one may obtain a negative phase velocity, which does not make sense. So, a large receptor cannot be used to measure the differential phase parameter. But, to characterize the signal transmission from a source to a large receptor, the integral phase velocity should be used.

At large distance $L \gg 1$ all solutions provide the same behavior of the phase parameter. According to Eq. (36) both parameters β_d and β_i slowly tend to zero. For very large distance L the parameters are related as $\beta_i = \frac{3}{2}\beta_d$.

So, in the free-molecular regime the differential phase velocity related to β_d and the integral one related to β_i are quite different quantities.

VII. CONCLUDING REMARKS

The exact solution of the free-molecular sound propagation was obtained from the collisionless Boltzmann equation. The solution is valid for a highly rarefied gas when the particles of the gas traveling from a sound source to a receptor do not undergo any intermolecular collision. It was shown that for small values of the source–receptor distance the influence of the receptor on the sound propagation is significant and must be taken into account in comparison of theoretical results with experimental data. The results when the influence of the receptor is taken into account qualitatively coincide with experimental data presented in Ref. 4. To perform a quantitative comparison, the same problem must be solved for a finite size of both source and receptor by applying the two-dimensional Boltzmann equation. This task could be carried out if the experimental papers would provide these characteristics.

To characterize the sound velocity, the differential and integral phase parameters were introduced. For a large receptor only the integral phase parameter makes physical sense, while the differential one gives a negative value of the phase velocity.

ACKNOWLEDGMENT

The authors, F.Sh. and G.M.K., thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil), for the support of the present work.

[1] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon, New York, 1989).
[2] M. Greenspan, "Propagation of sound in five monatomic gases," J. Acoust. Soc. Am. **28**, 644–648 (1956).
[3] G. Maidanik and M. Heckl, "Propagation and reflection of sound in rarefied gases. II. Experimental," Phys. Fluids **8**, 266–272 (1965).
[4] R. Schotter, "Rarefied gas acoustics in the noble gases," Phys. Fluids **17**, 1163–1168 (1974).
[5] C. S. Wang Chang and G. E. Uhlenbeck, "On the propagation of sound in monatomic gases," Stud. Stat. Mech. **5**, 43–75 (1970).
[6] L. Sirovich and J. K. Thurber, "Propagation of forced sound waves in rarefied gas dynamics," J. Acoust. Soc. Am. **37**, 329–339 (1965).
[7] J. K. Buckner and J. H. Ferziger, "Linearized boundary value problem for a gas and sound propagation," Phys. Fluids **9**, 2315–2322 (1966).
[8] W. Marques, Jr., "Dispersion and absorption of sound in monatomic gases: An extended kinetic description," J. Acoust. Soc. Am. **106**, 3282–3288 (1999).
[9] S. K. Loyalka and T. C. Cheng, "Sound-wave propagation in a rarefied-gas," Phys. Fluids **22**, 830–836 (1979).
[10] J. R. Thomas and C. E. Siewert, "Sound-wave propagation in a rarefied gas," Transp. Theory Stat. Phys. **8**, 219–240 (1979).
[11] N. G. Hadjiconstantinou and A. L. Garcia, "Molecular simulation of sound wave propagation in simple gases," Phys. Fluids **13**, 1040–1046 (2001).
[12] G. Maidanik, H. L. Fox, and M. Heckl, "Propagation and reflection of sound in rarefied gases. I. Theoretical," Phys. Fluids **8**, 259–265 (1965).

[13] D. Kahn and D. Mintzer, ''Kinetic theory of sound propagation in rarefied gases,'' Phys. Fluids **8**, 1090–1102 (1965).
[14] D. Kahn, ''Sound propagation in rarefied gases,'' Phys. Fluids **9**, 1867–1869 (1966).
[15] K. Toba, ''Kinetic theory of sound propagation in a rarefied gas,'' Phys. Fluids **11**, 2495–2497 (1968).
[16] K. Toba, ''Effect of gas–surface interaction on sound propagation,'' Phys. Fluids **11**, 507–514 (1968).
[17] F. B. Hanson and T. F. Morse, ''Free-molecular expansion polynominals and sound propagation in rarefied gases,'' Phys. Fluids **12**, 1564–1572 (1969).
[18] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972).

Scattering from a ribbed finite cylindrical shell with internal axisymmetric oscillators

Michel Tran-Van-Nhieu[a)]
Thales Underwater Systems, 7–9 rue des Mathurins, 92 221, Bagneux, France

(Received 18 December 2001; revised 2 May 2002; accepted 3 May 2002)

A theoretical formalism is elaborated to determine the acoustic field scattered from a finite cylindrical shell reinforced by a set of ribs with internal axisymmetric oscillators, the locations and properties of which can vary arbitrarily along the cylinder axis. Analytical expressions are derived for the scattered pressure and approximate expressions are proposed to calculate the backscattered field. Some applications have been carried out to investigate the problem of backscattering from a periodically ribbed finite cylindrical shell in the presence of rib aperiodicity or axisymmetric internal oscillators. The modifications of the main features associated with scattering from helical and Bloch–Floquet waves are examined and numerical results are presented with respect to the mean rib spacing and to the total mass of the internal structures. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1488659]

PACS numbers: 43.20.Fn, 43.40.Ey [LLT]

I. INTRODUCTION

The effect of stiffening members or reinforcing ribs on the scattering from elastic structures and particularly from finite cylindrical shell has been investigated in several papers.[1–4] It has been shown that when the ribs are periodically attached to the surface of the cylinder, the farfield scattered pressure displays a certain number of typical features associated with specular reflection and phase matching to the free waves that propagate on the shell, namely, the helical[5] and Bloch–Floquet waves.[1,2] Furthermore, it has been found that the farfield backscattering pattern of the Bloch–Floquet waves exhibits complex features related to the high azimuthal mode numbers.[2–4]

However, periodically ribbed finite cylindrical shells are rather simplified objects to model real marine structures, although they could be useful as a first approach to deal with the problem, since certain characteristics persist in real systems. Indeed naval underwater structures are often stiffened by ribs that are spaced more or less uniformly with weak deviations from a spatial periodicity. Furthermore, the elastic waves can interact with other discontinuities such as bulkheads,[6–8] and internal structures[9–13] so that the acoustic field could be greatly modified compared to that of the periodic case. Therefore it is of interest to derive a theoretical formalism that could predict the effects of irregularity and discontinuities on the above-mentioned features in the scattering process.

In an attempt to deal with the problem of scattering from a ribbed shell with internal oscillators and in order to simplify the theoretical analysis, it has been assumed in the present investigation that the discontinuities are axisymmetric and that their reactive forces interact with the surface of the shell only through their normal components. The present approach is an extension of a theory that was elaborated for a periodically ribbed finite cylindrical shell.[4] The theoretical formalism is presented in Sec. II and the expressions of the scattered pressure are derived in Sec. III. The obtained equations are applied in Sec. IV first to compute the farfield pressure backscattered from a finite cylindrical shell stiffened by a set of periodic ribs. Then the effects of rib aperiodicity and internal oscillators on the scattered field are, respectively, analyzed versus the deviations from the mean rib spacing and the mass of the oscillators. Finally, the principal results are summarized in Sec. V.

II. THEORY

A harmonic plane wave $p_i = e^{i(k_x x + k_z z)} e^{-i\omega t}$ is incident upon a long cylindrical shell that is submerged in water. Let a and L, respectively, represent the radius and the half-length of the cylinder. The interior of the shell is *in vacuo* and a set of ribs and/or internal oscillators are attached to the inside of the surface of the cylinder (see Fig. 1). In this section, the word ribs will be used to designate both ribs and/or internal structures for convenience.

In the present work, the analysis is made under the following assumptions: (H1) the harmonic time dependence is in $e^{-i\omega t}$; (H2) the cylindrical shell is slender: $a/L \ll 1$; (H3) the motion of the shell is described by Donnell's thin shell theory;[14] (H4) only the normal component of the reactive force applied by the ribs to the shell is considered; (H5) the positions of the different ribs on the z axis are given by $z = z_l$ for $l = 1:M$; (H6) the ends of the shell are assumed to be simply supported; and (H7) the acoustic effects of the end sections are neglected.

These assumptions are the same as in Ref. 4, except that concerning the ribs: indeed in the present analysis, their properties and locations can vary arbitrarily on the z axis though preserving the axisymmetry of the problem and, furthermore, it is assumed that their mechanical vibrations are completely determined by their modal impedances. The restrictions associated with the assumptions (H1)–(H7) are discussed elsewhere[4] and the reader can refer to this paper for further details.

a)Electronic mail: michel.tranvannhieu@fr.thalesgroup.com

 0001-4966/2002/112(2)/402/9/$19.00

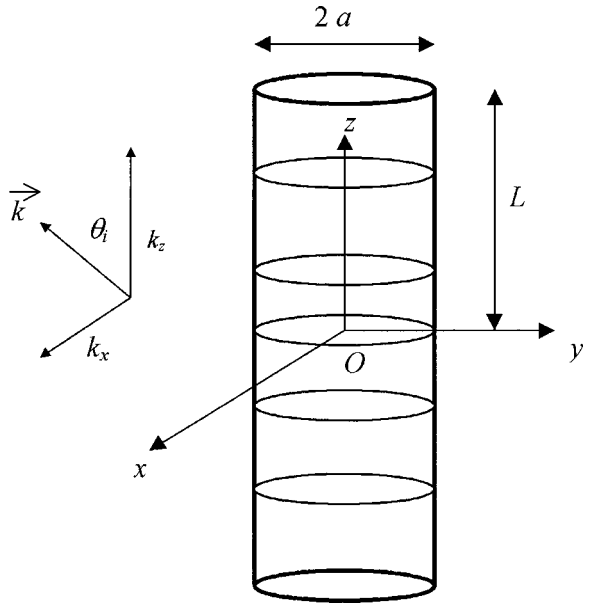


FIG. 1. Geometry of the ribbed cylindrical shell.

The motion of the shell is both driven by the sound pressure and the forcing terms due to the reinforcing ribs. The total pressure that is exerted on the cylinder by the fluid and by the ring supports is the sum of the incident pressure p_i, the scattered pressure p_s and the reactive force applied by the rings to the shell p_r.

Let the scattered pressure be split up into two components:

$$p_s = p_{\text{rig}} + p_{\text{res}}, \tag{1}$$

in which p_{rig} is the pressure scattered by the cylinder considered as a perfectly rigid object and p_{res} is due to the elasticity of the material. Proceeding as in Ref. 15, p_{rig} is expressed in cylindrical coordinates (r,φ,z) in the form

$$p_{\text{rig}} = -\frac{k}{\pi}\sum_{n=0}^{\infty} \epsilon_{\text{N}} i^n \frac{\cos n\varphi}{\sin^n \theta_i}\frac{J_n'}{H_n'}(k_x a) \times \int_{-L}^{L} h_n(k\rho)\left(\frac{r}{\rho}\right)^n e^{ik_z z_0}\, dz_0, \tag{2a}$$

while the second component of the scattered pressure p_{res} is sought as follows:

$$p_{\text{res}} = \frac{k}{\pi}\sum_{n=0}^{\infty}\sum_{p=1}^{\infty} \epsilon_{\text{N}} i^n \cos n\varphi\, b_{np} \times \int_{-L}^{L} h_n(k\rho)\left(\frac{r}{\rho}\right)^n \phi_p(z_0)\, dz_0. \tag{2b}$$

In the above equation, $k=\omega/c$ is the acoustic wave number, $k_x = k \sin\theta_i$, $k_z = k\cos\theta_i$, and the coefficients b_{np} are to be determined. J_n is the cylindrical Bessel function of order n, H_n and h_n are, respectively, the cylindrical and spherical Hankel function of order n of the first kind. The prime denotes the differentiation with respect to the argument, ϵ_{N} is the Neumann factor, $\epsilon_{\text{N}}=1$ for $n=0$ and $\epsilon_{\text{N}}=2$ for $n>0$, and $\rho=\sqrt{(z_0-z)^2+r^2}$.

The eigenfunctions ϕ_p and the axial wave numbers k_p of the finite cylindrical shell with simply supported ends are given by Eq. (3),

$$\phi_p(z) = \sin k_p(z+L), \quad \text{with } k_p = \frac{\pi p}{2L}, \quad \text{for } p = 1{:}\infty. \tag{3}$$

In order to simplify the analysis of the problem, it is assumed that the reactive force p_r interacts with the cylinder only through normal forces, whereas moments and shear forces are neglected. Thus, p_r may be written as

$$p_r = \sum_{n=0}^{\infty}\sum_{q=1}^{\infty}\sum_{l=1}^{M} \epsilon_{\text{N}} i^n \cos n\varphi\, Z_{nl}^r w_{nq} \phi_q(z_l)\,\delta(z-z_l), \tag{4}$$

where w_{nl} are the modal component of the normal displacement [see Eq. (6)] and $-Z_{nl}^r/i\omega$ is the modal impedance of the rib located at $z=z_l$.

Note that the results associated with shear and compressional wave scattering might change considerably if tangential forces are imposed at the attachments connecting the shell and the ribs.[6] However, it is expected that despite its simplified form, Eq. (4) enables one to grasp at the dominant physical features of the present problem with a limited amount of calculation.

The motion of the cylindrical shell is described by an approximate thin shell theory, namely, the Donnell equations.[14] This method that takes into account both the membrane and bending effects is used most widely in shell vibrations, and it is generally applied to study transverse motions of shells that are loaded normal to their surface. These equations may be written as

$$\begin{aligned}
&\frac{\partial^2 u}{\partial z^2} + \frac{\mu^-}{a^2}\frac{\partial^2 u}{\partial \varphi^2} + \frac{\mu^+}{a}\frac{\partial^2 v}{\partial z\,\partial\varphi} + \frac{\mu}{a}\frac{\partial w}{\partial z} + k_s^2 u = 0,\\
&\frac{\mu^+}{a}\frac{\partial^2 u}{\partial z\,\partial\varphi} + \mu^-\frac{\partial^2 v}{\partial z^2} + \frac{1}{a^2}\frac{\partial^2 v}{\partial\varphi^2} + \frac{1}{a^2}\frac{\partial w}{\partial\varphi} + k_s^2 v = 0,\\
&\frac{\mu}{a}\frac{\partial u}{\partial z} + \frac{1}{a^2}\frac{\partial v}{\partial\varphi} + \frac{w}{a^2} + \frac{h^2}{12}\left(\frac{\partial^2}{\partial z^2} + \frac{1}{a^2}\frac{\partial^2}{\partial\varphi^2}\right)^2 w - k_s^2 w\\
&\qquad = -\frac{p_i + p_{\text{rig}} + p_{\text{res}} + p_r}{\rho_s c_s^2 h},
\end{aligned} \tag{5}$$

where (u,v,w) are the axial, circumferential, and radial components of the shell displacements, $c_s = \sqrt{E_s/\rho_s(1-\mu^2)}$ is the longitudinal wave speed, μ is Poisson's constant, ρ_s is the shell density, E_s is the Young's modulus of the shell material, h is the shell thickness, $\mu^{\pm} = (1\pm\mu)/2$, and $k_s = \omega/c_s$.

In order to solve the above equations with the simply supported boundary conditions at the ends of the cylinder, the unknown components of the shell displacements are expanded in the eigenfunctions of the finite cylindrical shell:

$$u(z,\varphi) = \sum_{n=0}^{\infty}\sum_{p=1}^{\infty} \epsilon_{\text{N}} i^n \cos n\varphi\, \cos[k_p(z+L)]\, u_{np},$$

$$v(z,\varphi)=\sum_{n=0}^{\infty}\sum_{p=1}^{\infty}\epsilon_N i^n \sin n\varphi\ \phi_p(z)v_{np},\tag{6}$$

$$w(z,\varphi)=\sum_{n=0}^{\infty}\sum_{p=1}^{\infty}\epsilon_N i^n \cos n\varphi\ \phi_p(z)w_{np}.$$

The calculation is then similar to what has been done previously for a set of periodic ribs: first, the expressions of (p_i+p_{rig}) and p_{res} near the surface of the cylinder can be found by applying Eqs. (A3) and (A4) of Ref. 4. Then the expression of b_{np} occurring in the expansion of p_{res} is determined from the condition of continuity of the radial displacement at the interface between the shell and the fluid; this yields

$$b_{np}=\left(\frac{k}{\alpha_p}\right)^n\frac{\rho_f\omega^2 w_{np}}{\alpha_p H_n'(\alpha_p a)}.\tag{7}$$

Inserting Eq. (6) into Eq. (5) and using Eqs. (4) and (7), the following infinite linear system in the unknown modal amplitude W_{np} is found after some tedious calculation:

$$W_{np}+\sum_{l=1}^{M}D_{npl}S_{nl}=-A_{np}F(k_zL,k_pL),\tag{8}$$

for $p=1{:}\infty$ and $n=0{:}\infty$.

In the above-mentioned equation, W_{np} is related to the modal amplitude of the normal displacement w_{np} by the relationship

$$W_{np}=\left(\frac{2}{\pi k_x a H_n'(k_x a)}\right)^{-1}\rho_f\omega^2 a w_{np},\tag{9}$$

while D_{npl}, S_{nl} are defined as

$$\begin{aligned}&S_{nl}=\sum_{q=1}^{\infty}\phi_q(z_l)W_{nq},\\&D_{npl}=A_{np}A_{nl}^{r\,-1}\phi_p(z_l).\end{aligned}\tag{10}$$

In Eq. (8), the function $F(x,y)$ is given by

$$F(x,y)=\frac{\sin(x+y)}{x+y}e^{iy}-\frac{\sin(x-y)}{x-y}e^{-iy}\tag{11}$$

and

$$\begin{aligned}&A_{np}=[Z_n^f(\alpha_p a)-Z_n^s(k_p a)]^{-1};\\&Z_n^f(\alpha)=\frac{H_n(\alpha)}{\alpha H_n'(\alpha)};\quad A_{nl}^r=\left[\frac{Z_{nl}^r}{\rho_f\omega^2 aL}\right]^{-1}.\end{aligned}\tag{12}$$

The expression of Z_n^s that occurs in the definition of A_{np} is

$$Z_n^s(\alpha)=\frac{\rho_s h}{\rho_f a\Omega^2}\left\{\Omega^2-\beta_s^2(\alpha^2+n^2)^2-\frac{\Omega^4-\mu^-[\Omega^2n^2+(2\mu+3)\Omega^2\alpha^2-4\mu^-\mu^+\alpha^4]}{[\mu^-(\alpha^2+n^2)-\Omega^2](\alpha^2+n^2-\Omega^2)}\right\},\tag{13}$$

where $\beta_s=h/a\sqrt{12}$, $\Omega=\omega a/c_s$ is a dimensionless frequency parameter, and it is equal to the frequency divided by the ring frequency $f_r=c_s/2\pi a$.

Equation (8) can be solved directly by truncation by retaining a finite number of terms in the equations. However, an exact and more efficient method can be used to transform this infinite system into an M-by-M linear system. Indeed, let the coefficients $\overline{S_{nl}}$ be defined as $\overline{S_{nl}}=-A_{nl}^{r\,-1}S_{nl}$ for $l=1{:}M$; then multiplying both members of Eq. (8) by $\phi_p(z_j)$ and summing from $p=1{:}\infty$ for $j=1{:}M$ gives

$$\sum_{l=1}^{M}\Lambda_{njl}\,\overline{S_{nl}}=\overline{\Delta}_{nj},\tag{14}$$

where

$$\overline{\Delta}_{nj}=\sum_{p=1}^{\infty}A_{np}\phi_p(z_j)F(k_zL,k_pL),\tag{15a}$$

$$\Lambda_{njl}=\sum_{p=1}^{\infty}A_{np}\phi_p(z_j)\phi_p(z_l)+A_{nl}^r\delta_{jl},\tag{15b}$$

for $n=0{:}\infty$ and $j=1{:}M$. In the above equation, δ_{jl} is the Kronecker's symbol: $\delta_{jl}=0$ if $j\neq l$ and $\delta_{jl}=1$ if $j=l$.

The transformation of Eq. (8) to Eq. (14) constitutes the key point of the present approach since it allows one to carry out the calculation for $ka\gg 1$: indeed, in the case of a set of periodic or almost periodic spaced ribs as studied in the next section, the order of truncation of Eq. (8) that would be necessary for a good spatial resolution to describe the scattered field in the frequency range of interest, is usually very large compared to M (of the order of $8M$ at the highest frequencies) so that solving Eq. (8) would limit the computation to the low frequencies in practice. Furthermore, the above transformation enables one to avoid all the numerical difficulties associated with the resolution of very large linear systems.

By using Eq. (14) the expression of the scattered pressure can be determined from Eqs. (2), (7), and (8). Therefore, unlike the case of a set of periodic spaced ribs, where the solution is known explicitly, the present calculation requires the resolution of an algebraic linear system for each azimuthal mode number. More precisely, the inversion of the matrix, Λ_n the elements of which are Λ_{njl} defined by Eq. (15b) is necessary if the computation is carried out for several incidence aspect angles as the second member of Eq. (14) is a function of $k_z=k\cos\theta_i$. When the calculation concerns only a few incident angles, it would be better to solve this linear system directly: indeed, since the rank of Λ_n is equal to the number of the attached ribs, the computation load increases considerably when the number of ribs is large.

III. FARFIELD SCATTERED PRESSURE

The expression of the farfield scattered pressure is obtained by inserting the asymptotic expansion of spherical Hankel function h_n for large arguments into Eq. (2), it becomes

$$p_s=\frac{e^{ik\rho_0}}{\rho_0}(D_{\rm rig}+D_{\rm res}^{ucs}+D_{\rm res}^{r}), \tag{16}$$

when the distance from the observation point to the origin O denoted by ρ_0 tends to infinity.

In the above-mentioned equation, $D_{\rm rig}$, $D_{\rm res}^{ucs}$, and $D_{\rm res}^{r}$ are given by

$$D_{\rm rig}=\frac{2i}{\pi k}\frac{\sin[kL(\cos\,\theta_i-\cos\,\theta)]}{\cos\,\theta_i-\cos\,\theta}\times\sum_{n=0}^{\infty}\epsilon_{\rm N}\cos\,n\varphi\frac{J_n'}{H_n'}(k_xa)\left(\frac{\sin\,\theta}{\sin\,\theta_i}\right)^n, \tag{17a}$$

$$D_{\rm res}^{ucs}=\frac{2L}{\pi^2k_xa^2}\sum_{n=0}^{\infty}\sum_{p=1}^{\infty}\epsilon_{\rm N}\cos\,n\varphi\left(\frac{k\,\sin\,\theta}{\alpha_p}\right)^n\times\frac{F(-kL\,\cos\,\theta,k_pL)F(k_zL,k_pL)A_{np}}{H_n'(k_xa)\alpha_pH_n'(\alpha_pa)}, \tag{17b}$$

$$D_{\rm res}^{r}=-\frac{2L}{k_x\pi^2a^2}\sum_{n=0}^{\infty}\sum_{p=1}^{\infty}\sum_{l=1}^{M}\epsilon_{\rm N}\cos\,n\varphi\left(\frac{k\,\sin\,\theta}{\alpha_p}\right)^n\times\frac{A_{np}\phi_p(z_l)\overline{S}_{nl}F(-kL\,\cos\,\theta,k_pL)}{H_n'(k_xa)\alpha_pH_n'(\alpha_pa)}. \tag{17c}$$

The first two equations have been derived previously for a rigid and for a uniform finite thin cylindrical shell.[16] The last equation represents the contribution of the reinforcing ribs to the scattered pressure: it can be noticed that the components of the (n,p) modes in the presence of the ribs can be deduced from the component of a uniform cylinder by replacing the forcing term due to the incident pressure $F(k_zL,k_pL)$ in Eq. (17b) by $\Sigma_{l=1}^{M}\phi_p(z_l)\overline{S}_{nl}$, which is due to the action of the ribs on the shell.

From Eq. (17c) it can be seen that the amplitude of $D_{\rm res}^{r}$ is enhanced when one of the two terms A_{np} or $\overline{S}_{nl}$ is large: the first condition leads to the resonance condition for the modes associated with helical waves[4] whereas the second condition occurs at the natural frequencies of the ribbed cylinder, where the determinant of Λ_n is equal to zero.

The backscattered pressure is obtained from Eq. (17) by setting $\varphi=\pi$; $\theta=\pi-\theta_i$. In particular, $D_{\rm res}^{r}$ becomes, in the monostatic case,

$$D_{\rm res}^{r,m}=-\frac{2L}{k_x\pi^2a^2}\sum_{n=0}^{\infty}\sum_{p=1}^{\infty}\sum_{l=1}^{M}\epsilon_N\left(-\frac{k_x}{a_p}\right)^n\times\frac{A_{np}\phi_p(z_l)\overline{S}_{nl}F(k_zL,k_pL)}{H_n'(k_xa)\alpha_pH_n'(\alpha_pa)}. \tag{18}$$

A further simplification of Eq. (18) is possible and the numerical computation of $D_{\rm res}^{r,m}$ can be reduced significantly if we note that the function F defined by Eq. (11) can be approximated by

$$F(x,y)\approx\frac{\sin(x\pm y)}{x\pm y},\quad \text{for}\ |x\mp y|\gg1, \tag{19}$$

for large arguments so that the summations with index p occurring in Eqs. (15a) and (18) may be restricted to the axial wave numbers k_p close to k_z,

$$P_l=\{k_p<k:|k_z-k_p|L<\pi\},\quad \text{for}\ k_z>0. \tag{20}$$

Indeed, when the above condition is met, one has $\alpha_p\approx k_x$ so that Eqs. (17b) and (17c) become

$$D_{\rm res}^{ucs,m}=\frac{2L}{(\pi k_xa)^2}\sum_{n=0}^{\infty}\sum_{p\in P_l}\epsilon_{\rm N}(-1)^n\frac{F^2(k_zL,k_pL)A_{np}}{H_n'^2(k_xa)}, \tag{21a}$$

$$D_{\rm res}^{r,m}=-\frac{2L}{(k_x\pi a)^2}\sum_{n=0}^{\infty}\sum_{l=1}^{M}\epsilon_{\rm N}(-1)^n\frac{\overline{\Delta}_{nl}\overline{S}_{nl}}{H_n'^2(k_xa)}, \tag{21b}$$

where $\overline{\Delta}_{nl}$ and $\overline{S}_{nl}$ are determined from Eqs. (14) and (15) with $p\in P_1$.

Equation (21b) provides a practical expression to compute the contribution of the reinforcing ribs to the backscattered pressure. Although it is much simpler than Eq. (18), its calculation nevertheless requires a significant computation load when it is applied to shiplike structures, as the number of ribs is usually large. Finally, this equation constitutes a generalization of the one previously derived for a set of periodic ribs [see Eq. (33d) of Ref. 4].

IV. APPLICATIONS

To illustrate the above theoretical derivation, the obtained results were applied to investigate scattering from a ribbed finite cylindrical shell with and without internal structures. First, in order to check the accuracy of the present method, the farfield backscattered pressure D_m was calculated by using Eqs. (17a) and (21) for a set of periodic ribs and compared to previous experimental and numerical results. Next, the effect of rib aperiodicity in the framed-shell system on the acoustic response was analyzed with respect to the degree of irregularity of the rib spacing in the (aspect angle, frequency) domain. Then the effect of the internal mechanical oscillators on the scattered field was studied by varying the total mass of the oscillators and by assuming a random distribution of their resonance frequencies along the cylinder axis.

The computation was made for a ribbed cylinder that was previously investigated in Ref. 2. The following parameters are used for the target: radius $a=6.92$ cm, total length $2L=0.856$ m, shell thickness $h=0.635$ mm. The rings have length $l_r=0.533$ mm, thickness $h_r=5.7$ mm, and a rib spacing $d=1.05$ cm. The elastic parameters are that of nickel with $E_s=2.1\times10^{11}$ Pa, $\mu=0.3$ and $\rho_s=8800$ kg/m^3 whereas the sound speed and the density of the water are $c_f=1500$ m/s and $\rho_f=1000$ kg/m^3.

A. Scattering from a finite cylindrical shell reinforced by a set of periodic ribs

In the present calculations, the ribs are assumed to act like blocking masses so that Z_{nl}^{r} is given by

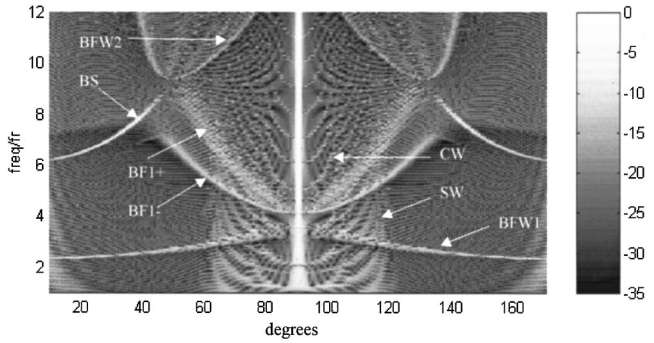


FIG. 2. Backscattering from a finite cylindrical shell stiffened with a set of periodic ribs. BFn: scattering of BFW of order n; BS: Bragg scattering; SW: shear wave scattering; CW: compressional wave scattering.

$$Z_{nl}^r = Z_n^r = -\rho_s A^r \omega^2, \tag{22}$$

where A^r denotes the cross-sectional area of the ring at $z = z_l$.

The farfield backscattered pressure D_m is shown in Fig. 2 as a function of the normalized frequency Ω and aspect angle. The computation was carried out at every degree for aspect angles varying between 10°–170° with a frequency increment $\Delta f = f_r/40$. The amplitude of the farfield is presented in black and white colors and it ranges between -35 and 0 dB. In the calculation of Λ_{njl}, the index p in Eq. (15b) varies between $1 \leqslant p \leqslant 8M$ and corresponds to a spatial resolution of magnitude $d/4$: that enables one to describe accurately scattering by Bloch–Floquet waves of second order that occurs at high frequencies.

As can be seen, the main features of the scattering process are clearly displayed: they are associated with specular reflection at broadside and with phase matching to helical shear and compressional waves and Bloch–Floquet waves (BFW) of first (resp., BFW1) and second order (resp., BFW2). It can be noted that Fig. 2 is very similar to Fig. 3 of Ref. 4, even if the ribs are considered in the present work to act like blocking masses instead of circular thin rods, as previously: this comparison clearly shows that a precise modeling of the ribs is not necessary for having the main aspects of the scattering pattern. As noticed previously, two distinct features denoted, respectively, by BF1− and BF1+ can be observed across a large range of aspect angles between $3 < \Omega < 9$. However, they are less curved toward the x axis than the two horizontal features that were found experimentally.[2]

The above-mentioned features may be explained by inspection of the BFW dispersion curves that are given by the resolution of the following equation:[4]

$$R_{\text{res}}^n(\chi_n) = \frac{\rho_f \omega^2 a d}{Z_n^r} + \sum_l \text{Real}\left(\frac{1}{Z_n^f(a\sqrt{k^2-\chi_{nl}^2}) - Z_n^s(\chi_{nl} a)} \right) = 0, \tag{23}$$

where l is an integer number associated with the order of BFW waves, $\chi_{nl} = \chi_n + 2\pi l/d$, and the unknown variable χ_n is the BFW wave number of the periodic shell related to the azimuthal mode n. When the interactions of BFW of different orders are weak, there is only one dominant term in the summation in the left member of Eq. (23) near the root of R_{res}^n.

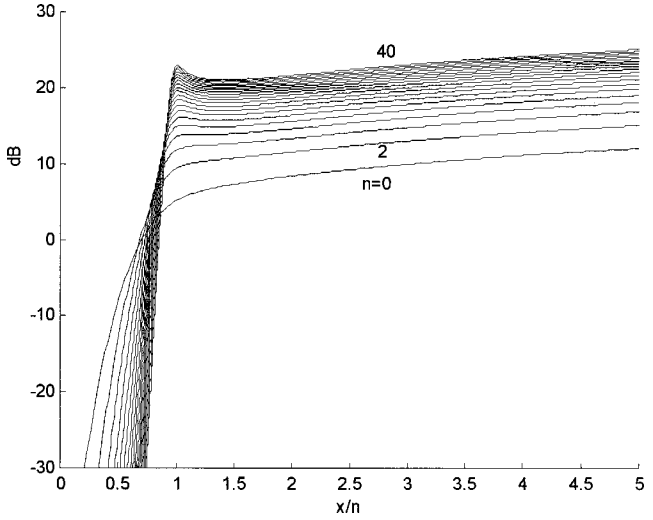


FIG. 3. Amplitude of $1/H_n'(x)$ vs x/n.

Using Eq. (22) and approximate expressions of Z_n^s and Z_n^f for $\Omega \gg 1$ and for large orders of n, Eq. (23) then becomes

$$\left(1 + \frac{\epsilon}{\tilde{\Omega}^{1/2}} \frac{1}{\sqrt{\eta - \tilde{\Omega}}} - \eta^2 \right) \frac{hd}{A^r} + 1 = 0, \tag{24a}$$

where $\epsilon = \rho_f c_s / \rho_s c_f \sqrt{12}$, $\tilde{\Omega} = c_s h \omega / c_f^2 \sqrt{12}$ is the radian frequency normalized by the coincidence frequency of the plate, and

$$\eta = ((\chi_n + 2\pi l/d)^2 a^2 + n^2)\beta_s / \Omega. \tag{24b}$$

It can be noted that Eq. (24a) is in agreement with Eq. (40) of Ref. 4 for $hd/A^r \gg 1$. Because of the presence of H_n' at the denominator in the expression of D_{res}^r, the amplitude of the scattered field is not significant when

$$ka \sin \theta_i \ll n, \tag{24c}$$

since $l/H_n'(x)$ decreases rapidly when $x/n \to 0$, as shown in Fig. 3 for $n = 0{:}40$ with an increment $\Delta n = 2$.

The locations of Bragg scattering and BFW of first and second order are displayed in Fig. 4 with respect to the azimuthal mode number n in the (angle, frequency) plane, where the aspect angle θ is defined as $\theta = \cos^{-1}(\chi_n/k)$. The BFW dispersion curves are computed by solving Eq. (24) with a simple routine to find the zeros of a real function of one variable with a starting value given by Eq. (43a) of Ref. 4, and only the portion of the curves verifying $ka \sin \theta_i \geqslant 0.8n$ are plotted in Fig. 4. As can be seen the locations of the BFW1 curves differ significantly with respect to n when $\Omega > 3$. These curves are located in a region bounded by BF1− (resp., BF1+) in its lower (resp., upper) part: the lower boundary BF1− is well defined whereas the upper boundary BF1+ is more diffuse because it is formed by portions of BFW1 dispersion curves meeting Eq. (24c). Inside this region, the acoustic field has a complex structure resulting from the interference of the different azimuthal modes of

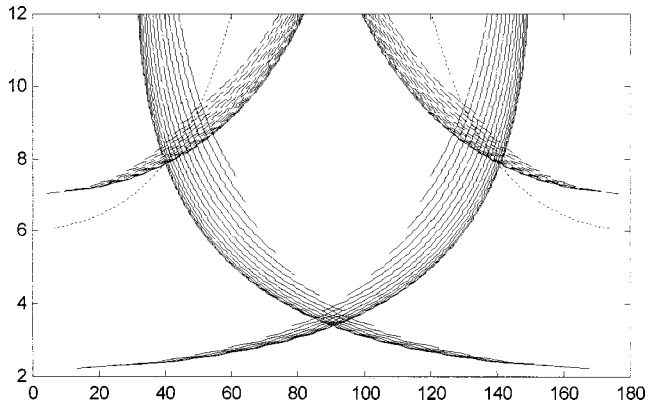


FIG. 4. Locations of Bragg scattering and Bloch–Floquet waves ($n=0{:}2{:}28$) computed by using Eq. (24) with $ka \sin\theta_i \geqslant 0.8n$; Bragg scattering ($\cdots$); Bloch–Floquet waves of first order (—) and second order (---).

BFW1 with an enhancement of the amplitude levels near the two boundaries BF1− and BF1+, as can be seen in Fig. 2. Note that at the intersections of BFW1 and BFW2 dispersion curves, the interactions between BFW waves are important so that Eq. (24) are no longer valid in the frequency range $7<\Omega<9$.

B. Effect of rib aperiodicity on backscattering

To analyze the effect of irregularity of the rib spacing on the backscattered pressure, let the positions of the rings be defined by

$$z_l = ld - L + \chi_l d, \quad \text{for } l = 1{:}M,$$

where $d=1.05$ cm is the mean spacing, χ_l is a random variable uniformly distributed in the interval $[-\tau/2, \tau/2]$.

In Fig. 5 the farfield backscattered pressure are shown for two values of τ ($\tau=0.1;0.2$) and for aspect angles varying between 10° and 90°. The presence of a small deviation of the ribs has the effect to broaden the features associated with BFW waves scattering for $\Omega>3$. It can be noted that the amplitude levels of the Bragg scattering, the scattering from shear and compressional waves, are almost unmodified, as can be seen in Fig. 6. In this figure, the backscattered pressure in the presence of irregularity is normalized by the backscattered pressure without irregularity ($\tau=0$) to display the differences in amplitude associated with the observed wave features. The broadening of BFW waves are accentuated when τ increases and a set of curves with significant amplitude levels can be observed in the upper part of the frequency band in Figs. 5(b)–6(b). They are more or less parallel to the Bragg peaks and tend to broaden the main features associated with the BFW scattering process while BF1+ and BFW2 are attenuated further and are beginning to be submerged in noise.

Therefore the scattered pressure does not fluctuate in a random manner in the presence of rib aperiodicity but displays coherent features that are related to phase matching to the low azimuthal mode number of Bloch–Floquet waves of first order, Bragg scattering and helical waves. However, the amplitude levels of these coherent features are not enhanced by irregularity as strongly as in the plate case.[17] The broadening and attenuation of the Bloch–Floquet waves of high

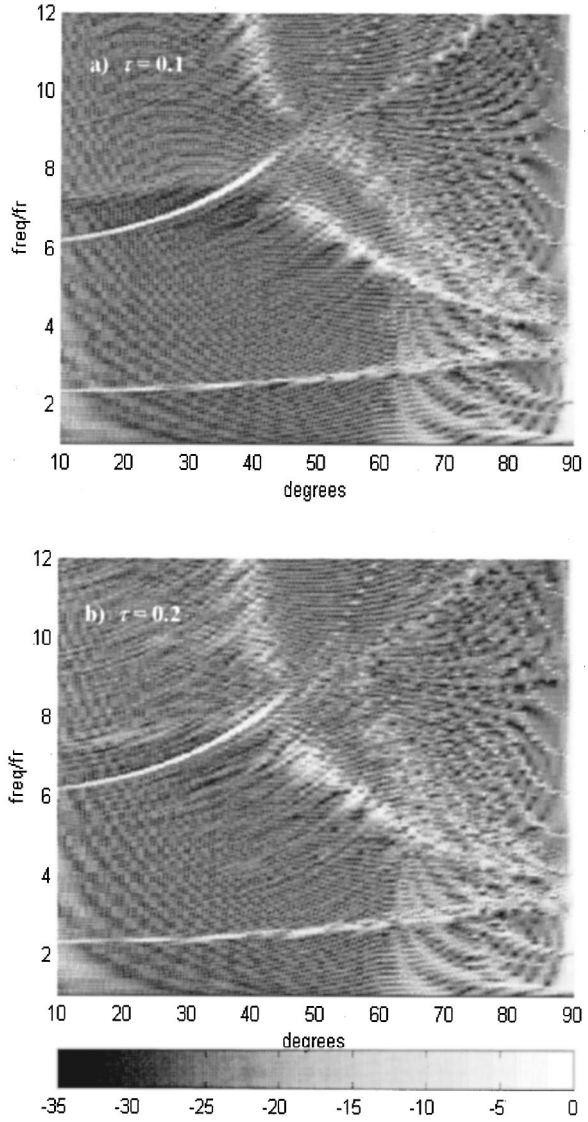


FIG. 5. Effect of the rib aperiodicity on the backscattered pressure for two degrees of irregularity τ.

azimuthal mode number might be due to Anderson localization effects[18–20] or incoherent fluctuations of the acoustic field.

C. Effect of internal mechanical oscillators on backscattering

Unlike the previous case, the presence of mechanical oscillators attached to the ribs for simulating internal structures can dramatically change the scattering pattern, as reported in Ref. 12. In order to investigate this problem, now let Z^r_{nl} be defined as the impedance of the rib–spring–mass system: then Z^r_{nl} is the sum of the impedance of the rib Z^r_n and that of a spring–mass combination:[21]

$$Z^r_{nl} = -\rho_s \omega^2 A^r - \rho_s \omega^2 A^r \xi \left(1 + \frac{1}{(f_l/f)^2 (1 - i\zeta_l) - 1} \right), \tag{25}$$

where ξ is the ratio of the mass of the oscillator to the mass of the rib in each periodic unit, f_l is the natural (or reso-

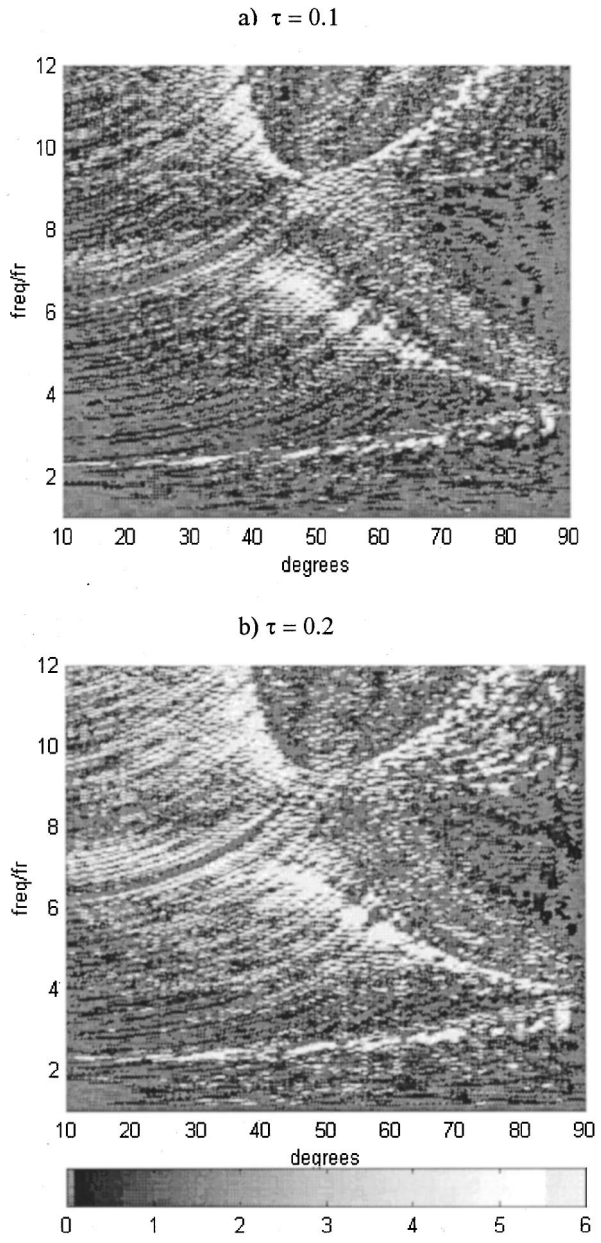


FIG. 6. Broadening and enhancement of Bloch–Floquet waves with respect to irregularity τ. Backscattered pressure for $\tau=0.1$ and 0.2 normalized by backscattered pressure for $\tau=0$.

nance) frequency, and ζ_l is the damping coefficient of the oscillator located at $z=z_l$. In Eq. (25), it has been assumed that all the oscillators have the same mass.

In order to compare the theoretical results to the experimental data reported in Ref. 12 in the frequency range 11–45 kHz, it remains in the framework of the present approach to determine f_l and ζ_l for $l=1:M$ such that when attached to the ribbed shell, the band of resonance frequencies of the internal oscillators corresponds to the "*in situ*" resonance frequency band (10–40 kHz). This is not straightforward because first the oscillators are assumed to be axisymmetric in the present modeling unlike in the experiments, and then the resonance frequency of the oscillator is shifted significantly when it is attached to the ribbed shell. Nevertheless, f_l can be determined approximately by considering the ribbed shell–spring–mass system like a mass–spring–mass system and using Eq. (25) with $A^r \to A^r + hd$ and $\zeta_l \to 0$, the resonance frequency of the total system f_s is given by the relationship $f_s = f_l\sqrt{1+\varpi}$, where $\varpi = \xi/(1+hd/A^r)$ is the ratio of the total mass of the attached oscillators to the mass of the ribbed shell. This gives $f_{\min}\approx 7$ kHz and $f_{\max}\approx 28$ kHz since $\varpi \approx 1$ in the experimental works; however, after several simulations, it turned out that better "coverage" could be realized with $f_{\min}\approx 8.5$ kHz and $f_{\max}\approx 30$ kHz. Note that these frequencies are quite different from the lowest ($f_{\min}\approx 3$ kHz) and highest ($f_{\max}\approx 20$ kHz) resonance frequencies of the beam–mass oscillators used in the experiments. Then in the following computations, it has been assumed that f_1 varies randomly between $f_{\min}\approx 8.5$ kHz and $f_{\max}\approx 30$ kHz. Furthermore, it has been supposed that ζ_l is a linear function of frequency $\zeta_l = 0.005f$ (in kHz): the value of the slope of the attenuation has been found empirically so that the effects of the internal oscillators are significantly damped above 40 kHz.

The modifications of the backscattered pressure with respect to ϖ are calculated by applying Eqs. (21) and (25) and they are shown in Figs. 7 and 8(a) for several values of ϖ. As can be seen, the structure of backscattering of the periodic ribbed shell plotted in Fig. 7(a) is modified when the mass of the oscillators increases and the appearance of scattering has been progressively replaced by a "Christmas tree" appearance, as can be seen in Figs. 7 and 8.

Theoretically there is no axial periodicity when $\varpi > 0$, since the mechanical impedances of the oscillators vary along the cylinder axis and, as a consequence, BFW scattering is strongly attenuated even if ϖ is small, as shown in Figs. 7(b) and (c). The motion of the total structure results from the interactions of the motions of its different components, namely the uniform shell, the ribs and the oscillators, and in the present case, the motion of the structure is almost completely driven by the motion of the uniform shell so that the backscattering pattern is mainly associated with helical waves scattering. As ϖ increases, the backscattered pressure loses its "chalice" appearance in the band of the resonance frequencies of the oscillators, where it is replaced by a new complex structure.

However, the interpretation of the new features by using Eq. (21) is difficult: indeed, although these equations are valid for a general model of ribbed axisymmetric cylindrical shell and are useful for numerical calculations, they do not provide a deep physical insight in the scattering process because of their complexities. So in order to improve the understanding of the physics of the problem, consider now the limiting case where the motions of the different components of the total structure are overwhelmed by the motions of the oscillators. It is then possible to neglect the first term in the right member of Eq. (15b), which is due to the motion of the submerged uniform shell. Therefore Λ_n is reduced to a diagonal matrix and from Eq. (14) $\overline{S_{nj}} = A_{nj}^{r\,-1}\overline{\Delta_{nj}}$ and Eq. (21b) becomes

$$D_{\text{res}}^{r,m} = -\frac{2L}{(k_x\pi a)^2}\sum_{n=0}^{\infty}\sum_{l=1}^{M}\epsilon_{\text{N}}(-1)^n\frac{A_{nl}^{r\,-1}\overline{\Delta_{nl}^2}}{H_n'^2(k_x a)}. \qquad (26)$$

The computation of the backscattered pressure with Eq.

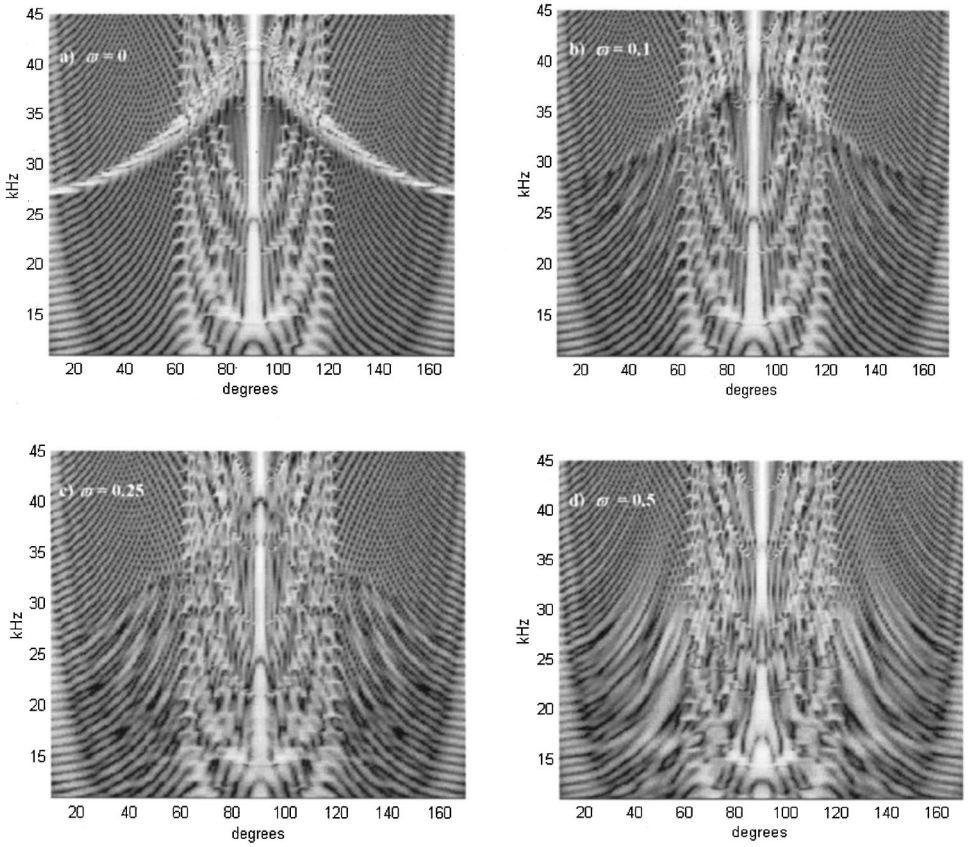


FIG. 7. Effect of the mechanical oscillators on the backscattered pressure for several ratios of the total mass of the internal oscillators to the mass of the ribbed shell (ϖ). The backscattered pressure is displayed using a black and white colors scale ranging from −35 to 0 dB, as shown in Fig. 8.

(26) is shown in Fig. 8(b) for $\varpi=1$: it is readily seen that it is very similar to Fig. 8(a), revealing that the motions of the internal oscillators prevails over the motions of the other components of the total structure when ϖ is close to unity ($\xi\approx 3.2$). That could explain the ''speckled'' appearance of scattering reported by Photiadis *et al.*, since in this experiment, the oscillators were randomly located on the surface of the shell without any average circumferential or axial periodicity.

The features associated with the ''Christmas tree'' appearance in Fig. 7 can now be interpreted by using Eq. (26). Indeed, the amplitude of $D_{\text{res}}^{r,m}$ is enhanced when Z_{nl}^{r} or Δ_{nl} are large: first, from Eq. (25), Z_{nl}^{r} is large when the frequency f of the incident plane wave is close to one of the natural frequencies of the oscillators f_l and then from Eq. (15a), Δ_{nl} is large when the spatial coincidence condition and/or the resonance condition for the helical waves are met. However, because of the presence of $\phi_p(z_l)$ in the expression of Δ_{nl}, this is only valid if the amplitude of ϕ_p associated with the p-axial mode meeting the spatial coincidence condition is not negligible at the location of the resonant oscillator.

When the above conditions are fulfilled simultaneously, the backscattering has the structure of a ''Christmas tree'' in the (angle, frequency) plane: the ''branches'' are associated with the coincidence condition $k_z=k\cos\theta_i=k_p$, whereas the regions of high-amplitude levels are related to the number, the locations, and the distribution of the resonance frequencies of the internal oscillators, but also to the resonances of the helical modes near the broadside.

Finally, to conclude this section, although the present simulations do not correspond exactly to the experimental conditions reported in Ref. 12, the same mean features could be reproduced in the backscattering pattern, except for the ''speckled'' appearance of scattering at grazing angles of incidence.

V. CONCLUSIONS

A theoretical formalism has been proposed to deal with the problem of scattering from an irregularly ribbed finite cylindrical shell with internal mechanical oscillators under a certain number of assumptions that permits us to simplify the derivation of the solution and to limit the amount of calculation. The approach is similar to that adopted for uniformly spaced ribs, except that the unknown variables are now expanded in the eigenfunctions of the finite cylindrical shell with simply supported ends, leading to an infinite algebraic linear system for each azimuthal mode. The key point of the present approach consists of transforming this infinite system into a finite linear system, the rank of which is equal to the number of the ribs. Approximate expressions are derived to compute the backscattered pressure by only retaining the contribution of the axial modes meeting approximately the spatial coincidence condition.

Some numerical simulations have been carried out to investigate backscattering from a periodically ribbed finite cylindrical shell as well as the effects of rib aperiodicity and internal structures on the scattering pattern. In the periodic case some results are presented to explain the complex structure of backscattering in the (aspect angle, frequency) plane that was found previously in the central frequency range. For weak deviations from the mean rib spacing, it has been

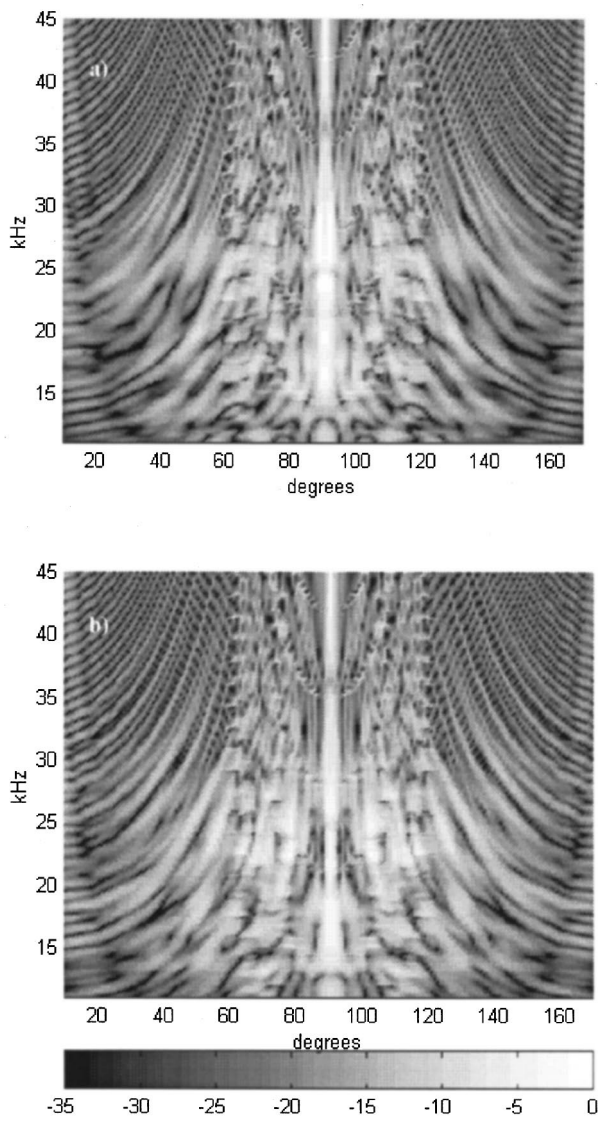


FIG. 8. Effect of the mechanical oscillators on the backscattered pressure when the total mass of the internal oscillators is equal to the mass of the ribbed shell ($\varpi = 1$): (a) Eq. (21b); (b) Eq. (26).

shown that the amplitude levels of the coherent features associated with phase matching to the low azimuthal mode number of Bloch–Floquet waves of first order, Bragg scattering and helical waves are enhanced by irregularity. Finally, in the presence of internal oscillators, the ''chalice'' structure of backscattering of the ribbed shell is progressively replaced by a ''Christmas tree'' appearance when the mass of the oscillators increases. In the present case of application where the total mass of the oscillators is equal to the mass of the ribbed shell, it has been found that the motion of the whole structure is almost completely driven by the motions of the internal oscillators.

[1] 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).
[2] 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).
[3] D. M. Photiadis, B. H. Houston, E. G. Williams, and J. A. Bucaro, ''Resonant response of complex shell structures,'' J. Acoust. Soc. Am. **108**, 1027–1035 (2000).
[4] M. Tran-Van-Nhieu, ''Scattering from a ribbed finite cylindrical shell,'' J. Acoust. Soc. Am. **110**, 2858-2866 (2001).
[5] 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).
[6] Y. P. Guo, ''Sound scattering by bulkheads in cylindrical shells,'' J. Acoust. Soc. Am. **95**, 2550–2559 (1994).
[7] J. M. Cuschieri and D. Feit, ''Acoustic scattering from a fluid-loaded cylindrical shell with discontinuities: Single plate bulkhead,'' J. Acoust. Soc. Am. **98**, 320–338 (1995).
[8] J. M. Cuschieri and D. Feit, ''Acoustic scattering from a fluid-loaded cylindrical shell with discontinuities: Double plate bulkhead,'' J. Acoust. Soc. Am. **98**, 339–352 (1995).
[9] 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).
[10] Y. P. Guo, ''Sound scattering from cylindrical shells with internal elastic plates,'' J. Acoust. Soc. Am. **93**, 1936–1946 (1992).
[11] A. Klauson and J. Metsaveer, ''Acoustic scattering by a submerged cylindrical shell stiffened by lengthwise ribs,'' J. Acoust. Soc. Am. **96**, 1575–1581 (1994).
[12] 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).
[13] J. A. Bucaro, J. A. Romano, A. Sarkissian, D. M. Photiadis, and B. H. Houston, ''Local admittance model for acoustic scattering from a cylindrical shell with many internal oscillators,'' J. Acoust. Soc. Am. **103**, 1867–1873 (1998).
[14] M. C. Junger and D. Feit, *Sound, Structures and Their Interaction*, 2nd ed. (MIT Press, Cambridge, MA, 1986).
[15] M. Tran-Van-Nhieu, M. Gensane, A. Tesei, S. Fiovaranti, and A. Maguer, ''Scattering by a fluid filled finite cylindrical shell: Comparison theory-experiments,'' *Proceedings of the 4th European Conference on Underwater Acoustics*, 1998, Vol. 2, pp. 819–824, edited by A. Alippi and G. B. Cannelli, CNR-IDAC, Rome, Italy.
[16] M. Tran-Van-Nhieu, ''Scattering from a finite cylindrical shell,'' J. Acoust. Soc. Am. **91**, 670–679 (1992).
[17] D. M. Photiadis, ''The effect of irregularity on the scattering of acoustic waves from a ribbed plate,'' J. Acoust. Soc. Am. **91**, 1897–1903 (1992).
[18] D. M. Photiadis, ''Localization of helical flexural waves by irregularity,'' J. Acoust. Soc. Am. **96**, 2291–2301 (1994).
[19] D. M. Photiadis and B. H. Houston, ''Anderson localization of vibration on a frame fluid-loaded cylinder,'' J. Acoust. Soc. Am. **106**, 1377–1391 (1999).
[20] H. M. Marcus, B. H. Houston, and D. M. Photiadis, ''Wave localization on a submerged cylindrical shell with rib aperiodicity,'' J. Acoust. Soc. Am. **109**, 865–869 (2001).
[21] M. Strasberg and D. Feit, ''Vibration damping of large structures induced by attached small resonant structures,'' J. Acoust. Soc. Am. **99**, 335–344 (1996).

Analysis of the time-reversal operator for scatterers of finite size

D. H. Chambers
Lawrence Livermore National Laboratory, PO Box 808, L-154, Livermore, California 94551

(Received 24 September 2001; revised 9 May 2002; accepted 11 May 2002)

Recently, it was shown that the time-reversal operator for a single, small spherical scatterer could have up to four distinguishable eigenstates [Chambers and Gautesen, J. Acoust. Soc. Am. **109**, 2616–2624 (2001)]. In this paper, that analysis is generalized for scatterers of arbitrary shape and larger size. It is shown that the time-reversal operator may have an indefinitely large number of distinguishable eigenstates, with the exact number depending on the nature of the scatterer and the geometry of the time-reversal mirror. In addition, the case of a multiple number of well-separated scatterers is investigated, with the result that the total spectrum is the direct combination of the eigenstates associated with each scatterer. As an example, the singular value spectrum of the time-reversal operator for a linear array is calculated explicitly for bubbles and hard rubber spheres of finite size. Both resonance peaks and apparent crossing points can be observed in the spectrum as the size of the scatterer increases. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1490362]

PACS numbers: 43.20.Fn, 43.60.Pt [LLT]

I. INTRODUCTION

The detection and identification of objects from their acoustic scattering characteristics is an important problem for nondestructive evaluation, underwater acoustics, and medical ultrasound. One approach to this problem is acoustic time reversal (see reviews by Fink *et al.*,[1] and Prada and Fink[2]). In a series of papers, Prada and Fink *et al.*[3–5] applied the concept of a time-reversal mirror (TRM) to the detection and identification of scatterers in a medium. A TRM is an array that transmits a pulse that scatters from objects in the medium, measures the scattered waveform, reverses the received pulse, then transmits it back into the medium. This process can be repeated until the transmitted pulses do not change significantly between iterations. Mathematically, the process has converged to an eigenstate of the time-reversal (TR) operator (typically the eigenstate with the largest eigenvalue). Prada *et al.*[3–5] showed how to determine all of the eigenstates (in principle) of the TR operator using a technique called D.O.R.T., which is equivalent to the singular value decomposition of the operator. They showed that for small, noninteracting spherical scatterers whose densities match the medium, each eigenstate of the time-reversal operator corresponded to a separate scatterer. Using the D.O.R.T. technique, each eigenstate could be identified and the number of scatterers determined. The eigenvalues of the TR operator are measures of the apparent strength of each scatterer, and the eigenvectors determine the driving pulses for each element of the TRM that create a field that focuses on each scatterer.

Recently, Chambers and Gautesen[6] have extended the analysis of the time-reversal operator to small spherical scatterers whose densities differ from the background medium. Those results showed that the time-reversal operator could have up to four eigenstates associated with each scatterer. These correspond to four orthogonal scattering moments induced by the TRM for a given scatterer. The relative magnitudes of the eigenvalues in the spectrum depended on the compressibility and density contrast between the scatterer and the medium. Thus the spectrum of the time-reversal operator provides information on the scatterer composition that might prove useful for characterization in a number of applications. For multiple spherical scatterers, the analysis of Chambers and Gautesen implies that the total number of possible eigenstates could be four times the number of scatterers. It also suggests that for more general scatterers of arbitrary shape or size, the number of eigenstates could be greater than four, depending on the number of different scattering moments that could be induced by an acoustic field radiated from a TRM.

Time-reversal experiments on extended objects show the presence of multiple eigenstates. Thomas, Roux, and Fink[7] observed multiple modes of focusing for time reversal applied to a hollow cylinder. Application of D.O.R.T. to this case by Komilikis, Prada, and Fink[8] showed a relationship between the eigenstates of the time-reversal operator and different modes of acoustic scattering from the cylinder. Later analysis by Prada and Fink[9] showed that six of the eigenstates were associated with different Lamb wave modes propagating around the cylinder. In an application of time reversal to large kidney and gallbladder stones, Thomas, Wu, and Fink[10] observed that time reversal focuses on small areas of an extended object that generate strong reflections. Multiple areas can exist on an extended object, creating eigenstates of the time-reversal operator associated with each area. These early experiments establish the general observation that multiple eigenstates of the time-reversal operator can exist for extended objects.

In this article, the analysis of Chambers and Gautesen is extended to scatterers of more general shapes and sizes but still small compared with the distance of the scatterer from the TRM. It explores the behavior of the time-reversal operator in the region between Rayleigh scattering from small objects (Ref. 6) and scattering from large extended objects (Refs. 7–10) where the curvature of the incident wavefront is

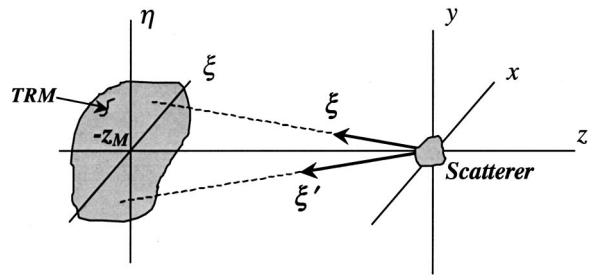


FIG. 1. Geometry of time-reversal mirror (TRM) and scatterer. The vectors $\boldsymbol{\xi}'$ and $\boldsymbol{\xi}$ indicate the positions of receiver and transmit elements on the TRM, respectively.

no longer negligible. It is shown that the number of eigenstates depends on the geometry of the TRM and scatterer shape, and generally increases with size. In addition, eigenvalues may cross or interchange identities (depending on the choice of ordering) with increasing size. For the case of multiple noninteracting scatterers, the total spectrum of eigenstates of the TR operator is composed of the collection of eigenstates from each individual scatterer. Though the analysis assumes the size of the scatterer is small compared with its distance to the TRM, it does not require that the size be smaller than the acoustic wavelength. The analysis starts with the relatively simple case of the TR operator for a single finite sphere in Sec. II, then derives the form for the nonspherical scatterer in Sec. III. In Sec. IV the singular value decomposition (SVD) of the time-reversal operator for the case of multiple noninteracting scatterers is partitioned into a set of SVDs for individual scatterers using a stationary phase argument. Finally, in Sec. V, numerical examples for spheres of different sizes are investigated. The spectrum of eigenvalues is calculated over a range of sphere sizes for a time-reversing array of 96 elements at 3 MHz in water.

II. TIME REVERSAL FOR A FINITE SPHERE

The analysis in this paper follows Prada and Ref. 6 in formulating an expression for the TR operator at a given frequency ω and performing an SVD to determine the properties of the eigenstates. In actual practice the TRM emits pulses, but for analytical convenience the analysis is performed for a single frequency. In principle, the broadband response can be obtained by Fourier synthesis. Following Cassereau and Fink,[11] the TRM in transmit (Fig. 1) is modeled as a planar region S in an infinite rigid baffle for which the normal velocity $U_M(\xi,\eta)$ is specified (ξ and η are the x- and y coordinates, respectively, of a point on the TRM). The TRM radiates an acoustic pressure field that scatters from objects in the medium and is measured back at each point of the TRM, $P_M(\xi',\eta')$. The pressure field emitted by the TRM at a given point (x,y,z) in the volume is[12]

$$p(x,y,z) = -\frac{i\rho\omega}{2\pi}\int_S U_M(\xi,\eta)\frac{e^{ikR}}{R}dS, \tag{1}$$

where $R=\sqrt{(z+z_M)^2+(x-\xi)^2+(y-\eta)^2}$, $k=\omega/c$, ρ is the density, and c is the sound speed in the medium. Since the sphere radius a is assumed to be small compared to the distance to the TRM ($a \ll z_M$), the incident pressure field on the sphere can be approximated by Eq. (1) evaluated at the sphere center (origin in Fig. 1). The resulting surface integral can be interpreted as a superposition of plane waves, each emitted from an area element $dS=d\xi\,d\eta$ at (ξ,η) on the TRM. The analysis of Ref. 6 for small, spherical scatterers can be extended to spheres of finite size by using the expression for the field scattered from a sphere centered at the origin and illuminated by a unit amplitude plane wave that is found in Morse and Ingard:[13]

$$P_S(r,\theta) = -\frac{1}{2}\sum_{n=0}^{\infty}(2n+1)i^n(1+R_n)P_n(-\cos\theta)h_n(kr), \tag{2}$$

where r is the radial distance, R_n are reflection coefficients, $P_n(x)$ are Legendre polynomials, and $h_n^{(1)}(x)$ are spherical Bessel functions. The polar angle θ is defined such that $\theta=0$ is the backscatter direction and $\theta=\pi$ is the forward-scatter direction (opposite to Ref. 13). The plane wave incident on the sphere from element dS of the TRM generates a scattered field given by Eq. (2), weighted by the amplitude of the plane wave as given by Eq. (1). If the scattered field is evaluated at a point (ξ',η') on the TRM and the contributions from each element dS are combined, the following expression is obtained for the pressure measured by the TRM:

$$P_M(\xi',\eta') = -\frac{i\rho c}{6\pi}(ka)^3\frac{e^{ikr'_M}}{r'_M}\sum_{n=0}^{\infty}B_nQ_n(-2ikr'_M) \times\int_S U_M(\xi,\eta)P_n(-\cos\theta)\frac{e^{ikr_M}}{r_M}dS, \tag{3}$$

where $r_M=\sqrt{\xi^2+\eta^2+z_M^2}$, $r'_M=\sqrt{\xi'^2+\eta'^2+z_M^2}$, $B_n=\{[3i(2n+1)]/[2(ka)^3]\}(1+R_n)$, θ is the angle between the vectors $\boldsymbol{\xi}=(\xi,\eta,-z_M)$ and $\boldsymbol{\xi}'=(\xi',\eta',-z_M)$ (see Fig. 1), and $h_n^{(1)}(x)=[e^{ikx}/(i^{1+n}x)]Q_n(-2ix)$, with $Q_n(z)$ an nth-order polynomial in $1/z$.[14] In the limit $ka\to 0$ the first two coefficients become $B_0\to(\kappa_s-\kappa)/\kappa$ and $B_1\to 3(\rho_s-\rho)/(2\rho_s+\rho)$, in agreement with Ref. 6 (κ and κ_s are the compressibilities of the medium and sphere, respectively, and ρ_s is the density of the sphere). The integral in Eq. (3) is over the surface S of the TRM.

Following Prada,[3] the TRM velocity distribution U_M and the received pressure distribution P_M are assumed to be proportional to a driving voltage $E(\xi,\eta)$ and an output voltage $R(\xi,\eta)$: $U_M(\xi,\eta)=cA_e(\omega)E(\xi,\eta)$, $R(\xi,\eta)=A_r(\omega)P_M(\xi,\eta)/\rho c^2$. The functions $A_e(\omega)$ and $A_r(\omega)$ model the frequency response of each transducer in transmit and receive modes. Incorporating these into the response of the TRM, the following expression for the output voltage is obtained:

$$R(\xi',\eta') = \int_S K_S(\xi',\eta';\xi,\eta)E(\xi,\eta)dS \equiv K_S\circ E. \tag{4}$$

The scattering kernel $K_S(\xi',\eta';\xi,\eta)$ is given by

$$K_S(\xi',\eta';\xi,\eta) = -\frac{i(ka)^3}{6\pi} A_r(\omega) A_e(\omega) \frac{e^{ik(r'_M + r_M)}}{r'_M r_M} \times \sum_{n=0}^{\infty} B_n P_n(-\cos\theta) Q_n(-2ikr'_M). \tag{5}$$

The scattering kernel combines the scattering physics of the sphere and the propagation geometry to describe the operation of the TRM, and is the basis for subsequent analysis of the TRM.

The condition of convergence of the time-reversal operation is equivalent to the following eigenvalues equation for the time-reversal operator.[1–6]

$$(K_s^* \circ K_s) \circ \Phi = \mu_{TR}^2 \Phi. \tag{6}$$

The kernel $(K_s^* \circ K_s)$ of the time-reversal operator is Hermitian, which guarantees that the eigenvalues are real and the eigenfunctions are orthogonal over the TRM surface S

$$\int_S \Phi_m^*(\xi,\eta)\Phi_n(\xi,\eta)\,dS \propto \delta_{mn}.$$

The eigenfunctions are the complex amplitude distributions on the TRM that are invariant under the time-reversal process. The eigenvalues are the strengths of each TR invariant mode. Together, they constitute the eigenstates of the TR operator and represent modes of coupling between the TRM and scatterers that are self-perpetuating, i.e., they produce the same fields after two applications of the time-reversal process. It is convenient to reduce the number of extraneous factors by defining

$$\phi(\xi,\eta) = \Phi(\xi,\eta) e^{ikr_M}, \quad \mu^2 = \frac{36\pi^2}{(ka)^6} \frac{\mu_{TR}^2}{|A_e A_r|^2}. \tag{7}$$

Thus, the time-reversal eigenvalue problem becomes

$$(K^* \circ K) \cdot \phi = \mu^2 \phi, \tag{8}$$

with the reduced scattering kernel

$$K(\xi',\eta';\xi,\eta) = \frac{1}{r'_M r_M} \sum_{n=0}^{\infty} B_n P_n(-\cos\theta) Q_n(-2ikr'_M). \tag{9}$$

The assumption that the distance to the scatterer is much greater than a wavelength implies $kr'_M \gg 1$ and $Q_n(-2ikr'_M) \to 1$ for all n that contribute significantly to the sum in (9).

In Ref. 6, the reduced scattering kernel was real and an eigenvalue problem equivalent to (8) could be obtained involving K alone. Here, for complex K, the singular value decomposition (SVD) for the reduced scattering kernel is given by (see the Appendix)

$$K \circ \phi = \mu \phi^*. \tag{10}$$

The squares of the singular values are the eigenvalues of the TR operator, and the singular value functions are identical to the eigenfunctions of Eq. (8). Thus, the SVD of the reduced scattering kernel is equivalent to the solution of the eigenvalue equation for the TR operator.

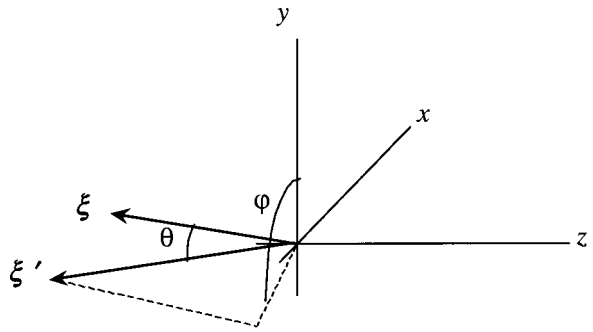


FIG. 2. Detail of the TRM geometry, showing the polar angle θ and azimuthal angle φ that specifies the position of $\boldsymbol{\xi}'$ relative to $\boldsymbol{\xi}$.

Suppose the scattered field from the sphere is well represented by the series (2) truncated at order N (total of $N+1$ terms). By replacing the $\cos\theta$ with the dot product of the vectors and rearranging terms, the reduced scattering kernel can be written

$$K(\xi',\eta';\xi,\eta) = \frac{1}{r_M r'_M} \sum_{n=0}^{N} A_n (-1)^n \left(\frac{\xi'\xi + \eta'\eta + z_M^2}{r'_M r_M} \right)^n, \tag{11}$$

where A_n are complex coefficients related to the previous coefficients B_n, and $Q_n(-2ikr'_M) \to 1$. This kernel is separable, with the total number of linearly independent terms equal to $(N+3)(N+2)(N+1)/6$. This is also the number of possible eigenvalues and eigenfunctions of the TR operator for this sphere. Each eigenfunction represents a complex amplitude distribution of the TRM that focuses on the sphere, with an apparent reflectivity given by the associated eigenvalue. The truncation order required for accurate representation of the scattered field [Eq. (2)] is on the order of the value of ka.[15] Higher-order terms decrease rapidly in size. Thus, the number of significant eigenvalues of the TR operator for a sphere should scale with size roughly as $(ka)^3/6$. Though the infinite number of terms in the representation of the scattered field [Eq. (2)] implies an infinite number of TR eigenvalues, the number of significant eigenvalues is expected to be finite.

III. TIME REVERSAL FOR AN ARBITRARY FINITE OBJECT

For our present purposes, the primary difference between an arbitrary object and the finite sphere is the form of the expansion for the scattered field [Eq. (2) in the previous section]. The spherical harmonics expansion for a general scattered field is

$$P_S(r,\theta,\varphi) = \sum_{n=0}^{\infty} \sum_{l=-n}^{n} C_{ln} P_n^l(-\cos\theta) e^{il\varphi} h_n(kr), \tag{12}$$

where $P_m^l(x)$ are the associated Legendre polynomials, and φ is the azimuthal angle coordinate of a spherical coordinate system oriented along the propagation direction of the incident field (see Fig. 2). The complex scattering coefficients C_{ln} are determined both by the scatterer shape and the propagation direction of the incident field. Thus, object orientation relative to the TRM as well as size and shape determines the

field scattered back to the TRM. The pressure distribution on the TRM is then

$$P_M(\xi',\eta') = -\frac{i\rho c}{6\pi}(ka)^3 \frac{e^{ikr'_M}}{r'_M} \sum_{n=0}^{\infty} i^{-n} Q_n(-2ikr'_M) \cdot \sum_{l=-n}^{n} \int_S U_M(\xi,\eta) B_{ln}(\xi,\eta) \times P_n^l(-\cos\theta) e^{il\varphi} \frac{e^{ikr_M}}{r_M} dS, \quad (13)$$

where the scattering coefficients B_{ln} ($\equiv 3C_{ln}/(ka)^3 i^{1+n}$) and angular coordinates θ and φ are functions of ξ and η. (The scattering coefficients have been scaled by an overall size factor a for comparison with the previous case.) This leads to the following form for the reduced scattering kernel of an arbitrary finite object:

$$K(\xi',\eta';\xi,\eta) = \frac{1}{r_M r'_M} \sum_{n=0}^{\infty} \sum_{l=-n}^{n} B_{ln}(\xi,\eta) \times Q_n(-2ikr'_M) P_n^l(-\cos\theta) e^{il\varphi}. \quad (14)$$

The behavior of the TRM for a general scatterer is determined by the variation of the scattering coefficients B_{ln} integrated over the surface of the TRM. The number of eigenvalues of the TR operator could be indefinitely large, depending on the exact nature of the scatterer. Even if a truncated expansion of (12) were used, the kernel may not be separable since both angular variables, θ and φ, and the scattering coefficients depend on ξ and η.

IV. TIME REVERSAL FOR MULTIPLE FINITE OBJECTS

It has been shown that the time-reversal operator for a single finite object has a complicated eigenvalue spectrum that depends on the shape, size, orientation, and position of the object relative to the TRM. In this section, the case of multiple scatterers is analyzed to see how the spectra from each scatterer combine to form the spectrum of the entire TRM system. Following Prada,[5] the scatterers are assumed to be well-separated, i.e., separation distances are much greater than a wavelength. Multiple scattering will also be ignored. Of course, these assumptions become less realistic when the scatterers are large. However, it is instructive to see how the multiple eigenstates associated with individual scatterers combine to form the spectrum of the full scattering kernel in the case where the analysis is tractable. With these assumptions, the pressure field returned from each scatterer to the TRM is given by (13) with the geometrical terms defined relative to each scatterer. The total pressure is a simple sum of each contribution, assuming no coupling between scatterers. The resulting expression for the scattering kernel for J objects is

$$K_S(\xi',\eta';\xi,\eta) = -\frac{i}{6\pi} A_r(\omega) A_e(\omega) \sum_{j=1}^{J} (ka_j)^3 \times K_j(\xi',\eta';\xi,\eta) e^{ik(r_j + r'_j)}, \quad (15)$$

$$K_j(\xi',\eta';\xi,\eta) \equiv \frac{1}{r_j r'_j} \sum_{n=0}^{\infty} \sum_{l=-n}^{n} B_{ln}^{(j)}(\xi,\eta) Q_n(-2ikr'_j) \times P_n^l(-\cos\theta_j) e^{il\varphi_j}, \quad (16)$$

where the subscript j denotes that the quantity refers to the jth object. In particular, the distances r_j and r'_j are given by $|\boldsymbol{\xi} - \mathbf{R}^{(j)}|$ and $|\boldsymbol{\xi}' - \mathbf{R}^{(j)}|$, respectively, where $\mathbf{R}^{(j)}$ is the position of the jth scatterer.

The SVD for the scattering kernel is

$$K_S \circ \Phi = \mu_{TR} \Phi^*. \quad (17)$$

In the case of a single object, the eigenfunctions of the full scattering kernel were written as a product of an amplitude and phase factor [Eq. (7)]. The form of the phase factor is determined by the requirement that the eigenfunction focus on the scatterer. This motivates looking for solutions in the case of multiple scatterers of the form

$$\Phi(\xi,\eta) = \sqrt{i\frac{A_e^* A_r^*}{|A_e A_r|}} \phi^{(q)}(\xi,\eta) e^{-ikr_q},$$

$$\mu_{TR} = \frac{(ka_q)^3}{6\pi} |A_e A_r| \mu^{(q)}, \quad (18)$$

where q refers to the qth scatterer. The SVD then becomes

$$\sum_{j=1}^{J} \left(\frac{a_j}{a_q}\right)^3 e^{ik(r'_j - r'_q)} \int_S K_j(\xi',\eta';\xi,\eta) e^{ik(r_j - r_q)} \phi^{(q)}(\xi,\eta) dS = \mu^{(q)} [\phi^{(q)}(\xi',\eta')]^*. \quad (19)$$

If the scatterers are well-separated, the complex exponential factor in the integrand is rapidly oscillating except for $j = q$, when it becomes one. Assuming the other factors in the integral vary slowly, the $j = q$ term dominates the sum. Thus, to leading order, Eq. (19) and the SVD becomes

$$K_q \cdot \phi^{(q)} = \mu^{(q)} [\phi^{(q)}]^*. \quad (20)$$

Since q was arbitrary, the original SVD for the scattering kernel decouples into J singular value decompositions, each equivalent to the single scatterer case, assuming no eigenvalues coincide. Thus, the eigenstates of the TR operator for multiple scatterers are the collection of eigenstates associated with each scatterer if the objects are well separated, no eigenvalues coincide, and the TRM is far from any given scatterer.

Prada *et al.*[5] showed, in the case of two scatterers with the same apparent reflectivity, that the eigenvalues associated with each scatterer coalesce into a single eigenvalue with two orthogonal eigenvectors. These eigenvectors do not necessarily focus on a single scatterer but possibly on both simultaneously. In fact, the choice of eigenvectors is somewhat arbitrary since any linear combination of the original two eigenvectors is also an eigenvector. If the scatterers are well-separated, it is possible to choose the eigenvectors so that each focuses primarily on a single scatterer, i.e., they have the form in Eq. (18). In this sense, it is still possible to regard the spectrum of the TR operator for well-separated multiple scatterers as a combination of the spectra of each scatterer alone, with the caveat that if eigenvalues coincide between

some scatterers, the associated eigenfunctions are not unique. Any linear combination of the eigenfunctions of coincident eigenvalues is also an eigenfunction of the full TR operator for multiple scatterers.

V. FINITE SPHERE EXAMPLE

Consider the example of the linear array of 96 elements used by Prada in Ref. 3. The center-to-center element spacing was 0.41 mm and it operated at a frequency of 3 MHz in water. Let a sphere be placed 90 mm in front of the midpoint of the array. Using the same approach shown in Ref. 6, the discrete array version of the singular value decomposition can be written

$$\mathbf{K}\cdot\boldsymbol{\varphi}=\boldsymbol{\mu}\boldsymbol{\varphi}^*, \tag{21}$$

where $\mathbf{K}$ is the reduced transfer matrix,[6] and $\boldsymbol{\varphi}$ is the singular vector which specifies the complex amplitudes for each array element associated with each singular state. For scattering from a finite sphere, $\mathbf{K}$ is given by

$$K_{pq}=\frac{z_M^2}{r_p r_q}\sum_{n=0}^{N} B_n P_n\left(-\frac{\xi_p\xi_q+z_M^2}{r_p r_q}\right), \quad r_p=\sqrt{\xi_p^2+z_M^2}, \tag{22}$$

where ξ_p is the position of the pth element from the array midpoint, and $Q_n(-2ikr_p)\rightarrow 1$. The coefficients B_n were given in Sec. II, with explicit expressions as functions of ka found in an article by Feuillade and Clay.[15] The number of terms in the sum (N) was set to the smallest integer greater than three plus the maximum ka of interest.[15] The reduced transfer matrix was evaluated and the singular value decomposition was performed for a range of ka values for the case of a spherical bubble and a hard rubber sphere in water ($c_0=1500$ m/s, $\rho_0=1$ gm/cm^3).

For the case of the bubble, the first 11 B_n coefficients were evaluated for ka in the range of zero to 2. The sound speed and density of the air were taken from Ref. 15 ($c_s=330$ m/s, $\rho_s=1.26$ mg/cm^3). The backscatter form function,[15] given by

$$F_\infty=\frac{2}{3}(ka)^2\left|\sum_{n=0}^{N}(-1)^n B_n\right|,$$

is shown in Fig. 3(a). The most prominent feature is the large monopole resonance at $ka\approx 0.0136$. This is reflected in the magnitude of the B_0 coefficient in Fig. 3(b). Other resonances can be seen in the higher-order coefficients and contribute to the peaks in the backscatter form function. The higher-order coefficients increase in magnitude relative to the first coefficient as ka increases. For $ka>1.2$ the dipole becomes the dominant contribution to the scattering. Note that only the monopole and dipole (B_1) term are nonzero in the Rayleigh scattering limit $ka\rightarrow 0$.

The SVD problem was solved using MATLAB;[16] the first 20 singular values for the time-reversal array are plotted as a function of ka in Fig. 3(c). Peaks associated with the resonances in the B_n coefficients can be seen, though not all singular values respond to a given resonance. This is consistent with the observation of Ref. 6 that only subsets of the terms in the expansion for the scattered field contribute to a given eigenstate of the time-reversal operator. The first singular value dominates the spectrum over the entire range of ka, even though the dipole term is greater than the monopole term for $ka>1.2$. The total number of eigenvalues is bounded by the number of elements in the array (96). In addition, the reduced transfer matrix [Eq. (22)] is separable

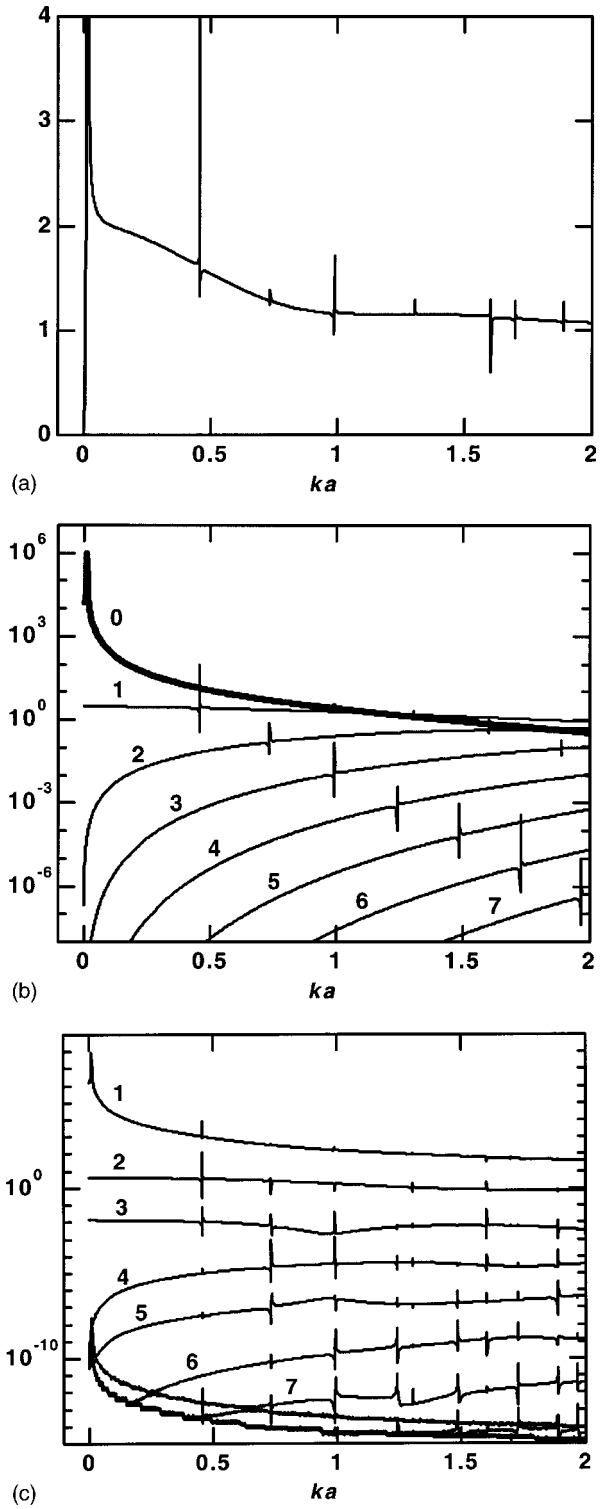


FIG. 3. Analysis of time reversal with a 96-element linear array and a bubble for a range of ka. (a) Backscatter form factor showing resonant peaks. (b) Magnitude of coefficients, $|B_n|$, in the expansion for the scattered field (labels indicate values of n). (c) Spectrum of the first 20 singular values μ_n (labels are shown for the first seven).

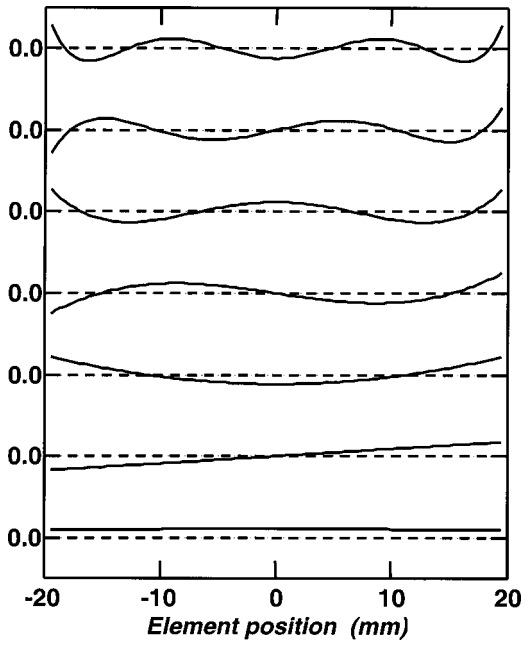


FIG. 4. The first seven singular vectors for a bubble at $ka=2$, starting with $n=1$ at the bottom and progressing to $n=7$ at the top. Solid line is the real part and dashed line is the imaginary part.

like the kernel of Eq. (11). The number of independent terms is then either $(N+1)(N+2)/2 \approx (ka)^2/2$, or the number of array elements (96), whichever is smaller. In fact, the number of singular values that exceed the noise floor created by the round-off error is much less than that predicted by this estimate. Apparently there are a substantial number of singular values too small to rise above the round-off error. These can be seen emerging from the noise floor in Fig. 3(c) as *ka* increases. Note that in the Rayleigh limit, only the first three singular values are finite. The ratios of these singular values in the limit $ka \to 0$ agree with the results in Ref. 6 for a linear array. Figure 4 shows the singular vectors for $ka=2$, which are ordered by number of zero crossings. This ordering can be useful in tracking the identities of singular values that may cross at particular values of *ka*.

The second case of a hard rubber sphere ($c_s = 2400$ m/s, $\rho_s = 1.1$ gm/cm^3) is shown in Fig. 5 for the range $0 < ka < 19$. Twenty-five coefficients were used to construct the scattered field. The backscatter form factor [Fig. 5(a)] shows a complicated oscillatory behavior caused by the successive dominance of higher-order terms in the scattered field [Fig. 5(b)]. The B_n coefficients show no prominent resonant peaks, but each has specific values of *ka* where it crosses zero. The first 20 singular values in Fig. 5(c) also have isolated zeros in their behavior with increasing *ka*. In addition to the zeros, some singular values appear to cross, exchanging identities, at certain values of *ka*. However, the isolated zeros and crossing points do not appear to be associated with the zeros of the B_n coefficients. Again, we note that the number of singular values above the noise floor is much smaller than the number of independent terms would suggest could exist. However, we do see more singular values rising above the floor as *ka* increases. Thus, the number of significant eigenvalues increases with *ka* but not as rapidly as the estimate in Sec. II, $(ka)^2/2$. Figure 6 shows the singular vectors for $ka=18$, ordered by zero crossings. Note that in spite of several points at which the singular values seem to cross before $ka=18$, the first six singular vectors are naturally ordered by zero crossings. Apparently, any rearranging of order through previous crossing points has been untangled before this value of *ka*.

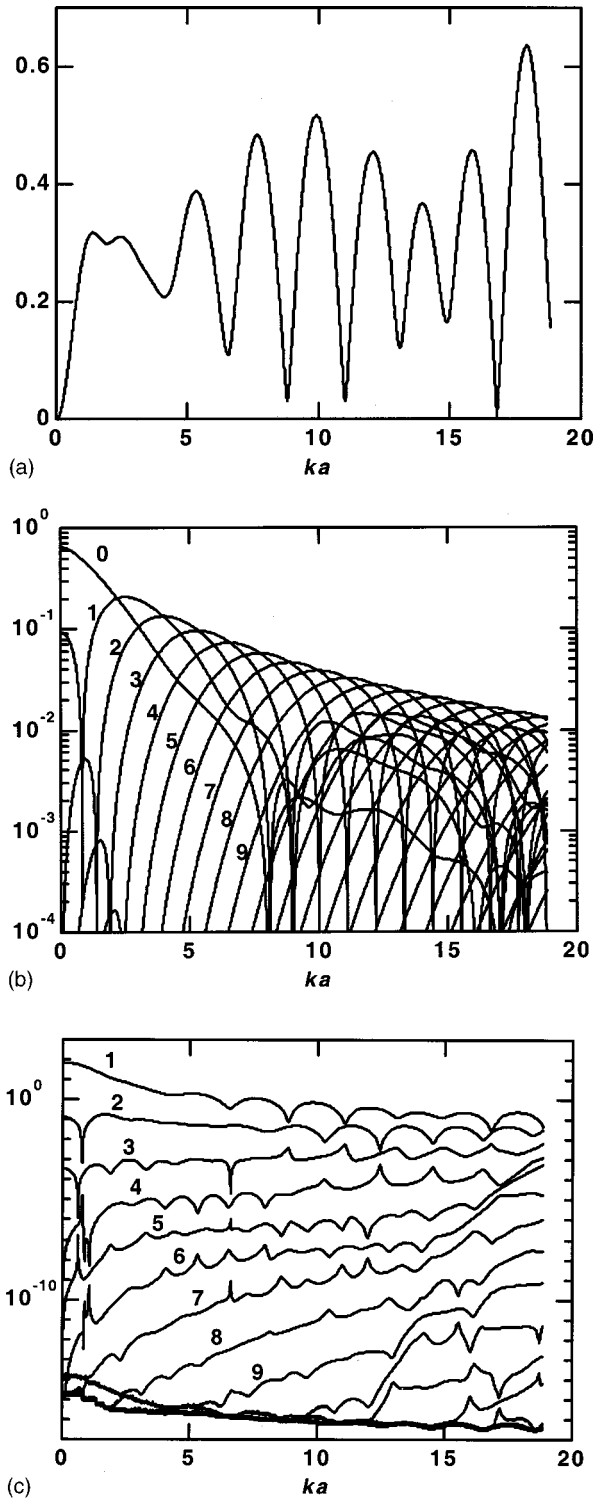


FIG. 5. Analysis of time reversal with a 96-element linear array and a bubble for a range of *ka*. (a) Backscatter form factor. (b) Magnitude of coefficients, $|B_n|$, in the expansion for the scattered field (labels indicate values of *n*). (c) Spectrum of the first 20 singular values μ_n (labels are shown for the first nine).

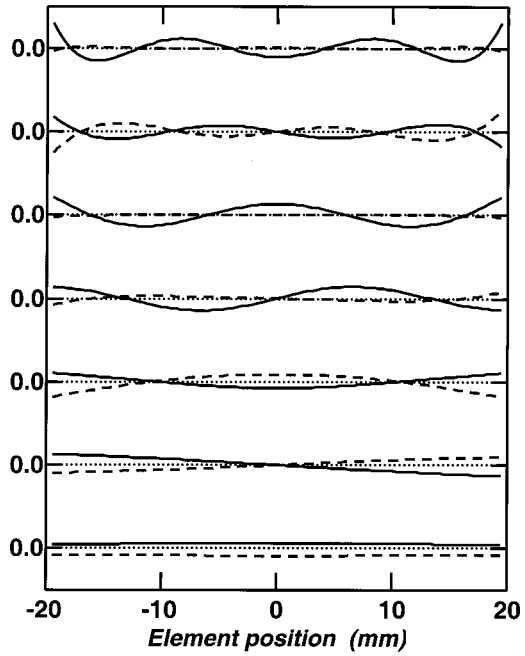


FIG. 6. The first seven singular vectors for a rubber sphere at $ka=18$, starting with $n=1$ at the bottom and progressing to $n=7$ at the top. Solid line is the real part and dashed line is the imaginary part.

It is interesting to investigate one of the points at which singular values appear to cross. Figure 7 shows the behavior of the second, third, and fourth singular values for the hard rubber sphere in the neighborhood of an apparent crossing near $ka=0.7618$. It is clear that the second and fourth singular values closely approach each other but do not cross. The third singular value is slowly varying and crosses the second singular value at two values of *ka*. Figure 8 shows the behavior of the singular vectors around the apparent crossing point. In these images of $-\log(|\boldsymbol{\varphi}^{(n)}|)$, $n=2,3,4$, the zeros of the singular vectors appear white on a dark background, making it easy to track the number and positions of the zeros as functions of *ka*. The point of closest approach for the singular values is a point where the second and fourth singular vectors go through an abrupt change in structure [Figs. 8(a) and (c)]. There is also a region between $ka=0.761$ and

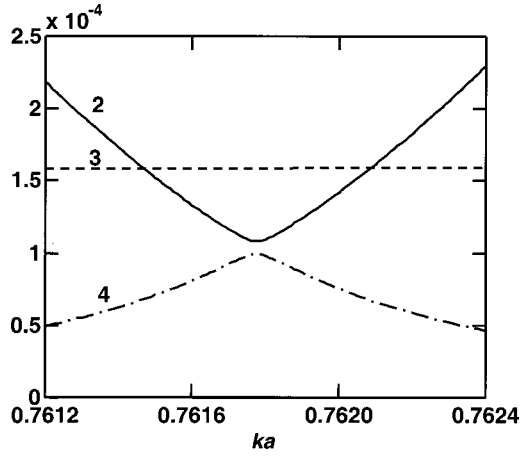


FIG. 7. Behavior of singular values μ_n, $n=2,3,4$ (labeled) around the apparent crossing point near $ka=0.7618$.

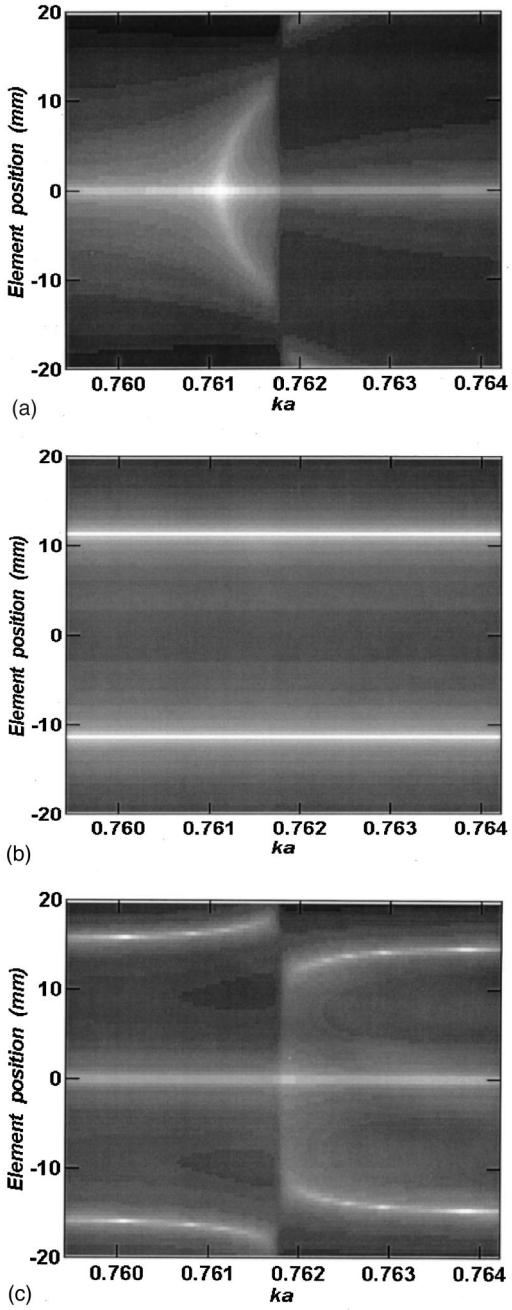


FIG. 8. Images of $-\log(|\boldsymbol{\varphi}^{(n)}|)$ for $n=2$ (a); 3 (b); and 4 (c), around the apparent crossing point near $ka=0.7618$. Zeros of the singular vectors appear as white against a dark background.

the closest approach point where the second vector has two minima in addition to the central zero. The minima occur where the imaginary part of the vector goes through zero but not the real part. Both real and imaginary parts go through zero at the center. All the singular vectors have the same number of zeros after the approach point as they did for $ka<0.761$. There is no permanent change of order in the singular values and singular vectors. In addition, even though the change in structure in the second and fourth singular vectors is rapid at the point of closest approach, it is not discontinuous, as Fig. 9 shows. The vectors vary continuously with *ka* through the approach point.

The question of whether the singular values cross at a

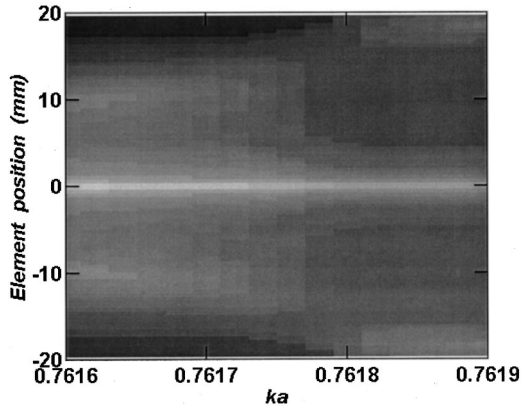


FIG. 9. Detail of $-\log(|\varphi^{(n)}|)$ for $n=2$ around $ka=0.7618$ showing the rapid but continuous change of structure near the point of closest approach with the fourth singular value.

given value of *ka* requires close examination of the behavior near that point. At the resolution of Fig. 5(c) ($\Delta ka=0.005$) there appear several crossing points for $ka<2$ and one near $ka=6.6$. Other points only appear to be crossing points, e.g., $ka=12.5$. Unfortunately, in this example the best candidates for crossing points correspond to very small spheres ($a<0.1$ mm) at a frequency of 3 MHz and would be difficult to observe experimentally. However, experimental investigation might be easier at lower frequencies in the audible range.

VI. SUMMARY AND CONCLUSIONS

By expanding the pressure field scattered from an object in spherical harmonics, an expression for the time-reversal operator can be obtained for scatterers of finite size. The analysis requires only that the scatterer be small compared with the distance to the TRM, not small compared with the wavelength. If the scatterer is a sphere, the kernel in the TR operator is separable and the number of eigenstates can be estimated from the number of terms required to accurately represent the scattered field. However, several of these eigenvalues may be smaller than the noise in a real system and would be unobservable. The number of observable eigenvalues increases with scatterer size. For more complicated scatterer shapes, the number of eigenvalues may be indefinitely large. For multiple noninteracting scatterers, the spectrum of the TR operator is the combination of the spectra associated with each individual scatterer.

The singular values and associated vectors for a spherical bubble and hard rubber sphere were calculated for the particular case of Prada's linear array.[3] For the bubble, the singular value spectrum shows peaks associated with resonances in the scattering. For the rubber sphere, there appear several values of *ka* where singular values appear to cross, exchanging their order. A detailed investigation of one such point shows two singular values approaching one another but not crossing, the ordering remaining the same. It is not clear at this time what determines this behavior, and further investigation is warranted.

This analysis explores the behavior of the time-reversal operator in the region between Rayleigh scattering from small objects (Ref. 6) and scattering from large extended objects (Refs. 7–10) where the curvature of the incident wavefront is no longer negligible. The number and magnitude of the eigenvalues depend on the TRM geometry and the shape and size of the scatterer. Each eigenstate of the TR operator can be considered to be a type of spatial resonance between the TRM and the scatterer. The field produced by the TRM couples with a characteristic scattering mode of the object that creates a scattered field that is proportional to the field originally radiated by the TRM. Since the number and relative amplitudes of the scattering modes depend only on scatterer characteristics, we would expect similar behavior for objects in media more complicated than that used here for the analysis. As such, the spectrum provides information that could be used for target characterization and classification in a variety of different media.

ACKNOWLEDGMENTS

The author would like to thank James Berryman, James Candy, and Art Gautesen for stimulating discussions and their suggestions in the preparation of the manuscript. The author would also to thank the reviewers for their helpful comments and observations. This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory, under Contract No. W-7405-ENG-48.

APPENDIX: SINGULAR VALUE DECOMPOSITION FOR COMPLEX SYMMETRIC KERNELS

From Smithies,[17] the singular value decomposition for a general complex valued kernel is

$$K\circ\phi=\mu\chi, \quad K^{\dagger}\circ\chi=\mu\phi, \tag{A1}$$

where the dagger superscript indicates the Hermitian conjugate, $K^{\dagger}(x,y)\equiv K^{*}(y,x)$, and the functions ϕ and χ are the eigenfunctions of the operators $K^{\dagger}\circ K$ and $K\circ K^{\dagger}$, respectively. The values of μ where (A1) has nontrivial solutions are real and occur in positive and negative pairs (unless zero). The non-negative values are called the singular values of the system. For the particular case that K is symmetric, though still complex valued, the Hermitian conjugate is equivalent to the usual complex conjugate. The second equation becomes $K^{*}\circ\chi=\mu\phi$, which is equivalent to the complex conjugate of the first equation if we identify $\chi=\phi^{*}$. This leads to the singular value decomposition

$$K\circ\phi=\mu\phi^{*}, \quad K^{*}\circ\phi^{*}=\mu\phi. \tag{A2}$$

Since the second equation is simply the complex conjugate of the first, it is omitted in the text. If the kernel is the sum of N separable terms, the integral equation pair of (A1) can be transformed into a matrix singular value decomposition of dimension N. This implies N possible singular values and eigenfunctions.

[1] M. Fink, D. Cassereau, A. Derode, C. Prada, P. Roux, M. Tanter, J-L. Thomas, and F. Wu, "Time-reversed acoustics," Rep. Prog. Phys. **63**, 1933–1995 (2000).
[2] M. Fink and C. Prada, "Acoustic time-reversal mirrors," Inverse Probl. **17**, R1–R38 (2001).

[3] C. Prada and M. Fink, ''Eigenstates of the time-reversal operator: A solution to selective focusing in multiple-target media,'' Wave Motion **20**, 151–163 (1994).
[4] 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).
[5] C. Prada, S. Manneville, D. Spoliansky, and M. Fink, ''Decomposition of the time-reversal operator: Detection and selective focusing on two scatterers,'' J. Acoust. Soc. Am. **99**, 2067–2076 (1996).
[6] D. H. Chambers and A. K. Gautesen, ''Time reversal for a single spherical scatterer,'' J. Acoust. Soc. Am. **109**, 2616–2624 (2001).
[7] J-L. Thomas, P. Roux, and M. Fink, ''Inverse scattering analysis with an acoustic time-reversal mirror,'' Phys. Rev. Lett. **72**(5), 637–640 (1994).
[8] S. Komilikis, C. Prada, and M. Fink, ''Characterization of extended objects with the D.O.R.T. method,'' Proceedings of the IEEE Ultrasonics Symposium (1996), Vol. 2, pp. 1401–1404.
[9] C. Prada and M. Fink, ''Separation of interfering acoustic scattered signals using the invariants of the time-reversal operator. Application to Lamb waves characterization,'' J. Acoust. Soc. Am. **104**, 801–807 (1998).
[10] J.-L. Thomas, F. Wu, and M. Fink, ''Time-reversal focusing applied to lithotripsy,'' Ultrason. Imaging **18**, 106–121 (1996).
[11] D. Cassereau and M. Fink, ''Focusing with plane time-reversal mirrors: An efficient alternative to closed cavities,'' J. Acoust. Soc. Am. **94**, 2373–2386 (1993).
[12] G. Harris, ''Transient field of a baffled planar piston having an arbitrary vibration amplitude distribution,'' J. Acoust. Soc. Am. **70**, 186–204 (1981).
[13] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1968).
[14] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (National Bureau of Standards, U.S. Department of Commerce, Washington, DC, 1972), p. 439.
[15] C. Feuillade and C. S. Clay, ''Anderson (1950) revisited,'' J. Acoust. Soc. Am. **106**, 553–564 (1999).
[16] MATLAB (The Math Works Inc., Boston, 1993).
[17] F. Smithies, *Integral Equations* (Cambridge University Press, Cambridge, UK, 1970).

Planar near-field acoustical holography in a moving medium[a)]

Richard J. Ruhala[b)]
University of Southern Indiana, Engineering Department, 8600 University Boulevard, Evansville, Indiana 47712

David C. Swanson[c)]
Graduate Program in Acoustics, The Applied Research Laboratory, The Pennsylvania State University, Room 9 ASB, P.O. Box 30, State College, Pennsylvania 16804

(Received 15 August 2000; revised 15 March 2002; accepted 15 March 2002)

Near-field acoustical holography (NAH) is a well-established method to study acoustic radiation near a stationary sound source in a homogeneous, stationary medium. However, the current theory of NAH is not applicable to moving sound sources, such as automobiles and trains. In this paper, the inclusion of a moving medium (i.e., moving source and receiver) is introduced in the wave equation and a new set of equations for plannar NAH is developed. Equations are developed for the acoustic pressure, particle velocity, and intensity when mean flow is either parallel or perpendicular to the hologram plane. If the source and the measurement plane are moving at the same speed, the frequency Doppler effect is absent, but a wave number Doppler effect exists. This leads to errors when reconstructing the acoustic field both towards and away from the source using static NAH. To investigate these errors, a point source is studied analytically using planar NAH with flow in one direction. The effect of the medium moving parallel to the hologram plane is noted by a shift of the radiation circle in wave number space (k-space). A k-space Green's function and a k-space filter are developed that include the effects of the moving medium. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1477931]

PACS numbers: 43.20.−f, 43.28.Py, 43.60.Sx [ANN]

I. INTRODUCTION

Near-field acoustical holography (NAH) is a well-known and extremely powerful methodology that facilitates reconstruction of the three-dimensional acoustic field from a two-dimensional measurement in the acoustical near-field of a sound source. The theory of NAH has been developed for Cartesian (planar),[1–3] cylindrical,[4,5] spherical,[6] and general[7] coordinate systems.

The motivation of this work was to improve the accuracy of using NAH when applied to measuring tire noise on a moving vehicle.[8–10] NAH has also been applied to tire noise on a moving vehicle by Rasmussen and Gade.[11] However, they did not include the effect of the moving sound source when using NAH, which may have led to additional errors when reconstructing the sound field closer to the tire. Other acoustical holography methods have been developed for application to a moving source that involve a stationary vertical array of microphones mounted very close to the path of a moving vehicle, including the moving frame technique for planar acoustic holography.[12–15] Using a stationary array has many advantages over an array attached to a moving vehicle, including a less complex measurement system. However, the moving frame technique is limited to sound with discrete frequencies and coherent bandlimited noise. Of course, using a step-by-step scanning array mounted very close to a moving tire (necessary for using the flow-inclusive NAH presented in this paper) creates other challenges, such as rejection of flow-induced noise and potential changes in the noise characteristics over a long measurement time. These issues are beyond the scope of this paper and are addressed elsewhere.[8,10] Also, the rotating effects of the tire and the possible temperature gradients are not addressed in this theory. The moving source (tire) and receivers (attached microphone array) are analogous to a stationary source and receiver in a moving medium. NAH theory in a moving medium has applications beyond tire noise to include other moving sound sources.

In this paper, a brief review of NAH in a stationary medium is presented in order to develop NAH in a moving medium. Next, the convected acoustic equations are presented. These equations are then used to develop the theory of NAH in a moving medium, including a new Green's function. Equations are presented for flow parallel and perpendicular to the NAH plane. Finally, analytical examples with a moving point source help to demonstrate the improvement of using the flow-inclusive NAH theory.

II. REVIEW OF PLANAR NAH IN A STATIONARY MEDIUM

A brief description of planar NAH in a stationary medium is presented in order to facilitate the development of NAH in a moving medium. For a complete derivation of NAH, the reader is referred to other papers.[2,3,8,16] The time convention used for the complex values in this paper is $e^{j\omega t}$, where $j=\sqrt{-1}$.

The fundamental equation in real space for NAH is

[a)]A part of this paper was presented at the 134th meeting of the Acoustical Society of America, 1–5 December 1997, San Diego, CA [R. J. Ruhala, D. C. Swanson, and C. B. Burroughs, "Extension of near-field acoustical holography to measurements in a moving medium," J. Acoust. Soc. Am. **102** (5-2), 3199(A) (1997)].

[b)]Electronic mail: rruhala@usi.edu
[c)]Electronic mail: dcs5@psu.edu

 0001-4966/2002/112(2)/420/10/$19.00

$$\hat{p}(x,y,z)=\int\int_{-\infty}^{+\infty}\hat{p}_H(x',y',z_H)\hat{G}'(x-x',y-y',z-z_H)\times dx'\,dy', \quad (1)$$

where $\hat{p}(x,y,z)$ is the complex pressure anywhere in volume V, $\hat{p}_H(x',y',z_H)$ is the measured complex pressure on the hologram plane, and $\hat{G}'(x-x',y-y',z-z_H)$ is the Green's function. The volume V is bounded by the source plane and contains no other acoustic sources or reflections.

Using the two-dimensional forward Fourier transform

$$\hat{p}_H(k_x,k_y,z_s)=\int_{-\infty}^{+\infty}\int\hat{p}_H(x,y,z_H)e^{jk_xx}e^{jk_yy}dx\,dy, \quad (2)$$

the convolution integral in Eq. (1) becomes

$$\hat{p}(k_x,k_y,z)=\hat{p}_H(k_x,k_y,z_H)\hat{G}'(k_x,k_y,d), \quad (3)$$

where $d=z-z_H$. The complex pressure in wave number space or k-space may be computed by multiplying the measured complex pressure on the hologram plane in k-space with the following Green's function in k-space:

$$\hat{G}'(k_x,k_y,d)=\begin{cases} e^{-jd|\hat{k}_z|} & \text{for } k_x^2+k_y^2\leqslant k^2 \\ e^{-d|\hat{k}_z|} & \text{for } k_x^2+k_y^2>k^2 \end{cases}, \quad (4)$$

where

$$\hat{k}_z=\begin{cases} \sqrt{k^2-k_x^2-k_y^2} & \text{for } k_x^2+k_y^2\leqslant k^2 \\ -j\sqrt{k_x^2+k_y^2-k^2} & \text{for } k_x^2+k_y^2>k^2 \end{cases}. \quad (5)$$

The term k is the acoustic wave number defined as

$$k=\frac{\omega}{c}, \quad (6)$$

where ω is the angular frequency in radians per second and c is the speed of sound in the fluid. The terms k_x and k_y are the wave number components in the x- and y direction, respectively.

This modified Green's function in k-space divides the pressure field into propagating and evanescent waves. It acts as the "propagator" in that it facilitates reconstruction of complex pressure on a parallel plane a distance d from the hologram plane. Pressure at wave numbers less than the acoustic wave number will propagate into the far field and contribute to the radiated sound power. But, at wave numbers greater than the acoustic wave number, the pressure decays exponentially as one moves away from the source, and does not contribute to the far-field power. These evanescent waves, however, increase exponentially during propagation back towards the source, and hence, contribute significantly to the identification of the acoustic sources. To measure the evanescent waves, the hologram plane must be in the acoustic near-field. Burns[16] has suggested that the hologram plan should be within a quarter wavelength of the noise source.

The pressure field may then be reconstructed at a plane parallel to the hologram plane, and transformed back to real space using the following two-dimensional inverse Fourier transform.

$$\hat{p}(x,y,z)=\frac{1}{4\pi^2}\int_{-\infty}^{+\infty}\int\hat{p}(k_x,k_y,z)e^{-jk_xx}e^{-jk_yy}dk_x\,dk_y. \quad (7)$$

By repeating this procedure to reconstruct the pressure on multiple planes, the three-dimensional acoustic pressure field may be reconstructed from a two-dimensional measurement.

In addition to reconstructing the acoustic pressure field, acoustic particle velocity and intensity may be reconstructed as well. Particle velocity is calculated using Euler's equation for momentum in a fluid (Ref. 17, p. 15). Assuming time-harmonic motion, Euler's equation takes the form

$$\hat{u}=\frac{j}{\omega\rho_o}\nabla\hat{p}, \quad (8)$$

where ρ_o is the mean density in the fluid. Combining Eqs. (7) and (8) yields the particle velocity components in the x-, y-, and z-directions

$$\hat{u}_x=\frac{1}{4\pi^2\omega\rho_o}\int_{-\infty}^{+\infty}\int k_x\hat{p}_H(k_x,k_y,z_H)e^{-j\hat{k}_zd}e^{-jk_xx}\times e^{-jk_yy}dk_x\,dk_y, \quad (9)$$

$$\hat{u}_y=\frac{1}{4\pi^2\omega\rho_o}\int_{-\infty}^{+\infty}\int k_y\hat{p}_H(k_x,k_y,z_H)e^{-j\hat{k}_zd}e^{-jk_xx}\times e^{-jk_yy}dk_x\,dk_y, \quad (10)$$

$$\hat{u}_z=\frac{1}{4\pi^2\omega\rho_o}\int_{-\infty}^{+\infty}\int \hat{k}_z\hat{p}_H(k_x,k_y,z_H)e^{-j\hat{k}_zd}e^{-jk_xx}\times e^{-jk_yy}dk_x\,dk_y, \quad (11)$$

such that

$$\mathbf{u}=u_x\mathbf{e}_x+u_y\mathbf{e}_y+u_z\mathbf{e}_z. \quad (12)$$

The active acoustic intensity components are determined from the complex particle velocity components and complex acoustic pressure.

$$I_m=\tfrac{1}{2}\,\mathrm{Re}\{\hat{p}\hat{u}_m^*\}, \quad (13)$$

where $m=x$, y, or z, and such that

$$\mathbf{I}=I_x\mathbf{e}_x+I_y\mathbf{e}_y+I_z\mathbf{e}_z. \quad (14)$$

(Note that * denotes a complex conjugate function.)

In theory, the hologram plane is infinite and continuous, but in practice, it is finite and discrete, and so it is called the hologram aperture. The complex pressure is sampled at equally spaced points in the vertical and horizontal positions along the grid-like hologram aperture, as shown in Fig. 1. The spacing between points in the x- and y-directions, Δx and Δy, determines the upper wave number/frequency limit, which occurs when the maximum of either Δx or Δy corresponds to a half wavelength

$$k_{\max}=\frac{\pi}{\max(\Delta x \text{ or } \Delta y)}, \quad (15)$$

$$f_{\max}=\frac{c}{2\max(\Delta x \text{ or } \Delta y)}. \quad (16)$$

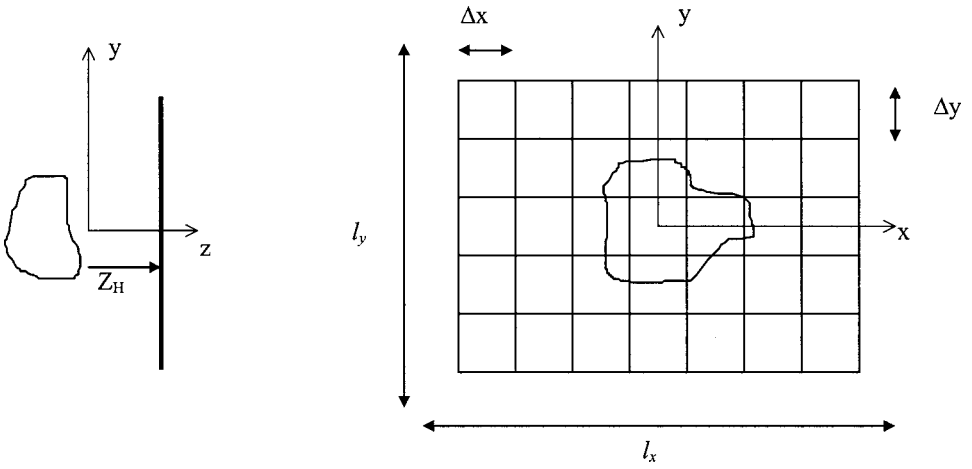


FIG. 1. Dimensions of the hologram aperture.

The size of the rectangular-shaped aperture should be much larger than the source, so that the pressures beyond the measured boundary can be neglected. The aperture dimensions are $l_x = n_x \Delta x$ and $l_y = n_y \Delta y$, where n_x and n_y are the number of measurement points in the x- and y-directions, respectively. The size of the hologram aperture determines the lower wave number/frequency limit

$$k_{\min} = \frac{\pi}{\min(l_x \text{ or } l_y)}, \tag{17}$$

$$f_{\min} = \frac{c}{2\min(l_x \text{ or } l_y)}. \tag{18}$$

The implementation of a spatially discrete hologram set may involve a two-dimensional spatial window to reduce leakage. Several windows have been used, such as the well-known Hanning and Hamming windows.[18] Kwon and Kim have developed a so-called "minimum error window" for use with NAH.[13,19] One problem using this window is that it becomes numerically unstable for large array sizes with double-precision computation. The authors of this work used a window similar to one applied by Burns.[16] It is near-unity for most of the aperture, but then tapers to near-zero at the edges. The window generating function in one dimension is

$$W(x) = \begin{cases} 1 - 0.5e^{(x/x_{\text{cut}} - 1)/\alpha} & \text{for } x \leqslant x_{\text{cut}} \\ 0.5e^{(1 - x/x_{\text{cut}})/\alpha} & \text{for } x > x_{\text{cut}}, \end{cases} \tag{19}$$

where $x_{\text{cut}} = 0.8\, x_{\max}$ and $\alpha = 0.2$. The window in the y direction, $W(y)$ is calculated in similar form. The 2D window is then calculated as $W(x,y) = W(x)W(y)$.

The array is zero padded to at least twice the original length in both dimensions prior to executing the 2D spatial discrete Fourier transform (DFT) to perform the circular convolution. The 2D array is now N_x by N_y points.

The 2D Fourier transform in Eq. (2) becomes the following forward DFT form:

$$\hat{p}_H[m_1,m_2] = \sum_{n_1=0}^{N_x-1} \sum_{n_2=0}^{N_y-1} \hat{p}_H[n_1,n_2] e^{j2\pi n_1 m_1/N_x} e^{j2\pi n_2 m_2/N_y}. \tag{20}$$

(Note the sign of the kernel is positive.) After the 2D DFT, the bins of the 2D complex pressure in k-space are ordered from 0 to $(N-1)$ in each direction. Next, the bins are *reordered* to range from $-N/2$ to $(N-2)/2$ so that the zero wave number bin is in the center. When reconstructing the acoustic field back towards the source, a k-space filter should be used to attenuate noise in the higher wave number components that increase exponentially in back propagation. Several k-spaced filters have been proposed, such as those by Veronesi and Maynard.[3] The filter used in this study was developed by Kwon (Ref. 13, p. 55), and is defined as

$$H(k_x,k_y) = \begin{cases} \begin{cases} 1 - 0.5e^{(k_r/k_c - 1)/\alpha} & \text{for } k_r \leqslant k_c \\ 0.5e^{(1 - k_r/k_c)/\alpha} & \text{for } k_r > k_c \end{cases} & \text{for } z_H \leqslant \dfrac{\lambda}{3} \\ \begin{cases} 1 & \text{for } k_r \leqslant k \\ 0 & \text{for } k_r > k \end{cases} & \text{for } z_H > \dfrac{\lambda}{3} \end{cases}, \tag{21}$$

where $\alpha = 0.2$, $k_r = \sqrt{k_x^2 + k_y^2}$, $k_c = 2\pi/3\, z_H$, and λ is the acoustic wavelength. The filtered complex pressure in k-space is then multiplied by the modified Green's function defined in Eq. (4) to calculate the complex pressure in k-space on a plane a distance d away form the source. The resulting bins of the complex array are *deordered* to range from 0 to $(N-1)$ before the reverse 2D DFT is performed using the following form:

$$\hat{p}[n_1,n_2] = \frac{1}{N_x N_y} \sum_{m_1=0}^{N_x-1} \sum_{m_2=0}^{N_y-1} \hat{p}[m_1,m_2] \times e^{-j2\pi/n_1 m_1/N_x} e^{-j2\pi/n_2 m_2/N_y}. \tag{22}$$

Acoustic particle velocity components are then computed using

$$\hat{u}_x[n_1,n_2] = \frac{1}{N_x N_y \omega \rho_o} \sum_{m_1=0}^{N_x-1} \sum_{m_2=0}^{N_y-1} k_x \hat{p}[m_1,m_2] \times e^{-j2\pi n_1 m_1/N_x} e^{-j2\pi n_2 m_2/N_y}, \tag{23}$$

$$\hat{u}_y[n_1,n_2] = \frac{1}{N_x N_y \omega \rho_o} \sum_{m_1=0}^{N_x-1} \sum_{m_2=0}^{N_y-1} k_y \hat{p}[m_1,m_2] \times e^{-j2\pi n_1 m_1/N_x} e^{-j2\pi n_2 m_2/N_y}, \tag{24}$$

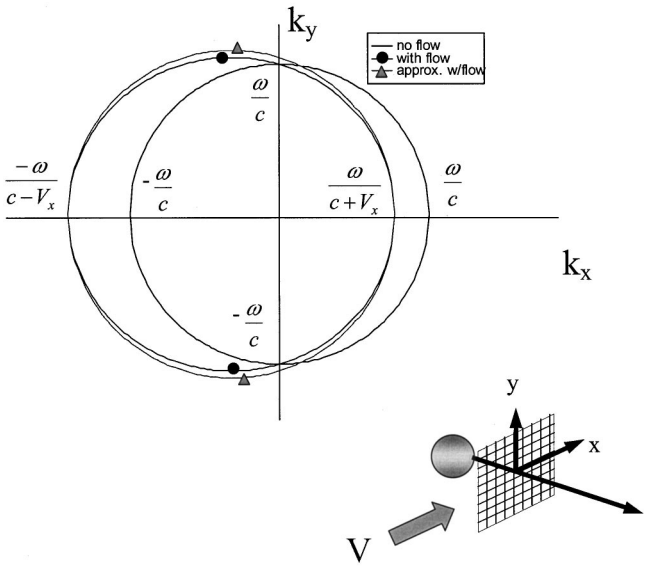


FIG. 2. Wave number Doppler effect of moving medium on radiation region. Medium is flowing in the positive x direction at Mach number of 0.3. Also shown is circular approximation of radiation function for moving medium using Eq. (42).

$$\hat{u}_z[n_1,n_2]=\frac{1}{N_xN_y\omega\rho_o}\sum_{m_1=0}^{N_x-1}\sum_{m_2=0}^{N_y-1}\hat{k}_z\hat{p}[m_1,m_2]\times e^{-j2\pi n_1m_1/N_x}e^{-j2\pi n_2m_2/N_y}. \tag{25}$$

The active intensity components are calculated as before in Eq. (13). The sound power flowing through the hologram aperture may be calculated by

$$\Pi=\Delta x\Delta y\sum_{n_i=1}^{n_x}\sum_{n_j=1}^{n_y}I_{zH}[n_i,n_j], \tag{26}$$

where I_{zH} is the intensity in the normal direction on the hologram aperture.

III. CONVECTED ACOUSTIC EQUATIONS

The theory of near-field acoustical holography (NAH) presented above is restricted to a stationary medium. The application of NAH proposed in this study is conducted in a moving medium, i.e., moving source and receiver in a stationary medium. Therefore, NAH theory is extended for application in a moving medium. The theory begins by including mean flow velocities into the wave equation, which then may be referred to as the convected wave equation. Morse and Ingard derived these equations using a transformation of variables procedure (Ref. 20, pp. 700–701). The convected wave equation is derived here based on physical quantities of pressure, velocity, and density.

The nonlinear form of the continuity equation is (Ref. 21, p. 102)

$$\frac{\partial\rho}{\partial t}+\nabla\cdot(\rho_t\mathbf{u}_t)=0, \tag{27}$$

where

$$\mathbf{u}_t=\mathbf{V}+\mathbf{u}, \tag{28}$$

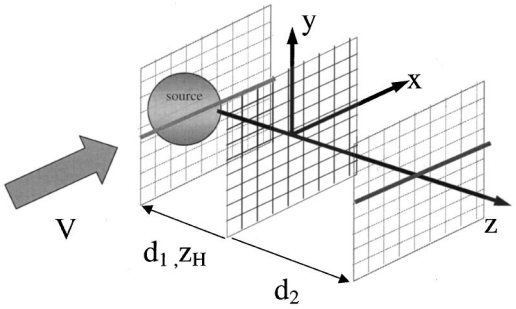
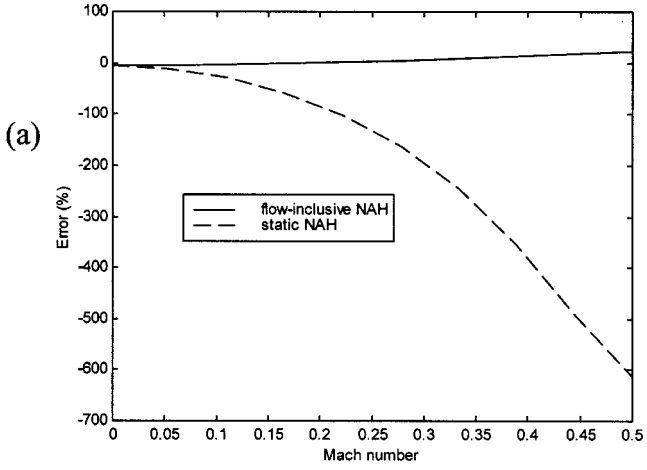


FIG. 3. Geometry for analytical study of the impact of moving medium for a point source and on NAH. Parameters: $\Delta x=\Delta y=z_H=0.05$ m, $n_x=n_y=32$, $c=343$ m/s, $\rho_o=1.2$ kg/m^3. Reconstruct at planes: $z=0$ m ($d_1=-0.05$ m) and $z=0.5$ m ($d_2=0.45$ m). Frequency is 1000 Hz. Dark lines on reconstructed planes indicate where acoustic pressure values are presented in Figs. 6 and 8.

and

$$\rho_t=\rho_o+\rho. \tag{29}$$

In the above equations, $\mathbf{u}_t$ is the total particle velocity vector consisting of the mean flow velocity, $\mathbf{V}$, and the acoustic



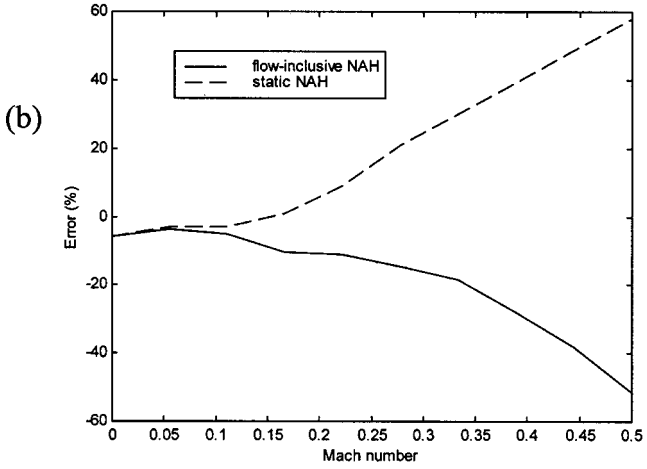


FIG. 4. Computation of the reconstruction error as a function of Mach number. Error is in term of the spatial-averaged-mean-squared pressure [Eqs. (53) and (54)] using NAH parameters noted in Fig. 3. Values for static and flow-inclusive NAH are shown (a) pressure reconstructed in the near-field on the source plane ($z=0$), and (b) pressure reconstructed in the farfield ($z=0.5$ m).

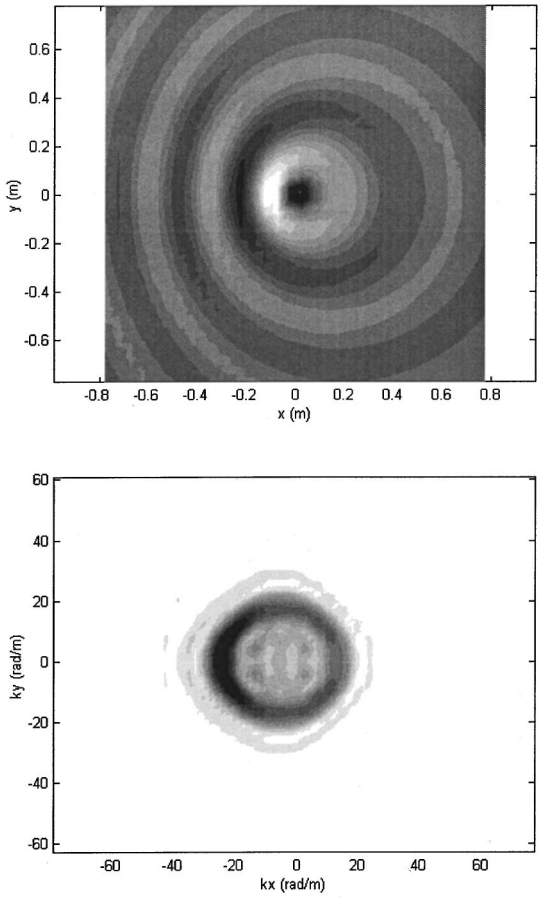


FIG. 5. (a) Real pressure on hologram aperture for point source in a moving medium. (b) Magnitude of complex pressure in 2D wave number space. Medium is flowing parallel to the hologram plane in the positive x direction at a Mach number of 0.3. Hologram plane is 0.05 m from the point source. Frequency is 1000 Hz.

particle velocity, $\mathbf{u}$. Likewise, ρ_t is the total density consisting of the static atmospheric density, ρ_o, and the fluctuating acoustic density, ρ. Assuming that $\mathbf{V}$ is constant, $|\mathbf{V}| \gg u$, $\rho_o \gg \rho$, and ignoring the nonlinear terms, Eq. (27) becomes the linear continuity equation with flow

$$\frac{\partial \rho}{\partial t} + \rho_o \nabla \cdot \mathbf{u} + \mathbf{V} \cdot \nabla \rho = 0. \tag{30}$$

The nonlinear form of Euler's equation is (Ref. 21, p. 104)

$$-\nabla p_t = \rho_t \frac{D\mathbf{u}_t}{Dt}, \tag{31}$$

where

$$p_t = p_o + \tfrac{1}{2}\rho_o |\mathbf{V}|^2 + p. \tag{32}$$

The derivative in Eq. (31) uses the total derivative so that $D\mathbf{u}_t/Dt = \mathbf{u}_t \cdot \nabla \mathbf{u}_t + (\partial \mathbf{u}_t/\partial t)$. In Eq. (32), p_t is the total pressure consisting of the static atmospheric pressure p_o, the acoustic pressure p, and the pressure head due to mean flow. Applying the same assumptions as with the continuity equation, Eq. (31) becomes the linear Euler's equation with flow.

$$-\nabla p = \rho_o \mathbf{V} \cdot \nabla \mathbf{u} + \rho_o \frac{\partial \mathbf{u}}{\partial t}. \tag{33}$$

The linearized equation of state is (Ref. 21, p. 105)

$$p = c^2 \rho, \tag{34}$$

which is the same with or without mean flow.

Combining Eqs. (30), (33), and (34) yields

$$\frac{1}{c^2}\left(\frac{\partial}{\partial t} + \mathbf{V} \cdot \nabla\right)^2 p = \nabla^2 p. \tag{35}$$

The above equation is the convected wave equation for real acoustic pressure, p, where $p = p(x,y,z,t)$. Note that if the mean velocity term is zero, Eq. (35) becomes the linear wave equation in a stationary medium for acoustic pressure.

With no loss in generality, the acoustic pressure in the above equations may be substituted with the complex acoustic pressure, $\hat{p}$. Further assuming time-harmonic motion so that $\hat{p}(x,y,z,t) = \hat{p}(x,y,z)e^{j\omega t}$, Eq. (35) reduces to

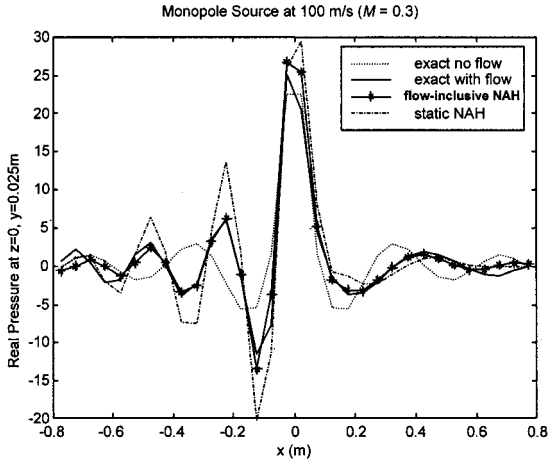
$$\nabla^2 \hat{p} + \left(\frac{\omega}{c}\right)^2 \left(1 - j\frac{1}{\omega}\mathbf{V} \cdot \nabla\right)^2 \hat{p} = 0, \tag{36}$$

and is called the convected Helmholtz equation (Ref. 20, p. 701). The complex pressure is now dependent only on space, where $\hat{p} = \hat{p}(x,y,z)$.

IV. MODIFIED EQUATIONS FOR PLANAR NAH IN A MOVING MEDIUM

Modified equations for planar NAH are derived here for mean flow parallel to the hologram plane and mean flow perpendicular to the hologram plane. These equations are developed based on the NAH theory in Sec. II and the convected acoustic equations in Sec. III.

(a)



(b)

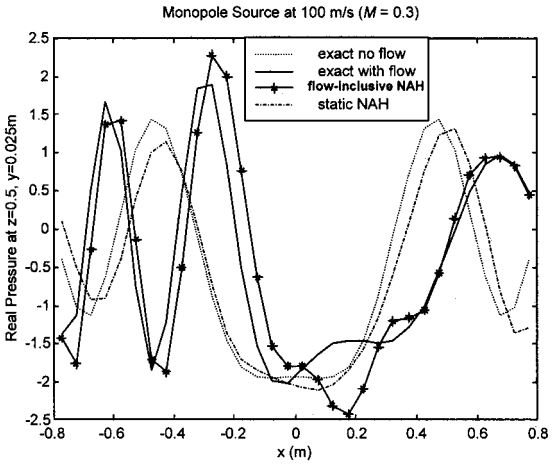


FIG. 6. Real pressure near monopole with mean velocity of 100 m/s ($M=0.3$). (a) Pressure along a line near the center from plane of pressure reconstructed in the near-field on the source plane ($z=0$). (b) Pressure along a line near the center of a plane in the far-field ($z=0.5$ m).

A. Flow parallel to NAH plane

Consider a planar hologram aperture in the $x-y$ plane in Cartesian coordinates. The flow in the fluid medium is moving in the positive x direction (and is therefore parallel to the hologram aperture) at a steady velocity, V_x. The z axis points away from the sound source as before.

The convected Helmholtz equation [Eq. (36)] simplifies to

$$\nabla^2\hat{p}+k^2\hat{p}=0, \tag{37}$$

where the wave number is

$$k=\frac{\omega}{c}\left(1-\frac{V_x k_x}{\omega}\right). \tag{38}$$

The new Green's function for NAH in wave number space is similar in form to the stationary condition [Eq. (4)], and is

$$\hat{G}'(k_x,k_y,d)=\begin{cases} e^{-jd|\hat{k}_z|} & \text{for } k_x^2+k_y^2\leqslant\left(\frac{\omega}{c}\right)^2\left(1-\frac{V_x k_x}{\omega}\right)^2 \\ e^{-d|\hat{k}_z|} & \text{for } k_x^2+k_y^2>\left(\frac{\omega}{c}\right)^2\left(1-\frac{V_x k_x}{\omega}\right)^2 \end{cases}, \tag{39}$$

where

$$\hat{k}_z=\begin{cases} \sqrt{\left(\frac{\omega}{c}\right)^2\left(1-\frac{V_x k_x}{\omega}\right)^2-k_x^2-k_y^2} \\ \qquad \text{for } k_x^2+k_y^2\leqslant\left(\frac{\omega}{c}\right)^2\left(1-\frac{V_x k_x}{\omega}\right) \\ -j\sqrt{k_x^2+k_y^2-\left(\frac{\omega}{c}\right)^2\left(1-\frac{V_x k_x}{\omega}\right)^2} \\ \qquad \text{for } k_x^2+k_y^2>\left(\frac{\omega}{c}\right)^2\left(1-\frac{V_x k_x}{\omega}\right) \end{cases}. \tag{40}$$

By setting $k_z=0$ above, the function separating the propagating waves from the evanescent waves is

$$k_y^2+\left(1-\frac{V_x^2}{c^2}\right)k_x^2+\frac{2V_x\omega k_x}{c^2}=\left(\frac{\omega}{c}\right)^2. \tag{41}$$

This function along with the radiation circle for a stationary medium is shown in Fig. 2. Note that with mean flow the function no longer is a circle. It crosses the k_x axis at $-\omega/(c-V_x)$ and $\omega/(c+V_x)$, and the k_y axis at $-\omega/c$ and ω/c. This agrees with the classic Doppler effect, since the phase speed upstream is reduced by the mean flow velocity, while downstream it is increased by the mean flow velocity. For low speeds, the new function may be approximated by a circle function by matching the points on the k_x axis. This approximation is

$$\left[k_x+\frac{\omega V_x}{(c^2-V_x^2)}\right]^2+k_y^2=\left[\frac{\omega c}{(c^2-V_x^2)}\right]^2. \tag{42}$$

The center of the circular function above is at

$$k_x=\frac{-\omega V_x}{c^2-V_x^2}, \quad k_y=0. \tag{43}$$

Because of the shift in the radiation circle, the wave number filter should not be centered at the origin as before, but at the wavenumber bin closest to Eq. 43 for $z_H\leqslant\lambda/3$. For the case when $z_H>\lambda/3$, the acoustic wave number, k, in Eq. (21) uses the form in Eq. (38) instead of Eq. (6). This shift in the radiation circle represents the wave number Doppler effect. Because the source and receiver are moving at the same velocity, there is no accompanying frequency Doppler effect.

The particle velocity in the x direction is affected by the mean flow according to the convected Euler's equation (Eq. 33).

$$\hat{u}_x=\frac{k_x\hat{p}}{\rho_o(\omega-k_x V_x)}. \tag{44}$$

The active and reactive intensities in the x direction will change accordingly. There are no changes in the calculations for the acoustic velocity and intensity components in the y- and z-directions.

The effect of a moving medium parallel to the hologram aperture will reduce the frequency range for NAH. The minimum and maximum frequencies (with mean flow in the positive x direction) are

$$f_{\min}=\frac{c+V_x}{2\ \min(l_x \text{ or } l_y)}, \tag{45}$$

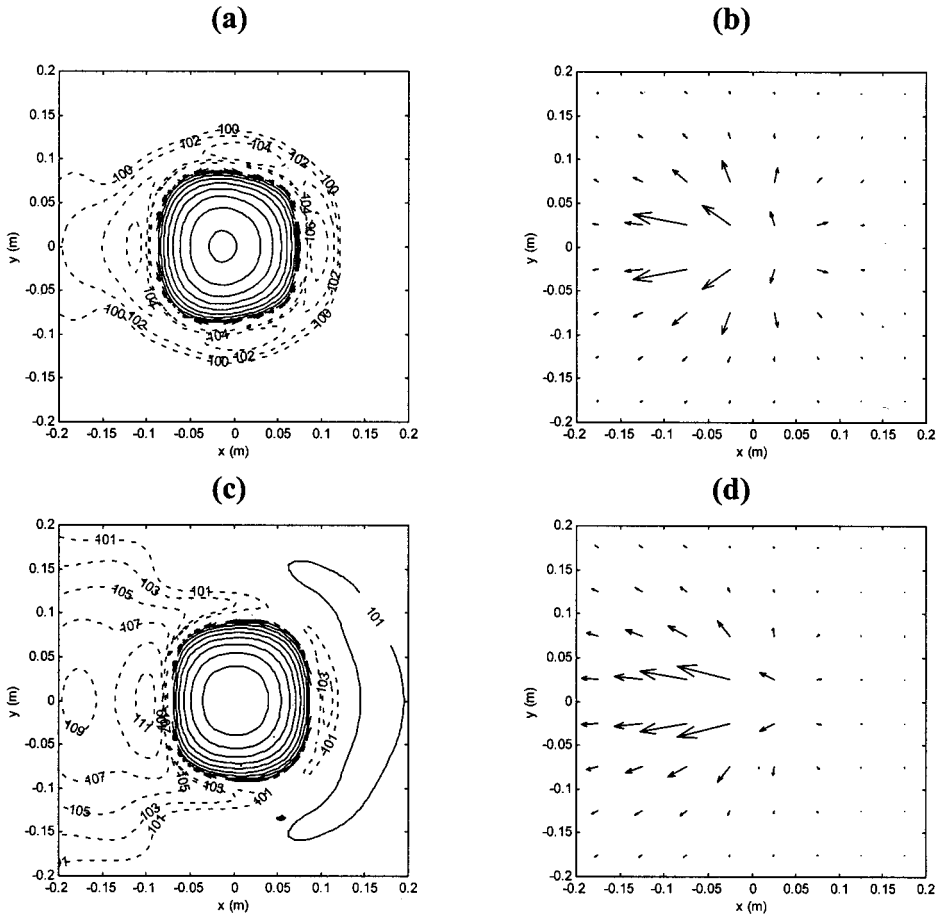


FIG. 7. Active intensity reconstructed on the source plane ($d=-0.05$ m) for monopole with mean velocity of 100 m/s ($M=0.3$). Contours of equal intensity in the normal direction using (a) flow-inclusive NAH procedure and (c) static NAH procedure. Contour lines are in 2-dB increments. Solid lines indicate positive intensity and dashed lines indicate negative intensity. 2D vectors for in-plane intensity components using (b) flow-inclusive NAH procedure and (d) static NAH procedure.

and

$$f_{\max}=\frac{c-V_x}{2\max(\Delta x \text{ or } \Delta y)}, \tag{46}$$

respectively.

B. Flow normal to the NAH plane

Not only does flow parallel to the NAH plane affect NAH calculations, but so will flow normal to the hologram plane. This affects the k_z term in the Green's function, which becomes

$$\hat{k}_z=\begin{cases}\left(\dfrac{c}{c+V_z}\right)\sqrt{\left(\dfrac{\omega}{c}\right)^2-k_x^2-k_y^2} & \text{for } k_x^2+k_y^2\leqslant\left(\dfrac{\omega}{c}\right)^2\\ -j\left(\dfrac{c}{c+V_z}\right)\sqrt{k_x^2+k_y^2-\left(\dfrac{\omega}{c}\right)^2} & \text{for } k_x^2+k_y^2>\left(\dfrac{\omega}{c}\right)^2,\end{cases} \tag{47}$$

and the particle velocity in the z direction as

$$\hat{u}_z=\frac{k_z\hat{p}}{\rho_o(\omega-k_zV_z)}. \tag{48}$$

For the case of a moving medium or flow normal to the NAH plane, there is no change in the radiation circle from the static or no-flow condition. Furthermore, there is no change in the particle velocity or intensity in the x- and y directions, nor is there any change in the frequency range from the static condition.

In Eqs. (47) and (48), it is assumed that the flow is directed in the positive z direction, thus away from the sound source, so that $\mathbf{V}=V_z\mathbf{e}_z$. This is the configuration for holographic measurements on the rear plane of a moving sound source. Hence, the moving medium has the effect of decreasing the wave number and increasing the propagating particle velocity and intensity in the downstream direction.

In the case where the mean flow velocity is directed in the negative z direction (towards the sound source), then $\mathbf{V}=-V_z\mathbf{e}_z$. Thus, V_z needs to be replaced with $-V_z$ in Eqs. (47) and (48). This is what occurs when measuring the front plane of a moving sound source. The effect of flow in this direction causes an increase in the wave number upstream of the moving source and a reduction in the particle velocity and intensity in the upstream direction.

V. EXAMPLES WITH A POINT SOURCE IN A MOVING MEDIUM

To verify the above theory and examine its impact, a moving point source is modeled. The convected Helmholtz equation [Eqs. (37) and (38)] is used to create an analytical solution for a point source in a moving medium. The solution for the pressure takes the form

$$\hat{p}=\frac{\hat{P}}{r}e^{j(\omega t-\hat{k}r)}, \tag{49}$$

where $\hat{P}$ is the complex amplitude and r is the radial distance from the point source. Equation (49) represents measured

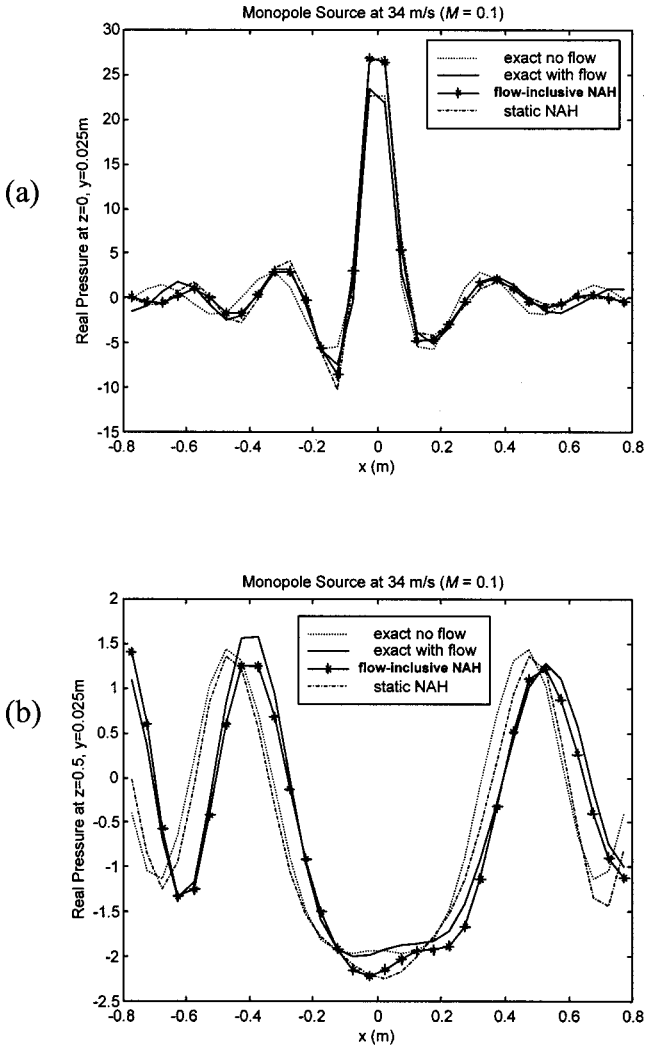


FIG. 8. Real pressure from monopole with mean velocity of 34 m/s ($M=0.1$). (a) Pressure along a line near the center from plane of pressure reconstructed in the near-field on the source plane ($z=0$). (b) Pressure along a line near the center of a plane in the far field ($z=0.5$ m).

sound field by an observer who is standing on the source coordinate. Solving for the complex wave number yields

$$\hat{k}=\frac{\omega r+jV_x\sin\theta}{(c+V_x\sin\theta)r}, \tag{50}$$

where θ is the angle from the z axis in the $x-z$ plane. Substituting Eq. (50) into Eq. (49), and separating the real and imaginary components yields

$$\hat{p}=\frac{\hat{P}}{r}e^{-k_I r}e^{j(\omega t-k_R r)}, \tag{51}$$

where

$$k_R=\frac{\omega}{c+V_x\sin\theta},\quad k_I=\frac{V_x\sin\theta}{(c+V_x\sin\theta)r}. \tag{52}$$

Note that the product $k_I r$ in Eq. (52) is actually independent of r.

The above equations are used as an exact solution to verify the flow-inclusive NAH procedures for application in a moving medium. Air is the medium with phase speed, c, of 343 m/s; and density, ρ_o, of 1.2 kg/m^3. The point source has amplitude of 1 Pa, and is radiating at a frequency of 1000 Hz. The hologram aperture consists of a 32- by 32-point array in the $x-y$ plane, and is 5 cm from the source ($z_H=0.05$ m). This distance may also be represented in nondimensional form as $(\omega/c)z_H=0.92$, or as 0.15 of a wavelength. The spacing between points in the hologram aperture is also 5 cm ($\Delta x=\Delta y=0.05$ m). The mean velocity is flowing parallel to the hologram plane in the positive x direction. The pressure field is reconstructed on the source plane ($z=0$, $d=-0.05$ m) for near-field analysis, and on a plane 0.5 m away from the source ($d=0.45$ m), for far-field analysis (Fig. 3). The results from NAH procedures for a stationary medium (static NAH) and procedures developed in this paper for a moving medium (flow-inclusive NAH) are both use the complex pressure on the hologram aperture from the exact solution with flow. Additionally, the exact solution without flow ($V_x=0$) is evaluated for comparison.

A. Error as a function of mean velocity

One means of quantifying the improvement in accuracy of NAH reconstruction using the flow-inclusive NAH verse the static NAH is to compute the spatial-averaged-mean-squared pressure

$$\overline{p^2_{\text{sp ave}}}=\frac{1}{n_x n_y}\sum_{n_i=1}^{n_x}\sum_{n_j=1}^{n_y}\overline{p^2[n_i,n_j]}, \tag{53}$$

The above term may be computed using the exact solution [Eqs. (51) and (52)], the static NAH method (Sec. II), and the flow-inclusive NAH method (Sec. IV). The error may then be computed

$$\%\ \text{error}=\frac{\overline{p^2_{\text{sp ave}}}(\text{exact})-\overline{p^2_{\text{sp ave}}}(\text{reconstructed})}{\overline{p^2_{\text{sp ave}}}(\text{exact})}. \tag{54}$$

Figure 4(a) shows that for near-field reconstruction using NAH, the flow-inclusive method was significantly more accurate than the static method. For high Mach number flows ($M>0.15$) the error when using the static NAH method was significant, as much as -612% for Mach number of 0.5. By comparison, using the flow-inclusive NAH theory presented in this paper, the error was less than 24% at a Mach flow of 0.5. For lower Mach number flows ($M<0.1$) the differences were much less but still apparent. At $M=0.1$, the error using the flow-inclusive method was -4.5% while the static method yielded an error of -25%.

For far-field reconstruction, based on the error computation in Eq. (54), there is no apparent advantage in using the flow-inclusive NAH method as seen in Fig. 4(b). Both methods were quite accurate up to a Mach flow of 0.2, but then worsened with increasing flow speed, but no more than $\pm 60\%$ error at $M=0.5$.

B. Example of NAH with mean velocity of 100 m/s

For this example the mean flow velocity, V_x, is 100 m/s. This corresponds to a Mach number of 0.29. This speed is not realistic for automobiles, but is examined only to exaggerate the effects of a moving medium on NAH.

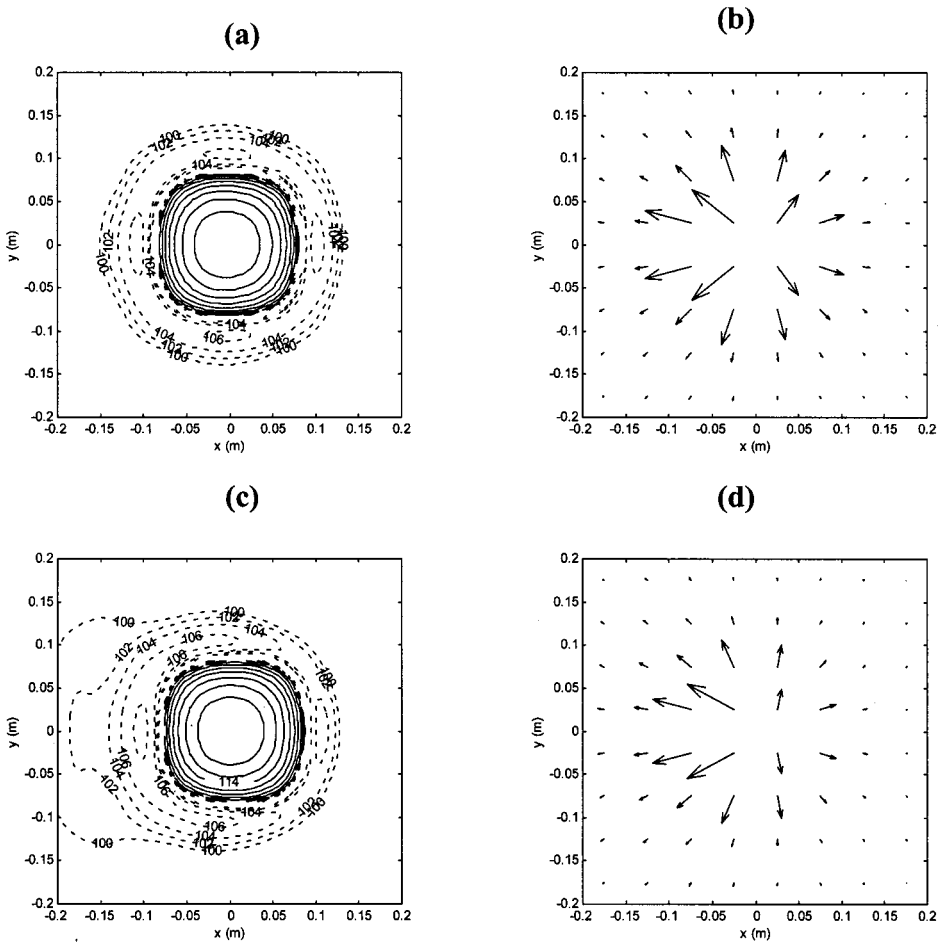


FIG. 9. Active intensity reconstructed on the source plane ($d=-0.05$ m) for monopole with mean velocity of 34 m/s ($M=0.1$). Contours of equal intensity in the normal direction using (a) flow-inclusive NAH procedure and (c) static NAH procedure. Contour lines are in 2-dB increments. Solid lines indicate positive intensity and dashed lines indicate negative intensity. 2D vectors for in-plane intensity components using (b) flow-inclusive NAH procedure and (d) static NAH procedure.

The acoustic pressure on the hologram aperture in real space and wave number space is shown in Fig. 5. Notice the shift in the wave number spectrum in the opposite direction of the flow (to the left). If one were to use the DFT forms that are common in many text books (Ref. 18, p. 516) and software programs such as MATLAB, along with the NAH theory presented here using the $e^{j\omega t}$ convention, then the 2D wave number spectrum would have been erroneously flipped along the k_y axis. To avoid this mistake, one must write a DFT subroutine that follows the kernel sign presented in Eqs. (20) and (22). However, the other DFT forms may be used if the sequence of the hologram array in real space is reversed just prior to the forward DFT (or FFT) and again after the inverse DFT necessary for reconstruction. This is because the forward transformation of the spatial dimension needs a DFT kernel of the opposite sign for the forward transformation and opposite sign for the inverse transformation. This has no effect on the real-space reconstruction when using the static NAH method though the k-space representation will still be reversed because the radiation circle is symmetrical. However, the correct from of DFT is necessary when using the flow-inclusive NAH due to the asymmetrical radiation region in k-space.

The real part of the complex pressure along a line in the reconstructed planes is shown in Fig. 6. Real pressure is utilized because it shows accuracy of both magnitude and phase. The actual effect of flow is seen in the exact solutions with and without flow. The flow amplifies and compresses the waves propagating upstream while reducing and expanding the waves propagating downstream. Using the static NAH procedures, the pressure magnitude is over amplified upstream in the nearfield reconstruction [Fig. 6(b)]. With far-field reconstruction, the static NAH procedure ignores the asymmetrical effect of flow almost completely. This is because the static NAH fails to identify a portion of the left-hand plan of the 2D wave number spectra as propagating waves (Fig. 2). Instead, these wave number components are wrongly classified as evanescent waves, where they decay in the far-field reconstruction (and intensify in near-field reconstruction). Contrariwise, the flow-inclusive NAH procedure reconstructs the acoustic pressure in the near-field and far field very accurately.

Results of active acoustic intensity data reconstructed on the source plane using static and flow-inclusive NAH procedures are shown in Fig. 7. Magnitudes of the intensity in the out-of-plane direction (z direction) are similar. The flow-inclusive NAH procedure more correctly locates the peak intensity at the point on the side of the source facing the direction of the flow while the static procedures falsely locates the peak intensity on the side of the source opposite the direction of the flow. In addition, the static NAH procedure exaggerates the large negative intensity region in front of the source. Furthermore, the correct location of the noise source at the origin is more accurately identified with the flow-inclusive NAH procedures when visualizing the vectors of the active intensity in the in-plane directions ($x-y$ plane).

The intensity vectors from the flow-inclusive method identify the sound source approximately 1 cm from the true source, while the static NAH method is approximately 4 cm off. The sound power through this reconstructed plane is 96.9 dB (*re*: 10^{-12} W) using the static NAH procedure and 98.8 dB using the flow-inclusive NAH procedure. The static NAH procedure erroneously groups some of the propagating energy as evanescent waves, thereby underestimating the sound power.

C. Example of NAH with mean velocity of 34 m/s

The same analysis is repeated with a mean velocity of 34 m/s. This corresponds to a Mach number of 0.1. This speed is more realistic of highway vehicles. The results of real pressure and active intensity reconstruction are shown in Fig. 8 and Fig. 9, respectively. Findings are similar to the previous example except that the flow at this velocity has a smaller effect on the acoustic field. Both NAH methods perform reasonably well in identifying the sound source and visualizing the acoustic field.

VI. CONCLUSIONS

Planar near-field acoustical holography (NAH) has been extended to include the effects of a moving medium. This theory may be applied to a moving sound source as long as the receivers (microphone array) are moving at the same constant speed. A new Green's function and modified wave number filter were developed based on the convected wave equation. The theory was verified using a simulation of a point source in a moving medium. The modified NAH algorithm proved more precise than existing NAH algorithms for both forward and backward propagation. This was most apparent for Mach flows greater than 0.15. One analysis at $M=0.5$ yielded reconstruction error, in terms of the spatial-averaged mean-squared pressure, of over 600% for the static NAH method compared to less than 25% error for the flow-inclusive NAH method. The flow-inclusive method was also much better at identifying the correct location of the sound source. The flow-inclusive method for far-field pressure reconstruction more accurately represented the correct phase of the acoustic pressure. The effect of the moving medium for low Mach number flows ($M<0.1$) was noticeable but not significant for acoustic propagation over short distances. Another outcome of the new procedure is the calculation of a valid frequency range for NAH in a moving medium. Furthermore, an improved understanding of the wave number Doppler effect is accomplished in this study.

ACKNOWLEDGMENTS

This work was partially funded by the Eisenhower Transportation Graduate Fellowship administered through the U.S. Department of Transportation. Other funding was provided by The Goodyear Tire & Rubber Company. Dr. Courtney B. Burroughs, Richard Ruhala's doctoral thesis advisor at The Pennsylvania State University, is recognized for his oversight of this work.

[1] E. G. Williams, J. D. Maynard, and E. Skudrzyk, "Sound source reconstruction using a microphone array," J. Acoust. Soc. Am. **68**, 340–344 (1980).
[2] J. D. Maynard, E. G. Williams, and Y. Lee, "Nearfield acoustic holography (NAH) I. Theory of generalized holography and the development of NAH," J. Acoust. Soc. Am. **78**, 1395–1413 (1985).
[3] W. A. Veronesi and J. D. Maynard, "Nearfield acoustic holography (NAH) II. Holographic reconstruction algorithms and computer implementation," J. Acoust. Soc. Am. **81**, 1307–1322 (1987).
[4] E. G. Williams and H. D. Dardy, "Generalized nearfield acoustical holography for cylindrical geometry: Theory and experiment," J. Acoust. Soc. Am. **81**, 389–407 (1987).
[5] C. Konzelman, "Near-field acoustical holography in cylindrical spaces," Ph.D. thesis, Graduate Program in Acoustics, The Pennsylvania State University, 1992.
[6] L. A. DeVries, J. S. Bolton, and J. C. Lee, "Acoustical holography in spherical coordinates for noise source identification," Proc. Noise-Con 92, pp. 935–940 (1992).
[7] W. A. Veronesi and J. D. Maynard, "Digital holographic reconstruction of sources with arbitrary shaped surfaces," J. Acoust. Soc. Am. **85**, 588–598 (1989).
[8] R. J. Ruhala, "A study of tire/pavement interaction noise using near-field acoustical holography," Ph.D. thesis, Graduate Program in Acoustics, The Pennsylvania State University, 1999.
[9] R. J. Ruhala and C. B. Burroughs, "Identification of sources of tire/pavement interaction noise," J. Acoust. Soc. Am. **103**(5-2), 2919(A) (1998).
[10] R. J. Ruhala and C. B. Burroughs, "Tire/pavement interaction noise source identification using multiplanar nearfield acoustical holography," Proc. 1999 SAE Noise and Vibration Conference, SAE paper no. 1999-01-1733.
[11] P. Rasmussen and S. Gade, "Tire noise measurement on a moving vehicle," Brüel & Kjær Application Note (1996).
[12] I. Sakamoto and T. Tanaka, "Interpretation of acoustic holography to measurement of noise on an operating vehicle," SAE Technical Paper Series, SAE paper no. 930199 (1993).
[13] H.-S. Kwon, "Sound visualization by using enhanced planar acoustic holographic reconstruction," Ph.D. thesis, Mechanical Engineering Department, Korea Advanced Institute of Science and Technology, 1997.
[14] H.-S. Kwon and Y.-H. Kim, "Moving frame technique for planar acoustic holography," J. Acoust. Soc. Am. **103**, 1734–1741 (1998).
[15] S.-H. Park and Y.-H. Kim, "An improved moving frame acoustic holography for coherent bandlimited noise," J. Acoust. Soc. Am. **104**, 3179–3189 (1998).
[16] T. H. M. Burns, "Measurement and visualization of instantaneous power flow in steady-state acoustic fields," Ph.D. thesis, Graduate Program in Acoustics, The Pennsylvania State University, 1995.
[17] A. D. Pierce, *Acoustics. An Introduction to its Physical Principles and Applications* (Acoustical Society of America, New York, 1989).
[18] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1989).
[19] H.-S. Kwon and Y.-H. Kim, "Minimization of bias error due to windows in planar acoustic holography using a minimum error window," J. Acoust. Soc. Am. **98**, 2104–2111 (1995).
[20] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968; first Princeton University Press edition, with errata page, Princeton, NJ, 1986).
[21] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York, 1982).

Resonance analysis of a 2D alluvial valley subjected to seismic waves

Juin-Fu Chai[a] and Tsung-Jen Teng
National Center for Research on Earthquake Engineering, Taipei, Taiwan

Chau-Shioung Yeh and Wen-Shinn Shyu
Department of Civil Engineering, National Taiwan University, Taipei, Taiwan

(Received 6 April 2001; revised 18 March 2002; accepted 22 May 2002)

The T-matrix formalism and an ultrasonic experiment are developed to study the scattering of in-plane waves for an alluvial valley embedded in a two-dimensional half-space. The solution of the in-plane scattering problem can be determined by the T-matrix method, where the basis functions are defined by the singular solutions of Lamb's problems with surface loading in both horizontal and vertical directions. In the experiment, a thin steel plate with a semicircular aluminum plate attached on the edge is used to simulate the two-dimensional alluvial valley in the state of plane stress. Based on the spectra of displacement signals measured at the free edge of the scatterer, the resonance frequencies where the peaks appear can be identified. It can be shown that the nondimensional resonance frequency is one of the characteristic properties of the scattering system. Furthermore, it is noted that the nondimensional resonance frequencies measured experimentally are in good agreement with those calculated theoretically.
[DOI: 10.1121/1.1494809]

PACS numbers: 43.20.Gp, 43.20.Ks, 43.35.Yb, 43.40.Ph [DEC]

I. INTRODUCTION

During an earthquake, more damage will occur to the structures at the top of the alluvial valley than to the structures at the top of the base rocks outside the alluvium. Therefore, the analysis of seismic response caused by the basin effects is an essential step in the aseismic design of important facilities. The most important and, however, the most difficult is the analysis of the resonant effects of the alluvium. The thorough and detailed understanding of the resonant effects, such as the resonance frequencies and the modal shapes of the alluvial valley subjected to seismic waves, is valuable to the strong motion seismology and earthquake engineering. There have been many theoretical studies on the resonant characteristics of various valleys.

In the investigation of resonant phenomenon of alluvial valley, Trifunac (1971) did not recognize the resonant nature of the response when he analyzed the surface motion of a semicircular cylindrical alluvial valley under the action of an incident plane SH-wave, although his computed transfer functions suggested the existence of resonance. Sanchez-Sesma and Esquivel (1979) have adopted the boundary element method to study the response of the cylindrical alluvial valley. They have shown the spatial variation of ground displacements at one or two low frequencies, but have not provided transfer functions nor pointed out the possibility of resonant response of the alluvial valley. Bard and Bouchon (1985) have investigated the two-dimensional resonance of a shallow sediment-filled cylindrical valley subjected to incident *SH*-, *SV*-, and *P* waves. They have shown that the fundamental resonance frequency is dependent on the geometric shape of the alluvium, but seems little affected by the impedance ratio. Jiang and Kuribayashi (1988) have considered the three-dimensional resonance of axisymmetrical sediment-filled basin, and shown that the resonance mode can be separated into two types: one is the shear mode due to shear wave and the other is the body mode due to longitudinal wave. Mossessian and Dravinski (1990) have analyzed the three-dimensional resonance motion of shallow, semiellipsoidal alluvial basins subjected to incident *P*- and *SV* waves. They have shown that the magnitude of the fundamental resonance frequency determined by the three-dimensional model is higher than that determined by the two-dimensional model, and the resonance is sensitive to the aspect ratio of the ellipsoid. However, the aforementioned analyses are based on the spectral-search method; the resonance frequencies are determined by the peaks of the spectrum of the ground response due to incident plane waves. Therefore, this analytic method is dependent on the incident plane waves, and is limited to determine only the fundamental resonance frequencies but not the whole resonance frequencies and their overtones.

Based on the in-phase condition, Rial (1989) and Rial *et al.* (1991) have adopted the ray method to determine the resonance frequencies of three-dimensional sedimentary basin. However, their method is applicable only to some special geometric shapes and the results are satisfied only at the center of the alluvium. Zhou and Dravinski (1994) have applied the eigenvalue method to analyze the resonance of two-dimensional deep sedimentary valley. By using the in-direct boundary element method, the resonance frequencies due to in-plane and antiplane seismic waves have been determined. It is noted that the aforementioned two methods have adopted the assumption of rigid base rocks. Therefore, the determined frequencies are only approximate values, and are somewhat different from the correct ones. In order to investigate the resonant nature of the surface above a soft semi-

[a]Electronic mail: chai@ncree.gov.tw

circular cylindrical alluvial valley, Wirgin (1995) has obtained the analytic solution for a bedrock–basin system subject to plane *SH* waves and shown that the resonances are manifestation of the excitation of normal modes of the system. However, his approach is restricted to the eigenfunctions for certain geometry.

Überall and his co-workers (Überall, 1978 and Flax *et al.*, 1981) have developed the background subtraction method by the far-field scattering form functions to determine the resonance frequencies. However, their method is applicable only to the special scatterers with simple geometric shapes such as the circular cylinder and sphere for which the scattering form function for every partial wave can be decoupled from each other. Furthermore, a background phase subtraction method is developed by Teng *et al.* (1998) to determine the resonance frequencies of a two-dimensional alluvial valley embedded in a half-space for general cases.

On the other hand, the T-matrix formalism determines the scattered and refracted coefficients from the incident wave directly, and is thus an efficient computational technique to determine the total field in any arbitrary alluvium and the surrounding half-space. In order to extend the application of T-matrix formalism to have the ability to solve the resonance problems, Yeh *et al.* (1998) have endeavored to develop a numerical technique to diagonalize the T-matrix and further to obtain the decoupled eigen partial waves for the scalar *SH* waves. Based on the eigen partial waves, the resonance frequencies and the modal shapes of normal modes and their overtones can be determined directly without any regard to the incident waves. However, the T-matrix formalism was not applied successfully to solve the scattering problems for in-plane waves, because the appropriate basis functions were not yet well defined.

Compared with the theoretical study on the resonance of scattering problems, scarce experimental data for earthquake engineering are available to verify the numerical results. Based on the viewpoint of nondestructive testing, it is expected to estimate the geometrical shape and physical properties of the scattering obstacles by the analysis of the scattered waves from an ultrasonic experiment. In general, a wideband ultrasonic pulse is exerted on the obstacle to generate the scattered waves, and the propagation properties of the scattered waves can be analyzed either from the measured transient signals in the time domain, or from the ultrasonic spectroscopy in the frequency domain (Gericke, 1970). Based on the ray theory, the wavefront and hence the arrival time can be identified from the time history of the transient waves to determine the velocity associated with a specified wave propagation type. However, for the complicated scattering problem, all of the modes of scattered waves are coupled together and one can hardly distinguish the wavefront and the associated wave propagation type to identify the properties of the scatterer. On the other hand, regular and distinguishable peak signals can be observed from the Fourier power spectrum of the transient waves in the frequency domain (Pao and Sachse, 1974); these peak signals correspond to the specified resonance modes of the scatterer and are dependent on the properties of the system.

In this paper, the T-matrix formalism is developed to determine the scattering of in-plane elastic waves in a two-dimensional half-space, and the basis functions are defined by the singular solutions of Lamb's problems with surface line loading in both horizontal and vertical directions. Furthermore, an ultrasonic experiment system is set up to measure the resonance frequencies of a two-dimensional alluvial valley. For convenience, a thin steel plate with a semicircular aluminum plate attached on the edge is used to simulate the alluvium in the state of plane stress. In the experiment, the transient waves are generated by a free-falling steel ball and then detected by the NBS conical transducer, and both the point source and point receiver are located on the free edge of the specimen. On the basis of the peaks in the Fourier power spectra of the measured transient signals, the resonance frequencies can be identified directly in the frequency domain. The equipment apparatus, experimental procedures, signal process methods, and the results are discussed in detail in this paper.

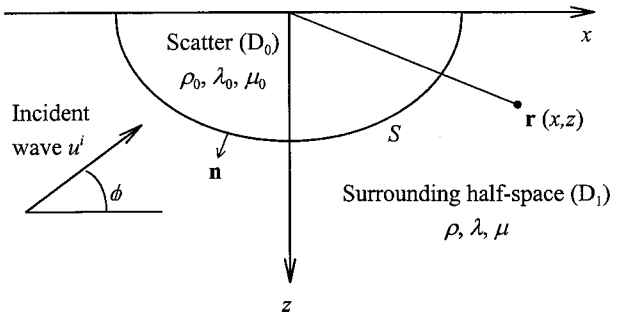


FIG. 1. The two-dimensional Cartesian coordinates and the model of an alluvial valley embedded in the surrounding half-space.

II. T-MATRIX FORMALISM

In this paper, the transition matrix (T-matrix) is developed to determine the scattering of in-plane elastic waves under the assumption of plane strain situation. Figure 1 shows the two-dimensional model of an alluvial valley beneath the free surface of a half-space, and the *x*- and *z* axes of the two-dimensional Cartesian coordinates are also defined by Fig. 1. It is noted that both the alluvial valley and the surrounding half-space are assumed to be elastic, isotropic, and homogeneous.

A time factor $\exp(i\omega t)$ is assumed for the in-plane waves with ω as the specified circular frequency. Considering only the time-independent parts, the *x*- and *z* components of the displacement vector $\mathbf{u}$ can be expressed by the potentials ϕ and ψ as

$$\mathbf{u}=\begin{Bmatrix} \partial\phi/\partial x-\partial\psi/\partial z \\ \partial\phi/\partial z+\partial\psi/\partial x \end{Bmatrix}. \tag{1}$$

The potentials ϕ and ψ will satisfy the Helmholtz equations corresponding to the longitudinal and shear wave numbers k_p and k_s, respectively, as written by

$$\nabla^2\phi+k_p^2\phi=0, \quad \nabla^2\psi+k_s^2\psi=0, \tag{2}$$

where ∇^2 is the Laplacian operator. Based on the specified circular frequency, the longitudinal and shear wave numbers k_p and k_s are expressed as

$$k_p=\frac{\omega}{C_p};\quad k_s=\frac{\omega}{C_s}. \tag{3}$$

Herein, C_p and C_s are the velocities of longitudinal and shear waves, respectively, and they are defined by

$$C_p=\sqrt{\frac{\lambda+2\mu}{\rho}};\quad C_s=\sqrt{\frac{\mu}{\rho}}, \tag{4}$$

where λ and μ are the Lamè constants, and ρ is the mass density of the elastic solid.

The key point in applying the T-matrix formalism is to define the complete set of basis functions that can satisfy both the wave equations and the free-surface boundary condition simultaneously, and hence the wave field can be expanded into a series of the specified basis functions. In this paper, for the in-plane elastic waves in a two-dimensional half-space, the basis functions are defined by the singular solutions of Lamb's problems with surface line loading in both horizontal and vertical directions.

Consider a two-dimensional half-space subjected to a surface line loading in the horizontal direction, and the loading function is defined by the *m*th derivative of the Dirac delta function $\delta(x)$ with respect to the spatial coordinate *x*. The associated boundary conditions at $z=0$ can be expressed as

$$\sigma_{zx}(x,z=0)=-\mu\frac{\partial^m\delta(x)}{\partial x^m};\quad \sigma_{zz}(x,z=0)=0. \tag{5}$$

On the other hand, the boundary conditions for the case of a surface loading in vertical direction can be expressed as

$$\sigma_{zx}(x,z=0)=0;\quad \sigma_{zz}(x,z=0)=-\mu\frac{\partial^m\delta(x)}{\partial x^m}. \tag{6}$$

Based on the Fourier transformation and its inverse between the spatial coordinate *x* and the horizontal wave number *k*, the general solutions of potentials ϕ and ψ in Eq. (2) can be determined; hence, the *x*- and *z* components of displacement vector $\mathbf{u}$ can be obtained from Eq. (1) as

$$\mathbf{u}=\begin{Bmatrix}\frac{1}{2\pi}\int_{-\infty}^{\infty}(ikBe^{-\upsilon z}+\upsilon' De^{-\upsilon' z})e^{-ikx}dk\\ \frac{1}{2\pi}\int_{-\infty}^{\infty}(\upsilon Be^{-\upsilon z}-ikDe^{-\upsilon' z})e^{-ikx}dk\end{Bmatrix}, \tag{7}$$

where υ and υ' are defined by

$$\upsilon=\sqrt{k^2-k_p^2};\quad \upsilon'=\sqrt{k^2-k_s^2}. \tag{8}$$

Therefore, the *m*th-order singular solution of the specified Lamb's problem can be defined by

$$\mathbf{u}_m^\alpha=\begin{Bmatrix}\frac{1}{2\pi}\int_{\Gamma_1}(ikB_m^\alpha e^{-\upsilon z}+\upsilon' D_m^\alpha e^{-\upsilon' z})e^{-ikx}dk\\ \frac{1}{2\pi}\int_{\Gamma_1}(\upsilon B_m^\alpha e^{-\upsilon z}-ikD_m^\alpha e^{-\upsilon' z})e^{-ikx}dk\end{Bmatrix}. \tag{9}$$

The superscript index $\alpha=h$ and υ, and *h* and υ denote the cases of horizontal loading ($\alpha=h$) and vertical loading ($\alpha=\upsilon$), respectively. Based on the boundary conditions as shown in Eq. (5) for the case of horizontal loading ($\alpha=h$), the associated coefficients B_m^h and D_m^h are determined as

$$B_m^h=(-ik)^m\frac{2ik\upsilon'}{F(k)};\quad D_m^h=(-ik)^m\frac{2k^2-k_s^2}{F(k)}. \tag{10}$$

On the other hand, the coefficients B_m^υ and D_m^υ can be determined from Eq. (6) for the case of vertical loading ($\alpha=\upsilon$) as

$$B_m^\upsilon=(-ik)^m\frac{2k^2-k_s^2}{F(k)};\quad D_m^\upsilon=-(-ik)^m\frac{2ik\upsilon}{F(k)}. \tag{11}$$

Herein, $F(k)$ denotes the Rayleigh function and can be expressed as

$$F(k)=(2k^2-k_s^2)^2-4k^2\upsilon\upsilon'. \tag{12}$$

The integral path Γ_1 in Eq. (9) is defined by Fig. 2(a) in the wave number domain (*k* domain), where k_R is the Rayleigh pole which will cause the Rayleigh function to be zero, i.e., $F(k_R)=0$. Furthermore, another set of the regular solutions $\hat{\mathbf{u}}_m^\alpha$ can be defined as

$$\hat{\mathbf{u}}_m^\alpha=\begin{Bmatrix}\frac{1}{2\pi}\int_{\Gamma_2}(ikB_m^\alpha e^{-\upsilon z}+\upsilon' D_m^\alpha e^{-\upsilon' z})e^{-ikx}dk\\ \frac{1}{2\pi}\int_{\Gamma_2}(\upsilon B_m^\alpha e^{-\upsilon z}-ikD_m^\alpha e^{-\upsilon' z})e^{-ikx}dk\end{Bmatrix}. \tag{13}$$

The coefficients B_m^α and D_m^α are defined by Eqs. (10) and (11) as the same for the case of singular solution $\mathbf{u}_m^\alpha$. However, the different integral path Γ_2 for regular solution $\hat{\mathbf{u}}_m^\alpha$ is defined as shown in Fig. 2(b). Based on the specified integral paths Γ_1 and Γ_2, the regular solution $\hat{\mathbf{u}}_m^\alpha$ can also be determined from the singular one as

$$\hat{\mathbf{u}}_m^\alpha=\mathbf{u}_m^\alpha-\overline{\mathbf{u}}_m^\alpha, \tag{14}$$

where $\overline{\mathbf{u}}_m^\alpha$ denotes the complex conjugate of $\mathbf{u}_m^\alpha$.

The singular solutions $\mathbf{u}_m^\alpha$ and the regular parts $\hat{\mathbf{u}}_m^\alpha$ with all of the non-negative integers *m* ($m=0,1,2,...,\infty$) are defined as the complete sets of basis functions for in-plane elastic waves in a two-dimensional half-space. With the assumed time factor $\exp(i\omega t)$, $\mathbf{u}_m^\alpha$ represents an outward-

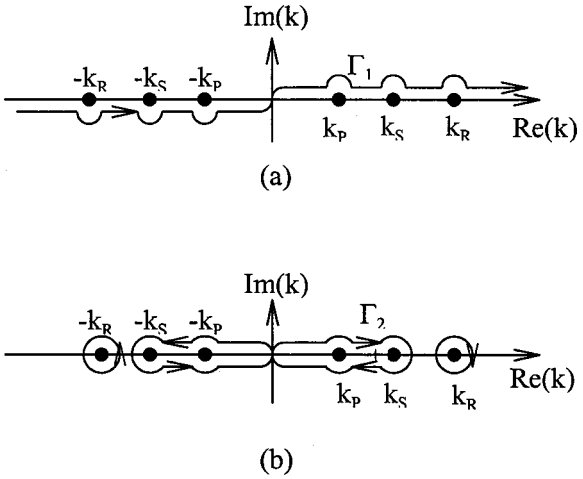


FIG. 2. (a) Integral path Γ_1 for the basis function $\mathbf{u}_m^\alpha$, and (b) integral path Γ_2 for the regular basis function $\hat{\mathbf{u}}_m^\alpha$.

propagating wave and $\hat{\mathbf{u}}_m^\alpha$ represents a standing wave. Therefore, the outgoing waves and the standing types can be expanded into a series of the basis functions $\mathbf{u}_m^\alpha$ and the regular part $\hat{\mathbf{u}}_m^\alpha$, respectively.

Consider the model as shown in Fig. 1; the wave field in the surrounding half-space (D_1) is called the exterior field, which can be separated into two parts: the free field and the scattered wave field. Far from the alluvial valley, the incident wave is reflected by the free surface of the half-space, and the reflected wave will interfere with the incident one to form the free field $\mathbf{u}^f$. It is noted that the free field is a standing wave and hence can be taken as a series of the regular basis functions $\hat{\mathbf{u}}_m^\alpha$. The scattered wave is propagating outward from the alluvium, and hence the corresponding displacement vector can be expanded into a series of the basis functions $\mathbf{u}_m^\alpha$. It is noted that the basis functions $\mathbf{u}_m^\alpha$ are regular in the surrounding half-space because the singular point is excluded outside the alluvium. Therefore, the displacement vector of the exterior field $\mathbf{u}^e$ can be expressed as

$$\mathbf{u}^e(x,z)=\mathbf{u}^f+\sum_{\alpha,m} C_m^\alpha \mathbf{u}_m^\alpha; \quad (x,z)\in D_1. \tag{15}$$

On the other hand, the free field can be refracted through the interface into the alluvial valley (D_0) to form the so-called interior field. The refracted wave will be reflected by the surface of the alluvial valley, and the standing wave appears due to the interference. Therefore, the displacement vector of the interior field $\mathbf{u}^0$ can be expanded into a series of the regular basis functions $\hat{\mathbf{u}}_{m(0)}^\alpha$ as

$$\mathbf{u}^0(x,z)=\sum_{\alpha,m} f_m^\alpha \hat{\mathbf{u}}_{m(0)}^\alpha; \quad (x,z)\in D_0. \tag{16}$$

The regular basis functions in the interior region $\hat{\mathbf{u}}_{m(0)}^\alpha$ can be determined from Eq. (13) by the Lamè constants λ_0 and μ_0 as well as the mass density ρ_0 of the alluvial valley. The coefficients C_m^α and f_m^α are called the scattered and refracted coefficients, respectively. Based on the conditions of continuity at the interface of the alluvium and the surrounding half-space, the scattered and refracted waves can be related to the free field by the T-matrix formalism.

The orthogonality conditions for the basis functions $\mathbf{u}_m^\alpha$ and $\hat{\mathbf{u}}_m^\alpha$ in the exterior region over the interface (denoted as S) of the alluvium and the surrounding half-space can be written as

$$\begin{aligned}
&\int_S (\mathbf{t}_m^\alpha\cdot\mathbf{u}_n^\beta-\mathbf{t}_n^\beta\cdot\mathbf{u}_m^\alpha)dS=0,\\
&\int_S (\hat{\mathbf{t}}_m^\alpha\cdot\hat{\mathbf{u}}_n^\beta-\hat{\mathbf{t}}_n^\beta\cdot\hat{\mathbf{u}}_m^\alpha)dS=0,\\
&\int_S (\mathbf{t}_m^\alpha\cdot\hat{\mathbf{u}}_n^\beta-\hat{\mathbf{t}}_n^\beta\cdot\mathbf{u}_m^\alpha)dS=E_{nm}^{\beta\alpha},
\end{aligned} \tag{17}$$

where $\mathbf{t}_m^\alpha$ and $\hat{\mathbf{t}}_m^\alpha$ denote the traction vectors on the interface S caused by the basis functions $\mathbf{u}_m^\alpha$ and $\hat{\mathbf{u}}_m^\alpha$, respectively. The proof of the orthogonality conditions is based on the Betti's third identity (Pao, 1978) and the asymptotic representation of the basis functions as $r=\sqrt{x^2+z^2}$ approaches infinity (Liao, 1997).

Let $\mathbf{u}^+$ and $\mathbf{t}^+$ denote the total displacement and its associated traction vectors at the interface S, respectively, approached from the surrounding half-space. Therefore, based on the representation of the exterior field and the orthogonality conditions for the basis functions as shown in Eqs. (15) and (17), we have

$$\int_S (\mathbf{t}^+\cdot\mathbf{u}_n^\beta-\mathbf{t}_n^\beta\cdot\mathbf{u}^+)dS=\int_S (\mathbf{t}^f\cdot\mathbf{u}_n^\beta-\mathbf{t}_n^\beta\cdot\mathbf{u}^f)dS=A_n^\beta, \tag{18}$$

and

$$\int_S (\mathbf{t}^+\cdot\hat{\mathbf{u}}_n^\beta-\hat{\mathbf{t}}_n^\beta\cdot\mathbf{u}^+)dS=\sum_{m,\alpha} E_{nm}^{\beta\alpha}C_m^\alpha, \tag{19}$$

It can be found from Eq. (18) that the coefficient A_n^β is determined from the free field without any regard to the scattered waves, and hence can be recognized as the given parameters for a specified incident wave. The detailed determination of the coefficients A_n^β corresponding to different types of incident waves is described in the Appendix.

On the other hand, let $\mathbf{u}^-$ and $\mathbf{t}^-$ denote the total displacement and its associated traction at the interface S, respectively, approached from the alluvium. Enforcing the conditions of continuity at the interface of the alluvium and surrounding half-space, $\mathbf{u}^+=\mathbf{u}^-$ and $\mathbf{t}^+=\mathbf{t}^-$, yields

$$\begin{aligned}
&\int_S (\mathbf{t}^+\cdot\mathbf{u}_n^\beta-\mathbf{t}_n^\beta\cdot\mathbf{u}^+)dS=\int_S (\mathbf{t}^-\cdot\mathbf{u}_n^\beta-\mathbf{t}_n^\beta\cdot\mathbf{u}^-)dS,\\
&\int_S (\mathbf{t}^+\cdot\hat{\mathbf{u}}_n^\beta-\hat{\mathbf{t}}_n^\beta\cdot\mathbf{u}^+)dS=\int_S (\mathbf{t}^-\cdot\hat{\mathbf{u}}_n^\beta-\hat{\mathbf{t}}_n^\beta\cdot\mathbf{u}^-)dS.
\end{aligned} \tag{20}$$

Based on Eqs. (18), (19), and the representation of the interior field as shown in Eq. (16), Eq. (20) can be written as

$$\begin{aligned}
&\sum_{m,\alpha} f_m^\alpha Q_{nm}^{\beta\alpha}=A_n^\beta,\\
&\sum_{m,\alpha} f_m^\alpha \hat{Q}_{nm}^{\beta\alpha}=\sum_{m,\alpha} E_{nm}^{\beta\alpha}C_m^\alpha,
\end{aligned} \tag{21}$$

where $Q_{nm}^{\beta\alpha}$ and its regular part $\hat{Q}_{nm}^{\beta\alpha}$ are defined as

$$\begin{aligned}
&Q_{nm}^{\beta\alpha}=\int_S (\hat{\mathbf{t}}_{m(0)}^\alpha\cdot\mathbf{u}_n^\beta-\mathbf{t}_n^\beta\cdot\hat{\mathbf{u}}_{m(0)}^\alpha)dS,\\
&\hat{Q}_{nm}^{\beta\alpha}=\int_S (\hat{\mathbf{t}}_{m(0)}^\alpha\cdot\hat{\mathbf{u}}_n^\beta-\hat{\mathbf{t}}_n^\beta\cdot\hat{\mathbf{u}}_{m(0)}^\alpha)dS.
\end{aligned} \tag{22}$$

Furthermore, Eq. (21) can be rewritten in the matrix form as

$$\mathbf{Qf}=\mathbf{A}, \quad \hat{\mathbf{Q}}\mathbf{f}=\mathbf{EC}. \tag{23}$$

The elements within the matrices and coefficient vectors are in order of the indices of $m=0,1,2...$, and $\alpha=h$ and v. It can be illustrated by the typical ones as

$$\mathbf{Q}=\begin{bmatrix} Q_{00}^{hh} & Q_{00}^{hv} & Q_{01}^{hh} & Q_{01}^{hv} & \cdots \\ Q_{00}^{vh} & Q_{00}^{vv} & Q_{01}^{vh} & Q_{01}^{vv} & \cdots \\ Q_{10}^{hh} & Q_{10}^{hv} & Q_{11}^{hh} & Q_{11}^{hv} & \cdots \\ Q_{10}^{vh} & Q_{10}^{vv} & Q_{11}^{vh} & Q_{11}^{vv} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}; \quad \mathbf{C}=\begin{Bmatrix} C_0^h \\ C_0^v \\ C_1^h \\ C_1^v \\ \vdots \end{Bmatrix}. \tag{24}$$

The matrix $\hat{\mathbf{Q}}$ is the regular part of $\mathbf{Q}$. Based on the expression of $\hat{\mathbf{u}}_m^{\alpha}$ in Eq. (14), the regular matrix $\hat{\mathbf{Q}}$ can be expressed by $\mathbf{Q}$ as

$$\hat{\mathbf{Q}}=\mathbf{Q}+\overline{\mathbf{Q}}. \tag{25}$$

Elimination of the refracted coefficient vector $\mathbf{f}$ in Eq. (23) yields

$$\mathbf{C}=\mathbf{TA}, \tag{26}$$

where the matrix $\mathbf{T}$ is defined as

$$\mathbf{T}=\mathbf{E}^{-1}\hat{\mathbf{Q}}\mathbf{Q}^{-1}, \tag{27}$$

and is called the transition matrix (T-matrix). It can be found from Eqs. (23) and (26) that the refracted and scattered coefficient vectors $\mathbf{f}$ and $\mathbf{C}$ can be determined straightforwardly from the vector $\mathbf{A}$, which is defined by the free field corresponding to the specified incident wave.

Finally, as long as the coefficient vectors $\mathbf{f}$ and $\mathbf{C}$ are determined, the exterior field $\mathbf{u}^e$ and interior field $\mathbf{u}^0$ can be determined by the matrix form as

$$\begin{aligned} &\mathbf{u}^e(x,z)=\mathbf{u}^f+\mathbf{UC}; \quad (x,z)\in D_1 \\ &\mathbf{u}^0(x,z)=\hat{\mathbf{U}}_{(0)}\mathbf{f}; \quad (x,z)\in D_0. \end{aligned} \tag{28}$$

Herein, the matrices $\mathbf{U}$ and $\hat{\mathbf{U}}_{(0)}$ are defined by the basis functions in the surrounding half-space (D_1) and the alluvium (D_0), respectively, as

$$\begin{aligned} &\mathbf{U}=[\mathbf{u}_0^h \quad \mathbf{u}_0^v \quad \mathbf{u}_1^h \quad \mathbf{u}_1^v \quad \cdots], \\ &\hat{\mathbf{U}}_{(0)}=[\hat{\mathbf{u}}_{0(0)}^h \quad \hat{\mathbf{u}}_{0(0)}^v \quad \hat{\mathbf{u}}_{1(0)}^h \quad \hat{\mathbf{u}}_{1(0)}^v \quad \cdots]. \end{aligned} \tag{29}$$

It is noted that the validity and accuracy of the T-matrix method can be verified by comparing the numerical results with the analytic solutions and the results determined by other numerical methods (Yeh *et al.*, 2002). Moreover, the complete sets of basis functions for elastic waves in a three-dimensional half-space can be defined similarly by the singular solutions of Lamb's problem; hence, the T-matrix method can be extended trivially to three-dimensional scattering problems.

III. EXPERIMENTAL APPARATUS

The purpose of this experiment is to measure the scattered transient waves and then to study the resonant effect of an alluvial valley. The experimental apparatus is shown in Fig. 3.

In order to simulate the alluvial valley embedded in a half-space, a semicircular aluminum plate is pasted with epoxy upon the concavity at the edge of a steel plate. It is noted that the thickness of the steel plate is only 0.2 cm, which is much less than the length (90 cm) and width (60 cm) of the plate. Therefore, the thin steel plate with an aluminum inclusion on the edge can be used to model a two-dimensional half-space in the state of plane stress.

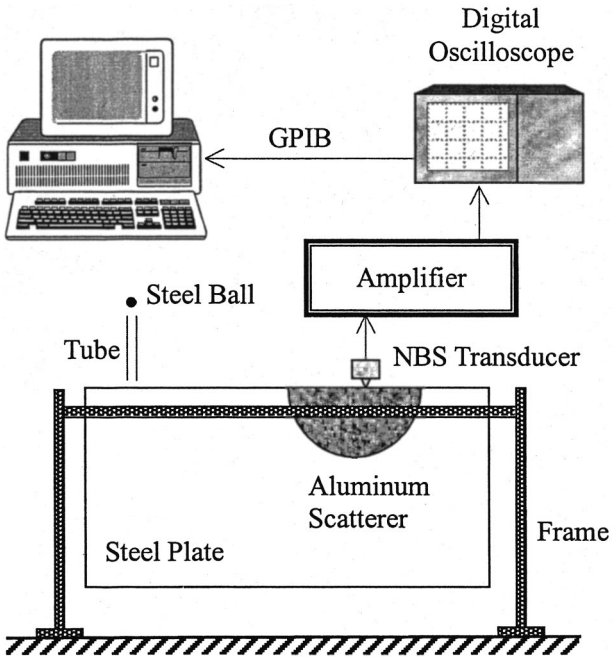


FIG. 3. Experimental apparatus to study the scattered waves.

The Lamè constants λ^* and μ^* under the assumption of plane stress are defined by λ and μ as

$$\lambda^*=\frac{2\lambda\mu}{\lambda+2\mu}; \quad \mu^*=\mu. \tag{30}$$

Therefore, the longitudinal and shear wave velocities C_p^* and C_s^* of a solid in the state of plane stress can be expressed as

$$C_p^{*2}=\frac{4C_s^2}{C_p^2}(C_p^2-C_s^2); \quad C_s^{*2}=C_s^2, \tag{31}$$

where C_P and C_S are the longitudinal and shear wave velocities of a solid in the state of plane strain, respectively, and can be easily measured by the ultrasonic pulse–echo method from the measured traveling time and the thickness of the specimen. Furthermore, the associated Rayleigh wave velocity C_R^* in the state of plane stress can be determined by the Rayleigh equation

$$\left(2-\frac{C_R^{*2}}{C_s^{*2}}\right)^2=4\sqrt{\left(1-\frac{C_R^{*2}}{C_p^{*2}}\right)\left(1-\frac{C_R^{*2}}{C_s^{*2}}\right)}. \tag{32}$$

Based on the pulse–echo experiment and Eqs. (31) and (32), the wave velocities of the steel and aluminum specimens in the state of plane stress can be determined; they are listed in Table I. The densities of steel and aluminum are 7800 and 2700 kg/m^3, respectively. Therefore, the impedance ratio of the aluminum scatterer relative to the surrounding steel half-space is less than 1.0; hence, the obstacle is recognized as a soft scatterer although the wave velocities of the scatterer are larger than the corresponding ones of the surrounding half-space. It is noted that two semicircular aluminum scatterers with radiuses of $a=3$ cm and $a=4$ cm are prepared in this experiment to study the effect of size of the alluvial valley.

TABLE I. The wave velocities in the state of plane stress and densities of steel and aluminum plates.

	Steel plate	Aluminum plate
Longitudinal wave	$C^*_{p,\text{steel}}=5270.2$ m/s	$C^*_{p,\text{Al}}=5561.6$ m/s
Shear wave	$C^*_{s,\text{steel}}=3131.1$ m/s	$C^*_{s,\text{Al}}=3179.6$ m/s
Rayleigh wave	$C^*_{R,\text{steel}}=2866.9$ m/s	$C^*_{R,\text{Al}}=2927.1$ m/s
Density	$\rho_{\text{steel}}=7800$ kg/m^3	$\rho_{\text{Al}}=2700$ kg/m^3

In this experiment, the point source to generate the transient waves is caused by a free-falling steel ball that impacts upon the edge of the surrounding steel half-space. The frequency bandwidth of the induced transient waves is dependent on the size of the steel ball, and the bigger steel balls result in the lower frequency content. According to the Hertz contact law, the source time function can be modeled by $\sin^{3/2}(\pi t/T_c)$ with $0\leqslant t\leqslant T_c$, where T_c is the impact duration that is proportional to the radius of the free-falling ball (Goldsmith, 1960). In order to study the effect of the frequency content of the source time function, two sizes of steel balls with diameters of $D=2.3$ mm and $D=1.0$ mm are prepared in this experiment. They are released freely from a height of 15 cm, and the impact duration can be estimated to be 7.43 and 3.23 μs for the big and small ones, respectively. The source time functions and the associated Fourier power spectra for these two sizes of steel balls are compared in Fig. 4. It can be observed that the frequency content caused by the big steel ball with $D=2.3$ mm is within the range less than 200 kHz, while the frequency content is up to 500 kHz for the small one with $D=1.0$ mm. In this experiment, the free-falling ball is guided by a glass tube to control the impact point. In order to study the effect of the location of the loading point, the distance between the impact point and the center of the scatterer is specified as $L=4a$, $L=5a$, and $L=6a$, where a is the radius of the semicircular scatterer.

The transient signals are detected by the NBS conical transducers, which are broadband point receivers. It is noted that the analog output from NBS conical transducers is linearly proportional to the vertical displacement on the observation point. As shown in Fig. 5, five observation points P_1 to P_5 which are located on the free edge of the specimen are specified in the experiment. The points P_1 and P_5 with coordinates of $x/a=-2.0$ and $x/a=1.0$ are located on the free edge of the surrounding steel half-space. On the other hand, the points P_2, P_3, and P_4 with coordinates of $x/a=-0.8$, -0.5, and 0.0, respectively, are located on the free edge of the semicircular aluminum plate. It is noted that the origin of the coordinate is coincident with the center of the scatterer, and the specified observation locations are normalized by the radius of the semicircular scatterer.

On the basis of the digital oscilloscope, the waveforms detected by the NBS conical transducers can be displayed on the screen, and further, the analog signals can be converted to digital signals and are recorded by PC through the GPIB interface. However, because the half-space is simulated by a finite specimen, only the signals before the arrival of the reflected longitudinal wave from the nearest boundary are recorded and analyzed.

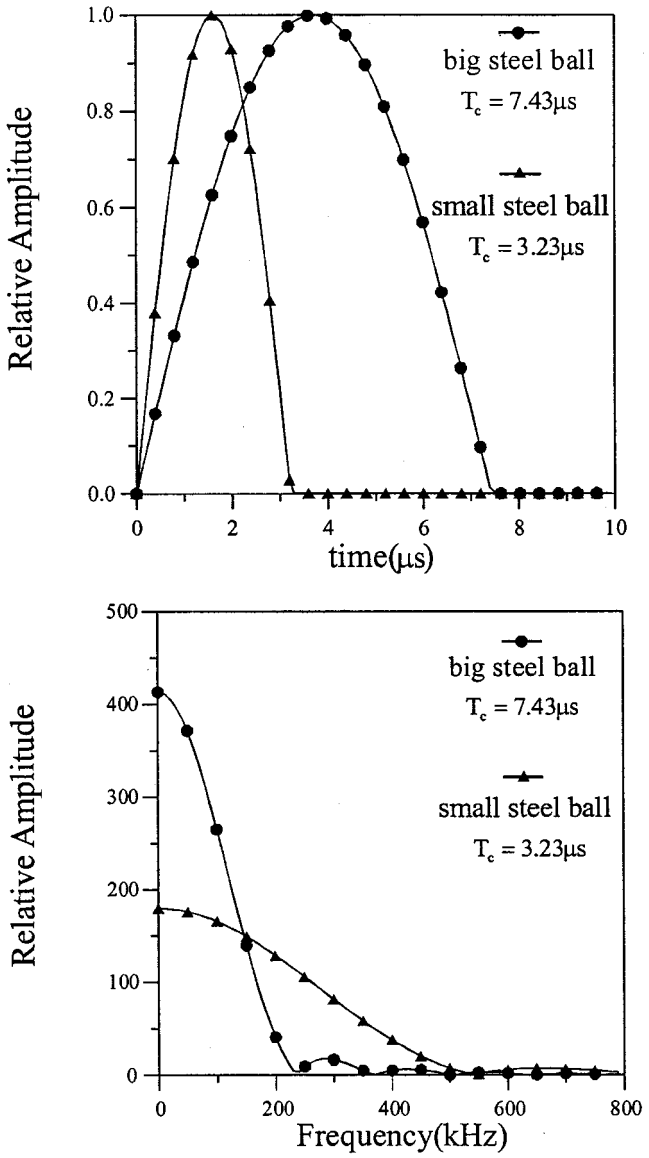


FIG. 4. Simulated source time functions and the associated Fourier power spectra.

IV. ANALYSIS OF THE MEASURED SIGNALS IN THE TIME DOMAIN

The initial time ($t=0$) is defined as the moment when the free-falling ball impacts upon the specimen to generate the transient waves. However, the digital oscilloscope is triggered when the amplitude of signal from the trigger channel is beyond a threshold, and then the signals from all channels are recorded simultaneously in the pretrigger mode. There-

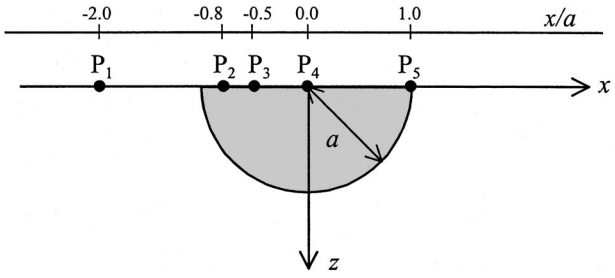


FIG. 5. Locations of the observation points.

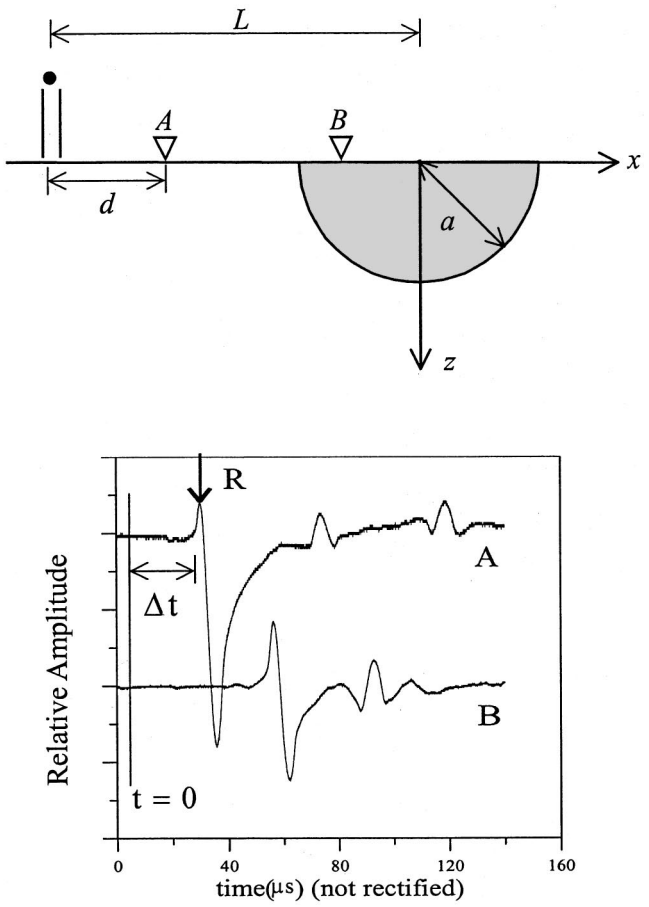


FIG. 6. Rectification of the initial time on the basis of Rayleigh wavefront.

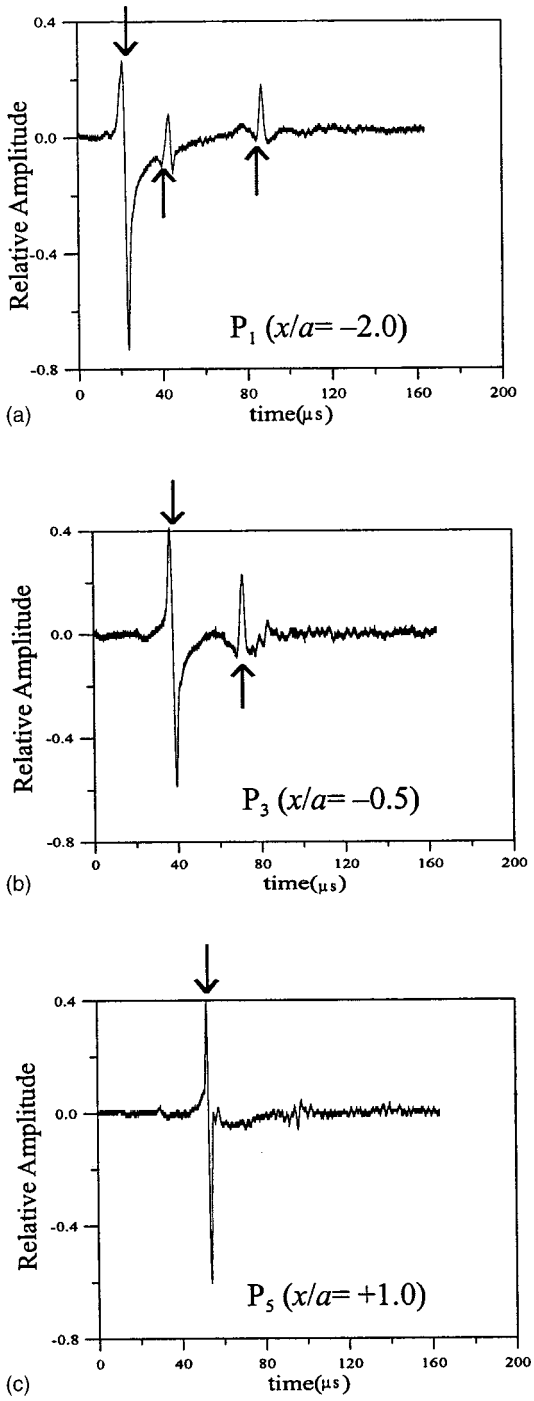


FIG. 7. Detected displacement signals at (a) P_1 ($x/a=-2.0$); (b) P_3 ($x/a=-0.5$); and (c) P_5 ($x/a=1.0$).

fore, the beginning time of the recorded signals is not the initial time as we defined. In order to identify the exact initial time, two NBS conical transducers are required. As shown in Fig. 6(a), transducer A is fixed on the free edge of the surrounding steel half-space, and the detected signal is specified to trigger the oscilloscope and then used to rectify the initial time. On the other hand, transducer B is located on each one of the observation points (P_1 to P_5) and used to record the transient displacement signals. The signals detected by transducer A are shown in Fig. 6(b); the sharp Rayleigh wavefront in the waveform of the Green's function for a half-space is observed clearly. This is because the detected signals are not yet affected by the scatterer before the arrival of the waves reflected from the front edge of the obstacle. The time interval between the Rayleigh wavefront and the initial time can be determined as $\Delta t = d/C^*_{R,\text{steel}}$, with d as the fixed distance between the impact point and the location of transducer A. Therefore, based on the clearly sharp Rayleigh wavefront detected by transducer A as well as the Rayleigh wave velocity of steel in the state of plane stress and the fixed traveling path, the initial time can be identified as shown in Fig. 6(b).

Consider the case of a semicircular aluminum scatterer with radius $a=3$ cm, where the transient waves are excited by a big steel ball (with diameter $D=2.3$ mm) impacting upon the surrounding steel half-space with $L=4a$. It is noted that all the detected signals have been rectified to begin at the moment that the transient waves are excited. Furthermore, the amplitude of every detected signal is processed by offsetting the initial value to zero and then scaling the amplitude to let the peak-to-peak value become 1.0.

Figure 7(a) shows the vertical displacement detected at the observation point P_1 ($x/a=-2.0$), which is located on the free edge of the surrounding half-space. Based on the ray theory, the arrival time of Rayleigh waves can be determined and compared with the detected transient signals. As the free-falling ball impacts the specimen, the incident Rayleigh wave propagates from the source toward the scatterer, it arrives at the observation point P_1 at $t_R = 2a/C^*_{R,\text{steel}} = 20.93\,\mu$s, and results in the first wavefront signal. The

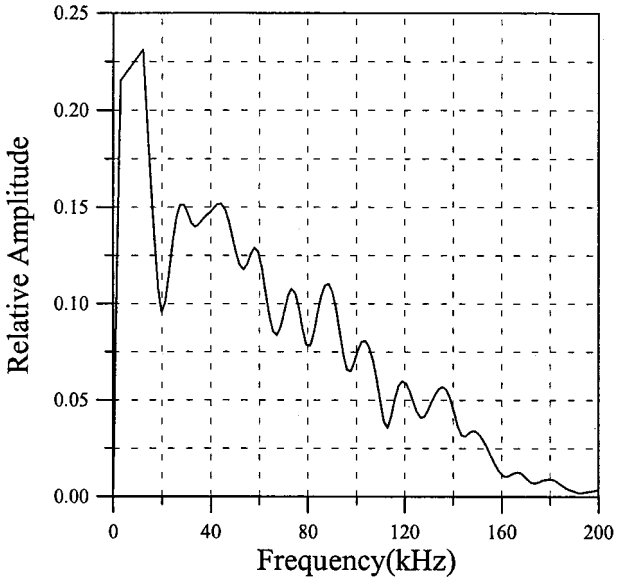


FIG. 8. Power spectrum of the signal detected at P_1 ($x/a=-2.0$).

incident Rayleigh wave will be reflected backward and refracted into the scatterer as it reaches the front edge of the obstacle. The reflected Rayleigh wave arrives at P_1 at $t_{R-R}=4a/C^*_{R,\text{steel}}=41.86\,\mu$s and results in the second wavefront signal. On the other hand, the refracted Rayleigh wave will reach the rear edge of the obstacle and be reflected backward and transmitted through the front edge toward P_1 at $t'_{R-R}=4a/C^*_{R,\text{steel}}+4a/C^*_{R,\text{Al}}=82.85\,\mu$s, hence resulting in the third wavefront signal.

Figure 7(b) shows the vertical displacement detected at the observation point P_3 ($x/a=-0.5$), which is located on the surface of the scatterer. Two Rayleigh wavefronts at 36.48 and 68.78 μs can be observed clearly, which correspond to the arrival times of the refracted Rayleigh wave and the one reflected back from the rear edge to P_3, respectively. It is noted that the arrival times can also be verified by ray theory as $t_R=3a/C^*_{R,\text{steel}}+0.5a/C^*_{R,\text{Al}}=36.51\,\mu$s and $t_{R-R}=3a/C^*_{R,\text{steel}}+3.5a/C^*_{R,\text{Al}}=67.26\,\mu$s. The displacement signals detected at the observation point P_5 ($x/a=1.0$) are shown in Fig. 7(c). Again, the arrival time of the refracted Rayleigh wave can be determined by ray theory as $t_R=3a/C^*_{R,\text{steel}}+2a/C^*_{R,\text{Al}}=53.87\,\mu$s, and at that time a clear Rayleigh wavefront appears. It is noted that the determined arrival times of Rayleigh waves based on the ray theory are in good agreement with the ones observed from the measured transient signals in the time domain.

V. ANALYSIS OF THE MEASURED SIGNALS IN THE FREQUENCY DOMAIN

The Fourier power spectrum of the signal detected at the observation point P_1 is shown in Fig. 8. It is found that the energy is within the range less than 200 kHz, and it is in accordance with the frequency content of the source time function corresponding to the impact of the big steel ball as shown in Fig. 4. Furthermore, the dips can be observed at some special frequencies caused by the resonance effect of the soft scatterer. In fact, the scatterer can be recognized as a filter and have the property of selective transmission or reflection (Pao and Mow, 1976). The total transmission occurs as the frequency of the incident wave coincides with the resonance frequency. Therefore, as the normal mode is excited, no energy is scattered backward because of the total transmission, and sharp dips will appear in the power spectra of the backward-scattered waves. On the other hand, the amplitude of the displacement within the scatterer should be amplified by the resonance effect and cause the peaks in the

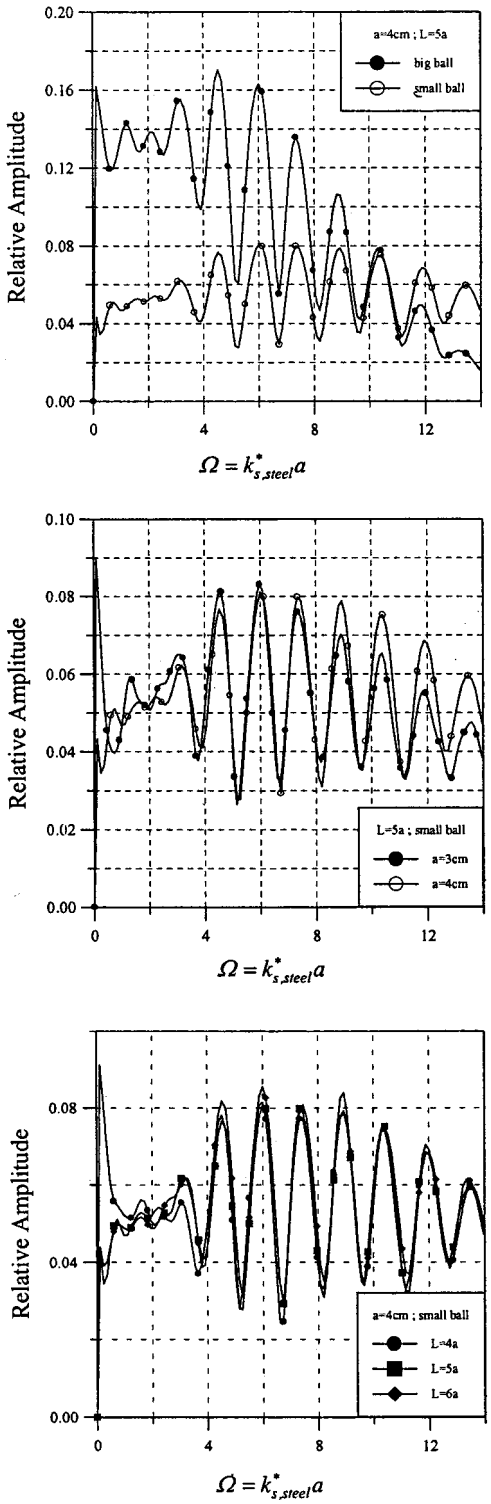


FIG. 9. Comparison of the power spectra of signals detected at P_2 ($x/a=-0.8$) under different experimental conditions.

TABLE II. The nondimensional resonance frequencies observed at P_2 and P_4.

			P_2 ($x/a=-0.8$)					P_4 ($x/a=0.0$)		
$a=$	$D=$	$L=$	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=1$	$n=2$	$n=4$
3 cm	2.3 mm	$4a$	3.12	4.50	5.79	7.17	8.63	2.94	4.68	7.16
		$5a$	3.12	4.50	5.97	7.35	8.73	2.94	4.68	7.26
		$6a$	3.12	4.50	5.97	7.35	8.82	2.94	4.68	7.16
	1.0 mm	$4a$	3.12	4.50	5.79	7.26	8.73	2.94	4.68	7.35
		$5a$	3.12	4.60	5.97	7.35	8.91	3.03	4.78	7.35
		$6a$	3.12	4.59	5.97	7.35	8.82	2.94	4.68	7.26
4 cm	2.3 mm	$4a$	3.06	4.53	5.88	7.35	8.82	2.94	4.90	7.35
		$5a$	3.18	4.53	6.00	7.35	8.82	3.06	4.90	7.47
		$6a$	3.18	4.53	6.00	7.47	8.82	3.06	4.90	7.47
	1.0 mm	$4a$	3.06	4.53	6.00	7.35	8.94	3.06	4.90	7.35
		$5a$	3.18	4.53	6.00	7.35	8.94	3.06	4.90	7.59
		$6a$	3.18	4.53	6.00	7.47	8.94	3.06	4.90	7.47

power spectrum at these resonance frequencies. In order to study the resonance frequencies systematically, the nondimensional frequency is defined by the radius of the semicircular scatterer as

$$\Omega = \frac{2\pi f a}{C^*_{s,\text{steel}}} = k^*_{s,\text{steel}} a, \tag{33}$$

where $k^*_{s,\text{steel}}$ is the shear wave number of the surrounding steel half-space in the state of plane stress.

On the basis of the random selection among the radius of the semicircular scatterer ($a=3$ and 4 cm), the size of the steel ball (big and small), and the location of impact point ($L=4a$, $5a$, and $6a$), there is a total of 12 different experimental conditions. Consider the observation point P_2 ($x/a=-0.8$); the Fourier power spectrum of the measured displacement versus the nondimensional frequency can be developed one by one under the 12 different experimental conditions. Based on the peaks in the developed power spectrum, the nondimensional resonance frequencies can be identified. Some typical examples to demonstrate the effects of the steel-ball size, the scatterer radius, and the source location on the power spectra are compared in Fig. 9. It can be found that the nondimensional resonance frequencies are almost identical to each other under different experimental conditions. Furthermore, more than one nondimensional resonance frequency can be identified, which are corresponding to the different normal modes and their overtones. All of the nondimensional resonance frequencies for these 12 experimental conditions are listed in Table II. Similarly, the Fourier power spectra of the transient signals detected at the observation point P_4 ($x/a=0.0$) are compared in Fig. 10, and the associated nondimensional resonance frequencies are listed in Table II.

From Table II, it can be found that all of the nondimensional resonance frequencies observed at P_4 can also be observed at P_2, but not vice versa. It is because that the observation point P_4 ($x/a=0.0$) is just the node of the antisymmetric modes and hence the contribution of the antisymmetric modes at P_4 are vanished. Therefore, the resonance frequencies corresponding to the antisymmetric modes or their overtones cannot be identified from the displacement signals detected at the observation point P_4. Based on the experimental results, it can be shown that the nondimensional resonance frequencies are characteristic constants for the scattering system; they are independent of the radius of the semicircular alluvium, the bandwidth of the source time function, and the location of the source point.

VI. NUMERICAL RESULTS

Because the point source and point receiver are both located on the free surface, the contribution of Rayleigh wave is much larger than that of longitudinal and shear waves. Therefore, based on the consideration of Rayleigh wave incidence, the associated power spectra of the scattered displacement on P_2 ($x/a=-0.8$) and P_4 ($x/a=0.0$) within the nondimensional frequency range of $1.0 \leqslant \Omega \leqslant 6.0$ are determined numerically by the T-matrix method. In order to determine the wave fields that can be compared with the experimental results, the Lamè constants λ and μ in the numerical calculation should be replaced by λ^* and μ^* to satisfy the assumption of plane stress. Figure 11 shows the results; it can be observed that the nondimensional resonance frequencies can be identified to be $\Omega=2.9$ and 4.8 for observation point P_4 ($x/a=0.0$), and $\Omega=3.1$, 4.6 and 5.9 for P_2 ($x/a=-0.8$), respectively. They are in good agreement with the experimental results as listed in Table II.

VII. CONCLUSION

The T-matrix formalism is developed in this paper to determine the scattering of in-plane elastic waves in a two-dimensional half-space, and the basis functions are defined by the singular solutions of Lamb's problems with surface line loading in both horizontal and vertical directions. An ultrasonic experiment is set up to study the resonance frequencies of scattered waves from a two-dimensional alluvial valley embedded in a half-space. In time domain, the identified wavefronts of the Rayleigh waves can be verified from the arrival times determined by ray theory. On the basis of the power spectra of the measured displacement signals, the resonance frequencies where the peaks appear can be identified. It can be shown from the experiment results that the nondimensional resonance frequency is one of the characteristic properties of the scattering system. They are constants

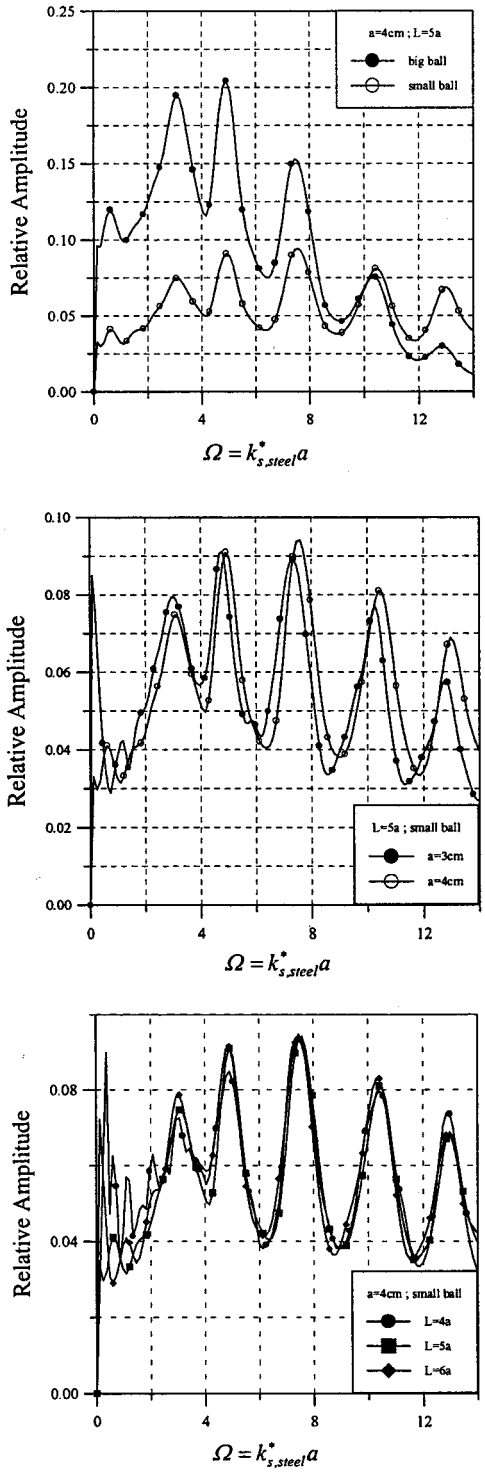


FIG. 10. Comparison of the power spectra of signals detected at P_4 ($x/a = 0.0$) under different experimental conditions.

and independent of the radius of the semicircular alluvium, the bandwidth of the source time function, and the location of the source point. It is noted that the resonance frequencies measured experimentally are in accordance with those calculated theoretically, and hence can be recognized as the efficient parameters to identify the alluvium system.

APPENDIX: DETERMINATION OF VECTOR A

Based on Eq. (18), the component A_n^β is defined by the integration of free field in the exterior region over the interface S of the scatterer and the surrounding half-space. In this Appendix, the evaluation of the component A_n^β corresponding to different incidence is described.

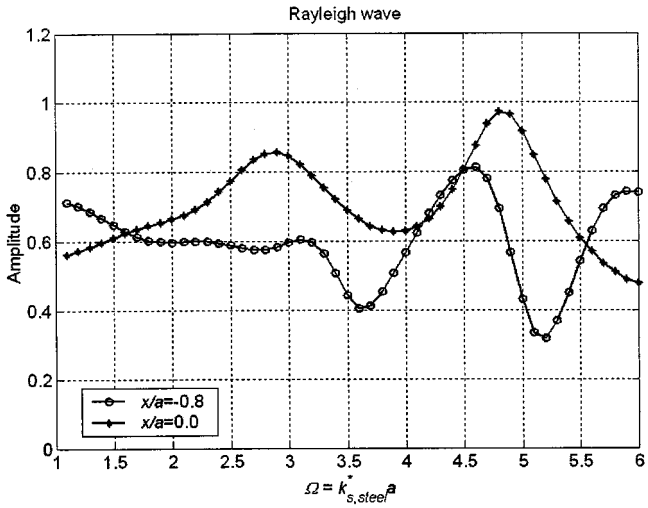


FIG. 11. Power spectra at P_2 ($x/a = -0.8$) and P_4 ($x/a = 0.0$) determined numerically by T-matrix formalism for the case of an incident Rayleigh wave.

Considering the Betti's third identity (Pao, 1978) for any two solutions **u** and **v** of the in-plane wave equations in an arbitrary two-dimensional volume bounded by a closed boundary S', we have

$$\oint_{S'} [\mathbf{t}(\mathbf{u}) \cdot \mathbf{v} - \mathbf{t}(\mathbf{v}) \cdot \mathbf{u}] dS = 0, \tag{A1}$$

where $\mathbf{t}(\mathbf{u})$ and $\mathbf{t}(\mathbf{v})$ are the corresponding traction vectors at the closed boundary S' due to the displacements **u** and **v**, respectively. Therefore, as shown in Fig. 12, Eq. (18) can be rewritten as

$$A_n^\beta = \int_{S_f} (\mathbf{t}^f \cdot \mathbf{u}_n^\beta - \mathbf{t}_n^\beta \cdot \mathbf{u}^f) dS = -\int_{S_f} (\mathbf{t}_n^\beta \cdot \mathbf{u}^f) dS. \tag{A2}$$

Herein, S_f denotes the segment at the free surface where $\mathbf{t}^f = 0$. On the other hand, based on the boundary conditions as defined by Eqs. (5) and (6) for the basis functions at S_f, Eq. (A2) can be rewritten as

$$A_n^h = \mu (i\xi)^n u_{x,0}^f; \quad A_n^v = \mu (i\xi)^n u_{z,0}^f, \tag{A3}$$

where ξ is the horizontal wave number, and can be defined for different types of incident waves as

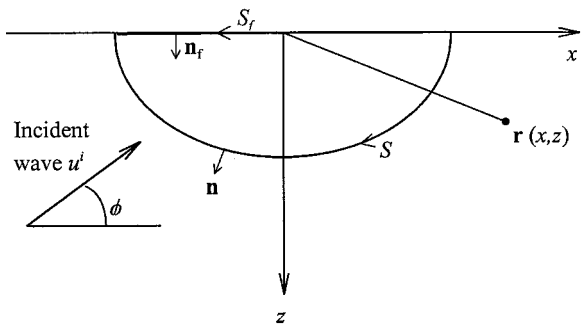


FIG. 12. Closed boundary and corresponding unit normal for Betti's third identity.

$$\xi = \begin{cases} k_p \cos\phi; & \text{for } P\text{-wave incidence} \\ k_s \cos\phi; & \text{for } S\text{-wave incidence} \\ k_R; & \text{for Rayleigh wave incidence.} \end{cases} \quad \text{(A4)}$$

The notations $u^f_{x,0}$ and $u^f_{z,0}$ are the x- and z components of the free-field displacement evaluated at $(x,z)=(0,0)$. It is noted that the expression of the free-field displacements corresponding to the longitudinal, shear, and Rayleigh incident waves in a two-dimensional half-space can be found elsewhere (Achenbach, 1973) and hence are not derived here.

Achenbach, J. D. (**1973**). *Wave Propagation in Elastic Solids* (North-Holland, New York).

Bard, P. Y., and Bouchon, M. (**1985**). "The two-dimensional resonance of sediment-field valley," Bull. Seismol. Soc. Am. **75**, 519–541.

Flax, L., Gaunaurd, G. C., and Überall, H. (**1981**). "Theory of resonance scattering," in *Physical Acoustics XV*, edited by W. P. Mason and R. N. Thurston (Academic, New York), pp. 191–294.

Gericke, O. R. (**1970**). *Research Technique in Non-Destructive Testing*, edited by R. S. Sharpe (Academic, London), Chap. 2, pp. 31–61.

Goldsmith, W. (**1960**). *Impact: The Theory and Physical Behavior of Colliding Solids* (Edward Arnold, London).

Jiang, T., and Kuribayashi, E. (**1988**). "The three-dimensional resonance of axisymmetric sediment-fields," Soils Found. **28**, 130–146.

Liao, W.-I. (**1997**). "The response of an obstacle included half-space subjected to elastic waves," Ph.D. thesis, National Taiwan University, Taipei.

Mosseian, T. K., and Dravinski, M. (**1990**). "Resonance motion of three-dimensional alluvial basins," 4th U.S. Nat. Conference on Earthquake Eng., 1, 525–534.

Pao, Y.-H., and Sachse, W. (**1974**). "Interpretation of time records and power spectra of scattered ultrasonic pulse in solids," J. Acoust. Soc. Am. **56**, 1478–1486.

Pao, Y. H., and Mow, C. C. (**1976**). "Theory of normal modes and ultrasonic spectral analysis of the scattering of waves in solids," J. Acoust. Soc. Am. **59**, 1046.

Pao, Y.-H. (**1978**). "The transition matrix for the scattering of acoustic waves and elastic waves," in *Proceedings of the IUTAM Symposium on Modern Problems in Elastic Wave Propagation*, edited by J. Miklowitz and J. Achenback (Wiley, New York), pp. 123–144.

Rial, J. A. (**1989**). "Seismic wave resonance in 3D sedimentary basins," Geophys. J. Int. **99**, 81–90.

Rial, J. A., Saltzman, N. G., and Ling, H. (**1991**). "Computation of normal modes of three-dimensional resonators by semiclassical and variational methods: Seismological implications," Wave Motion **14**, 377–398.

Sanchez-Sesma, F. J., and Esquivel, J. A. (**1979**). "Ground motion on alluvial valley under the incident plane *SH* waves," Bull. Seismol. Soc. Am. **69**, 1107–1120.

Teng, T.-J., Chai, J.-F., and Yeh, C.-S. (**1998**). "Determination of resonance frequency of two-dimensional alluvial valley by background phase subtraction method," Chin. J. Mech., **14**, No. 1, 1–15.

Trifunac, M. D. (**1971**). "Surface motion of a semi-cylindrical alluvial valley for incident plane *SH* waves," Bull. Seismol. Soc. Am. **61**, 1755–1770.

Überall, H. (**1978**). "Modal and surface wave resonance in acoustic-wave scattering from elastic objects and in elastic wave scattering from cavities," in *Proceedings of the IUTAM Symposium: Modern Problems in Elastic-Wave Propagation*, edited by J. Achenbach and J. Miklowitz (Wiley-Interscience, New York), p. 239.

Wirgin, A. (**1995**). "Resonance response of a soft semicircular cylindrical basin to an *SH* seismic waves," Bull. Seismol. Soc. Am. **85**, 285–299.

Yeh, C.-S, Teng, T.-J., and Chai, J.-F. (**1998**). "On the resonance of two-dimensional alluvial valley," Geophys. J. Int. **134**, No. 3, 787–808.

Yeh, C.-S, Teng, T.-J., Liao, W.-I., and Chai, J.-F. (**2002**). (unpublished).

Zhou, T., and Dravinski, M. (**1994**). "Resonance prediction of deep sediment valleys through an eigenvalue method," Geophys. J. Int. **117**, 749–762.

Acoustic streaming related to minor loss phenomenon in differentially heated elements of thermoacoustic devices

Mikhail Mironov
N. N. Andreev Acoustics Institute, Russian Academy of Sciences, 4, Shvernik str., 117036 Moscow, Russia

Vitalyi Gusev[a)]
Laboratoire de Physique de l'Etat Condensé, UMR-CNRS 6087, Faculté des Sciences, Université du Maine, av. O. Messiaen, 72085 Le Mans, France and Laboratoire d'Acoustique, UMR-CNRS 6613, Faculté des Sciences, Université du Maine, av. O. Messiaen, 72085 Le Mans, France

Yves Auregan, Pierrick Lotton, and Michel Bruneau
Laboratoire d'Acoustique, UMR-CNRS 6613, Faculté des Sciences, Université du Maine, av. O. Messiaen, 72085 Le Mans, France

Pavel Piatakov
N. N. Andreev Acoustics Institute, Russian Academy of Sciences, 4, Shvernik str., 117036 Moscow, Russia

(Received 15 November 2001; accepted for publication 13 May 2002)

It is demonstrated that the differentially heated stack, the heart of all thermoacoustic devices, provides a source of streaming additional to those associated with Reynolds stresses in quasi-unidirectional gas flow. This source of streaming is related to temperature-induced asymmetry in the generation of vortices and turbulence near the stack ends. The asymmetry of the hydrodynamic effects in an otherwise geometrically symmetric stack is due to the temperature difference between stack ends. The proposed mechanism of streaming excitation in annular thermoacoustic devices operates even in the absence of thermo-viscous interaction of sound waves with resonator walls. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1490361]

PACS numbers: 43.25.Nm [MFH]

The First International Workshop on Thermoacoustics, held recently,[1] has clearly indicated a growth of interest in possible manifestations in thermoacoustic devices of such nonlinear phenomena as acoustic streaming. An important (and even dominant under some conditions) role of acoustic streaming in heat transport in annular thermoacoustic devices was discovered experimentally both in quasi-isothermal[2,3] and quasi-adiabatic[4] regimes of interaction between acoustic and thermal waves inside the stack. These observations correlate with available theoretical analysis.[2,5] An important point here is the existence in annular thermoacoustic devices of a closed-loop path for steady streaming, which carriers a nonzero mass flow through a cross section of a thermoacoustic resonator. Moreover, it was demonstrated[6] that suppression of streaming leads to a significant increase in the efficiency of the device performance. At the same time it was proposed[6] to use for the suppression of the time-averaged mass flux through a thermoacoustic device another hydrodynamic effect related to a possible asymmetry of so-called ''minor'' losses. Minor losses (resulting from changes in form of flow cross section) are traditionally attributed to possible excitation of vortices and turbulence.[7–11] Classical examples are the formation of a jet in the fluid motion through a sudden enlargement of tube cross section [in the phase of outflow, Fig. 1(a)] and the formation of vortices in the region of vena–contracta just after a sudden contraction in tube cross section [in the phase of inflow, Fig. 1(b)]. In both cases additional losses in flow energy (called ''minor,'' ''shock,'' ''local,'' or ''head'' losses) are physically associated with the formation of vortices and, perhaps, of turbulence with additional dissipation of energy in them.[7,8] Importantly, these additional losses are accompanied by an additional stress difference across the transition region. If the total hydrodynamic stress difference provided by the regions of minor losses (along close loop of an annular thermoacoustic device) is nonzero, then it induces a streaming additional to a classical one[2,5,9] excited by Reynolds stresses in the straight parts of the resonator. Additional streaming may increase or decrease the efficiency of the thermoacoustic device depending on its magnitude and direction relative to classical streaming. To get a nonzero stress difference associated with the regions of minor loss in an oscillating gas flow there should be an asymmetry in the system. The asymmetry in minor losses may be achieved, for example, by geometrically asymmetric installations inside the resonator.[6]

In the present short communication we describe one more possible mechanism of the mass flow excitation in thermoacoustic devices, which, to the best of our knowledge, has never been analyzed before. The physical nature of this new source of streaming (or, more generally, of hydrodynamic stresses) is related to minor losses. However, an asymmetry of minor losses is related not to a geometrical asymmetry of the installation but to a temperature gradient along otherwise geometrically symmetric installation. Each thermoacoustic device contains this source of stress difference because the thermoacoustic stack (Fig. 2) is exactly the installation with required properties. We demonstrate that stress differences

[a)]Author to whom correspondence should be addressed. Electronic mail: vitali.goussev@univ-lemans.fr

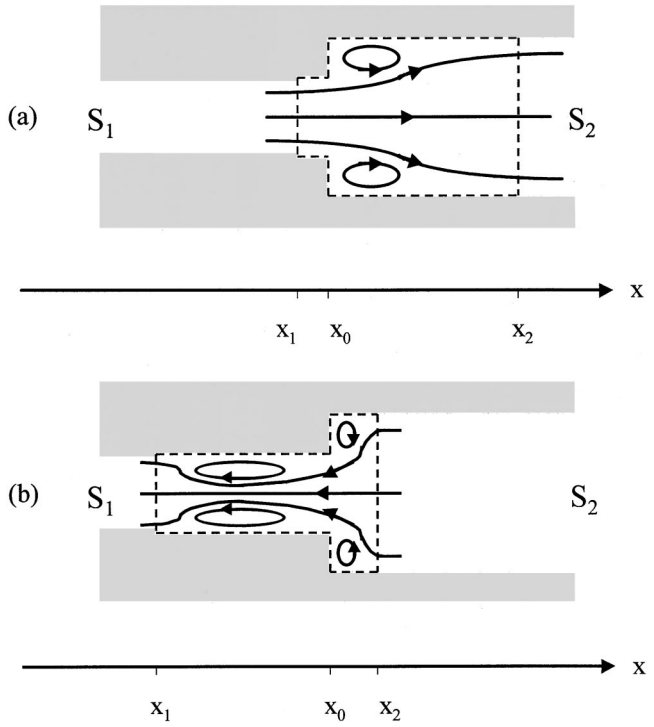


FIG. 1. Qualitative picture of streamlines for the outflow (a) and inflow (b) through a sudden enlargement in tube cross section area from S_1 to S_2. Dashed lines are used to show control volume where conservation laws (of momentum, energy, and mass) are analyzed.

associated with minor losses at the cold and hot ends of thermoacoustic stack (or regenerator) are not mutually compensated on average over the oscillation period because they take place at different temperatures. We estimate that this additional source of streaming might provide a valuable contribution to the total mass flow in annular thermoacoustic prime movers.

From a physics point of view, an additional source of streaming (or an additional stress difference) in the region of sudden enlargement or contraction of tube cross section can be attributed to a loss of momentum flux in oscillatory fluid motion. The stress transferring momentum from oscillatory flow to dc flow (acoustic streaming), due to formation of vortices and nonlinear processes in them, can be found by

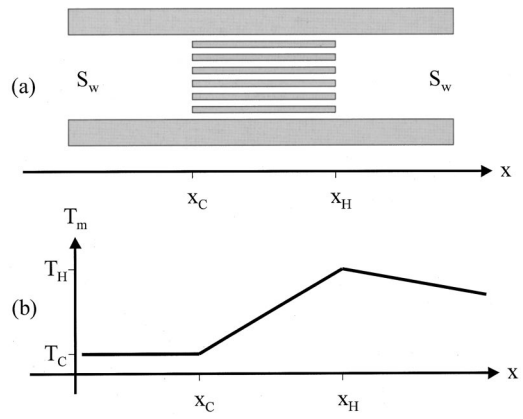


FIG. 2. Simplified presentation of thermoacoustic stack (region $x_C<x<x_H$) inside the tube of cross section area S_W (a) and of temperature distribution along it (b).

combining the equations of mass, momentum, and energy conservation[7,10,11] for the control volume, which includes the region of cross-section enlargement ($x_1<x<x_2$ in Fig. 1). The additional stress related to minor loss phenomenon is traditionally presented in terms of additional abrupt pressure variation Δp^{ml} in the region of abrupt change in tube cross section,[10,11]

$$\Delta p^{ml}_{\rm out}(x_0)\equiv p^{ml}_{\rm out}(x_0+0)-p^{ml}_{\rm out}(x_0-0)$$
$$\approx -K_{\rm out}\tfrac{1}{2}\rho_m v_x^2(x_0-0), \quad (1)$$
$$\Delta p^{ml}_{\rm in}(x_0)\equiv p^{ml}_{\rm in}(x_0+0)-p^{ml}_{\rm in}(x_0-0)$$
$$\approx K_{\rm in}\tfrac{1}{2}\rho_m v_x^2(x_0-0). \quad (2)$$

Here the notation "$x_0\mp 0$" indicates that the function has to be evaluated left/right of x_0 in its close proximity, ρ_m stands for the mean density, and $v_x(x)$ denotes the axial component of the flow velocity. Positive coefficients $K_{\rm out}>0$ and $K_{\rm in}>0$ of minor losses for the phase of outflow and inflow, respectively, describe a debit of energy in Bernouilli equation.[7,10,11] The coefficients $K_{\rm out}$ and $K_{\rm in}$ depend on the ratio of the tube cross sections (S_1/S_2) before and after its sudden enlargement (Fig. 1) and on the Reynolds number if viscosity is taken into account.

We assume that, if the gas displacement amplitude in the acoustic field is much greater than the tube diameter, the flow at any instant has little memory of its past history, so that the acoustic behavior can be deduced from careful time averaging of high-Reynolds-number steady-flow phenomena predicted by Eqs. (1) and (2).[6,12] Considering that each of Eqs. (1) and (2) holds during a half-period of a sinusoidal oscillation, an expression for the time-averaged linear density of stresses acting in the region of sudden variation in flow cross section is derived:[6]

$$\langle \Delta p^{ml}(x_0)\rangle \equiv -\tfrac{1}{8}(K_{\rm out}-K_{\rm in})\rho_m|\tilde{v}_x(x_0-0)|^2. \quad (3)$$

Here $\langle\cdots\rangle$ stands for time averaging over a period of acoustic oscillation and $|\tilde{v}_x|$ denotes the amplitude of harmonic (sinusoidal) velocity oscillation.

Additional stress due to minor losses [Eq. (3)] can be directly incorporated in available theories of streaming in thermoacoustic devices containing closed loops for mass flow.[2,5] In this case mass flow is proportional to the ratio of integral hydrodynamic stress over the closed loop and the total hydrodynamic resistance of the resonator.[5] Pressure jumps in Eq. (3) (related to minor losses) contribute to the stress across thermoacoustic stack [Fig. 2(a)] even if the stack is geometrically symmetric [that is, if $K_{\rm out}(x=x_H)=K_{\rm out}(x=x_C)$ and $K_{\rm in}(x=x_H)=K_{\rm in}(x=x_C)$, where x_H and x_C denote in Fig. 2 the coordinates of hot and cold ends of the stack]:

$$S^{ml}\equiv -[\langle \Delta p^{ml}(x_H)\rangle - \langle \Delta p^{ml}(x_C)\rangle]$$
$$=\tfrac{1}{8}(K_{\rm out}-K_{\rm in})\{\rho_m(x_H)|\tilde{v}_x(x_H-0)|^2$$
$$-\rho_m(x_C)|\tilde{v}_x(x_C+0)|^2\}. \quad (4)$$

In Eq. (4) it is assumed that the minor loss effects at the ends of the stack are mutually independent, that is vorticity generated at one end of the stack dissipates before being convected to its other end.[11] In accordance with Eq. (4) the

magnitude and the sign of the stress S^{ml} is controlled not only by the asymmetry of minor losses at each end of the stack (described by the multiplier $K_{\text{out}}-K_{\text{in}}$), but also by a possible difference in average dynamic pressures ($\propto \rho_m|v_x|^2$) at hot and cold ends of the stack. For an ideal gas $\rho_m \propto 1/T_m$, where T_m is the mean temperature and stress S^{ml} takes the form

$$S^{ml}=\tfrac{1}{8}(K_{\text{out}}-K_{\text{in}})\rho_m(x_C)|\tilde{v}_x(x_C+0)|^2 \times\left[\frac{T_C|\tilde{v}_x(x_H-0)|^2}{T_H|\tilde{v}_x(x_C+0)|^2}-1\right]. \quad (5)$$

In accordance with Eq. (5) the magnitude and the sign of S^{ml} depend on the cold to hot temperature ratio (T_C/T_H) [see qualitative presentation of temperature distribution in Fig. 2(b)] and on the ratio of particle velocities at the ends of the stack. The latter itself depends, in general, on (T_C/T_H) and also, importantly, on all other elements inside thermoacoustic resonator contributing to formation of a particular spatial distribution of counter-propagating acoustic waves. Even in the traveling-wave devices the role of a nonzero standing-wave component may be of crucial importance. So the magnitude and the sign of the expression in square brackets in Eq. (5) should be carefully evaluated in each particular situation. Here, to proceed a little bit further in a quasi-adiabatic analysis, we assume additionally that the stack may be approximated as infinitely thin and that Rott's boundary condition $v_x(x_H-0)=v_x(x_C+0)\equiv v_x$ holds at the δ-localized temperature jump.[13] Note that in accordance with the theory[5] the condition of velocity continuity holds in annular prime mover with quasi-adiabatic stack of a finite length H under the requirement $(T_H-T_C)/T_C \geqslant k(T_C)H$, where $k(T_C)$ is the acoustic wave number evaluated at cold temperature [see Eq. (29) in Ref. 5]. Then the stress S^{ml} across the thermoacoustic stack is approximated by

$$S^{ml}=\frac{1}{8}(K_{\text{in}}-K_{\text{out}})\left(1-\frac{T_C}{T_H}\right)\rho_m(T_C)|\tilde{v}_x|^2. \quad (6)$$

Consequently, in this limiting case the effect under consideration is caused by differential heating of the stack. Solution (6) predicts that additional stress S^{ml} related to minor loss phenomena is created across thermoacoustic stack only if $T_H \neq T_C$. When the sign of the temperature gradient is fixed [as, for example, in Fig. 2(b), where $T(x_H)>T(x_C)$], then the sign of S^{ml} is controlled by the sign of the difference $K_{\text{in}}-K_{\text{out}}$.

Empirical relations for the case of high Reynolds numbers ($\text{Re}\geqslant 10^4$) predict the following variation of minor loss coefficients with increasing porosity: $K_{\text{out}}\approx(1-S_S/S_W)^2$ and $K_{\text{in}}\approx 0.5(1-S_S/S_W)$ or, more precisely, $K_{\text{in}}\approx 0.5(1-S_S/S_W)^{3/4}$ (see Ref. 11, Chap. 4). Here S_W is the total cross sectional area available for the flow in the wave guide [see Fig. 2(a)] and (S_S/S_W) is the porosity of the stack. In this limit $K_{\text{in}}\geqslant K_{\text{out}}$ already in the stacks with porosity exceeding 50%. This result is also in accordance with the other available data.[8,10] So, the direction of streaming excited by S^{ml} in typical quasi-adiabatic thermoacoustic stacks (which have significantly higher porosity) coincides with the direction of streaming excited in quasi-adiabatic (QA) annular prime-mover by Reynolds stresses. However, it should be mentioned that rounding of the transition region is commonly known to influence (to diminish) K_{in} much stronger than K_{out} (see, for example, the case $(S_1/S_2)\to 0$ in Ref. 6). Consequently, it is possible that for the stack with rounded pore edges the direction of the streaming related to asymmetry of minor losses in the quasi-adiabatic stack will be opposite to that of classical streaming.

Using for the Reynolds stress S^{QA} the solution derived in Ref. 5, Eq. (35), and approximating the ratio of acoustic pressure amplitude $|\tilde{p}|$ to particle velocity amplitude $|\tilde{v}_x|$ by its traveling wave value $\rho(T_C)a(T_C)$ [where $a(T_C)$ is the adiabatic speed of sound at cold temperature], we get the following estimate for the ratio of stresses contributing to streaming excitation:

$$\frac{S^{ml}}{S^{QA}}\approx\frac{\frac{1}{8}(1-T_C/T_H)}{[(c+0.5d)(\beta+1)/3](\delta_{\nu c}/D_S)\ln(T_H/T_C)} \times(K_{\text{in}}-K_{\text{out}}). \quad (7)$$

Equation (7) is valid for a stack with square pores where the volume of boundary layers is approximately twice as large as in the case of the parallel plates stack considered in Ref. 5. Thus S^{QA} from Eq. (35) in Ref. 5 was doubled in Eq. (7). We use Eq. (7) to estimate a possible role of S^{ml} in streaming excitation under the conditions of the experiment in Ref. 4 (where $\text{Re}\equiv|\tilde{v}_x|D_S/\nu\propto 10^4$). We use the values $c\approx 0.74$ and $d\approx 0.27$ for the Kirchhoff and Kramer constants of air, $\beta\approx 0.73$ for the phenomenological parameter describing the temperature dependence of the cinematic viscosity ν ($\nu\propto T_m^{\beta+1}$). The ratio of the viscous boundary layer thickness at cold temperature $\delta_{\nu c}$ to characteristic width D_S of a single pore was about 0.19 in Ref. 4, $T_H/T_C\approx 2$ and the porosity $S_S/S_W\approx 0.8$. The result of the estimate is $S^{ml}/S^{QA}\approx 0.1$. Consequently, temperature-induced asymmetry in minor loss phenomena at the thermoacoustic stack edges might be responsible for about 10% of mass flow observed in the experiment in Ref. 4.

In general, it is clear from Eq. (7) that in the quasi-adiabatic regime the relative importance of source S^{ml} increases when the parameter of adiabaticity $(\delta_{\nu c}/D_S)\ll 1$ diminishes. In principle, the magnitude of the investigated source of streaming is nonzero and does not depend on viscosity when $\text{Re}\to\infty$ (though the final stage of shock loss phenomenon is, surely, viscous dissipation of energy in vortices or in turbulence). So the predicted mechanism of streaming excitation exists even for the adiabatic regime of sound propagation inside the thermoacoustic resonator when the source S^{QA} of classical acoustic streaming disappears ($S^{QA}\to 0$).

The dependence of minor loss phenomena on the Reynolds number in the region $\text{Re}<10^4$ is significant for all values of porosity. For example, from the results presented in Ref. 11, Chap. 4, it follows that for $(S_S/S_W)=0.6$ even the sign of S^{ml} depends on Re ($K_{\text{in}}>K_{\text{out}}$ for $\text{Re}\geqslant 4\times 10^3$ and $\text{Re}\leqslant 10^2$, $K_{\text{in}}<K_{\text{out}}$ for $10^2\leqslant\text{Re}\leqslant 2\times 10^3$). The general tendency for the regimes with $\text{Re}\leqslant 10^4$ is the increase of both K_{in} and K_{out} with diminishing Reynolds number (Ref. 11,

Chap. 4). For example, for $(S_S/S_W)=0.6$ the coefficients K_{in} and K_{out} increase approximately four and five times, respectively, when Re diminishes from $\text{Re}\geqslant 10^4$ to $\text{Re}\propto 10^2$. Unfortunately, we have failed to find any empirical data for higher porosity [and, in particular, for $(S_S/S_W)=0.8$ corresponding to our experiment[4]]. So, currently, we are unable to predict for the quasi-adiabatic stacks with high porosity and $10<\text{Re}\leqslant 10^4$ not just the magnitude but even the direction of the additional streaming associated with minor loss phenomena.

However, from the data presented in Ref. 11, it follows that the dependence of K_{in} and K_{out} on porosity gradually disappears when Re approaches the magnitudes $\text{Re}\leqslant 10$. Assuming that the data for $\text{Re}\leqslant 10$ in Ref. 11 can be applied to the high-porosity thermoacoustic stack, we are able to make the formal estimates for the quasi-isothermal regime of the fluid/stack interaction (usually corresponding to relatively low Reynolds numbers inside the stack[3]).

In the quasi-isothermal regime the incompressible fluid flow inside the stack is unidirectional. From the mass conservation law it follows that $\tilde{v}_x\propto 1/\rho(x)\propto T(x)$. In an annular prime mover with quasi-isothermal stack of a finite length H the latter relation holds under the requirement $(T_H-T_C)/T_C\geqslant k(T_C)H$ [see Eq. (29) in Ref. 5]. Under this condition the approximate solution for the pressure variation is $\partial\tilde{p}/\partial x\propto T^{\beta+1}$ [see Eq. (44) in Ref. 5]. Consequently, from the balance of viscous stress and pressure gradients in the flow, it follows once again that $\tilde{v}_x\propto[1/\nu(x)\rho(x)]\partial\tilde{p}/\partial x\propto T(x)$. Using $\tilde{v}_x\propto T(x)$, we deduce from Eq. (5) that, due to this dependence of the particle velocity on temperature, the stress S^{ml} across the quasi-isothermal stack differs from the result presented in Eq. (6) by an additional multiplier $(-T_H/T_C)$. Consequently, the source of streaming related to temperature-induced asymmetry of minor loss phenomena has opposite directions in the quasi-isothermal and quasi-adiabatic stacks. The factor $(-T_H/T_C)$ is included in all the estimates presented below.

In the case of a traveling wave thermoacoustic engine the solution for the Reynolds stress S^{QI} in the quasi-isothermal regime has been derived in Ref. 5, Eq. (46). Using this solution and approximating the ratio of acoustic pressure amplitude $|\tilde{p}|$ to particle velocity amplitude $|\tilde{v}_x|$ by its traveling wave value $\rho(T_C)a(T_C)$, we get the following estimate for the ratio of stresses contributing to streaming excitation:

$$\frac{S^{ml}}{S^{QI}}\approx-\frac{70}{8(17+8\beta)\sigma}(K_{\text{in}}-K_{\text{out}})\approx-\frac{1}{2}(K_{\text{in}}-K_{\text{out}}) \quad (8)$$

in air ($\sigma\approx 0.71$ is the Prandtl number). In accordance with Ref. 11 the difference in minor loss coefficients $K_{\text{in}}-K_{\text{out}}\approx 2$ for $\text{Re}=10$ and briefly diminishes with diminishing Reynolds number when the difference between inflow and outflow progressively disappears. Consequently, from Eq. (8) it follows that in a traveling wave thermoacoustic engine in quasi-isothermal regime with $\text{Re}\equiv|\tilde{v}_x|D_S/\nu\leqslant 10$ and $T_H/T_C\geqslant 2$, temperature-induced asymmetry in minor loss phenomena at the thermoacoustic stack edges might significantly diminish the mass flow.

To estimate the role of the considered mechanism in a thermoacoustic Stirling engine,[3] where the ratio $|\tilde{v}_x|/|\tilde{p}|\ll 1/\rho(T_C)a(T_C)$ [contrary to traveling wave thermoacoustic engine,[5] where $|\tilde{v}_x|/|\tilde{p}|\approx 1/\rho(T_C)a(T_C)$], we first found the cross-sectional average mass flow M^{ml} due to minor losses. This is achieved by taking the ratio of the stress S^{ml} [Eq. (6) multiplied by $(-T_H/T_C)$] to total hydrodynamic resistance $R\equiv\oint(2/D)^2\nu\,dx$ of a closed-loop device [here $D=D(x)$ is the channel width].[5] We estimated that the hydrodynamic resistance is dominated by the contribution from the quasi-isothermal stack (regenerator), we neglected the dependence of cinematic viscosity on temperature $\nu(x)\approx\text{const}$, and we approximated the resistance by $R\approx 2[1+(T_H/T_C)]\nu H_S/D_S^2$ (where H_S denotes the stack length). We compared the derived mass flow M^{ml} with the solution proposed by Gedeon:[2] $|M^G|=(\frac{1}{8})[(1/\gamma)-1]\rho(T_C)|\tilde{v}_x||\tilde{p}|/p_m$ (here γ is the ratio of specific heats, i.e., is the adiabatic parameter, and p_m is the mean pressure in the system). The result of this comparison is

$$\frac{|M^{ml}|}{|M^G|}\equiv\frac{|S^{ml}/R|}{|M^G|}\approx\frac{1}{2}\left(\frac{\gamma}{\gamma-1}\right)\frac{[(T_H/T_C)-1]}{[(T_H/T_C)+1]}\left(\frac{D_S}{H_S}\right)\frac{\text{Re}}{(|\tilde{p}|/p_m)}\times|K_{\text{in}}-K_{\text{out}}|. \quad (9)$$

The ratio of the mass flows in Eq. (9) depends on the level of acoustic oscillations only implicitly through the dependence of the minor loss coefficients on the Reynolds number. For the system described in Ref. 3 we have estimated that $|M^{ml}|/|M^G|\approx|K_{\text{in}}-K_{\text{out}}|/20$. The reported thermoacoustic Stirling heat engine[3] approaches the regime $\text{Re}\propto 10$ inside the stack only in the most powerful operating point (where acoustic Mach number approaches the value 0.1). In this regime temperature-induced asymmetry in minor loss phenomena at the thermoacoustic stack edges might be responsible for about 30% of the mass flow ($K_{\text{in}}-K_{\text{out}}\approx 2$). For lower Reynolds numbers the expected relative contribution of minor loss phenomenon to mass flow is less.

In discussion we would like to mention the following important point. It should not be forgotten that we have used for the estimates the values of K_{in} and K_{out} measured for a sudden transition in tube cross section (Fig. 1), while in the thermoacoustic device the transition takes place between the stack of tubes (pores) and a single tube of much larger diameter [Fig. 2(a)]. The structure of the flow in wider regions [S_2 in Fig. 1 and S_W in Fig. 2(a)] might be quite different in the case of a moderate Reynolds number. To overcome the above described restriction it looks attractive to use somehow the empirical data on hydraulic resistance of barriers uniformly distributed over the channel cross section (Ref. 11, Chap. 8), because the thermoacoustic stack is exactly such a barrier. However, it is not obvious at all how to incorporate in these results differential heating of the barrier (stack).

In conclusion, the effect of streaming excitation in annular thermoacoustic devices due to combined action of differential heating of the stack and of asymmetry of inflow and outflow at each end of the stack is predicted theoretically.

[1] *Book of Abstracts, First International Workshop on Thermoacoustics*, 22–25 April 2001, 's-Hertogenbosch, The Netherlands.
[2] D. Gedeon, ''DC gas flows in Stirling and pulse-tube cryocoolers,'' in *Cryocoolers*, edited by R. G. Ross (Plenum, New York, 1997), Vol. 9, pp. 385–392.
[3] S. Backhaus and G. W. Swift, ''A thermoacoustic Stirling heat engine,'' Nature (London) **399**, 335–338 (1999).
[4] S. Job, V. Gusev, P. Lotton, and M. Bruneau, ''On the velocity of acoustic streaming in annular thermoacoustic prime-mover,'' J. Acoust. Soc. Am. (submitted).
[5] V. Gusev, S. Job, H. Bailliet, P. Lotton, and M. Bruneau, ''Acoustic streaming in annular thermoacoustic prime-movers,'' J. Acoust. Soc. Am. **108**, 934–945 (2000).
[6] G. W. Swift, D. L. Gardner, and S. Backhaus, ''Acoustic recovery of lost power in pulse tube refrigerators,'' J. Acoust. Soc. Am. **105**, 711–724 (1999).
[7] G. K. Betchelor, *An Introduction to Fluid Dynamics* (Cambridge U.P., Cambridge, 1994).
[8] B. S. Massey, *Mechanics of Fluids* (Chapman & Hall, London, 1996).
[9] J. R. Olson and G. W. Swift, ''Acoustic streaming in pulse tube refrigerators: Tapered pulse tubes,'' Cryogenics **37**, 769–776 (1997).
[10] V. L. Streeter, *Handbook of Fluid Dynamics* (McGraw–Hill, New York, 1961).
[11] I. E. Idelchik, *Handbook of Hydraulic Resistance*, 2nd ed. (Hemisphere, 1986).
[12] G. R. Thurston, ''Nonlinear acoustic properties of orifices of varied shapes and edge conditions,'' J. Acoust. Soc. Am. **30**, 452–455 (1958).
[13] N. Rott, ''Thermoacoustics,'' Adv. Appl. Mech. **20**, 135–175 (1980).

A linearized Eulerian sound propagation model for studies of complex meteorological effects

Reinhard Blumrich[a] and Dietrich Heimann[b]
Institut für Physik der Atmosphäre, DLR-Oberpfaffenhofen, D-82234 Weßling, Germany

(Received 3 January 2001; revised 4 April 2002; accepted 22 April 2002)

Outdoor sound propagation is significantly affected by the topography (including ground characteristics) and the state of the atmosphere. The atmosphere on its part is also influenced by the topography. A sound propagation model and a flow model based on a numerical integration of the linearized Euler equations have been developed to take these interactions into account. The output of the flow model enables the calculation of the sound propagation in a three-dimensionally inhomogeneous atmosphere. Rigid, partly reflective, or fully absorptive ground can be considered. The linearized Eulerian (LE) sound propagation model has been validated by means of four different scenarios. Calculations of sound fields above rigid and grass-covered ground including a homogeneous atmosphere deviate from analytic solutions by $\leqslant 1$ dB in most parts of the computed domain. Calculations of sound propagation including wind and temperature gradients above rigid ground agree well with measured scale model data. Calculations of sound propagation over a screen including ground of finite impedance show little deviations to measured scale model data which are probably caused by an insufficient representation of the complex ground impedance. Further calculations included the effect of wind on shading by a screen. The results agree well with the measured scale model data. [DOI: 10.1121/1.1485971]

PACS numbers: 43.28.Fp, 43.28.Gq, 43.28.Js [LCS]

I. INTRODUCTION

The propagation of sound outdoors is influenced by the predominant topography and the state of the atmosphere. The topography includes ground characteristics, ground-based obstacles, and terrain features which cause reflection and diffraction. The atmosphere affects sound propagation by gradients of wind and temperature which lead to refraction and by turbulent fluctuations which cause scattering. Temperature and humidity determine the degree of absorption of the acoustic wave energy. Overviews of outdoor sound propagation were given, for example, by Piercy *et al.*[1] and Embleton.[2] Ground characteristics were discussed, for example, by Attenborough.[3]

Topography and atmosphere, however, do not act independently. Terrain features, obstacles, and roughness elements modify the exchange of momentum, energy (sensitive and latent heat, radiation), and mass (of water) between ground and air and thus determine the behavior of atmospheric parameters. Hence, the topography exerts a direct and indirect (via the atmosphere) influence on the sound field.

Several models exist for a frequency-dependent calculation of outdoor sound propagation. Almost all are based on the acoustic wave equation. The various algorithms use either the parabolic wave equation, ray-tracing methods, or a Hankel transform of the Helmholtz equation. They consider topographic and atmospheric influences to some extent. Different algorithms for refraction in the presence of flat, even ground were reviewed in 1995.[4] Beyond that, more complicated situations of sound propagation have been discussed, as for example multiple reflection and diffraction in a nonrefracting atmosphere in 1997,[5] or diffraction and refraction due to inhomogeneous atmosphere above plane ground in 1998.[6] The combined effect of a barrier and wind gradients has been investigated by means of measured wind speed profiles.[7,8]

Recent studies use input data from either meteorological mesoscale models or fluid dynamics models which simulate the atmospheric environment influenced by the topography. These models provide consistent atmospheric data for the whole considered volume. The atmosphere is seen to be stationary while the sound propagates through the model domain. The acoustic models, however, still approximate the atmospheric influences on the sound waves. The particle model of Heimann *et al.*[9] based on ray-tracing methods, for instance, requires us to parametrize diffraction. The approach of Hole *et al.*[10] based on a fast-field program assumes range-independent acoustical properties of the atmosphere. The acoustic models based on the parabolic equation (PE) need to combine the effects of wind and temperature gradients in one parameter, viz., the vertical gradient of the effective speed of sound. As a consequence, the three-dimensionality of the wind vector is neglected.[11–13]

Our investigations aim at a simulation of sound propagation through an inhomogeneous atmosphere including as few approximations as possible. For this, a grid-based algorithm derived from the Euler equations was chosen as a basis for a sound propagation model. The model uses input data from a meteorological mesoscale model or a flow model. A frozen atmospheric state is considered, as in the case of the recent studies mentioned above. The linearized Eulerian (LE) sound propagation model is able to consider flow features

[a]Electronic mail: reinhard.blumrich@dlr.de
[b]Electronic mail: d.heimann@dlr.de

and other acoustical properties of the atmosphere in three dimensions and dependent on the range. Diffraction behind obstacles is explicitly simulated, hence, a parametrization is not necessary. The calculations can be carried out for two or three dimensions, for single or multiple frequencies.

Our approach allows determining the contribution of single atmospheric influences to the whole effect on sound waves. Thus, the error of neglecting specific influences, e.g., the vertical wind speed component, can be estimated and requirements for simplified, efficient algorithms can be defined. The error of neglecting specific influences will be discussed in a separate paper.

The present article introduces the flow model and the LE sound propagation model and discusses the validation. Results of calculations are compared with either analytical calculations or measured data and other sound propagation models. The measured data were derived from scale model measurements. Inhomogeneous atmosphere is either simulated by the flow model (here: 2D) or prescribed by constant wind and temperature gradients.

II. MODEL DESCRIPTION

The flow model and the sound propagation model are based on the governing equations of a compressible and adiabatic gaseous medium in a nonrotating system. These are the equation of motion, the equation of continuity, and the first law of thermodynamics for adiabatic processes [Eqs. (1)–(3)] (gravity is neglected):

$$\frac{\partial \mathbf{u}}{\partial t}+(\mathbf{u}\cdot\nabla)\mathbf{u}=-\frac{1}{\rho}\nabla p, \tag{1}$$

$$\frac{\partial \rho}{\partial t}+\mathbf{u}\cdot\nabla\rho=-\rho\nabla\cdot\mathbf{u}, \tag{2}$$

$$\frac{c_p}{\rho}\frac{d\rho}{dt}=\frac{c_v}{p}\frac{dp}{dt}. \tag{3}$$

We use $\mathbf{x}=x\ \mathbf{i}+y\ \mathbf{j}+z\ \mathbf{k}$ as the position vector and $\mathbf{u}=u\ \mathbf{i}+v\ \mathbf{j}+w\ \mathbf{k}$ as the three-dimensional velocity vector. Here **i**, **j**, and **k** are the unity vectors in the *x*, *y*, and *z* directions, respectively. **k** points to the vertical. With *t*, ρ, and *p* we denote the scalar quantities time, density, and pressure, respectively. c_p and c_v are the specific heats of air at constant pressure and constant volume, respectively.

With the help of Eq. (3) the density ρ can be eliminated from Eq. (2), resulting in a new equation for the pressure *p*:

$$\frac{\partial p}{\partial t}+\mathbf{u}\cdot\nabla p=-\kappa p\nabla\cdot\mathbf{u} \tag{4}$$

with $\kappa=c_p/c_v$.

In the following the atmospheric variables $\phi=(\mathbf{u},p,\rho)$ are split up into their meteorological, turbulent, and acoustic parts:

$$\phi=\overline{\phi}+\phi'+\phi''. \tag{5}$$

The overbar variables $\overline{\phi}$ denote mean wind $\overline{\mathbf{u}}$, air pressure $\overline{p}$, and air density $\overline{\rho}$. The single prime variables ϕ' indicate the turbulent deviations from the mean meteorological values. Finally, the double prime variables ϕ'' signify the deviations from $\overline{\phi}$ according to acoustic waves, in particular sound pressure p'' and particle velocity $\mathbf{u}''$. The time scale of the acoustic fluctuations is smaller than that of the turbulent fluctuations.

A. Flow model

The flow model is based on the nonlinear Eqs. (1) and (4). It provides numerical solutions for $\overline{\mathbf{u}}$ and $\overline{p}$ in a neutrally stratified atmosphere in which buoyancy does not play a role (note: gravity is neglected). Since subgrid-scale turbulent fluctuations cannot be explicitly calculated, their effect on the mean meteorological variables is parametrized. For this reason a further equation is introduced by which the turbulent kinetic energy $\overline{E}=0.5\overline{|\mathbf{u}'|^2}$ is calculated. Following Yamada,[14] the temporal derivative of $\overline{E}$ is given by

$$\frac{\partial \overline{E}}{\partial t}+\overline{\mathbf{u}}\cdot\nabla E=K_M|\nabla\overline{\mathbf{u}}|^2-\frac{c_E\overline{E}^{3/2}}{\ell}. \tag{6}$$

The two terms on the right side describe production and dissipation of turbulent kinetic energy. The production term only accounts for shear production, because buoyant accelerations are not described by Eqs. (1) and (4) anyway. $c_E=1.2$ is a constant. ℓ is a mixing length which is calculated in analogy to Blackadar[15] by

$$\ell=\frac{kd}{1+kd/\ell_\infty} \tag{7}$$

with *d* being the distance to the closest rigid surface. $k=0.4$ is the von Karman constant. ℓ_∞ is set to 25 m. K_M is the turbulent diffusion coefficient for momentum. It is determined from the local turbulent kinetic energy by the Prandtl–Kolmogorov relation[16]

$$K_M=\ell\sqrt{0.2\overline{E}}. \tag{8}$$

The equation of motion of the flow model is completed by a turbulent diffusion term which accounts for the turbulence effects on the mean flow field:

$$\frac{\partial \overline{\mathbf{u}}}{\partial t}+(\overline{\mathbf{u}}\cdot\nabla)\overline{\mathbf{u}}=-\frac{1}{\overline{\rho}}\nabla\overline{p}+(\nabla\cdot K_M\nabla)\overline{\mathbf{u}}. \tag{9}$$

The pressure equation of the flow model reads

$$\frac{\partial \overline{p}}{\partial t}+\overline{\mathbf{u}}\cdot\nabla\overline{p}=-\kappa\overline{p}\nabla\cdot\overline{\mathbf{u}}. \tag{10}$$

Equations (6), (9), and (10) are numerically integrated in time on a staggered (Arakawa C-type)[17] orthogonal grid with grid cells of the size $\Delta x\times\Delta y\times\Delta z$. The velocity components $\overline{u}$, $\overline{v}$, and $\overline{w}$ are defined on the side walls of the grid cells which are perpendicular to the respective component. The scalars $\overline{p}$ and $\overline{\rho}$ are defined in the center of the grid cells.

The time step Δt must obey the Courant–Friedrich–Levi criterion[18] $\Delta t\leqslant 0.5\Delta x c^{-1}$ where *c* denotes the fastest

signal velocity. In the framework of Eqs. (9) and (10) this is the speed of sound, i.e., $c=\sqrt{\kappa\overline{p}/\overline{\rho}}$. In general, explicit forward-in-time and centered-in-space differences are employed. As an exception, the diffusion term is solved by an implicit time-scheme.

Ground and obstacles are defined by ''solid'' grid cells, i.e., the velocity components (normal and tangential) on their side walls are set to zero. This corresponds to a no-slip boundary condition at material surfaces and prevents the air from penetrating either ground or obstacles. For special studies, however, model options can be invoked which account for free-slip boundary conditions or semi-permeable obstacles such as, for example, hedges. The semi-permeability can be simulated by a fractional damping of the flow velocity which acts as a momentum sink.[19]

At the time of initialization, flat ground and permeable obstacles are assumed. Accordingly, the initial flow is horizontal and uniform. The friction at ground is accounted for by a logarithmic vertical wind speed profile:

$$|\overline{\mathbf{u}}|=\frac{u_*}{k}\ln\left(\frac{z+z_o}{z_o}\right) \tag{11}$$

with the friction velocity u_*, the von Karman constant $k=0.4$, and the roughness length z_o. Friction velocity and roughness length have to be prescribed.

After initialization the permeability of elevated ground and obstacles is successively reduced to zero within a certain span of time. This procedure helps to avoid the generation of spurious distortions due to initial inconsistencies between flow and rigid boundaries.

Towards the upper and lateral inflow boundaries of the numerical grid domain the solution for $\overline{\mathbf{u}}$ is forced to adapt to the corresponding values of the initial flow with increasing emphasis (nudging technique).[20]

B. LE sound propagation model

Although the physical formulation of the flow model includes sound waves, numerical reasons suggest the use of a separate acoustic model. The problem originates from the large difference in magnitude between the speed of the mean airflow ($\approx 10^1$ m/s) and the particle velocity associated with sound waves ($\ll 10^{-2}$ m/s). As a consequence, numerical noise that inevitably emerges from the air flow solution superimposes and distorts the wave solution.

In the following we describe a sound propagation model which is based on prognostic equations of $\mathbf{u}''$ and p''. The model equations are deduced from Eqs. (1) and (4) with $\mathbf{u}=\overline{\mathbf{u}}+\mathbf{u}''$ (velocity), $p=\overline{p}+p''$ (pressure), and $\rho^{-1}=\alpha=\overline{\alpha}+\alpha''$ (specific volume). Since the atmosphere is seen to be stationary, the turbulent parts ϕ' are disregarded here. Turbulence effects on the mean flow field $\overline{\mathbf{u}}$ are included (see previous subsection).

Taking the stationarity of the mean state into account ($\overline{\mathbf{u}}$, $\overline{p}$, and $\overline{\alpha}$ are constant in time) and after linearization with respect to the mean state, i.e., neglecting the nonlinear terms $(\mathbf{u}''\cdot\nabla)\mathbf{u}''$ and $\mathbf{u}''\cdot\nabla p''$, the model equations read as

$$\frac{\partial\mathbf{u}''}{\partial t}+(\overline{\mathbf{u}}\cdot\nabla)\mathbf{u}''+(\mathbf{u}''\cdot\nabla)\overline{\mathbf{u}}=-\overline{\alpha}\nabla p''-\alpha''\nabla\overline{p}+\nu\nabla^2\mathbf{u}'',$$

$$\text{with}\quad \overline{\alpha}=\frac{1}{\overline{\rho}}\ \text{and}\ \ \alpha''=-\frac{1}{\kappa}\frac{p''}{\overline{p}}\frac{1}{\overline{\rho}}, \tag{12}$$

$$\frac{\partial p''}{\partial t}+\overline{\mathbf{u}}\cdot\nabla p''+\mathbf{u}''\cdot\nabla\overline{p}=-\kappa\overline{p}\nabla\cdot\mathbf{u}''-\kappa p''\nabla\cdot\overline{\mathbf{u}}. \tag{13}$$

The last term on the right side of Eq. (12), which is a diffusion term, was added in order to simulate the effect of atmospheric absorption. It can be understood as an artificial viscosity which takes into consideration classical absorption as well as molecular absorption. The diffusion coefficient ν is empirically determined as a function of temperature and relative humidity such that the frequency-dependent attenuation due to air absorption is simulated according to atmospheric attenuation coefficients given by ISO 9613-1.[21] These attenuation coefficients cannot be used directly because the length of the propagation path (including refraction and reflections) is not given explicitly by the simulation.

The prognostic model equations are numerically solved on an orthogonal staggered grid. The numerical scheme conforms to that of the flow model except that the explicit forward-in-time scheme is also used for the diffusion term. The spatial distribution of $\overline{\mathbf{u}}$, $\overline{p}$, and $\overline{\rho}$ is taken from the results of the flow model. The values of $\mathbf{u}''$ and p'' are initially set to zero.

Sound waves are triggered by a prescribed harmonic oscillation of the pressure p'' at the location of the source

$$p''_{\text{source}}(t)=\sum_{i=1}^{n_f}(p''_i\cos[2\pi f_i t]), \tag{14}$$

with p''_i being the pressure amplitude at frequency f_i and n_f the total number of frequencies. In order to better represent the directivity of a sound source, grid points neighboring to the source can be defined to be solid.

In order to avoid spurious reflections from the lateral and top boundaries an impedance boundary condition is applied. The relation between particle velocity and sound pressure is given by

$$|\mathbf{u}''|=\frac{p''}{Z_a}, \tag{15}$$

where $Z_a=\overline{\rho c}$ denotes the impedance of air. At the boundaries, only the perpendicular component of the particle velocity has to be determined. For example, the component u'' (note: $\mathbf{u}''=u''\mathbf{i}+v''\mathbf{j}+w''\mathbf{k}$) is calculated at the lateral boundaries in the x direction by

$$u''=\sqrt{\left(\frac{p''}{Z_a}\right)^2-v''^2-w''^2}. \tag{16}$$

III. VALIDATION

The quality of the LE sound propagation model has been proved by comparisons of model results with analytical solutions or experimental data from scale model measurements. The comparisons focus on the reproduction of ground characteristics, refraction due to wind and temperature gra-

TABLE I. Cases chosen for validation of the LE sound propagation model. The grid spacing Δx ranges between $\frac{1}{10}$–$\frac{1}{12}$ λ, except for the reproduction of ground characteristics ($\frac{1}{24}$ λ). The time step amounts to $\Delta t \approx 0.45 \cdot \Delta x/c$. With multi-frequency runs, grid spacing and time steps were adapted to the center or the maximum frequency.

Scenario	Obstacle	Ground	Wind	Temperature (°C)	Frequency (range) (Hz)	Validation data	Grid spacing Δx (m)	Time step Δt (μs)
Ground characteristics	none	rigid	none	15	100	analytic results	0.14	200
	none	grass-covered	none	15	100	analytic results	0.14	200
Wind and temperature gradients	none	rigid	gradient	gradient	42–990	measured data	0.03	35
	none	rigid	gradient	20	112–990	measured data	0.03	35
Shading by screen	screen	cotton and felt	none	15	120–1260	measured data	0.028	37.2
Influence of wind on shading effect	screen	rigid	gradient	15	185–375	measured data	0.10	140
	screen	rigid	gradient	15	370–750	measured data	0.05	70
	screen	rigid	gradient	15	740–1500	measured data	0.025	35

dients, shading effect of obstacles, and the influence of wind on the shading effect. These scenarios have been chosen to validate the LE model especially with respect to short- and medium-range propagation including topographic and atmospheric influences. An overview of the considered scenarios is given in Table I. The details are discussed in the following subsections. In case of a comparison with scale model measurements, all numbers given refer to the full-scale context.

The investigated applications are restricted to frequencies lower than 1.5 kHz and ranges shorter than 110 m. In these cases, atmospheric absorption remains below 1 dB (according to ISO 9613-1,[21] temperature 20 °C, relative humidity 70%). Therefore, absorption is neglected and the diffusion coefficient ν was consequently set to zero. Turbulent scattering of sound waves is not considered here. All presented calculations were carried out in two dimensions, i.e., the x-z plane.

The grid spacing Δx used in the LE model usually ranges between $\frac{1}{10}$ and $\frac{1}{12}$ of the considered wave length λ. In the case of ground of finite impedance, however, the ground layer (see Sec. II A) required a minimum spatial resolution. Thus, for low frequencies a grid spacing smaller than $\frac{1}{12}$ λ has to be used (e.g., $\frac{1}{24}$ λ with $f = 100$ Hz). The time step of the numerical integration was set to $\Delta t \approx 0.45 \cdot \Delta x/c$. In the case of multi-frequency runs [i.e., $n_f > 1$ in Eq. (14)], grid spacing and time steps were adapted to the center or the maximum frequency. The respective values of Δx and Δt are given in Table I.

A. Reproduction of ground characteristics

The reproduction of specific ground characteristics was investigated on the basis of reflection from a rigid, i.e., totally reflecting, surface and a representation of grass-covered ground. To be able to attribute effects on the sound field unambiguously to the ground response, the calculations were carried out without obstacles and atmospheric influences. This configuration allows us to compare the numerical calculations with analytical solutions based on the Weyl–van der Pol equation (a general description is given by Sutherland and Daigle,[22] the specific implementation of the equation is described in a previous paper[23]).

Rigid ground was reproduced simply by setting the vertical component of the particle velocity at the ground surface to zero. The presented sound field above rigid ground ($f = 100$ Hz) shows the interference pattern due to the superposition of direct and reflected waves (Fig. 1). The numerical results agree well with the analytic solution. Over a large range the deviations are lower than 1 dB (Fig. 1). Near the interference minima, the deviations are higher since their positions are slightly shifted.

The reproduction of ground of complex, finite impedance is not straightforward within the LE model. Since the calculations are based on real numbers, a complex impedance cannot be explicitly prescribed. Instead, a ground layer of specific acoustic properties is required. The density ρ_g of the ground layer is set to be higher than the density of the air $\overline{\rho}$. This corresponds to a lower sound speed within the layer. Additionally, the wave is attenuated exponentially with depth within the layer. Every nth grid-column was set to be rigid, in order to enforce local reaction. By varying the parameters ρ_g, n, and layer thickness, the ground response can be tuned to reproduce complex ground impedances.

As a reference for the characteristics of grass-covered ground, the best fit of a two-parameter impedance model[24] to measured data[23] was taken. The best fit was found for an

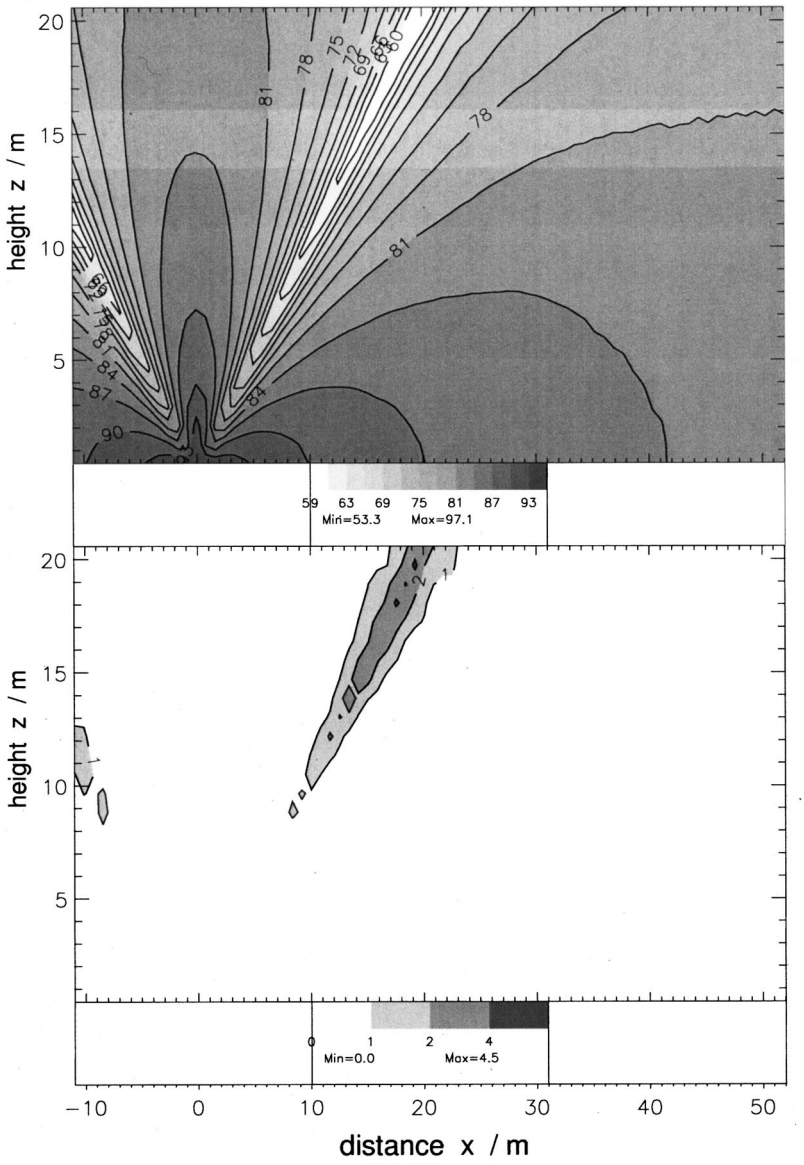


FIG. 1. Top: Sound pressure level above rigid ground (dB *re* 20 μPa, $f=100$ Hz, source at $x=0$ m, $z=1.2$ m) calculated by the LE model. Bottom: Absolute deviation of the calculated sound field from the corresponding analytic solution in dB. The vertical scales are stretched by a factor of 2.

effective flow resistivity $\sigma_e=100$ kPa s m^{-2} and an effective decrease of porosity with depth $\alpha_e=9$ m^{-1}. The set of parameters ρ_g, n, and layer thickness which reproduces such a ground the best was determined empirically by a comparison with the analytic solution.

The interference pattern of the sound field above grass-covered ground ($f=100$ Hz, Fig. 2) is less pronounced compared to the case of total reflection. The relatively high sound pressure level close to the ground is caused by the ground wave. The ground wave appears especially at frequencies below approximately 200 Hz (for the considered ground and geometry). A surface wave, which also raises the sound pressure at low frequencies close to the ground, appears usually only at larger distances.[2] According to an estimation [Eq. (17) in Sutherland and Daigle[22]], for the given ground impedance and source height surface waves should arise at distances larger than 150 m. The deviations from the analytic solution are again less than 1 dB over a large range (Fig. 2). In the area above the sound source and around the interference minima larger deviations appear. However, sound propagation simulations normally aim at the far field up to 15 m above the ground. In this area, the numerical and the analytic solution agree almost as well as in the case with total reflection.

B. Wind and temperature gradients

Atmospheric inhomogeneities were considered in the form of two different configurations including wind and temperature gradients above rigid, plane ground. The results of the calculations were compared with data from scale model measurements which were carried out by Gabillet *et al.*[25] Additionally, Gabillet *et al.* provide Gaussian beam solutions.

Since the experimental data were available for a whole frequency range, multi-frequency runs were carried out in order to compare frequency spectra. The frequencies ranged from $f=42$ to 990 Hz in steps of 14 Hz ($n_f=69$) in the first case and from $f=112$ to 990 Hz (steps: 14 Hz, $n_f=64$) in the second case. The frequency spectra were determined

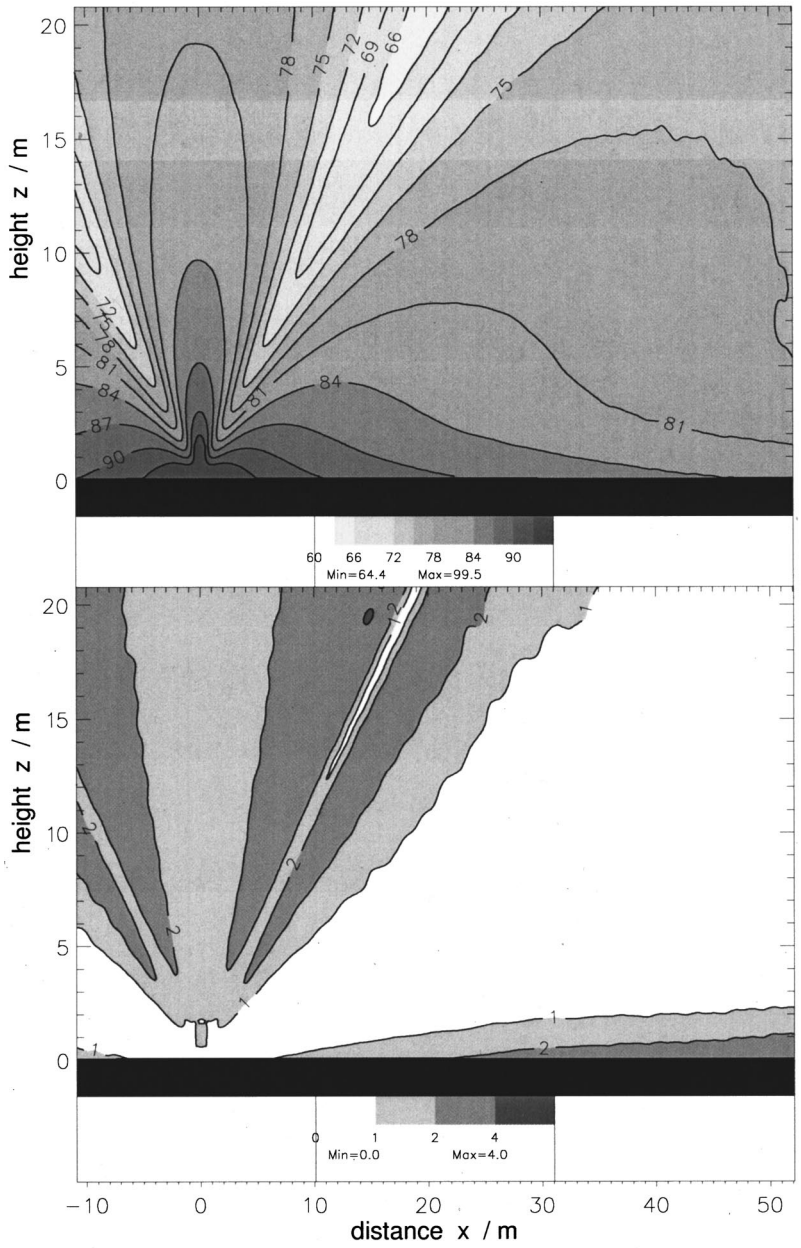


FIG. 2. Top: Sound pressure level above grass-covered ground (dB *re* 20 μPa, $f=100$ Hz, source at $x=0$ m, $z=1.10$ m, ground layer thickness$=1.68$ m, calculated by LE model). Bottom: Absolute deviation of the calculated sound field from the analytic solution (two-parameter impedance model, $\sigma_e=100$ kPa s m^{-2}, $\alpha_e=9$ m^{-1}) in dB.

from the time domain data at the receiver position by means of a fast Fourier transform. Source and receiver positions are given below.

The first case considers downwind propagation including a temperature gradient. Wind speed (in m/s) and temperature (in °C) profiles above $z=0.4$ m are given by Eq. (13) in Ref. 25:

$$u(z)=1.826\,\lg\left(\frac{z}{m}\right)+4.16, \quad T(z)=0.744\,\lg\left(\frac{z}{m}\right)+20 \qquad (17)$$

(lg denotes the logarithm to base 10). Below $z=0.4$ m u and T are extrapolated linearly down to the ground by means of the gradient at $z=0.4$ m. Source and receiver are located 2 and 5 m above the ground, respectively. The separation distance amounts to 100 m. The LE model result is compared with the measured data and a Gaussian beam solution (Fig. 3). The minimum sound pressure level of the LE result appears at the same frequency as measured, the level itself,

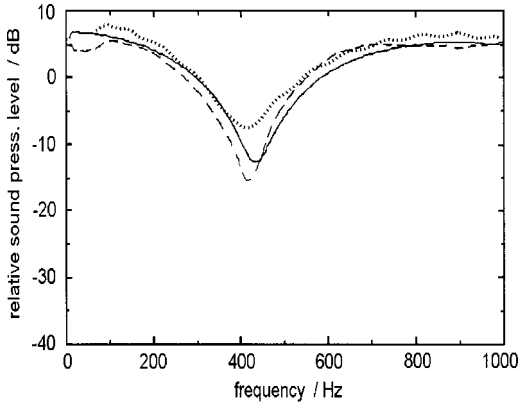


FIG. 3. Frequency spectra of downwind propagation including a temperature gradient above rigid, plane ground (sound pressure level *re* free field). $h_s=2$ m, $h_r=5$ m, $d=100$ m. Dashed line: measured data, dotted line: LE result, solid line: Gaussian beam solution. Gaussian beam solution, measured data, and graph adapted from Ref. 25, Fig. 11(a).

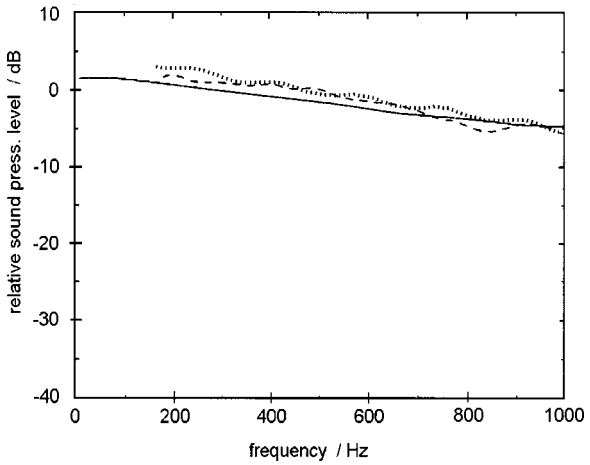


FIG. 4. Frequency spectra of upwind propagation above rigid, plane ground (no temperature gradient, sound pressure level *re* free field). $h_s = h_r = 1$ m, $d = 80$ m. Dashed line: measured data, dotted line: LE result, solid line: Gaussian beam solution. Gaussian beam solution, measured data, and graph adapted from Ref. 25, Fig. 13.

however, is overestimated. Over the whole spectrum, theagreement of the LE result with the measured data, and the Gaussian beam solution as well, is satisfactory.

The second case considers upwind sound propagation in an isotherme atmosphere. The wind speed profile (in m/s) is given by Eq. (15) in Ref. 25:

$$u(z) = -4.85\left(\frac{z}{3m}\right)^{0.14} \tag{18}$$

Source and receiver are located 1 m above the ground at a distance of 80 m. Figure 4 shows the frequency spectra of the measured data, a Gaussian beam solution, and the LE result. Between 300 and 700 Hz, the LE result agrees perfectly with the measured data. Both, LE result as well as the Gaussian beam solution, deviate from the measured data by around 1 dB, on average, over the whole frequency range.

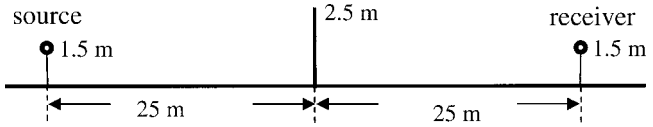


FIG. 5. Setup of the measurements of Rasmussen (full scale, not true to scale) and the corresponding calculations (see Figs. 6 and 7). The vertical line in the middle represents the totally reflecting screen.

C. Shading effect of a screen

The shading effect of a screen was first calculated for a homogeneous resting atmosphere. The results of the calculations were compared with data from a scale model measurement which was carried out by Rasmussen. A principal description of the measurement is given by Rasmussen and Arranz.[7] Geometry and ground impedance differed slightly from this description. The setup of the measurement, and thus for the calculations as well, is presented in Fig. 5.

The ground consisted of two different layers of cotton and felt of which the acoustical characteristics are described by a two-parameter impedance model.[26] The area in front of the screen (source side) is described by an effective flow resistivity of $\sigma_e = 150$ kPa s m^{-2} and an effective rate of porosity decrease with depth of $\alpha_e = 350$ m^{-1}. Behind the screen (receiver side) the values amount to $\sigma_e = 7$ kPa s m^{-2} and $\alpha_e = 125$ m^{-1}, respectively (note: these numbers refer to the full-scale frequencies).

A multi-frequency run was carried out with frequencies ranging from $f = 120$ to 1260 Hz in steps of 30 Hz ($n_f = 39$). The spectrum was calculated for the position of the receiver. However, only one set of ground parameters can be used for one run, thus, the representation of the ground impedances is optimized for only one (mean) frequency.

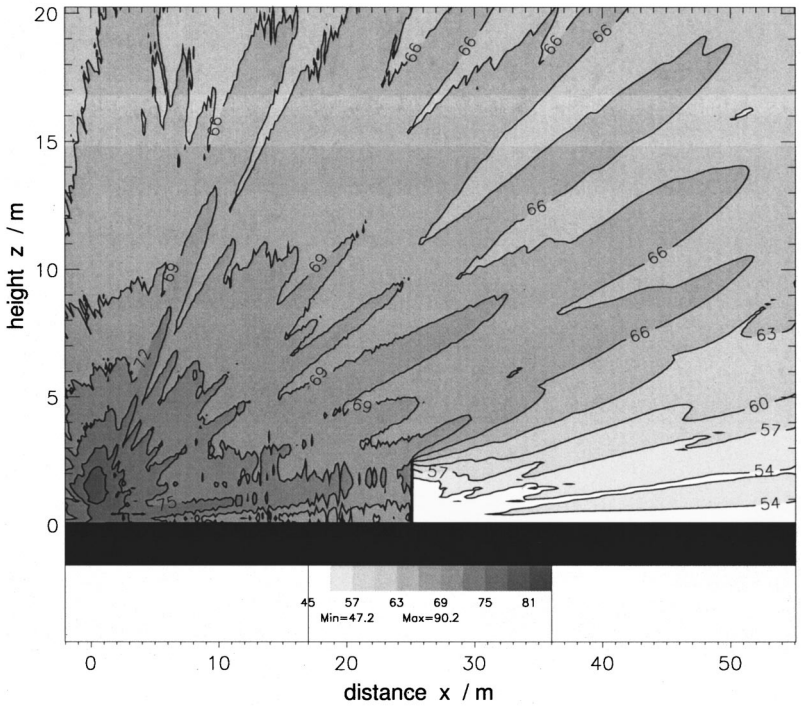


FIG. 6. Calculated sound pressure level (dB *re* 20 μPa) in the presence of a screen ($f = 120$–1260 Hz, steps: 30 Hz, source at $x = 0$ m, setup see Fig. 5, ground impedance source side: $\sigma_e = 150$ kPa s m^{-2}, $\alpha_e = 350$ m^{-1}, receiver side: $\sigma_e = 7$ kPa s m^{-2}, $\alpha_e = 125$ m^{-1}). Atmospheric influences have been excluded.

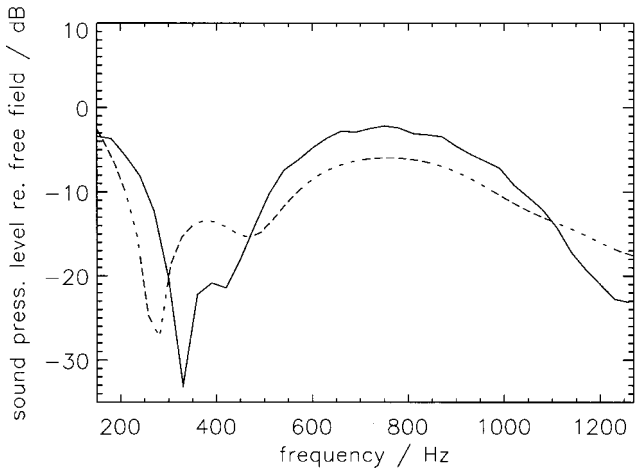


FIG. 7. Comparison of calculated (solid line) and measured (dashed line) frequency spectrum (dB *re* free field, same calculation as in Fig. 6) at the receiver position (see Fig. 5).

In the presence of a screen the shading effect is evident (Fig. 6). The remaining sound behind the screen is caused by diffraction. In front of the (totally reflecting) screen, interferences appear due to reflections. Figure 7 presents the calculated and the measured frequency spectrum, both normalized to free-field propagation. The calculated spectrum generally resembles the measured one. The double dip structure around 280 and 480 Hz is not exactly matched in terms of dip frequencies and levels. There are also positive and negative biases at higher frequencies. The deviations are probably due to the simplified representation of the ground, which was optimized for one frequency only.

D. Effect of wind on the shading effect

The last considered scenario deals with the presence of a screen in an inhomogeneous moving atmosphere. Again, the results of the LE model were compared with results from scale model measurements, in this case measurements by Salomons.[11] The measurements focused on the effect of screens on wind gradients which in turn impair the efficiency of screens. The setup is given in Fig. 8. The ground surface is assumed to be rigid and plane.

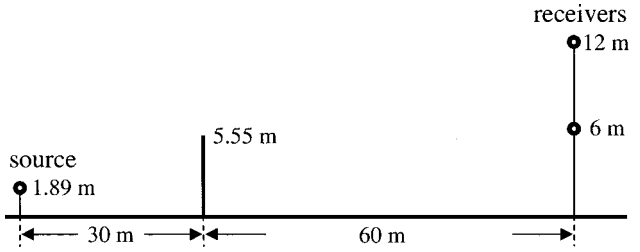


FIG. 8. Setup of the measurements of Salomons (full scale, not true to scale) and the corresponding calculations (see Figs. 9 and 10). Both receivers are not in line of sight to the source, the ground surface is rigid and plane.

The wind fields were calculated by the flow model described in Sec. I A. The initial, undisturbed wind speed profile is given by

$$u(z)=\frac{u_9}{\ln(1+z/z_0)}\ln\left(1+\frac{z_9}{z_0}\right) \tag{19}$$

with $z_9=9$ m and the roughness length $z_0=0.001$ m (see Ref. 11). The wind speed at a height of 9 m u_9 was varied from 0 to 14 m/s in steps of 2 m/s. The flow model was integrated over 30 s until a steady state was achieved.

Figure 9 shows a representation of the wind field for $u_9=10$ m/s. The deformation of the wind field due to the screen is clearly visible. Above the screen the wind gradients are much stronger than in the undisturbed case. In the wake

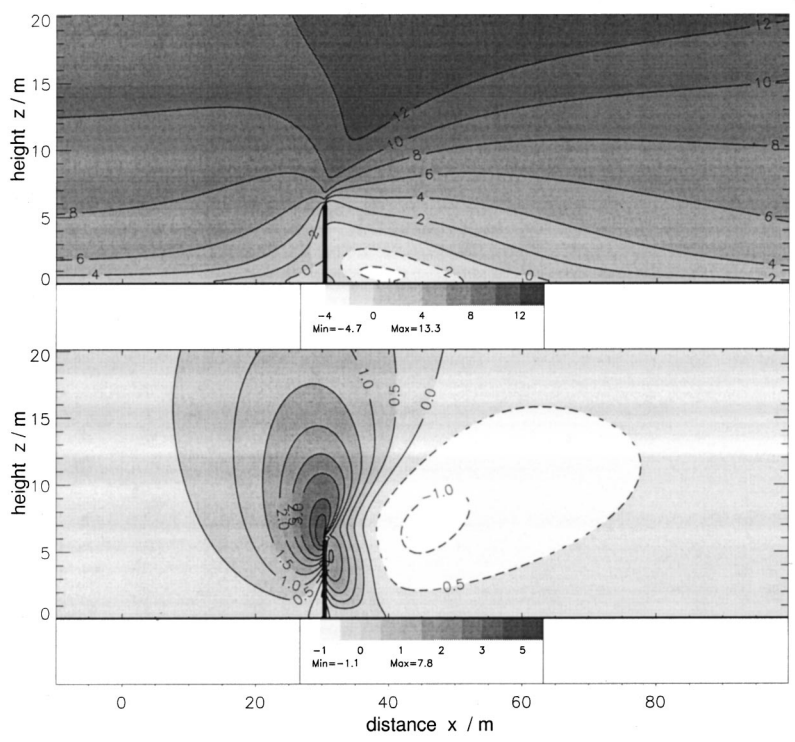


FIG. 9. Example of a wind field [$u_9=10$ m/s, see Eq. (19)], calculated by the flow model (top: horizontal component $\overline{u}$, bottom: vertical component $\overline{w}$). The combination of positive and negative (dashed lines) values of both components indicates a vortex behind the screen.

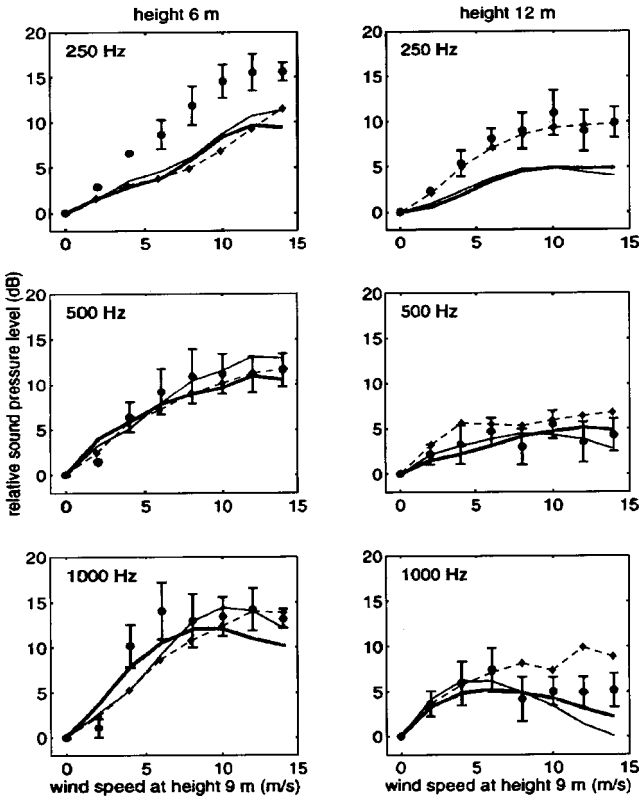


FIG. 10. Sound pressure level as a function of u_9 relative to the no-wind case ($u_9=0$) at a receiver height of 6 m (left column) and 12 m (right column) for the three different octave bands. Dashed line: LE model; other lines: PE model for different wind gradients; dots with error bars: measurements. PE model results, measurements, and graph adapted from Ref. 11.

of the screen a vortex forms as the negative values of the horizontal component indicate.

The simulations were performed for eight different values of u_9 and three different octave bands. Each octave band was represented by 20 equidistant frequencies (250-Hz band: 185 to 375 Hz, 500-Hz band: 370 to 750 Hz, 1000-Hz band: 740 to 1500 Hz). From each simulation, the sound pressure levels at the receiver positions (Fig. 8) were taken and normalized with respect to the sound pressure level of the no-wind case ($u_9=0$).

The results are presented in Fig. 10 together with the measurements of Salomons. In addition, the results of a parabolic equation (PE) model from Salomons are shown. The left column shows the relative sound pressure level at a receiver height of 6 m for the three different octave bands, the right column at a receiver height of 12 m.

At 6-m height the results of the LE model are very similar to the PE model results by Salomons. They also fit the measured data well, except for the 250-Hz case. In the latter case, both models yield similar deviations. The calculations for the receiver height of 12 m also show satisfactory results. Especially for 250 Hz, the results of the LE model agree perfectly with the measured data. For 500 Hz the relative sound pressure level is slightly too high, which means the effect of the wind is slightly overestimated. In the case of 1000 Hz, at lower wind speeds ($u_9 \leqslant 6$ m/s) the LE model agrees well with the measured data. For $u_9 \geqslant 8$ m/s the wind effect is predicted to be higher than measured.

IV. CONCLUSIONS

The LE model turns out to be a suitable, fairly accurate numerical tool for calculating sound propagation under various topographic and atmospheric conditions. Input data from meteorological mesoscale models or flow models can be used to allow for a three-dimensionally inhomogeneous atmosphere. Single- or multi-frequency runs are possible in two or three dimensions.

The comparison of model results with analytic solutions for rigid and grass-covered ground shows that such ground characteristics can be reproduced. The representation of a finite ground impedance, however, requires a costly determination of ground parameters. The calculations of sound propagation above rigid, plane ground including wind and temperature gradients result in frequency spectra which agree satisfactorily with the measured results and a Gaussian beam solution.

LE model results for shading by a screen above ground of finite impedance show reasonable agreement with measurements. The calculated frequency spectrum resembles the measured one; however, deviations are visible. A frequency-dependent representation of the ground characteristics would presumably improve the reasonable agreement with the measurements. This would require one calculation for each frequency. The calculations concerning the effect of wind on diffraction behind a screen agree well with the measured data except for the 250-Hz octave band and 6-m receiver height. In the latter case, the calculations of two different sound propagation models show the same deviations. This suggests that the deviations are not caused by errors of the algorithms.

The crucial treatment of finite ground impedance and the time-consuming calculations (e.g., NEC-SX4 single processor: around 10 000 cpu-seconds for a 4600×1200 point grid and a integration time of 2.3 s including atmospheric influences) are the two main disadvantages of the LE model. On the other hand, this model enables a simulation of sound propagation through inhomogeneous atmosphere without major approximations. The wind vector can be considered in three dimensions, diffraction behind obstacles is explicitly included, and the representation of topography is limited only by the spatial resolution of the grid. Thus, the LE model can be used as a tool for specific situations which can be considered by other models only in a simplified way. The effect of approximations and isolated atmospheric influences can be estimated and hereby specific requirements for simplified, efficient algorithms can be defined.

ACKNOWLEDGMENTS

The authors would like to thank K. B. Rasmussen for providing experimental data of his scale model measurements and answering patiently our questions about the data. Thanks go also to E. M. Salomons for helpful discussions and to E. Premat for providing information about measurement data.

[1] J. E. Piercy, T. F. W. Embleton, and L. C. Sutherland, ''Review of noise propagation in the atmosphere,'' J. Acoust. Soc. Am. **61**, 1403–1418 (1977).
[2] T. F. W. Embleton, ''Tutorial on sound propagation outdoors,'' J. Acoust. Soc. Am. **100**, 31–48 (1996).
[3] K. Attenborough, ''Review of Ground Effects on Outdoor Sound Propagation from Continuous Broadband Sources,'' Appl. Acoust. **24**, 289–319 (1988).
[4] K. Attenborough, S. Taherzadeh, H. E. Bass, X. Di, R. Raspet, G. R. Becker, A. Güdesen, A. Chrestman, G. A. Daigle, A. L'Espérance, Y. Gabillet, K. E. Gilbert, Y. L. Li, M. J. White, P. Naz, J. M. Noble, and H. A. J. M. Hoof, ''Benchmark cases for outdoor sound propagation models,'' J. Acoust. Soc. Am. **97**, 173–191 (1995).
[5] E. M. Salomons, ''Sound Propagation in Complex Outdoor Situations with a Non-Refracting Atmosphere: Model Based on Analytical Solutions for Diffraction and Reflection,'' Acust. Acta Acust. **83**, 436–454 (1997).
[6] K. M. Li, V. E. Ostashiev, and K. Attenborough, ''The Diffraction of Sound in a Stratified Moving Atmosphere Above an Impedance Plane,'' Acust. Acta Acust. **84**, 607–615 (1998).
[7] K. B. Rasmussen and M. G. Arranz, ''The insertion loss of screens under the influence of wind,'' J. Acoust. Soc. Am. **104**, 2692–2698 (1998).
[8] N. Barriere and Y. Gabillet, ''Sound Propagation Over a Barrier with Realistic Wind Gradients. Comparison of Wind Tunnel Experiments with GFPE Computations,'' Acust. Acta Acust. **85**, 325–335 (1999).
[9] D. Heimann and G. Gross, ''Coupled simulation of meteorological parameters and sound level in a narrow valley,'' Appl. Acoust. **56**, 73–100 (1999).
[10] L. R. Hole and H. M. Mohr, ''Modeling of sound propagation in the atmospheric boundary layer: Application of the MIUU mesoscale model,'' J. Geophys. Res. D **104**, 11891–11901 (1999).
[11] E. M. Salomons, ''Reduction of the performance of a noise screen due to screen-induced wind-speed gradients. Numerical computations and wind-tunnel experiments,'' J. Acoust. Soc. Am. **105**, 2287–2293 (1999).
[12] M. West and Y. Lam, ''Prediction of sound fields in the presence of terrain features which produce a range dependent meteorology using the generalised terrain parabolic equation (GT-PE) model,'' Proc. Internoise, 2/943 (2000).
[13] E. M. Salomons and K. B. Rasmussen, ''Numerical computation of sound propagation over a noise screen based on an analytic approximation of the wind speed field,'' Appl. Acoust. **60**, 327–341 (2000).
[14] T. Yamada, ''Simulations of nocturnal drainage flows by a q^2l turbulence closure model,'' J. Atmos. Sci. **40**, 91–106 (1983).
[15] A. K. Blackadar, ''The vertical distribution of wind and turbulent exchange in a neutral atmosphere,'' J. Geophys. Res. **67**, 3095–3102 (1962).
[16] See, for example, R. B. Stull, *An Introduction to Boundary Layer Meteorology* (Kluwer Academic, Dordrecht, 1988), Chap. 6.
[17] A. Arakawa and V. R. Lamb, ''Computational design of the basic dynamic process of the UCLA general circulation model,'' Methods Comput. Phys. **17**, 173–265 (1977).
[18] D. R. Durran, *Numerical Methods for Wave Equations in Geophysical Fluid Dynamics* (Springer, New York, 1999), Chap. 2.2.3.
[19] N. R. Wilson and R. H. Shaw, ''A higher order closure model for canopy flow,'' J. Appl. Meteorol. **16**, 1197–1205 (1977).
[20] A principal description of the technique is given by H. C. Davies and R. H. Turner, ''Updating prediction models by dynamical relaxation: an examination of the technique,'' Q. J. R. Meteorol. Soc. **103**, 225–245 (1977).
[21] International Standards Organization, ''Acoustics-Attenuation during propagation outdoors—Part I: Method of Calculation of the Attenuation of Sound by Atmospheric Absorption,'' ISO/DIS 9613-1, September 1990.
[22] L. C. Sutherland and G. A. Daigle, ''Atmospheric sound propagation,'' in *Encyclopedia of Acoustics*, edited by M. J. Crocker (Wiley, New York, 1997), Sec. 4.3, Vol. I, Chap. 32, pp. 341–365.
[23] R. Blumrich and J. Altmann, ''Ground Impedance Measurement Over a Range of 20 m,'' Acust. Acta Acust. **85**, 691–700 (1999).
[24] K. Attenborough, ''Acoustical Impedance Models for Outdoor Ground Surfaces,'' J. Sound Vib. **99**, 521–544 (1985), Eq. (31).
[25] Y. Gabillet, H. Schroeder, G. A. Daigle, and A. L'Esprance, ''Application of the Gaussian beam approach to sound propagation in the atmosphere: Theory and experiments,'' J. Acoust. Soc. Am. **93**, 3105–3116 (1993).
[26] K. Attenborough, ''Ground parameter information for propagation modeling,'' J. Acoust. Soc. Am. **92**, 418–427 (1992), Eq. (12).

Breaking wind waves as a source of ambient noise

Pavlo Tkalich and Eng Soon Chan
Tropical Marine Science Institute, The National University of Singapore, Singapore 119223

(Received 9 August 2000; revised 22 March 2002; accepted 7 May 2002)

A theoretical model for the prediction of ambient noise level due to collective oscillations of air bubbles under breaking wind waves is presented. The model uses a budget of the energy flux from the breaking waves to quantify acoustic power radiation by a bubble cloud. A shift of the noise spectra to lower frequency due to collective bubble oscillation is assumed. The model derives good estimates of the magnitude, slope, and frequency range of the noise spectra using the wind speed or height of breaking waves. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1489436]

PACS numbers: 43.28.Ra, 43.30.Nb, 43.50.Ed, 43.50.Rq [SAC-B]

I. INTRODUCTION

The classification of underwater sound sources requires a detailed knowledge of the ambient noise structure under different hydro-meteorological conditions. Since Knudsen *et al.*'s (1948) empirical study, showing the dependence of ambient noise on the wave height and wind velocity, several approaches have been made to develop qualitative and quantitative descriptions of the phenomenon. One of the earliest attempts to link the Knudsen spectrum and wave action was made by Marsh (1963). Using the pressure spectrum of the Neumann–Pierson sea-surface spectra, the noise intensity (I) was evaluated as a function of wave height (H) and acoustic frequency (f), i.e., $I \sim H^{6/5} f^{-5/3}$. Kerman (1984), combining various aspects of breaking-wave and air-bubble dynamics, developed a simple algebraic formula for calculation of the acoustic noise generated by a single bubble cloud subjected to wind shear. In terms of wind speed (U) the derived relationship, $I \sim U^{1.9}$, agrees well with observations for large U. The general concept that wind is indirectly responsible for the ambient noise generation (via breaking wind waves), was supported by the experiments of Medwin and Beaky (1989), who correlated laboratory acoustic spectral slope in the absence of wind to the Knudsen spectra at sea. Loewen and Melville (1991a) experimentally found that the fraction of dissipated wave energy radiated as sound is of the order 10^{-8}.

There are two major points of view on the type of air-bubble oscillation in a bubble cloud under breaking waves—individual (Medwin and Daniel, 1990; Loewen and Melville, 1991b) and collective (Carey and Bradley, 1985; Prosperetti, 1985). The difference lies in answering the question as to whether the amount of large-size (a few mm) bubbles is enough to produce the observed low-frequency noise within the individual mode, or whether the introduction of the collective mode is required. Exploring the idea on collective oscillations, Prosperetti (1988) and Carey *et al.* (1993) quantitatively estimated the source level from a bubble cloud near the ocean surface to obtain observed values. Oguz (1994) applied a series solution to the Helmholtz's equation to investigate the influence of bubble clouds and whitecaps on low-frequency ambient noise. In the model, the acoustic radiation from an individual bubble cloud was assumed to be independent of wind speed, and the observed increase in sound level with wind speed was explained by the increase in the number of whitecaps. Though the model was successful in explaining the noise due to distribution of breaking waves, it did not predict spectra from individual waves that are found experimentally. Means and Heitmeyer (2001) proposed an alternative model of sound generation by the resonant bubble cloud beneath the individual wave. While the assumption of collective bubble oscillation appears to be able to predict the main features of ambient noise, the alternative approach also seems appropriate, as claimed by Medwin and Daniel (1990) and Loewen and Melville (1991b). Although Crum (1995) and Kolaini *et al.* (1995) suggested that, in a transition from a less violent to more violent type of wave breaking the bubbles may change their oscillation mode from individual to collective, quantitatively, this transition has yet to be understood. Since both mechanisms contribute to the sound spectra independently, the phenomena can be studied separately.

In this paper, the hypothesis of the collective oscillation of air bubbles under breaking waves is examined, and a theoretical model of ambient noise generation is developed. Following available experimental evidence and theoretical considerations, we assume that wind is the source of the Knudsen spectra only through the mediation of breaking waves, which are themselves the sources of the sound-radiating bubbles. We also assume that the shifting of ambient noise to lower frequencies is due to the collective oscillation of the air bubbles in the cloud. Although many of the dominant physical phenomena are parametrized, most of the empirical parameters can be found in publications. The model predicts the intensity spectra using wind speed or height of the respective breaking waves as an environmental input. The model is able to explain the main observed features of ambient nose, including the magnitude, slope, and frequency range of the spectra. Since the model characterizes sound spectra in the source vicinity, other important for sound propagation parameters, such as proximity of seafloor or anisotropic ocean environment, are out of this paper's range.

II. AIR-BUBBLE POPULATION

A. Bubble spectra

According to Wu (1988), an empirical shape of air-bubble size spectra is defined by $\hat{n}(b) \sim b^{-s}$, where $\hat{n}$ is the number of the bubbles per unit water volume, b is the bubble radius, and s is the slope of the spectra. Farmer and Vagle (1989) reported the spectral slope of about 4 to 6. According to Baldy (1988), the slope increases from the value of 2.5 to 4 at a depth of 0.5 to 1 times the significant wave height, and then remains constant at a value of 4 for the deeper part of the water column. Data of Deane (1997) show a transition from $b^{-2.5}$ for radii less than 1 mm to $b^{-4.5}$ for the greater radii. Lamarre and Melville (1994) attributed the discrepancy to the variation in environmental conditions, sensitivity of integration technique, and sound scattering due to the water surface. The size spectra have a well-defined peak at about 30–50 μm (Baldy, 1988), and the size range of the bubbles depends on wind and wave conditions, showing an averaged magnitude $(b_{\min}, b_{\max}) \approx (30\ \mu\text{m}, 1000\ \mu\text{m})$ (Wu, 1988). Medwin and Daniel (1990) found the spectra peak at 150 μm and traced a few bubbles as large as 7.4 mm. However, Prosperetti (1988) doubted that the amount of such large bubbles was sufficient to make an impact on the ambient noise. To study exclusive influence of collective bubble oscillations on sound generation, and to minimize the possible effect of individual oscillations, only radii less than 1 mm are considered in this paper. Generalizing experimental evidence, the bubble-size spectra (number of bubbles per unit water volume) in the bubble cloud near the water surface can be expressed by

$$\hat{n}(b) = n_0 / b^s, \tag{1}$$

where n_0 is the normalization function. Using Eq. (1), the void (air volume) fraction spectra can be defined as

$$\hat{\beta}(b) = \frac{4\pi}{3} \frac{n_0}{b^{s-3}}. \tag{2}$$

Taking into account that $\nu_{\max}$ and $\nu_{\min}$ are the volumes of the maximal and minimal bubbles from the spectra, and ν $(=4\pi b^3/3)$ is the volume of a single bubble, one obtains the total void fraction B as

$$B = \int_{\nu_{\min}}^{\nu_{\max}} \hat{n}(b)\, d\nu = \frac{4\pi}{3} \frac{n_0}{B_{s-3}}, \quad (s \neq 3), \tag{3}$$

where

$$B_i = \frac{i}{3} \frac{b_{\max}^i b_{\min}^i}{b_{\max}^i - b_{\min}^i}. \tag{4}$$

The normalization function n_0 can be found from Eq. (3) as

$$n_0 = \frac{3}{4\pi} B B_{s-3}. \tag{5}$$

Using Eq. (5), the normalized void fraction spectra can be rewritten as

$$\frac{\hat{\beta}(b)}{B} = \frac{B_{s-3}}{b^{s-3}}, \tag{6}$$

and the mean bubble radius of the entire bubble population becomes

$$b_m = \frac{1}{n_{\max} - n_{\min}} \int_{n_{\min}}^{n_{\max}} b\, d\bar{n} = \frac{B_s}{B_{s-1}}, \tag{7}$$

where $n_{\max}$ and $n_{\min}$ are the amount of the maximal and minimal bubbles from the spectra.

B. Bubble cloud characteristics

To estimate the characteristics of a bubble cloud generated by a breaking wave, one can evaluate the potential energy exchange in the ''breaking wave–bubble cloud'' system. According to Komen *et al.* (1994), the mean wave energy per unit of horizontal area is

$$E_0 = \rho g M_0, \tag{8}$$

where g is the gravity acceleration, ρ is the water density, and M_0 is the zero-order moment of the wave spectrum, related to significant wave height H as

$$M_0 = H^2/16. \tag{9}$$

According to Lamarre and Melville (1991), during wave breaking as high as 30% to 50% of the total amount of dissipated wave energy $(\kappa_d E_0)$ is expended to entrain the bubbles against buoyancy E_b, that is

$$E_b \approx \kappa_b \kappa_d S E_0, \tag{10}$$

where S is the characteristic horizontal cross section of the bubble cloud, $\kappa_b \sim 0.3$–0.5 is the fraction of wave energy expanded for bubble cloud entrainment, and κ_d is the fraction of dissipated wave energy. The energy due to buoyancy of the bubble cloud, occupying a volume V within the water column, is given by

$$E_b \approx \rho g\, \delta V B, \tag{11}$$

where δ is the characteristic depth (of center mass) of the cloud.

We assume that the bubble cloud has a spherical shape due to the following reasons. First, it was indicated (Loewen, 1991) that the shape of the bubble cloud has very little effect on the prediction of the resonant frequency at the lowest mode. Second, since the maximum sound level occurs in the middle of the bubble cloud in spite of a location of applied pressure (Omta, 1987; Oguz, 1994), the actual shape of boundaries, consisting of sparse bubbles, presumably is not very important. Last, there is little information available on the bubble cloud geometry dynamics. Taking into account Eqs. (8), (9), and (11), one can solve Eq. (10) to determine the radius of the bubble cloud, R, yielding

$$R = \kappa H^2/(B\delta), \tag{12}$$

where $\kappa = (3/64)\kappa_b \kappa_d$.

III. BUBBLE ACOUSTICS

A. Frequency spectra of bubble cloud oscillation

The lowest eigenfrequency of a single bubble oscillation is given by Minnaert (1933) as

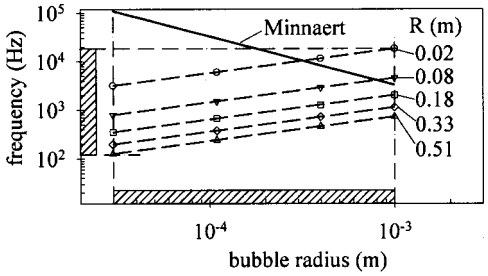


FIG. 1. Frequency spectra for bubble oscillation in a single mode (solid line) and in a collective mode within the cloud of the radius R (dash-symbol lines). $B=0.01$, $s=4$.

$$f_b=\frac{1}{2\pi b}\sqrt{\frac{3\gamma P_0}{\rho}}, \tag{13}$$

where γ is the polytropic index (1.4 for air), P_0 is the ambient pressure. According to the Minnaert formula, a bubble having a radius within the Wu's (1988) interval (30 μm, 1000 μm), must have a natural frequency within 3–100 kHz (see Fig. 1). However, this contradicts available data on the frequency range. By contrast, Medwin and Daniel's (1990) bubble radii range (150 μm, 7400 μm) gives reasonable frequencies, consistent with available data. Considering that the few bubbles with radii greater than 1 mm cannot affect ambient noise significantly, Carey and Bradley (1985) and Prosperetti (1985) suggested that the bubble cloud is capable of pulsating according to collective modes. Using the bubble mixture theory, Omta (1987), Prosperetti *et al.* (1993), and Carey *et al.* (1993) found that the frequency spectra of the emitted sound from the mixture has a peak at a frequency much lower than the eigenfrequency of an individual bubble. The expression for the lowest eigenfrequency of the cloud is given by Carey *et al.* (1993) as

$$f\approx\frac{1}{2\pi R}\sqrt{\frac{3\gamma P_0}{\beta\rho}}, \tag{14}$$

where $\gamma=1$.

It is expected that within the bubble cloud, a subgroup of bubbles with similar radius b and total void fraction $\hat{\beta}(b)$ may have its own resonant oscillation frequency $\hat{f}(b)$. Therefore, a subset of the rest of the bubbles from the cloud [a complement of $\hat{\beta}(b)$], having void fraction $\hat{\beta}_r$, will affect the mixture density according to $\hat{\rho}=(1-\hat{\beta}_r)\rho+\hat{\beta}_r\rho_a$, where ρ_a is air density. Similar to Eq. (14), one obtains from the mixture theory that

$$\hat{f}(b)\approx\frac{1}{2\pi R}\sqrt{\frac{3\gamma P_0}{\hat{\beta}\hat{\rho}}}. \tag{15}$$

Equation (15), within $O(\hat{\beta}_r)$ order of accuracy, can be rewritten as

$$\hat{f}^2(b)=f_0^2 B/\hat{\beta}(b), \tag{16}$$

where

$$f_0\approx\frac{1}{2\pi R}\sqrt{\frac{3\gamma P_0}{B\rho}}. \tag{17}$$

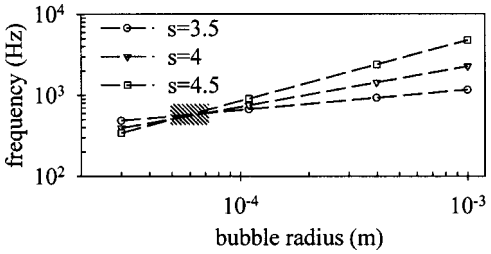


FIG. 2. Computed frequency spectra for bubble oscillation in a collective mode for different slopes of the bubble size spectra. $B=0.05$, $R=0.074$.

One may notice that the value f_0, given by the relationship (17), is a combination of several empirical parameters γ, P_0, R, and B. Even though the range of the parameters is discussed extensively in the literature (Kerman, 1984; Wu, 1988; Deane, 1997), there are no reliable theoretical formulas yet to predict them exactly. Therefore, if there are no available empirical data for the typical wave-breaking scenarios, the quantity f_0 should be considered as the reference frequency.

Equations (6) and (16) predict that for the given void fraction B, as the slope of bubble size spectra increase, so does the range of number of bubbles in the spectra and, correspondingly, the range of bubble oscillation frequencies (as in Fig. 2). Substituting the following parameters—void fraction $B=0.01$, bubble size spectra slope $s=4$, and bubble size range 30–1000 μm in Eq. (16), the frequency spectrum is plotted in Fig. 1 for different radii of the bubble cloud. The plots depict a gradual transition from relatively high-frequency acoustic emissions of a gently spilling breaker (smaller R) to relatively low-frequency emissions of plunging breakers (larger R). All the frequencies are within the observed range—from 100 Hz to 20 kHz.

In the real stochastic ocean with multiple breaking events and a variety of bubble clouds having different bubble size spectra, a well-defined local maximum of sound spectra level is observed in the vicinity of 0.3–0.5 kHz (Kerman, 1984). The predicted frequency spectra for different slopes of the bubble size spectra (Fig. 2) shows that all the curves cross in a small area at the low-frequency range. If a summation of the acoustic energy of the clouds is assumed, the area is a likely candidate for the noise spectra maximum.

B. Acoustic intensity of bubble cloud

We introduce a polar coordinate system with an origin at the center of the spherical bubble cloud. Since the radius of the cloud is much smaller than the usual distance from the cloud to the hydrophone, the acoustic source from the group of the bubbles can be projected into the origin of the radial coordinate, r. The acoustic power, radiated by a single bubble from the cloud at a distance r ($\gg R$) from the bubble center, is given by the generalized relationship (Kinsler *et al.*, 1982)

$$\Pi_b(b)\approx 4\pi b^2 I_b(\hat{D}(b)b/r)^2.$$

Here, $I_b=p_b^2/2\rho c$ is the acoustic intensity of a single bubble, p_b is the pressure amplitude of radiated acoustic wave, c is the speed of sound in water, and D is the damping coeffi-

cient. Kerman (1984) assumed that the resulting power of the cloud originates from an individual bubble, b, weighted by its number of occurrences, $\hat{n}(b)$. However, not all the bubbles in the cloud equally radiate acoustic energy—the maximum sound level occurs in the middle of the bubble cloud (Omta, 1987; Oguz, 1994). Therefore, it is assumed that only the acoustically most active bubbles, located within a sphere of radius χR ($\chi \leqslant 1$) in the middle of the bubble cloud, are contributing to the acoustic power of the cloud, i.e.,

$$\hat{\Pi}(b) = \tfrac{4}{3}\pi(\chi R)^3 \hat{n}(b)\Pi_b(b). \tag{18}$$

The acoustic intensity of the cloud, $\hat{I}(b) = \hat{\Pi}(b)/4\pi R^2$, with respect to Eqs. (1), (2), and (18), yields

$$\hat{I}(b) \approx \chi^3 I_b \frac{R b_m}{r^2} \hat{D}^2(b)\hat{\beta}(b). \tag{19}$$

Equation (19) is an important result of the paper, showing that acoustic spectra of the bubble cloud are proportional to the cloud radius, void fraction spectra, and the damping coefficient. Sound damping in the bubble cloud occurs mainly due to the combination of scattering and absorption cross sections of the bubbles, of which the scattering cross section is dominant (d'Agostino and Brennen, 1988). It is assumed that in the cloud, while one subgroup of the bubbles is oscillating with the collective frequency, $\hat{f}(b)$, the complementary bubbles are damping the acoustic energy of the subgroup. Thus, the damping coefficient, $\hat{D}(b)$, is different for each of the subgroup, i.e.,

$$\hat{D}(b) = \exp(-\hat{A}(b)R). \tag{20}$$

Here, the attenuation coefficient, $\hat{A}(b)$, is calculated as the total scattering cross section of the complementary bubbles

$$\hat{A}(b) = 0.5\left(\int_{\sigma_{\min}}^{\sigma_{\max}} \hat{n}(b)\, d\overline{\sigma} - \hat{n}(b)\sigma \right),$$

where σ and $\overline{\sigma}$ are the scattering cross sections of a single bubble; $\sigma_{\max}$ and $\sigma_{\min}$ are the scattering cross sections of the maximal and minimal bubbles from the spectra. Taking into account that the scattering cross section is $\sigma = 4\pi b^2 \varepsilon$, the attenuation coefficient is simplified to

$$\hat{A}(b) = 0.5\varepsilon B B_{s-3}\hat{a}(b), \tag{21}$$

where $\varepsilon \sim (f/f_b)^4$ for the frequencies f well below the individual resonant frequency f_b (Kinsler *et al.*, 1982), and

$$\hat{a}(b) = \left(\frac{2}{B_{s-2}} - \frac{3}{b^{s-2}} \right). \tag{22}$$

C. Characteristic depth of the bubble cloud

The characteristic depth of the bubble cloud, δ, can be eliminated from Eqs. (19) and (22) using an empirical hypothesis. There are few reliable data on δ; however, several empirical relationships are available for the bubble entrainment depth, Δ ($\Delta \approx 2\delta$). Some researchers scale Δ with a wind speed as following:

$$\Delta[m] = \begin{cases} 0.4(U-2.5), & \text{Thorpe (1982);} \\ \begin{bmatrix} 0.4, & U<7 \text{ m/s} \\ 0.4+0.12(U-7), & 7 \leqslant U \leqslant 12 \text{ m/s} \end{bmatrix}, & \text{Wu (1988);} \\ 0.42\exp(0.16U), & \text{Plueddemann } et\ al. \text{ (1996).} \end{cases}$$

An alternative approach is to use the significant wave height for the scaling. Donelan (1978) reported that the immediate mixing of surface waters due to turbulence, produced by a whitecap, penetrates to depths of the order of the wave height. Hwang *et al.* (1990), analyzing most available literature data, scaled the vertical entrainment of bubbles with a significant height of waves according to the relationship $\Delta = \alpha H$, where α is order of unity and appears to decrease with fetch. Terrill and Melville (1997), using sound-speed measurements in field conditions, suggested the expression $\Delta = 0.31 + 0.26H$. The data discrepancy can be attributed to the difference in measurement techniques, as well as to the different types of wave breaking. For instance, whitecapping as a less violent type of wave breaking cannot force bubbles down into the water column as efficiently as a plunging breaker. A comparison of the different empirical hypotheses is presented in Fig. 3. Since no detailed wave information was provided by some of the authors, we estimated the significant wave height from wind-speed measurements using the "fully developed sea" condition (Neumann and Pierson, 1966)

$$H \approx 0.2U^2/g. \tag{23}$$

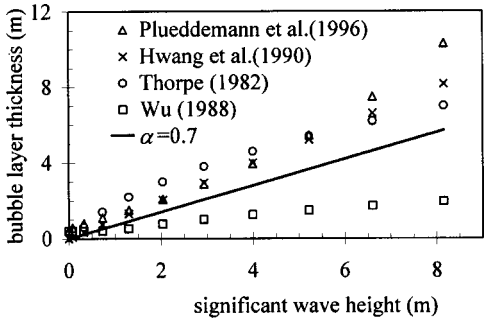


FIG. 3. The bubble layer thickness versus height of breaking waves. Comparison of different empirical formulas using the fully developed sea condition (symbols) and the fitted curve (solid line).

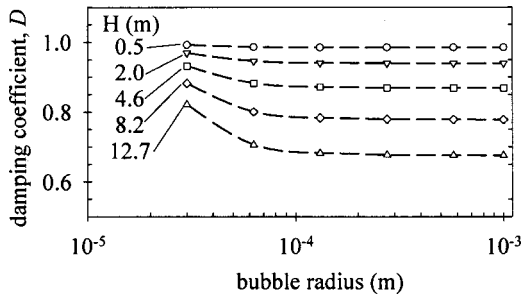


FIG. 4. Computed damping coefficient, $\hat{D}(b)$, for the different height of breaking wave, H. $\varepsilon=0.003$, $s=4$, $k_b=0.3$, $k_d=0.01$, $\alpha=0.7$.

The relationship, $\Delta=\alpha H$, was used for fitting the coefficient, α, to the empirical curves. The value $\alpha=0.27$ respects Wu's (1988) data, while $\alpha=1.2$ corresponds to Plueddemann *et al.* (1996). The fitted value $\alpha=0.7$ corresponds to the median curve for the different hypotheses. This is also consistent with the finding of Terray *et al.* (1996) that "breaking depth" is of the order of $0.6H$. Finalizing the analysis, the characteristic depth of the bubble cloud for $U>5$ m/s is estimated as

$$\delta=0.5\alpha H \quad (\alpha=0.7). \tag{24}$$

D. Wave height dependence of ambient noise

Equations (19) and (20), with respect to relationships (12), (21), and (24), yield

$$\hat{D}(b)=\exp(-2.5\varepsilon\mu B_{s-3}\hat{a}(b)H), \tag{25}$$

$$\hat{I}(b)=5\mu\lambda I_b\frac{\hat{\beta}(b)}{B}\frac{\hat{D}^2(b)H}{r}, \tag{26}$$

where the single parameter $\mu=(\frac{3}{160})\kappa_b\kappa_d/\alpha$ scales the contribution of wave-breaking energy to the acoustic noise generation, and $\lambda=\chi^3 b_m/r$. According to Eq. (25), the efficiency of damping of the acoustic energy increases with the height of the breaking wave (Fig. 4). This result corresponds to data of Wille and Geyer (1984) on noise reduction due to bubble screening. Taking into account Eq. (16), Eq. (26) can be rewritten as

$$\hat{I}(b)=5\mu\lambda I_b\left(\frac{f_0}{\hat{f}(b)}\right)^2\frac{\hat{D}^2(b)H}{r}. \tag{27}$$

Equation (26) suggests that the acoustic intensity spectrum of the bubble cloud is proportional to the dissipated wave energy, wave height, the normalized void fraction spectra, damping coefficient, source strength of individual air bubbles, and amount of acoustically active bubbles. According to the model [and previous findings of Oguz (1994)], the intensity level is not as sensitive to the total void fraction as to the bubble-size distribution. Equation (27) is in good agreement with empirical data and previously developed relationships. The similar dependence, $I\sim H$, has been empirically found by Knudsen *et al.* (1948), and later confirmed by Wille and Geyer (1984) for frequencies of a few kHz. The derived relationship, $I\sim f^{-2}$ (-6 dB per octave f for intensity level spectra) follows empirical data more closely than

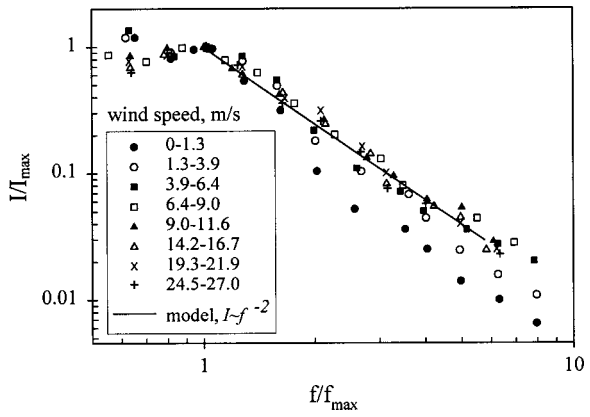


FIG. 5. Comparison of the present model (solid line) with the data of Perrone (1969) (symbols).

$I\sim f^{-5/3}$ (-5 dB/octave) (Kerman, 1984). The general form of Eq. (27) is similar to the theoretical expression of Marsh (1963), $I\sim H^{6/5}f^{-5/3}$, for the Neumann–Pierson wave spectra. Comparison of the present model with the data of Perrone (1969), interpreted by Kerman (1984), shows a good agreement in terms of the slope of the predicted acoustic spectra (Fig. 5).

It follows from the model that the main reasons for the observed -6-dB/octave slope of the intensity level are that the intensity spectra $\hat{I}(b)$ is proportional to the normalized void fraction spectra, $\hat{\beta}(b)/B$, and the latter is inversely proportional to the square of the normalized frequency spectra, $\hat{f}(b)/f_0$. In spite of the -6-dB/octave slope of the noise spectra for each of the single events of wave breaking, a summation of multiple-breaking events may reduce the slope down to the often-observed value -5 dB/octave. This is illustrated in Fig. 6, where the calculated intensity levels for the different breaking waves are plotted as a function of frequency. It is clear from the figure that the acoustic intensity spectra, generated by the individual waves, can accurately be modeled using the relationship $I\sim f^{-2}$. Summation of noise from the individual waves, leads to the values denoted in the figure with the symbol "+." For the frequencies above 400 Hz, the points can be approximated by a curve $I\sim f^{-5/3}$ (5 dB/octave) with the correlation coefficient as high as 0.96. Figure 6 shows that as the height of the breaking wave increases, so does the intensity level, and the maximum of the

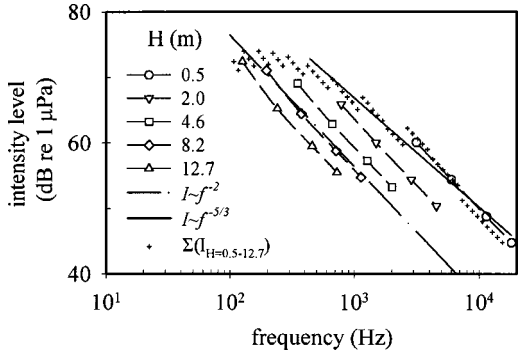


FIG. 6. Computed intensity level spectra for different height of breaking waves (dash-symbol lines). Summated noise from the individual breaking waves (symbol "+") and the respective fitted curve (solid line). $\varepsilon=0.003$, $s=4$, $k_b=0.3$, $k_d=0.01$, $\alpha=0.7$, $r=100$ m, B$=0.01$, $\chi=0.1$

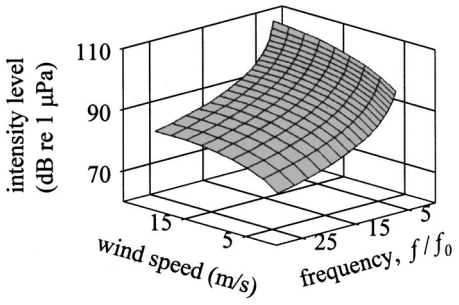


FIG. 7. An example of computed ambient noise intensity level, $\hat{S}$, versus wind speed, U, and bubble oscillation frequency f/f_0. $\varepsilon=0.002$, s=4.5, $k_b=0.3$, $k_d=0.02$, $\alpha=0.7$, $r=100$ m, $B=0.05$, $\chi=1$.

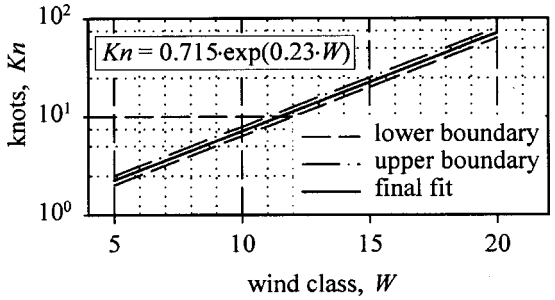


FIG. 8. Fitted transfer function, connecting knots (kn) and wind class (W). The lower boundary—dash line, the upper boundary—dash-dot line, and the mean values—solid line.

acoustic emission may shift to the very low frequencies (<200 Hz). Kolaini *et al.* (1995), observing similar phenomena in laboratory conditions, made the assumption that during the transition from weak breaker toward plunging breaker, the void fraction increases, forcing the bubbles to change their oscillation mode from individual (high frequency) to collective (low frequency). The present model predicts the frequency decrease within the collective mode due to the increase in radius of the bubble cloud. Kolaini *et al.* (1995) observed a widening of the frequency spectra for larger breakers. This phenomenon has a clear explanation using the present model: an increasing breaker produces larger bubbles, thus expending the radii range and, respectively, the oscillation frequency range.

E. Wind speed dependence of ambient noise

To follow the tradition of ambient noise scaling with wind speed, one has to substitute a transfer function between the wind speed and the significant height of the respective wind waves into Eq. (27). The general relationship is $H \sim U^{1-2}$ (Komen *et al.*, 1994). However, any particular case, defined by a type of dominant wind-wave spectra, has to be considered separately (see the discussions in the next section). In the case of the ''fully developed sea'' state, characterized by Eq. (23), Eqs. (25) and (27) become

$$\hat{D}(b)=\exp\left(-0.5\frac{\varepsilon\mu}{g}B_{s-3}\hat{a}(b)U^2\right), \tag{28}$$

$$\hat{I}(b)=\mu\lambda I_b\left(\frac{f_0}{\hat{f}(b)}\frac{\hat{D}(b)U}{U_0}\right)^2, \tag{29}$$

where $U_0=\sqrt{gr}$ or $U_0=(gr)^{1/2}$ is the scale parameter. The intensity level for Eq. (29) can be written as

$$\hat{S}(b)=S_0+20\log(\hat{D}(b)U/U_0)-20\log(\hat{f}(b)/f_0), \tag{30}$$

where

$$S_0=10\log(\mu\lambda I_b/I_{\mathrm{ref}}).$$

A typical example of computed noise spectra using Eq. (30) is shown in Fig. 7. The similar dependence ($I\sim U^2$) was first reported by Knudsen *et al.* (1948) for wind speed greater than 5 m/s, then confirmed by Piggott (1964) for shallow water ($I\sim U^{2.4}$), and by Crouch and Burt (1971) for deep water. Wille and Geyer (1984) obtained a similar power relationship from measurements at the North and Baltic Seas. To plot the results, they used the Piggott's (1964) wind class (W), that is related to knots (kn) via a transfer table. For convenience of comparison, we fit the transfer function, Kn $\approx 0.715\exp(0.23W)$ (Fig. 8), using values from the table. To compare the present model with the measurements of Wille and Geyer (1984) in the German Bay (North Sea), two sets of model parameters are considered: the first, for breaking waves at the distance 100 m from the hydrophone, and the second, for the waves as far as 500 m. To reproduce a variety of breaking events, the bubble cloud radius and void fraction are varied for both sets so that the reference frequency f_0 corresponds to 405 and 2863 Hz, respectively. Plotted in Fig. 9, the first case is represented by a series of three upper curves, and the second by the three lower curves. The comparison of the predicted noise and the data demonstrates good performance of the developed model. The model predicts correctly a decrease in the high-frequency part of the intensity level spectra due to the sound damping by the bubbles.

Oguz (1994) assumed that acoustic radiation from an individual bubble cloud is independent of wind speed, and that the observed wind dependence of the sound level could be explained by the increase in the number of the clouds (breaking events). In the present model, the summation of the noise intensities from multiple bubble clouds is possible

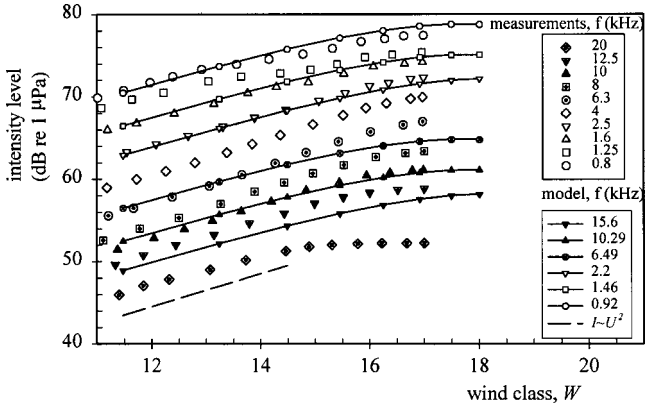


FIG. 9. Computed (dash-symbol lines) and measured (symbols) ambient noise intensity level, $\hat{S}$, versus wind class, W. Dash line corresponds to 2-dB/wind class ($I\sim U^2$). Measurements are adopted from Wille and Geyer (1984). $\varepsilon=0.02$, $s=2.5$, $k_b=0.3$, $k_d=0.01$, $\alpha=0.7$, $r_1=100$ m, $r_2=500$ m, $(f_0)_1=405$ Hz, $(f_0)_2=2863$ Hz, $\chi=0.23$.

TABLE I. Dependence of the bubble cloud acoustic intensity on the wind speed.

	Wind speed (m/s)	Acoustic intensity, I (present model)	Acoustic intensity, I (measurements; Wilson, 1983)	Likely wave state during acoustic measurements
I	$U<5$	$\sim U^{0-2}$	0	Capillary
II	$5<U<10$	$\sim U^{1-3}$	$\sim\{0.004U^3-0.049U^2+0.463U-1.5\}$	Gravity-capillary
III	$10<U<15$	$\sim U^2$		Fully developed sea
IV	$U>15$	$\sim U^{1-2}$	$\sim U^{1.5}$	Limited duration of wave growing

as well. However, the typical wind dependence of noise, $I \sim U^2$, is developed within a single bubble cloud, as result of energy flux from a breaking wave. The major contributor to the intensity level is parameter χ, characterizing the acoustically effective fraction of the bubble cloud. Assuming $\chi \sim 1$, one obtains the magnitude of the intensity level within 70–100 dB, which corresponds to upper values of observed ambient noise (Wille and Geyer, 1984; Kolaini *et al.*, 1994). In the case of $\chi \sim 0.1$, the intensity levels correspond to most observations, i.e., 50–70 dB.

IV. REASONS FOR THE DATA DISCREPANCY

Measurements have suggested a power relationship, $I \sim U^{\phi}$, for the wind-speed dependence of ambient noise. Although Wille and Geyer (1984) showed that $\phi \approx 2$ for most of their measurements in the vicinity of 1 kHz, the power varied as $\phi = 0-4$ for different ranges of the wind speed. Kerman (1984) defined two subranges, which, in terms of the wind speed may be rewritten as: $I \sim U^{3.8}$ for low wind speed, $I \sim U^{1.9}$ for high wind speed. Shang and Anderson (1986), combining data from several authors, showed the transition from the poorly defined power $\phi \approx 1-4$ at the low wind speed ($U<5$ m/s) to the well-defined power $\phi \approx 2$ at the high wind speed ($U>10$ m/s). Felizardo and Melville (1995) obtained the relationship, $I \sim U^3$, for $U \leqslant 10$ m/s. To understand the reasons for the exponent ϕ variation in the measurements, we consider that part of Eq. (19), which is related to the wind speed, namely,

$$I \sim \frac{H^2}{\delta}. \tag{31}$$

Here, we have not included sound attenuation by bubble screening, which is significant only for high wind speed ($U \geqslant 20$ m/s), where according to the model and observations, the power ϕ varies within the interval $(-1, 1)$ for the different frequencies.

Due to the different growing stages of wind-wave spectra, the wave height can be evaluated as $H \sim U$ for limited fetch or duration, and $H \sim U^2$ for fully developed sea (*Shore Protection Manual*, 1984). The minimum time needed to get the fully developed sea state increases drastically at very small and very high wind speeds, for example, for wind speed less than 5 m/s or higher than 15 m/s, the time may exceed 24 h (Pierson *et al.*, 1955). In these cases, measurements at the lower and higher wind speeds are likely made in the duration-limited conditions where $H \sim U$. Less probable are the cases of the fully developed sea, $H \sim U^2$, and swell domination, $H \sim U^0$. For the characteristic depth of the bubble cloud, the assumption chosen initially in Eq. (24) as $\delta \sim H$ is valid, strictly speaking, for high (>10 m/s) wind speeds. For lower velocities, the relationship of Wu (1988) has to be used, namely: $\delta \sim U^0$ for $U<7$ m/s, and $\delta \sim U$ for $7 \text{ m/s} < U < 12$ m/s. A likely reason for the metamorphosis from $\delta \sim U^0$ to $\delta \sim U$ and $\delta \sim H$ in the interval $5 \text{ m/s} \leqslant U \leqslant 10$ m/s, is a transition from low-energy capillary waves to high-energy gravity waves. Since the wind speed must exceed the wave velocity to allow wind energy to be absorbed by the waves, a wind from the interval can generate only relatively small gravity-capillary waves.

The above considerations are summarized as follows: $\delta \sim U^0$ for $U<5$ m/s (modified Wu, 1988); $\delta \sim U$ for $5 \text{ m/s} < U < 10$ m/s (modified Wu, 1988); $\delta \sim H \sim U^2$ for $10 \text{ m/s} < U < 15$ m/s (fully developed sea); $\delta \sim H \sim U^{1-2}$ for $U > 15$ m/s (limited duration of wave growing). Using the most probable cases of the numerator and denominator in Eq. (31), one obtains acoustic intensities for the different wind-speed intervals as in Table I. The relationships fit most available empirical data. Similar ranges and power relationships were suggested by Carey *et al.* (1993) for frequency less than 500 Hz: $\phi = 0-2$ at the "noise-limited" region (wind range I), $\phi = 3-6$—at the "transition" region (range II), $\phi = 2-3$—for the wind $10 \text{ m/s} \leqslant U \leqslant 20$ m/s (range III), and $\phi = -1-0$—for the wind $U \geqslant 20$ m/s (range IV). The negative slope for the high wind speed at the Carey *et al.*'s (1993) data is due to sound attenuation by bubbles, which we exclude from the scale analysis. The relationships agree also with the empirical conclusion of Wilson (1983), that measured noise level is directly proportional to the "whitecap index," leading to the formulas presented in Table I. The scale analysis suggests that the duration-limited wave spectra at the wind ranges I, II, and IV; influence of the capillary waves at the ranges I and II; and sound attenuation by bubbles at wind speed $U \geqslant 20$ m/s, are the likely reasons for the deviation of the exponent ϕ from the basic value $\phi = 2$ in the relationship $I \sim U^{\phi}$.

V. CONCLUSIONS

A theoretical model has been developed to explain the wave and wind dependence of underwater ambient noise. For the model closure, various theoretical and semiempirical relationships of wind-wave kinetics and air-bubbles entrainment are employed. Although most of the empirical parameters used have well-defined ranges, some, concerning energy flux from breaking waves to oscillating bubbles, still require extensive experimental calibration. The results are given as simple formulas for the wave height and wind-speed

dependence of the noise intensity spectra, Eq. (27) and Eq. (29), respectively. The developed model, using the hypothesis on collective oscillation of bubbles in the cloud, is able to simulate correctly the main features of ambient noise spectra, including the magnitude, slope, and frequency range. The model suggests the basic wind-speed dependence of intensity level as being $I \sim U^2$, and identifies the most possible reasons for deviations from the power law.

According to the model, the bubble-size spectra and size of the bubble cloud define the observed noise frequency range, 0.1–20 kHz. The noise intensity level is sensitive to the bubble cloud radius and amount of acoustically effective bubbles. The slope, $I \sim f^{-2}$ (-6 dB/octave), being an expression of the hypothesis on collective bubble oscillation in a single bubble cloud, does not depend on wave height or wind speed. However, the summation of intensity spectra of multiple-breaking events may reduce the slope down to the often-observed value of -5 dB/octave.

Baldy, S. (**1988**). ''Bubbles in the close vicinity of breaking waves: Statistical characteristics of the generation and dispersion mechanism,'' J. Geophys. Res. **93**(C7), 8239–8248.

Carey, W. M., and Bradley, M. P. (**1985**). ''Low-frequency ocean surface noise sources,'' J. Acoust. Soc. Am. **78**, S1–S2.

Carey, W. M., Fitzgerald, J. W., and Browning, D. G. (**1993**). ''Low frequency noise from breaking waves,'' in *Proceedings of II Int. Meeting on Natural Physical Sources of Underwater Sound "Sea Surface Sound'92," Cambridge, UK, 3–6 July 1990*, edited by B. R. Kerman (Kluwer Academic, Dordrecht), pp. 277–304.

Crouch, W. W., and Burt, P. L. (**1971**). ''The logarithmic dependence of surface generated ambient-sea-noise spectrum level on wind speed,'' J. Acoust. Soc. Am. **51**, 1066–1072.

Crum, L. A. (**1995**). ''Unresolved issues in bubble-related ambient noise,'' in *Proceedings of III Int. Meeting on Natural Physical Processes Related to Sea Surface Sound "Sea Surface Sound'94," University of California, Lake Arrowhead, 7–11 March 1994*, edited by M. J. Buckingham and J. R. Potter (World Scientific, Singapore), pp. 243–269.

d'Agostino, L., and Brennen, C. E. (**1988**). ''Acoustical absorption and scattering cross sections of spherical bubble cloud,'' J. Acoust. Soc. Am. **84**, 2126–2134.

Deane, G. B. (**1997**). ''Sound generation and air entrainment by breaking waves in the surf zone,'' J. Acoust. Soc. Am. **102**, 2671–2689.

Donelan, M. A. (**1978**). ''Whitecaps and momentum transfer,'' *NATO Conference on Turbulent fluxes through the sea surface, wave dynamics and prediction*, edited by A. Favre and K. Hasselmann (Plenum, New York), pp. 273–287.

Farmer, D. M., and Vagle, S. (**1989**). ''Waveguide propagation of ambient sound in the ocean-surface bubble layer,'' J. Acoust. Soc. Am. **86**, 1897–1908.

Felizardo, F. C., and Melville, W. K. (**1995**). ''Ambient surface noise and ocean surface waves,'' in *Proceedings of III Int. Meeting on Natural Physical Processes Related to Sea Surface Sound "Sea Surface Sound'94," University of California, Lake Arrowhead, 7–11 March 1994*, edited by M. J. Buckingham and J. R. Potter (World Scientific, Singapore), pp. 202–319.

Hwang, P. A., Hsu, Y. H. L., and Wu, J. (**1990**). ''Air bubbles produced by breaking wind waves: A laboratory study,'' J. Phys. Oceanogr. **20**, 19–28.

Kerman, B. R. (**1984**). ''Underwater sound generation by breaking wind waves,'' J. Acoust. Soc. Am. **75**, 149–165.

Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**1982**). *Fundamentals of Acoustics* (Wiley, New York).

Knudsen, V. O., Alford, R. S., and Emling, J. W. (**1948**). ''Underwater ambient noise,'' J. Mar. Res. **3**, 410–429.

Kolaini, A. R., Crum, L. A., and Roy, R. A. (**1994**). ''Bubble production by capillary-gravity waves,'' J. Acoust. Soc. Am. **95**, 1913–1921.

Kolaini, A. R., Dandenault, P., and Ruxton, A. D. (**1995**). ''Passive and active acoustical measurements of laboratory breaking waves,'' in *Proceedings of III Int. Meeting on Natural Physical Sources of Underwater Sound "Sea Surface Sound'94," University of California, Lake Arrowhead, 7–11 March 1994*, edited by M. J. Buckingham and J. R. Potter (World Scientific, Singapore), pp. 229–240.

Komen, G. J., Cavaleri, L., Donelan, M., Hasselmann, K., Hasselmann, S., and Janssen, P. A. E. M. (**1994**). *Dynamics and Modelling of Ocean Waves* (Cambridge University Press, Cambridge).

Lamarre, E., and Melville, W. K. (**1991**). ''Air entrainment and dissipation in breaking waves,'' Nature (London) **351**, 469–472.

Lamarre, E., and Melville, W. K. (**1994**). ''Sound-speed measurements near the ocean surface,'' J. Acoust. Soc. Am. **96**, 3605–3616.

Loewen, M. R., and Melville, W. K. (**1991a**). ''Microwave backscatter and acoustic radiation from breaking waves,'' J. Fluid Mech. **224**, 601–623.

Loewen, M. R., and Melville, W. K. (**1991b**). ''A model of the sound generated by breaking waves,'' J. Acoust. Soc. Am. **90**, 2075–2080.

Loewen, M. R. (**1991**). ''Laboratory measurements of the sound generated by breaking waves,'' Ph.D. thesis, Ocean Engineering, Massachusetts Institute of Technology.

Marsh, H. W. (**1963**). ''Origin of the Knudsen spectra,'' J. Acoust. Soc. Am. **35**, 409–410.

Means, S. L., and Heitmeyer, R. M. (**2001**). ''Low-frequency sound generation by an individual open-ocean breaking wave,'' J. Acoust. Soc. Am. **110**, 761–768.

Medwin, H., and Beaky, M. M. (**1989**). ''Bubble sources of the Knudsen sea noise spectra,'' J. Acoust. Soc. Am. **86**, 1124–1130.

Medwin, H., and Daniel, A. C. (**1990**). ''Acoustical measurements of bubble production by spilling breakers,'' J. Acoust. Soc. Am. **88**, 408–412.

Minnaert, M. (**1933**). ''On musical air bubbles and the sound of running water,'' Philos. Mag. **16**, 235–248.

Neumann, G., and Pierson, W. J. (**1966**). *Principles of Physical Oceanography* (Prentice-Hall, Englewood Cliffs, NJ).

Oguz, H. N. (**1994**). ''A theoretical study of low-frequency oceanic ambient noise,'' J. Acoust. Soc. Am. **95**, 1895–1912.

Omta, R. (**1987**). ''Oscillations of a cloud of bubbles of small and not so small amplitude,'' J. Acoust. Soc. Am. **82**, 1018–1033.

Perrone, A. J. (**1969**). ''Deep-ocean ambient-noise spectra in the north-west Atlantic,'' J. Acoust. Soc. Am. **46**, 762–770.

Pierson, W. J., Neumann, G., and James, R. (**1955**). *Practical Methods for Observing and Forecasting Ocean Waves by Means of Wave Spectra and Statistics* (H.O. Pub. 603, U.S Navy Hydrographic Office, Washington D.C.).

Piggott, C. L. (**1964**). ''Ambient sea noise at low frequencies in shallow water of Scotian shelf,'' J. Acoust. Soc. Am. **36**, 2152–2163.

Plueddemann, A. J., Smith, J. A., Farmer, D. M., Weller, R. A., Crawford, W. R., Pinkel, R., Vagle, S., and Gnanadesikan, A. (**1996**). ''Structure and variability of Langmuir circulation during the Surface Waves Processes Program,'' J. Geophys. Res. **101**, 3525–3543.

Prosperetti, A. (**1985**). ''Bubbles related ambient noise in the ocean,'' J. Acoust. Soc. Am. **78**, S2.

Prosperetti, A. (**1988**). ''Bubble-related ambient noise in the ocean,'' J. Acoust. Soc. Am. **84**, 1042–1054.

Prosperetti, A., Lu, N. Q., and Kim, H. S. (**1993**). ''Active and passive acoustic behavior of bubble clouds at the ocean's surface,'' J. Acoust. Soc. Am. **93**, 3117–3127.

Shang, E. S., and Anderson, V. C. (**1986**). ''Surface-generated noise under low wind speed at kilohertz frequencies,'' J. Acoust. Soc. Am. **79**, 964–971.

Shore Protection Manual (**1984**). Dept. of the Army, Waterways Experiment Station, Corps of Engineers, V.1.

Terray, E. A., Donelan, M. A., Agrawal, Y. C., Drennan, W. M., Kahma, K. K., Williams, A. J., Hwang, P. A., and Kitaigorodskii, S. A. (**1996**). ''Estimates of kinetic energy dissipation under breaking waves,'' J. Phys. Oceanogr. **26**, 792–807.

Terrill, E., and Melville, W. K. (**1997**). ''Sound-speed measurements on the surface-wave layer,'' J. Acoust. Soc. Am. **102**, 2607–2625.

Thorpe, S. A. (**1982**). ''On the clouds of bubbles formed by breaking wind waves in deep water, and their role in air–sea gas transfer,'' Philos. Trans. R. Soc. London, Ser. A **304**, 155–210.

Wille, P. C., and Geyer, D. (**1984**). ''Measurements on the origin of the wind-dependent ambient noise variability in shallow water,'' J. Acoust. Soc. Am. **75**, 173–185.

Wilson, J. H. (**1983**). ''Wind-generated noise modeling,'' J. Acoust. Soc. Am. **73**, 211–216.

Wu, J. (**1988**). ''Bubbles in the near-surface ocean: A general description,'' J. Geophys. Res. **93**(C1), 587–590.

Wavefronts and waveforms in deep-water sound propagation

Chris T. Tindle[a)]
Acoustics Division, Applied Research Laboratory, Penn State University, P.O. Box 30, State College, Pennsylvania 16804

(Received 30 October 2000; revised 4 October 2001; accepted 7 May 2002)

A new method of calculating waveforms in underwater sound propagation is presented. The method is based on a Hankel transform-generalized Wentzel–Kramers–Brillouin (WKB) solution of the wave equation. The resulting integral leads to a new form of ray theory which is valid at relatively low frequencies and allows evaluation of the acoustic field on both the illuminated and shadow sides of caustics and at cusps where two caustics meet to form a focus. The integral is evaluated by stationary phase methods for the appropriate number of stationary points. Rays of nearby launch angle which have a travel time difference less than a quarter period must be considered together. The description of all other ray arrivals corresponds to simple ray theory. The phase, amplitude, and travel time of broadband acoustic pulses are obtainable directly from a simple graph of ray travel time as a function of depth at a given range. The method can handle range dependence but is illustrated here in long-distance propagation in deep water where the ray paths do not pass close to surface or bottom. The method is fast and gives close agreement with normal-mode calculations. The field on the shadow side of a caustic is properly given in terms of rays with complex launch angles, but good approximations can be obtained without the need to find complex rays. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1489437]

PACS numbers: 43.30.Cq, 43.30.Dr [SAC-B]

I. INTRODUCTION

In this paper we present a new approach to the calculation of waveforms in underwater acoustics based on a Hankel transform-WKB solution of the wave equation. The method has many elements of ray theory, but leads to an integral expression for the acoustic field which is valid at cusps and caustics and in shadow zones. The approximations are different from those of conventional ray theory and the results are valid at relatively low frequencies. The method is fully range dependent, but is validated here by comparison with a broadband normal-mode solution for long-distance propagation in deep water.

The ray approach to the solution of propagation problems in underwater acoustics is well established and useful because of the simple, intuitive picture of energy propagation along ray paths. However, ray theory has well-known difficulties associated with the formation of caustics and diffraction into shadow zones, and various improvements of simple ray theory have been devised. Seckler and Keller[1,2] showed that the field in the geometric shadow of a caustic could be obtained in terms of complex rays and asymptotic approximations to exact solutions. Pedersen and Gordon[3] used the caustic formulas of Brekhovskikh[4] to find the propagation loss in a simple shadow region. White and Pedersen[5] represented the field near a caustic in terms of Airy functions and continued the field into the shadow using complex rays. Even though these methods for describing the field near a caustic are available, the large number of caustics and cusps which occur in deep water sound propagation[6] has thus far made it difficult to incorporate caustic corrections into ray-based computer codes.

In the present work the ray theory solution is obtained by introducing a WKB approximation for the depth-separated part of the wave equation into a Hankel transform solution. This method was used originally in underwater acoustics by Sachs and Silbiger[7] and gives the field as an integral over horizontal wave number. Sachs and Silbiger showed that stationary points of the integral corresponded to acoustic rays and that the field near a caustic could be evaluated in terms of Airy functions. Batorsky and Felsen[8] also used the WKB approximation in an integral expression for the field, and identified the lateral displacement associated with turning points near boundaries. Sachs[9] used the method to find transient waveforms near caustics and obtained qualitative agreement with experimental results. Murphy and Davis[10] and Davis[11] used a generalization of the WKB approximation to show the phase changes of a ray with a turning point near a boundary or sound-speed maximum. Weinberg[12] summarized the earlier work and combined the generalized WKB approach with numerical evaluation of integrals to show that ray theory could handle caustics and accurately describe propagation loss in general cases.

Interest in long-range ray calculations was boosted with the proposal of Munk and Wunsch[13] in 1979 to observe large-scale oceanographic features using acoustic transmissions. They showed that identification of individual ray arrivals and monitoring of relative arrival times would allow inversion to monitor sound-speed changes and the inference of ocean properties. The early ray arrivals on a deep-water vertical array clearly show wavefronts passing each hydrophone in succession as shown, for example, in the results of Worcester *et al.*[14]

[a)]On leave from Physics Dept., Univ. of Auckland, Bag 92019, Auckland, New Zealand.

 0001-4966/2002/112(2)/464/12/$19.00

In order to analyze receptions from impulsive sources, Brown *et al.*[15] introduced the ''depth–time diagram,'' which shows the folded wavefront arising from a ray calculation of travel time and depth at a fixed range. The depth–time diagram corresponds closely to the waveforms from a pulsed source observed on a vertical array. Brown[16] went on to show that the use of WKB Green's functions was able to give a good description of pulsed arrivals in deep water, including the vicinity of caustics. However, the formulation had difficulty when source or receiver was near the turning point of a geometric ray. Brown[17] used a time-domain wavefield representation to describe the transient field near caustics of all orders. The method reduces to ray theory when geometric ray theory is valid, but Brown suggests it is a replacement for ray theory rather than a modification.

In the present work we show that the waveforms expected at a hydrophone can be calculated directly from the depth–time diagram of simple geometric ray theory. To illustrate the method, we consider the situation corresponding to the long-range acoustic experiments designed to detect and monitor global warming.[18] In these experiments the important arrivals correspond to ray paths which do not interact with the ocean surface or the ocean bottom. Ray paths which repeatedly interact with surface or bottom rapidly lose intensity by scattering and absorption and are not important at long ranges. By considering deep-water rays which turn well clear of surface and bottom, we also avoid the ''lateral shift'' [8] which occurs when rays turn near a boundary, and obtain a simple expression which can be evaluated to give the phase and amplitude of pulse arrivals. The lateral shift of rays which turn near boundaries is interesting in its own right and its inclusion will be discussed in future work.

The wavefront method of the present work is valid everywhere, including the vicinity of caustics and cusps and in shadow zones. Details of the phase are important and different from earlier ray theory formulations. Rays of nearby launch angle which arrive less than a quarter period apart must be treated as pairs near caustics and triplets near cusps. The method can be applied to range-dependent cases, but here we consider range-independent examples in order to compare the results with broadband normal-mode calculations.

In the examples to follow, the depth–time diagram is found from simple numerical ray tracing and gives the amplitude, phase, and travel time of all real rays. The depth–time diagram can be extended through caustics into shadow regions using complex rays. However, the search for complex rays can be avoided and the field near caustics and cusps can be found from numerical approximations to the phase function.

In Sec. II we derive the ray theory solution directly from the wave equation by inserting the WKB approximation into the Hankel transform integral and expanding the result as a power series. The ray expressions are obtained by evaluating the resulting integral at its saddle points by the method of stationary phase or steepest descents. Similar derivations have appeared elsewhere,[8,19] but here we consider only deep-water rays which do not interact with surface or bottom, and obtain new expressions for the field with precise details of the phase. In Sec. III we consider caustics and cusps by evaluating the integrals of Sec. II for pairs and triplets of nearby saddle points. In Sec. IV we show how the depth–time diagram and wavefront modeling can be used to determine the waveforms expected at a series of hydrophones. The results are compared with broadband normal-mode calculations, and there is good agreement. Discussion and conclusions are given in Sec. V.

II. SOLUTION OF THE WAVE EQUATION

A. Hankel transform solution of the wave equation

The acoustic wave equation for a point harmonic source of angular frequency ω in a cylindrically symmetric situation can be written[20]

$$\frac{\partial^2 p}{\partial r^2}+\frac{1}{r}\frac{\partial p}{\partial r}+\frac{\partial^2 p}{\partial z^2}+\frac{\omega^2}{c^2}p=-\frac{2}{r}Q\,\delta(r)\,\delta(z-z_s), \tag{1}$$

where c is the sound speed and $p(r,z)$ is the acoustic pressure at range r and depth z. The source is at zero range and depth z_s, and the source strength Q is unity if the acoustic pressure has unit amplitude 1 m from the source.

If the sound speed c is a function of depth only, and $F(k,z)$ is the Hankel transform of $p(r,z)$, we obtain

$$\frac{\partial^2 F}{\partial z^2}+\left(\frac{\omega^2}{c^2(z)}-k^2\right)F=-2Q\,\delta(z-z_s), \tag{2}$$

where k is the horizontal wave number.

We now assume that functions $U(z)$ and $V(z)$ are linearly independent solutions of

$$\frac{\partial^2 U}{\partial z^2}+\gamma^2 U=0, \tag{3}$$

where

$$\gamma^2=\frac{\omega^2}{c^2(z)}-k^2, \tag{4}$$

and that the functions $U(z)$ and $V(z)$ satisfy boundary conditions above and below the source, respectively.

A solution of Eq. (2) can then be constructed and inserted into the inverse Hankel transform to yield the following integral solution of the wave equation:[19,20]

$$p(r,z)=-Q\int_{-\infty}^{\infty}\frac{U(z)V(z_s)}{W}H_0^{(1)}(kr)k\,dk, \quad z<z_s \tag{5}$$

$$=-Q\int_{-\infty}^{\infty}\frac{U(z_s)V(z)}{W}H_0^{(1)}(kr)k\,dk, \quad z>z_s \tag{6}$$

where the Wronskian W is given by

$$W=U(z)V'(z)-U'(z)V(z). \tag{7}$$

B. WKB solutions

If the depth functions $U(z)$ and $V(z)$ have turning points a and b, respectively, within the water column and well clear of the surface or bottom, their WKB approximations are

$$U(z)\approx B|\gamma(z)|^{-1/2}\sin\left(\int_a^z \gamma(z')dz'+\pi/4\right) \tag{8}$$

$$V(z)\approx C|\gamma(z)|^{-1/2}\sin\left(\int_z^b \gamma(z')dz'+\pi/4\right), \tag{9}$$

so the Wronskian W becomes

$$W=-BC\sin\left(\int_a^b \gamma(z')dz'+\pi/2\right), \tag{10}$$

where we have neglected terms in $d\gamma/dz$ to be consistent with the WKB assumptions of slow parameter variation with respect to wavelength.

By changing the sine function to the sum of exponentials, the Wronskian can be written

$$W=-\frac{BC}{2i}\exp\left\{-i\left(\int_a^b \gamma(z')dz'-\pi/2\right)\right\}\times\left[1-\exp 2i\left(\int_a^b \gamma(z')dz'-\pi/2\right)\right]. \tag{11}$$

C. Expansion of the denominator

The following expansion is based on the work of Batorsky and Felsen.[8] A similar expansion is derived by Frisk.[19] The Wronskian is in the denominator of the integral expressions (5) and (6) for the field, so we note the following expansion:

$$[1-x]^{-1}=1+x+x^2+x^3+\cdots=\sum_{n=0}^{\infty}x^n. \tag{12}$$

If the expansion is used to replace the square bracket term in the Wronskian, and the result is put into the expression for the pressure field, we obtain

$$p(r,z)=Q\int_{-\infty}^{\infty}|\gamma(z_s)|^{-1/2}\sin\left(\int_a^{z_s}\gamma(z')dz'+\pi/4\right)\times|\gamma(z)|^{-1/2}\sin\left(\int_z^b\gamma(z')dz'+\pi/4\right)\times 2i\exp\left[i\left(\int_a^b\gamma(z')dz'-\pi/2\right)\right]\times\sum_{n=0}^{\infty}\left\{\exp\left[2i\left(\int_a^b\gamma(z')dz'-\pi/2\right)\right]\right\}^n\times H_0^{(1)}(kr)k\,dk,\quad z>z_s, \tag{13}$$

with a similar expression for $z<z_s$.

We now write the sine functions in terms of exponentials and use the asymptotic form of the Hankel function, where

$$H_0^{(1)}(kr)\sim(2/\pi kr)^{1/2}\exp[i(kr-\pi/4)]. \tag{14}$$

The expansion of Eq. (13) becomes

$$p(r,z)=\sum_{n=0}^{\infty}\sum_{j=1}^{4}p_{nj}, \tag{15}$$

where n is as in Eq. (13), and p_{nj} can be written

$$p_{nj}(r,z)=\int_{-\infty}^{\infty}Q|\gamma(z_s)|^{-1/2}|\gamma(z)|^{-1/2}(1/2i)^2 2i(2/\pi kr)^{1/2}\times k\exp(i\phi_{nj})\exp(-i\pi/4)\exp(i\pi)dk, \tag{16}$$

which simplifies to

$$p_{nj}(r,z)=Q(2\pi r)^{-1/2}\exp(i\pi/4)\times\int_{-\infty}^{\infty}k^{1/2}|\gamma(z_s)\gamma(z)|^{-1/2}\exp(i\phi_{nj})dk. \tag{17}$$

The factor $\exp(i\pi)$ in Eq. (16) is inserted for convenience, and the phase ϕ_{nj} is given by

$$\phi_{nj}=\phi_{0j}+2n\left(\int_a^b\gamma(z')dz'-\pi/2\right), \tag{18}$$

where

$$\phi_{01}=\int_a^{z_s}\gamma(z')dz'+\pi/4+\int_z^b\gamma(z')dz'+\pi/4+\int_a^b\gamma(z')dz'-\pi/2+kr-\pi, \tag{19}$$

$$\phi_{02}=\int_a^{z_s}\gamma(z')dz'+\pi/4-\int_z^b\gamma(z')dz'-\pi/4+\int_a^b\gamma(z')dz'-\pi/2+kr-\pi+\pi, \tag{20}$$

$$\phi_{03}=-\int_a^{z_s}\gamma(z')dz'-\pi/4+\int_z^b\gamma(z')dz'+\pi/4+\int_a^b\gamma(z')dz'-\pi/2+kr-\pi+\pi, \tag{21}$$

$$\phi_{04}=-\int_a^{z_s}\gamma(z')dz'-\pi/4-\int_z^b\gamma(z')dz'-\pi/4+\int_a^b\gamma(z')dz'-\pi/2+kr-\pi+\pi+\pi. \tag{22}$$

The extra $+\pi$ phase terms in Eqs. (20), (21), and (22) come from the minus signs when the sine functions in Eq. (13) are expanded in terms of exponentials.

The integral terms in the expression for ϕ_{01} correspond to the phase accumulated along ray paths from source to upper turning point, lower turning point to receiver, and from upper turning point to lower turning point, respectively. The entire expression in Eq. (19) therefore represents the phase accumulated by a ray propagating from source to receiver with one upper turning point and one lower turning point. The residual part of the phase is $-\pi$ and corresponds to a phase change of $-\pi/2$ for each turning point.

In a similar fashion, the expressions for ϕ_{02} and ϕ_{03} correspond to the phase along a path with one upper turning point and one lower turning point, respectively. Each has a

residual phase of $-\pi/2$ corresponding to each having one turning point. The expression for ϕ_{04} corresponds to the direct path and has zero residual phase.

Equation (18) shows that in general the phase contains n complete cycles and the term ϕ_{0j} describes the remaining incomplete cycle. Each complete cycle adds a phase change of $2\times(-\pi/2)$, and so there is a phase of $-\pi/2$ for each turning point along the ray.

Expressions for ray contributions to the field similar to Eq. (17) have appeared elsewhere.[8,9,16,19,21] Our result differs in two important details. First, it makes the turning point phases explicit in Eqs. (19)–(22) and second, the modulus of $\gamma(z_s)\gamma(z)$ appears in the denominator of Eq. (17) rather than the signed value. These differences allow us to compute absolute phase and to give a new interpretation of the phase changes along a ray path.

D. Saddle points

In order to evaluate the integral over k, the saddle points are first found by solving

$$\frac{d\phi_{nj}}{dk}=0. \tag{23}$$

Now, from Eq. (4) we have

$$\frac{d\gamma}{dk}=\frac{-k}{\sqrt{\omega^2/c^2-k^2}}=\frac{-k}{\gamma}. \tag{24}$$

Therefore, the saddle-point condition for $n=0$, $j=1$ can be rearranged to give

$$r=\int_a^{z_s}\frac{k}{\gamma(z')}dz'+\int_z^b\frac{k}{\gamma(z')}dz'+\int_a^b\frac{k}{\gamma(z')}dz'. \tag{25}$$

The well-known relations between the wave numbers and equivalent ray of angle θ with respect to the horizontal can be written

$$k=(\omega/c)\cos\theta, \tag{26}$$

$$\gamma=(\omega/c)\sin\theta. \tag{27}$$

Therefore, Eq. (25) can be written as

$$r=\int_a^{z_s}\cot\theta\,dz'+\int_z^b\cot\theta\,dz'+\int_a^b\cot\theta\,dz', \tag{28}$$

which is clearly the equation of the ray between source and receiver which has one upper ($z=a$) and one lower ($z=b$) turning point. The phase ϕ_{01} along this path was given in Eq. (19).

Other terms in the expansion of Eq. (13) correspond to the other ray paths linking source and receiver. The $n=0$, $j=2,3,4$ terms correspond to rays with one upper turning point, one lower turning point, and zero turning points, respectively and their phases are given in Eqs. (20)–(22).

The method of steepest descents (also called the saddle-point method) is used to find analytic approximations to integrals of the form

$$I=\int e^{i\phi(z)}F(z)dz. \tag{29}$$

If $\phi(z)$ is real, the method is called the method of stationary phase.

If $F(z)$ is slowly varying, the main contribution to the integral comes from regions near saddle points, i.e., where $d\phi/dz=0$. If z_0 is such a point, the function ϕ can be approximated by a quadratic in z about the point, and the resulting integral can be evaluated to give

$$I\approx\left[\frac{2\pi i}{\phi''(z_0)}\right]^{1/2}F(z_0)\exp[i\phi(z_0)], \tag{30}$$

where the prime indicates d/dz. It should be noted that if $\phi''(z_0)$ is real, the square bracket term will lead to a phase of $\pm\pi/4$ depending on the sign of $\phi''(z_0)$.

The saddle-point result can be used to find the phase and amplitude of the ray contribution. To evaluate the integral in Eq. (17) by the saddle-point method, it is convenient to define a parameter $r^*(k)$ as follows:

$$r^*(k)=\int_{z_s\sim}^{z_r}\frac{k}{\gamma(z')}dz', \tag{31}$$

where z_r is the receiver depth and the notation $z_s\sim$ indicates that the integral is taken through the sequence of upper and lower turning points appropriate to the particular term of the integrand in Eq. (13) which is being considered.

With this notation, we can now write

$$\frac{d\phi}{dk}=-r^*(k)+r, \tag{32}$$

and therefore

$$\frac{d^2\phi}{dk^2}=-\frac{dr^*}{dk}. \tag{33}$$

Now, evaluating the integral in Eq. (17) using Eq. (30) leads to

$$\begin{aligned}p_{n,1}(r,z_r)=&(2\pi r)^{-1/2}\exp(i\pi/4)\left(\frac{-2\pi i}{dr^*/dk}\right)^{1/2}\\&\times k^{1/2}|\gamma(z_r)\gamma(z_s)|^{-1/2}\\&\times\exp\left(i\left[\int_{z_s\sim}^{z_r}\gamma(z')dz'-m\pi/2+kr\right]\right),\end{aligned} \tag{34}$$

where m is the number of turning points.

Simplifying Eq. (33), we obtain

$$\begin{aligned}p_{n,1}(r,z_r)=&Qr^{-1/2}\left[\frac{1}{dr^*/dk}\right]^{1/2}k^{1/2}|\gamma(z_r)\gamma(z_s)|^{-1/2}\\&\times\exp\left(i\left[\int_{z_s\sim}^{z_r}\gamma(z')dz'-m\pi/2+kr\right]\right).\end{aligned} \tag{35}$$

The equivalent ray relation of Eqs. (26) and (27) give

$$\frac{dr^*}{dk}=\frac{dr^*}{d\theta}\frac{d\theta}{dk}=\frac{dr^*}{d\theta}\frac{1}{-(\omega/c)\sin\theta}. \tag{36}$$

It is simple to show that the travel time T_{nj} along a ray path is given by

$$T_{nj}(r,z_r,\theta_s)=(1/\omega)\left(\int_{z_s\sim}^{z_r}[\omega/c(z')]\sin\theta(z')dz' + (\omega/c)\cos\theta_s r\right), \quad (37)$$

and using Eqs. (36) and (37) we can write Eq. (35) as

$$p_{n,1}(r,z_r)=Qr^{-1/2}\left[\frac{-\sin\theta_s}{dr^*/d\theta_s}\right]^{1/2}(\cos\theta_r)^{1/2}|\sin\theta_r\sin\theta_s|^{-1/2} \times\exp[i(\omega T_{nj}-m\pi/2)], \quad (38)$$

where quantities evaluated at source and receiver are identified with subscripts *s* and *r*, respectively, and we have used the fact that *k* has the same value at source and receiver through Snell's law.

Equation (38) can be further simplified by defining a phase δ to give

$$p_{n,1}(r,z_r)=Qr^{-1/2}|dr^*/d\theta_s|^{-1/2}|\cot\theta_r|^{1/2} \times\exp[i(\omega T_{nj}-m\pi/2+\delta)], \quad (39)$$

where

$$\delta=(\pi/4)[1+\text{signum}(\theta_s\times dr^*/d\theta_s)]. \quad (40)$$

Equation (39) is precisely the field due to a conventional ray except for the extra phase terms $m\pi/2$ and δ. These terms arise naturally in the foregoing derivation, and show that

(a) there is a phase change of $-\pi/2$ at each turning point, and
(b) there is an extra phase term of $+\pi/2$ if θ_s and $dr^*/d\theta_s$ have the same sign.

These two terms are entirely equivalent to the conventional phase change of $-\pi/2$ at caustics as follows. Very close to the source, before any rays have turning points, $dr^*/d\theta_s$ is negative for θ_s positive, and vice versa. Therefore, $dr^*/d\theta_s$ and θ_s have opposite signs and $\delta=0$. This means there is no residual phase change along a path with no turning points. If the ray now turns before touching a caustic, $dr^*/d\theta_s$ will change sign and the $\pi/2$ arising from δ will cancel the $-\pi/2$ arising from the turning point to give no residual phase as expected. When the ray eventually touches the caustic, $dr^*/d\theta_s$ changes sign again to give $\delta=0$ and the residual phase is $-\pi/2$ arising from the turning point.

Therefore, as the ray propagates it apparently accumulates a residual phase of $-\pi/2$ at each caustic.

E. Fixed range

For propagation to fixed range, it is convenient to define $z^*(\theta_s)$ as the depth at range *r* of a ray with launch angle θ_s. It is simple to show

$$dr^*/d\theta_s=-\cot\theta_r\, dz^*/d\theta_s, \quad (41)$$

so that Eq. (38) becomes

$$p_{nj}(r,z)=Qr^{-1/2}|dz^*/d\theta_s|^{-1/2}\exp[i(\omega T_{nj}-m\pi/2+\delta)], \quad (42)$$

and the phase δ is given by

$$\delta=(\pi/4)[1-\text{signum}(\theta_r\times\theta_s\times dz^*/d\theta_s)]. \quad (43)$$

Equations (42) and (43) are particularly simple, and can be used as the basis of a computer code to find the phase, amplitude, and travel time of each ray arrival. The function $z^*(\theta_s)$ is found by tracing a fan of rays to a fixed range *r* and noting their depths. The value of $dz^*/d\theta_s$ is found by numerical differentiation.

It is worth noting that the expression for the field given by Eq. (42) is valid at all field points, even though the WKB expressions of Eqs. (8) and (9) are singular when source or receiver are at a turning point. The cancellation of terms in $\sin\theta_s$ and $\sin\theta_r$ in Eqs. (38) and (42) is fortuitous, and one of the difficulties associated with use of the WKB functions is avoided.

The ray theory described above is based on the assumption that saddle points are isolated. It gives an accurate description of the acoustic field except when rays from nearby angles arrive at the field point separated by less than a quarter wavelength.

III. CAUSTICS AND CUSPS

Ray arrivals cannot be considered as isolated in the vicinity of caustics and cusps. In such cases the pair or sometimes triplet of rays must be treated together and the integral of Eq. (17) evaluated accordingly.

A. Caustics—Insonified zone

Near caustics, in the insonified zone, there are two ray arrivals with similar angles θ_1 and θ_2 and travel times T_1 and T_2, respectively. This occurs because two saddle points of the contour integral are close together. The two saddle points become coincident on the caustic. In such a case the two rays cannot be treated as isolated, and the phase function ϕ_{nj} of Eq. (17) must be approximated by a cubic. The resulting uniform asymptotic expansion was derived by Chester *et al.*[22] and described by Felsen and Marcuvitz.[23] The expansion is well behaved as the caustic is approached and $T_2\to T_1$.

Assuming $T_2\to T_1$, the combined field p_{ca} near the caustic can be written

$$p_{ca}(r,z_r)=Qr^{-1/2}\pi^{1/2}[S_+Ai(-x)+iS_-Ai'(-x)] \times\exp\{i[\omega(T_1+T_2)/2-m\pi/2+\delta+\pi/4]\}, \quad (44)$$

where

$$x=[3\omega(T_2-T_1)/4]^{2/3}, \quad (45)$$

$$S_+=x^{1/4}[|dz^*/d\theta_1|^{-1/2}+|dz^*/d\theta_2|^{-1/2}]. \quad (46)$$

$$S_-=x^{-1/4}[|dz^*/d\theta_1|^{-1/2}-|dz^*/d\theta_2|^{-1/2}]. \quad (47)$$

The functions Ai and Ai′ are the Airy function and its derivative. The phase term δ is the value appropriate to the second arrival at time T_2. This ray has already touched the caustic.

For a harmonic source, Eq. (44) gives the amplitude and phase of the field at the point *r,z* close to the caustic. Superficially, Eq. (44) appears to represent a contribution to the

field which arrives at time $(T_1+T_2)/2$. However, the parameter x also depends on the time, so the situation is more complicated.

To describe pulse propagation, it is necessary to rewrite Eq. (44) as

$$p_{ca}(r,z_r)=A_1\exp(i\omega T_1)+A_2\exp(i\omega T_2), \tag{48}$$

where

$$A_1=(Q/2)\pi^{1/2}r^{-1/2}x^{1/4}|-dz^*/d\theta_1|^{-1/2}\times[\mathrm{Ai}(-x)-ix^{-1/2}\mathrm{Ai}'(-x)]\times\exp\{i[+\omega(T_2-T_1)/2-m\pi/2+\delta+\pi/4]\}, \tag{49}$$

and

$$A_2=(Q/2)\pi^{1/2}r^{-1/2}x^{1/4}|+dz^*/d\theta_2|^{-1/2}\times[\mathrm{Ai}(-x)+ix^{-1/2}\mathrm{Ai}'(-x)]\times\exp\{i[-\omega(T_2-T_1)/2-m\pi/2+\delta+\pi/4]\}. \tag{50}$$

The field given by Eqs. (48)–(50) is algebraically identical to that given by Eqs. (44)–(47), but Eq. (48) can now be interpreted as describing two pulses with travel times T_1 and T_2, respectively. Equations (49) and (50) give the complex amplitudes of these two pulses.

Further from the caustic, the asymptotic form of the Airy functions can be used to show that the time dependence of A_1 and A_2 cancels out and Eq. (48) is equivalent to two terms of the form of Eq. (42). The formulas in Eqs. (44)–(47) are equivalent to those given by White and Pedersen.[6]

B. Caustics—Shadow zone

On the shadow side of a caustic, the phase function ϕ_{nj} of Eq. (17) has two nearby saddle points at complex values of the wave number k. As the caustic is crossed from the insonified zone to the shadow zone, the two saddle points for real k merge to become coincident on the caustic and then separate into two complex saddle points. The formulas in Eqs. (16)–(28) and (31)–(37) may all be continued analytically into regions of complex k and complex angle θ, and we can consider complex rays.

A complex ray is a ray with a complex launch angle which arrives at a real depth at a real range. A complex ray can be found by taking the ray angle corresponding to the caustic and adding an imaginary part. The ray of complex angle is then traced as before using Snell's law by computing its depth and travel time as the range is incremented. The range is real, but the depth and travel time will usually be complex for a complex starting angle. The real part of the launch angle is now varied iteratively until the depth of the ray at the desired range is real. The real depth is always in the shadow zone and the distance into the shadow zone increases as the imaginary part of the launch angle is increased. For each value of the imaginary part of the angle, it is necessary to iterate to find the corresponding real part which gives a real arrival depth.

After iteration the complex ray is obtained as a ray which arrives at a real depth and range with a complex travel time. In fact, the complex conjugate launch angle satisfies the same conditions, so we obtain two rays with complex conjugate launch angles and complex conjugate travel times which arrive at the same real depth for the given real range.

The combined field for these two complex rays could be given by Eqs. (44)–(47) with a more complicated definition of the parameter x. However, since the ray angles and arrival times are complex conjugates, it is convenient to assume $\mathrm{Im}(T_2)=-\mathrm{Im}(T_1)>0$ and rewrite the expressions as follows:

$$p_{ca}(r,z_r)=r^{-1/2}\pi^{1/2}[S_+\mathrm{Ai}(-x)-S_-\mathrm{Ai}'(-x)]\times\exp\{i[\omega\,\mathrm{Re}(T_1)-m\pi/2+\delta+\pi/4]\}, \tag{51}$$

where

$$x=-[3\omega\,\mathrm{Im}(T_2)/2]^{2/3}, \tag{52}$$

$$S_+=|x|^{1/4}\,\mathrm{Re}\{[-i/(dz^*/d\theta_1)]^{1/2}\}, \tag{53}$$

$$S_-=|x|^{-1/4}\,\mathrm{Im}\{[-i/(dz^*/d\theta_1)]^{1/2}\}. \tag{54}$$

It is clear from Eq. (51) that the combined field of the two complex rays has a real travel time equal to the real part of the two complex conjugate travel times. The parameter x is now real and negative, and increases with distance into the shadow zone. The Airy functions, and therefore the field, decay exponentially in the shadow zone. The phase δ is the same as on the insonified side of the caustic.

The derivative $dz^*/d\theta_1$ is approximately pure imaginary in the shadow zone because Δz^* is real and $\Delta\theta$ is approximately pure imaginary. Therefore, the square bracket terms in Eqs. (53) and (54) are approximately real and S_- is much smaller than S_+.

Expressions (51) and (52) are equivalent to those given by White and Pedersen,[6] but Eqs. (53) and (54) differ somewhat. The expressions above are derived here from first principles from the wave equation.

Complex rays enable extension of the depth–time diagram into the acoustic shadow. The two complex rays combine to give a single field contribution which has a real travel time. The expressions presented above enable calculation of the phase and amplitude of the shadow zone field. The search for complex rays is one-dimensional because the imaginary part of the angle is selected and then the real part is iterated to give a real arrival depth.

C. Cusps

To evaluate the field p_{cu} in the vicinity of a cusp, it is convenient to rewrite Eq. (17) as an integral over angle using Eqs. (26) and (27); we obtain

$$p_{cu}(r,z_r)=Q(2\pi r)^{-1/2}\exp(i\pi/4)\times\int_{-\infty}^{\infty}[\omega/c(z_s)](\cos\theta_s|\sin\theta_s|)^{1/2}\times|\omega\sin\theta_r/c(z_r)|^{-1/2}\exp(i\phi)d\theta_s, \tag{55}$$

where $\phi(\theta_s)$ can be written using Eqs. (18)–(22) and the notation of Eq. (31) as

$$\phi(r,z_r,\theta_s)=\int_{z_s\sim}^{z_r}[\omega/c(z')]\sin\theta(z')dz' + [\omega/c(z_s)]\cos\theta_s r - m\pi/2. \quad (56)$$

In the vicinity of a cusp there are two merging caustics and the phase function of Eq. (56) has three nearby saddle points. The integral in Eq. (55) can be evaluated approximately by expanding the phase function as a quartic about $\theta_s=\theta_0$, where $\phi'''(\theta_0)=0$ and the prime denotes $d/d\theta_s$.

The canonical form of the integral is the Pearcey integral,[24] defined by

$$Pc(p,q)=\int_{-\infty}^{\infty}\exp[i(pt+qt^2+t^4)]dt, \quad (57)$$

and the resulting field $p_{cu}(r,z)$ near the cusp can be written

$$p_{cu}(r,z)=Q(2\pi r)^{-1/2}\exp(i\pi/4)\frac{\omega}{c(z_s)}\left(\frac{\cos\theta_s|\sin\theta_s|}{\omega|\sin\theta_r|/c(z_r)}\right)^{1/2} \times\alpha\exp[i(\phi_0+\delta_0)]Pc(\alpha\phi_0',\alpha^2\phi_0''/2), \quad (58)$$

where

$$\alpha=(\phi_0''''/24)^{-1/4}, \quad (59)$$

and the subscript zero indicates values at $\theta_s=\theta_0$.

In the numerical work to be discussed later, the first step in the computation of the waveforms is to trace a fan of rays to the desired range. The depths and travel times yield the functions $z^*(r,\theta_s)$ and $T(r,z^*,\theta_s)$. Comparing Eqs. (37) and (56), we see that ϕ_s is given by

$$\phi(r,z_r,\theta_s)=\omega T(r,z^*,\theta_s)+\int_{z^*}^{z_r}[\omega/c(z')]\sin\theta(z')dz' - m\pi/2. \quad (60)$$

The phase function in the vicinity of a cusp is readily found using Eq. (60). The numerical values of ϕ_0, ϕ_0', etc., are found by curve fitting $\phi(\theta_s)$ around θ_0, and Eq. (58) can be evaluated.

Equation (60) is rather trivial if a real ray at angle θ_s passes through the point (r,z_r) and it allows the phase to be found from the already-calculated phase at (r,z^*) without having to trace the ray path again. However, Eq. (60) allows the phase to be found for all values of (r,z_r,θ_s) whether or not a real ray path exists. This is most important in the vicinity of cusps, as it allows $\phi(r,z_r,\theta_s)$ to be expanded in a power series about θ_0 for all field points (r,z_r). In the numerical work discussed later, Eq. (60) is approximated by

$$\phi(r,z_r,\theta_s)=\omega T(r,z^*,\theta_s)+(z_r-z^*)\{\omega/c[(z_r+z^*)/2]\} \times\sin\theta_r-m\pi/2. \quad (61)$$

D. Caustics—Shadow zone approximation

The search for complex rays requires an iteration and can be computation intensive. It is therefore convenient to have an approximation to the field in the caustic shadow. As noted above, the integral in Eq. (17) has been evaluated near a caustic by expanding the phase function ϕ_{nj} as a cubic. The resulting Airy function expressions in Eqs. (44)–(47) were then given in terms of the ray travel times and the derivatives of the depth as a function of angle.

The parameters of the Airy functions can be equally well obtained by the same procedure used for a cusp. The phase function of Eq. (56) is expanded as a cubic and the canonical integral is the definition of the Airy function, i.e.,

$$\mathrm{Ai}(-x)=(2\pi)^{-1}\int_{-\infty}^{\infty}\exp[i(t^3/3+xt)]dt. \quad (62)$$

When the phase function ϕ_{nj} of Eq. (17) has two nearby saddle points, the field integral can be written as in Eqs. (55) and (56) and the phase $\phi(\theta_s)$ can be expanded as a cubic about the point $\theta_s=\theta_0$, where $\phi''(\theta_0)=0$. Use of Eq. (62) then leads to

$$p_{ca}(r,z)=Q(2\pi/r)^{1/2}\exp(i\pi/4)\frac{\omega}{c(z_s)}\left(\frac{\cos\theta_s|\sin\theta_s|}{\omega|\sin\theta_r|/c(z_r)}\right)^{1/2} \times\beta\exp[i(\phi_0+\delta_0)]\mathrm{Ai}(-|\phi_0'\beta|), \quad (63)$$

where

$$\beta=|\phi_0'''/2|^{-1/3}. \quad (64)$$

In some of the numerical work to follow, Eq. (63) is used to find the phase, amplitude, and arrival time of a pulse on the shadow side of a caustic. The results will be compared with those obtained using Eq. (51).

IV. RESULTS

A. Sound-speed profile

In order to test the present version of ray theory, we consider a deep-water Munk[25] sound-speed profile as follows:

$$c(z)=c_0\{1+\varepsilon[\exp(-\eta)+\eta-1]\}, \quad (65)$$

where

$$\eta=2(z-z_0)/b, \quad (66)$$

and the parameter values are taken as $c_0=1478.5$ m/s, $z_0=800$ m, $b=576$ m, and $\varepsilon=0.003$.

B. Wavefronts

Figure 1 shows the depth–time diagram obtained by tracing rays using Snell's law in the above profile to a range of 80 km. A source depth of $z_s=800$ m is assumed and the range step is 100 m. The solid curve is obtained by tracing a fan of rays in the range $\pm$ 15° with an angular increment of 0.1°. The circles show the individual values. The upper-right corner of Fig. 1 is reproduced in Fig. 2. The density of circles in Figs. 1 and 2 gives an indication of ray amplitudes which are determined by $dz^*/d\theta_s$ in Eq. (42). Close, overlapping circles correspond to large amplitude arrivals. At the same time as the depth–time diagram is generated, a file containing *m* the number of turning points *m* is kept in order to determine the phases of the pulse arrivals.

It is convenient to refer to the solid curve in Figs. 1 and 2 as a wavefront because it represents the arrival structure of a wavefront from a distance point source passing a vertical

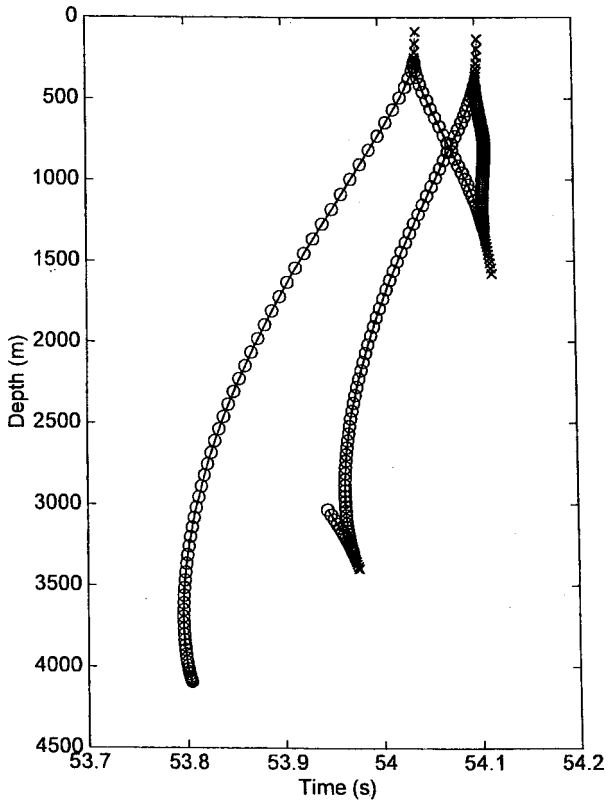


FIG. 1. Depth–time diagram at 80 km for a source at the minimum in a Munk sound-speed profile. The circles are for source ray angles between $\pm 15°$ in 0.1° increments. The crosses are for complex ray angles in 0.1° increments in the imaginary part.

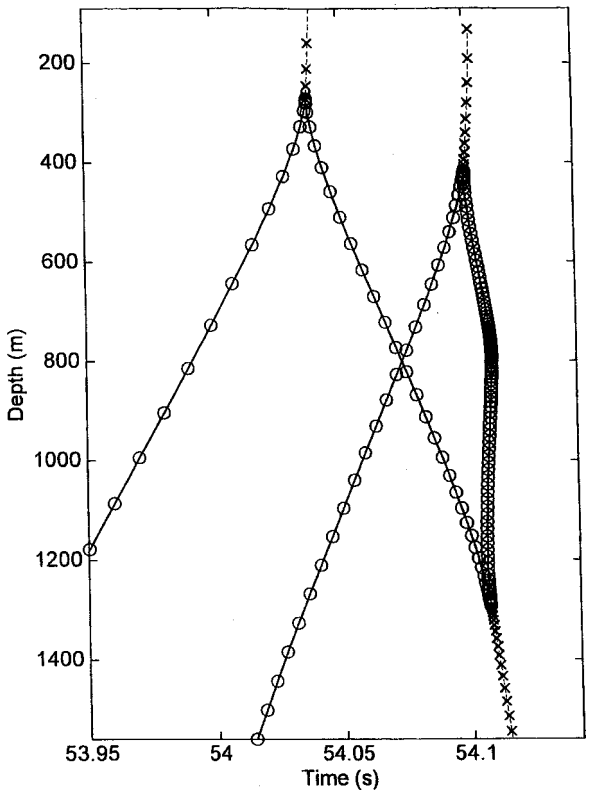


FIG. 2. Depth–time diagram at 80 km for a source at the minimum in a Munk sound-speed profile. The circles are for source ray angles with 0.1° increments. The crosses are for complex ray angles in 0.1° increments in the imaginary part. This figure is an expanded section of Fig. 1.

hydrophone array. The sharp turning points in the wavefront are where the corresponding rays touch a caustic.

The crosses in Figs. 1 and 2 show the extension of the wavefront beyond the caustics into the acoustic shadows using complex angles. The value at a caustic is found by adding a very small imaginary part (0.0001°) to the real ray angle nearest the caustic and iterating on the real part of the angle until the arrival depth is real. The other crosses are found by incrementing the imaginary part by 0.2° and iterating for real depth again. This procedure extends the wavefront into the shadow region. The field falls off rapidly in the shadow and the extrapolation into the shadow can be stopped when the parameter x of Eq. (52) exceeds a value of 3, after which the field in the shadow is negligible.

C. Complex rays in shadow zone

Waveforms calculated for a vertical array of hydrophones at a range of 80 km are shown in Figs. 3 and 4. The depth and time windows correspond to those of Fig. 2. The source pulse is a two-cycle cosine wave at center frequency f_0 filtered with a passband of $f_0/2$ to $3f_0/2$, where f_0 is 75 Hz and 120 Hz in Figs. 3 and 4, respectively. The first pulse on the waveforms at 800-m depth is an inverted replica of the source pulse.

The heavier lines in Figs. 3 and 4 are the waveforms obtained by the present wavefront method using complex rays in the shadow zones. The lighter lines are reference solutions found using the normal-mode code ORCA.[26]

The wavefront solutions in Figs. 3 and 4 were generated directly from the data of Fig. 2 by placing pulses at the appropriate arrival times. For a given receiver depth the pulse arrival times are found by interpolation of the curves of Fig. 2 at the desired depth. The numerical derivative of these curves gives $dz^*/d\theta_s$ and determines the pulse amplitude. The phase is determined from the number of turning points of the ray and the parameter δ of Eq. (43).

The particular formulas used to construct the waveforms depend on whether or not individual arrivals can be considered as isolated. For example, a horizontal line at 1000-m depth cuts the wavefront in four places. The time separation exceeds a quarter period in all cases, and all arrivals can be treated as isolated. Equation (42) is used to find the amplitudes and phases of the pulses, and they are placed at the appropriate times to give the waveform at 1000 m in Figs. 3 and 4. The first two pulses are well separated. The first has a phase of $-\pi$ and the second has a phase of $-\pi/2$ relative to the source pulse. This is because the first arrival has touched two caustics and the second has touched one. The third and fourth arrivals at 1000 m overlap in time and cannot be distinguished. The fourth arrival has large amplitude, as can be seen from the closeness of the circles, and dominates the final part of the waveform.

At a depth of 1400 m there are two arrivals in the time frame of Fig. 2. The first is an isolated pulse which is found using Eq. (42). The second arrival at 1400 m is in the shadow zone of the caustic at about 1300 m. The phase and amplitude of the caustic shadow pulse are found using Eq. (54).

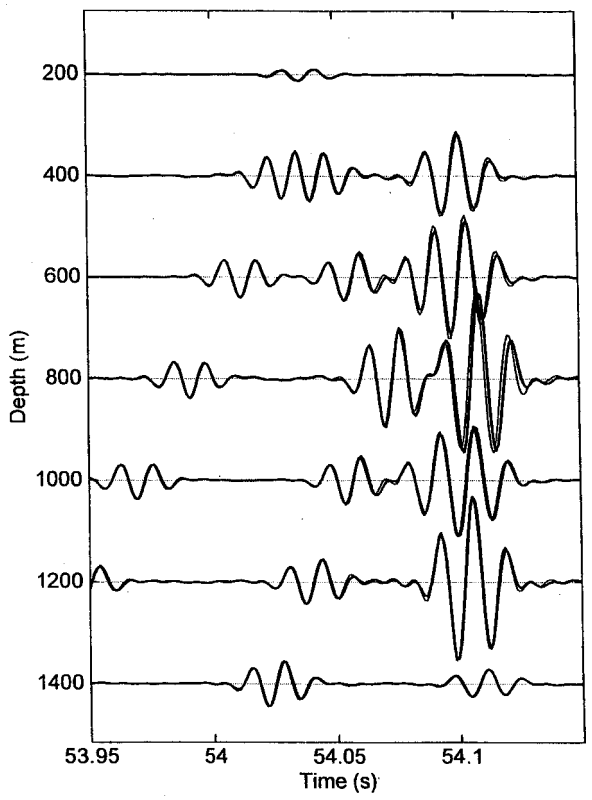


FIG. 3. Waveforms for hydrophones at the depths shown and a filtered two-cycle source pulse at 75 Hz. The depth-and time windows correspond to Fig. 2. The heavier line is the present wavefront model; the lighter line is the normal-mode reference solution.

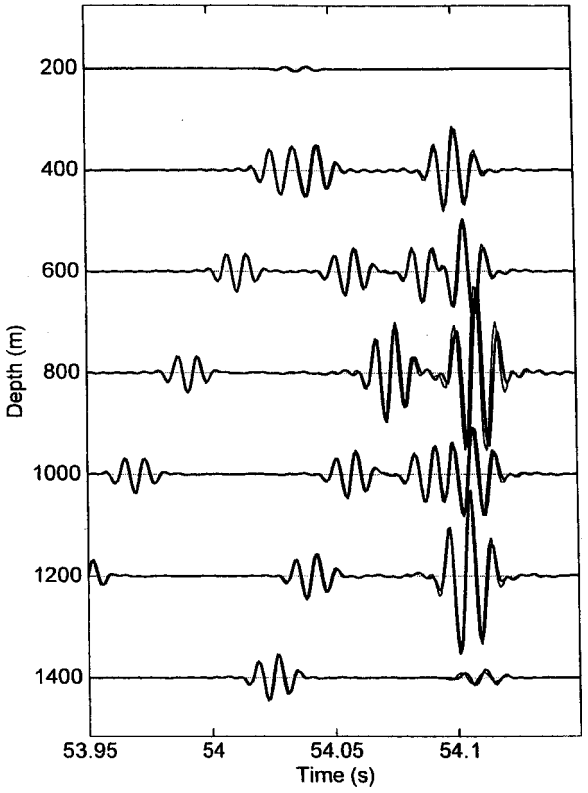


FIG. 4. Waveforms for hydrophones at the depths shown and a filtered two-cycle source pulse at 120 Hz. The depth-and time windows correspond to Fig. 2. The heavier line is the present wavefront model; the lighter line is the normal-mode reference solution.

It is interesting to compare Figs. 3 and 4. The arrival times of the pulses are the same in both figures, as expected, since the arrival time information provided by the depth–time diagram of Fig. 2 is the same at all frequencies. The amplitude information provided by Fig. 2 is also the same at all frequencies, and the amplitudes of corresponding real ray pulses are the same in Figs. 3 and 4. There is a difference in the shadow zones, however, because the parameter x of Eq. (52) increases with frequency. Therefore, the shadow zone field given by Eqs. (51)–(54) falls off faster in the shadow zone at higher frequencies. This effect can be seen by comparing relative amplitudes of the pulse at 200 m and the second pulse at 1400 m between Figs. 3 and 4. The effect is not seen in the second pulse at 400 m because it is very close to the caustic and the value of x of Eq. (52) is very small.

D. Approximate shadow calculations

The details of the transition from insonified zone to shadow zones can be seen in Figs. 5 and 6. Figure 5 shows a small section of the depth–time diagram of Fig. 2 in the vicinity of the caustic at a depth of 416 m and Fig. 6 shows the calculated waveforms at 120 Hz for the same depth–time window.

The caustic at 416-m depth has a shadow region which gives pulse arrivals in the top three waveforms of Fig. 6. The waveforms in this case were calculated using the approximate shadow formulas of Eqs. (61)–(64). The derivatives of the phase function ϕ of Eq. (61) were found by fitting a cubic to the function evaluated at six points bracketing the value at the caustic. The waveform at a depth of 400 m can be compared with that of Fig. 4. It is clear that the approximate shadow calculation used for Fig. 6 gives essentially the same results as the more correct complex ray calculation used for Fig. 4. Since the approximate shadow calculation is simpler and much faster, it will be used in future calculations.

A comparison of Figs. 5 and 6 shows that there are two real ray arrivals at a depth of 360 m and four real ray arrivals at a depth of 540 m. The first two arrivals at each depth in Fig. 6 are modeled by isolated single pulses. The two pulses strongly overlap as the first pulse at 360-m depth and then gradually separate into two distinct pulses at 540-m depth. The source cosine pulse has a positive peak at the center of the pulse, whereas inspection of the first two pulses at 540 m shows they correspond to $-$cosine and $+$sine pulses, respectively. Therefore, their phases are $-\pi$ and $-\pi/2$, respectively, as the first corresponding ray has touched two caustics and the second has touched one caustic.

The final arrival in the waveforms of Fig. 6 shows a smooth transition from the shadow through the caustic to the insonified region. The first three waveforms involve the caustic shadow. The next at 420 m has two real rays very close together within 1/50 of a period and described by Eqs. (44)–(47). The waveforms at 440, 460, and 480 m have two real rays arriving within a quarter period and are treated as a pair using Eqs. (48)–(50). For the waveforms at 500, 520, and 540 m, the two rays are treated independently using Eqs.

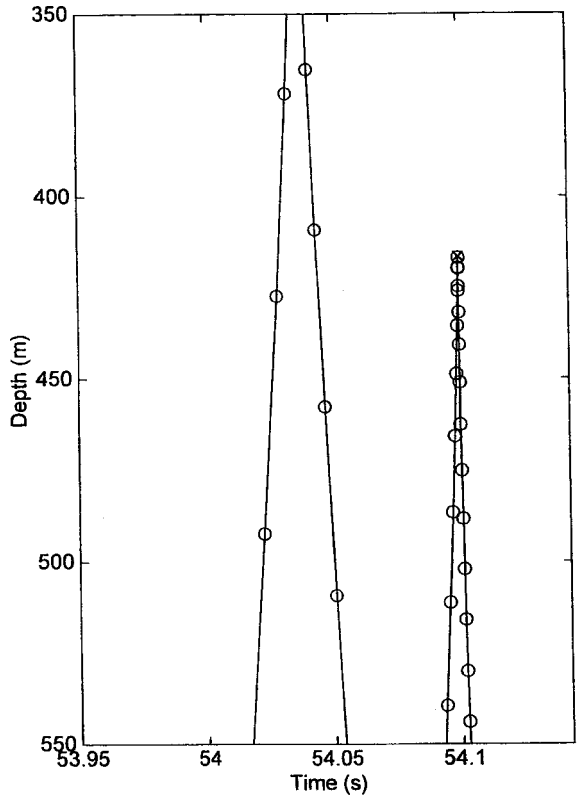


FIG. 5. Depth–time diagram at 80 km for a source at the minimum in a Munk sound-speed profile. The circles are for source ray angles in 0.1° increments. This figure is an expanded section of Figs. 1 and 2, but complex rays have been omitted.

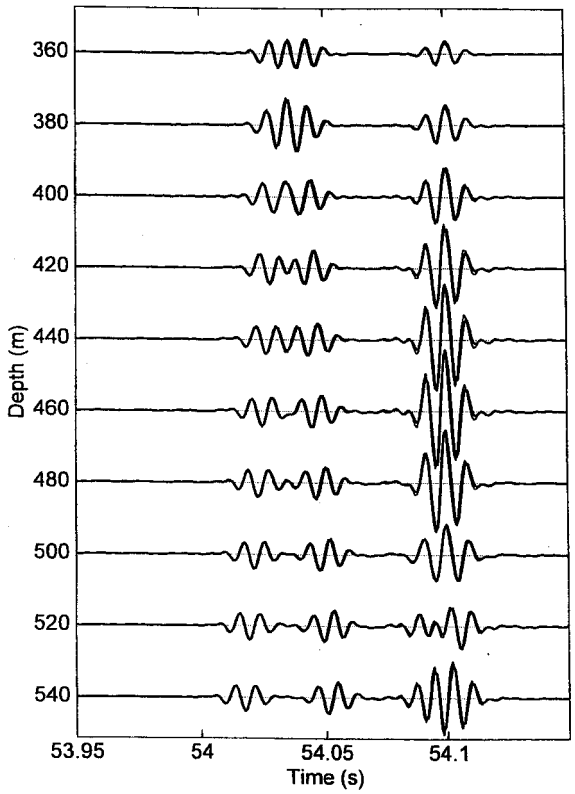


FIG. 6. Waveforms for hydrophones at the depths shown and a filtered two-cycle source pulse at 120 Hz. The depth-and time windows correspond to Fig. 5. The heavier line is the present wavefront model; the lighter line is the normal-mode reference solution.

(42) and (43), even though their pulses strongly overlap, interfering to give low amplitude at 520 m and high amplitude at 540 m.

E. Cusp

At 80 km the final ray arrivals at 800 m in Figs. 1 and 2 are converging to form a cusp. At a range of 83 km the cusp is passed; the depth–time diagram near the axis is shown in Fig. 7. The wavefront has folded over to produce two nearby caustics. The waveforms corresponding to the time window of Fig. 7 are shown in Fig. 8 for a source pulse at 75 Hz. As before, the heavier line is the wavefront result and the lighter line is the normal-mode reference solution. There are some slight differences in phase and amplitude, but otherwise the agreement is good.

The first part of the waveforms of Fig. 8, up to a time of 56.11 s, are formed by two isolated pulses arriving close together in time so that their waveforms overlap and interfere. They do not need to be treated as a pair because they do not correspond to rays with nearby launch angles. The two rays arrive simultaneously at 800 m and interfere to give a maximum. They arrive about half a period apart at 840 m and 760 m and interfere destructively to give overall low amplitude.

The second half of the waveforms of Fig. 8, after a time of 56.11 s at depths 760–900 m, were calculated using the cusp formulas of Eqs. (58), (59), and (61). The derivatives of the phase function ϕ of Eq. (61) were found by fitting a quartic to ten values of ϕ bracketing the angle θ_0 for which $\phi''(\theta_0)=0$. The cusp formula is only necessary for arrivals of nearby launch angle which arrive within a quarter period of each other. For the waveforms at 700-, 720-, and 740-m depth, the caustic shadow arrival is separated from the earlier real ray arrival by more than a quarter period and the approximate caustic shadow formulas were used to calculate the waveforms.

V. DISCUSSION AND CONCLUSIONS

The usual development of ray theory begins by assuming separation of phase and amplitude in the wave equation and leads to the eikonal and transport equations.[20] The theoretical development described in Secs. II and III above is an alternative derivation of the equations of ray theory.

The derivation in Secs. II and III proceeded by solving the wave equation using the Hankel transform and the WKB approximation followed by asymptotic evaluation of the resulting integral by the method of steepest descents. The advantage of the Hankel transform and WKB approach is that it leads to a summation over integral expressions for the field. Each term in the summation contains the phase and amplitude information in a single expression which can be evaluated for the appropriate number of saddle points.

The two approximations associated with this alternative derivation of the equations of ray theory are readily identified. The first is the use of the WKB expressions for the depth-separated solution of the wave equation. This is well known to be a good approximation when the change of

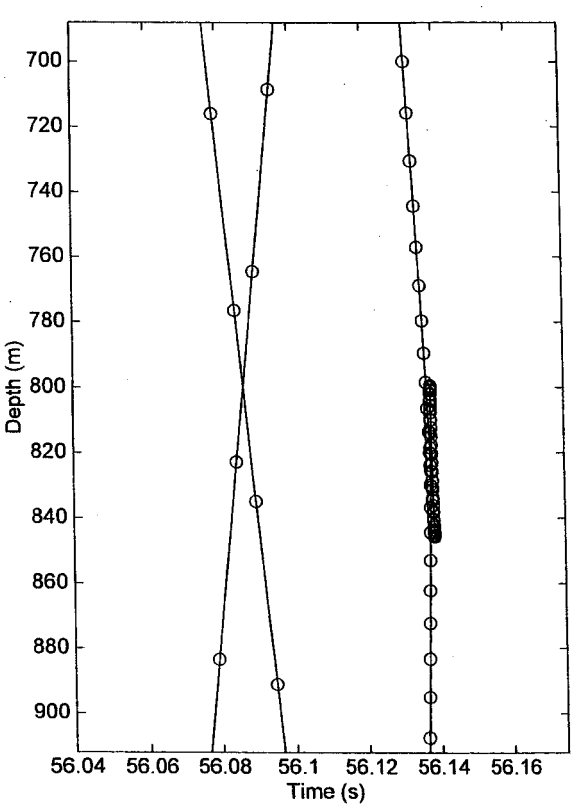


FIG. 7. Depth–time diagram at 83 km for a source at the minimum in a Munk sound-speed profile. The circles are for source ray angles with 0.1° increments. The wavefront folds back on itself between 800-and 844-m depth.

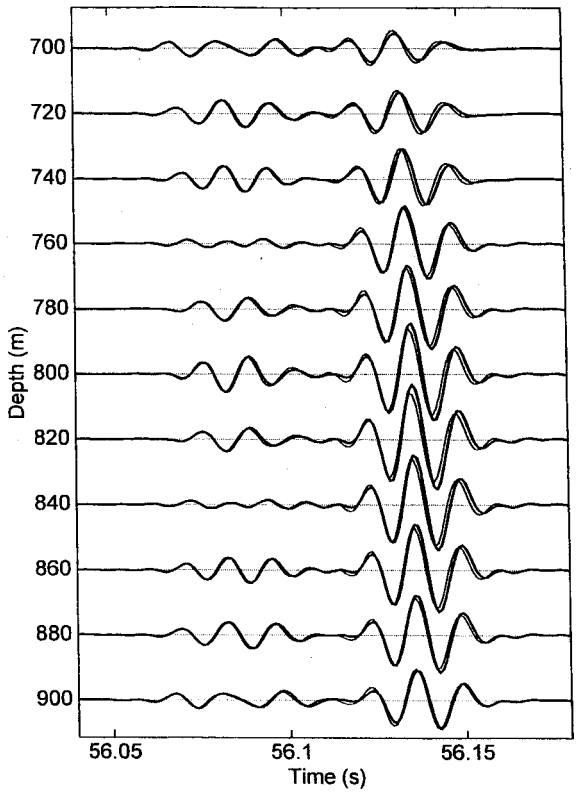


FIG. 8. Waveforms for hydrophones at the depths shown and a filtered two-cycle source pulse at 75 Hz. The depth-and time windows correspond to Fig. 7. The heavier line is the present wavefront model: the lighter line is the normal-mode reference solution.

physical parameters is small compared to the wavelength. The second approximation arises from the asymptotic evaluation of the integrals. It is also a good approximation because we use one, two, or three saddle points as appropriate for isolated rays, caustics, and cusps, respectively. Both approximations actually work well in underwater acoustics, and the resulting wavefront calculations are accurate to relatively low frequencies.

The results in Figs. 3, 4, 6, and 8 show that there is generally good agreement between the wavefront calculations and a normal-mode reference solution. The results show that ray theory and the wavefront method are able to accurately describe the acoustic field everywhere, including the vicinity of caustics and cusps.

The wavefront method has two distinct advantages. The first advantage is that the depth–time diagram gives an informative and physically intuitive picture of the arrival structure expected on a vertical hydrophone array. It shows at a glance where pulses are expected, and the density of the points on the wavefront gives an indication of the pulse amplitudes.

The second advantage is speed for broadband calculations. The ray trace leading to the depth–time diagram is simple and fast. There is no need to find eigenrays, and the iteration associated with eigenrays is avoided by simple interpolation on the depth–time curve. In contrast, the normal-mode solution is computation intensive because it requires finding the normal modes at the appropriate range of frequencies and a Fourier transform to obtain the waveforms.

The examples of the present paper are restricted to a range-independent case and rays which have turning points well clear of surface and bottom. This restriction avoids the lateral displacement of rays with turning points near boundaries[8] and allows an unambiguous comparison with a normal mode reference solution. The wavefront method is readily applied to range-dependent problems. Range dependence and the description of rays which turn near boundaries will be considered in a future paper.

We conclude that wavefront modeling is a fast, accurate method of finding acoustic waveforms.

ACKNOWLEDGMENTS

This work was partially supported by ONR Contract No. N00014-00-1-0138. The author is grateful for the hospitality of Dave Bradley and the Acoustics Division, ARL, Penn State while on sabbatical leave, and thanks Finn Jensen for supplying reference solutions in the early stage of this work.

[1] B. D. Seckler and J. B. Keller, "Geometrical theory of diffraction in inhomogeneous media," J. Acoust. Soc. Am. **31**, 192–205 (1959).
[2] B. D. Seckler and J. B. Keller, "Asymptotic theory of diffraction in inhomogeneous media," J. Acoust. Soc. Am. **31**, 206–216 (1959).
[3] M. A. Pedersen and D. F. Gordon, "Normal mode and ray theory applied to underwater acoustic conditions of extreme downward refraction," J. Acoust. Soc. Am. **51**, 323–368 (1972).
[4] L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1960).
[5] D. W. White and M. A. Pedersen, "Evaluation of shadow zone fields by uniform asymptotics and complex rays," J. Acoust. Soc. Am. **69**, 1029–1059 (1981).

[6] R. J. Urick, ''Caustics and convergence zones in deep water sound transmission,'' J. Acoust. Soc. Am. **38**, 348–358 (1965).
[7] D. A. Sachs and A. Silbiger, ''Focusing and refraction of harmonic sound and transient pulses in stratified media,'' J. Acoust. Soc. Am. **49**, 824–840 (1971).
[8] D. V. Batorsky and L. B. Felsen, ''Ray-optical calculation of modes excited by sources and scatterers in a weakly inhomogeneous duct,'' Radio Sci. **6**, 911–923 (1971).
[9] D. A. Sachs, ''Sound propagation in shadow zones,'' J. Acoust. Soc. Am. **51**, 1091–1097 (1972).
[10] E. L. Murphy and J. A. Davis, ''Modified ray theory for bounded media,'' J. Acoust. Soc. Am. **56**, 1747–1760 (1974).
[11] J. A. Davis, ''Extended modified ray theory in bounded and unbounded inhomogeneous media,'' J. Acoust. Soc. Am. **57**, 276–286 (1975).
[12] H. Weinberg, ''Application of ray theory to acoustic propagation in horizontally stratified oceans,'' J. Acoust. Soc. Am. **58**, 97–109 (1975).
[13] W. H. Munk and C. Wunsch, ''Ocean acoustic tomography: A scheme for large scale monitoring,'' Deep-Sea Res., Part A **26**, 123 (1979).
[14] P. F. Worcester, B. D. Cornuelle, J. A. Hildebrand, W. S. Hodgkiss, Jr., T. F. Duda, J. Boyd, B. M. Howe, J. A. Mercer, and R. C. Spindel, ''A comparison of measured and predicted broadband acoustic arrival patterns and travel time–depth coordinates at 1000-km range,'' J. Acoust. Soc. Am. **95**, 3118–3128 (1994).
[15] M. G. Brown, W. H. Munk, J. L. Spiesberger, and P. F. Worcester, ''Long-range acoustic transmission in the Northwest Atlantic,'' J. Geophys. Res. **85**, 2699–2703 (1980).
[16] M. G. Brown, ''Application of WKBJ Green's functions to acoustic propagation in horizontally stratified oceans,'' J. Acoust. Soc. Am. **71**, 1427–1432 (1982).
[17] M. G. Brown, ''The transient wave fields in the vicinity of the cuspoid caustics,'' J. Acoust. Soc. Am. **79**, 1367–1384 (1986).
[18] P. F. Worcester, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, B. M. Howe, J. A. Mercer, R. C. Spindel, J. A. Colosi, K. Metzger, T. G. Birdsall, and A. B. Baggeroer, ''A test of basin-scale acoustic thermometry using a large-aperture vertical array at 3250-km range in the eastern North Pacific Ocean,'' J. Acoust. Soc. Am. **105**, 3185–3201 (1999).
[19] G. V. Frisk, *Ocean and Seabed Acoustics* (Prentice-Hall, Englewood Cliffs, NJ, 1994).
[20] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (Springer, New York, 1993).
[21] L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1980).
[22] C. Chester, B. Friedman, and F. Ursell, ''An extension of the method of steepest descents,'' Proc. Cambridge Philos. Soc. **53**, 599–611 (1957).
[23] L. B. Felsen and N. Marcuvitz, *Radiation and Scattering of Waves* (Prentice-Hall, Englewood Cliffs, NJ, 1973).
[24] T. Pearcey, ''The structure of an electromagnetic field in the neighborhood of a cusp of a caustic,'' Philos. Mag. **37**, 311–317 (1946).
[25] W. H. Munk, ''Sound channel in an exponentially stratified ocean with applications to SOFAR,'' J. Acoust. Soc. Am. **55**, 220–226 (1974).
[26] E. K. Westwood, C. T. Tindle, and N. R. Chapman, ''A normal mode model for acousto-elastic ocean environments,'' J. Acoust. Soc. Am. **100**, 3631–3645 (1996).

A two-way parabolic equation that accounts for multiple scattering

Joseph F. Lingevitch,[a] Michael D. Collins, and Michael J. Mills
Naval Research Laboratory, Washington, DC 20375

Richard B. Evans
Science Applications International Corporation, Mystic, Connecticut 06355

(Received 8 February 2002; revised 3 May 2002; accepted 6 May 2002)

A two-way parabolic equation that accounts for multiple scattering is derived and tested. A range-dependent medium is divided into a sequence of range-independent regions. The field is decomposed into outgoing and incoming fields in each region. The conditions between vertical interfaces are implemented using rational approximations for the square root of an operator. Rational approximations are also used to relate fields between neighboring interfaces. An iteration scheme is used to solve for the outgoing and incoming fields at the vertical interfaces. The approach is useful for solving problems involving scattering from waveguide features and compact objects. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1490364]

PACS numbers: 43.30.Dr, 43.30.Gv [WMS]

I. INTRODUCTION

Many wave propagation and scattering problems can be solved efficiently with parabolic equation techniques,[1] which are based on one-way wave equations. This approach is widely used to solve propagation problems in media that vary gradually with range. For this application, the operator in the elliptic wave equation is factored into a product of incoming and outgoing operators, outgoing energy is assumed to dominate the incoming energy, and a parabolic wave equation for the outgoing field is obtained. Parabolic equation techniques can also be used to solve scattering problems. For this application, the normal derivative on a scattering surface is approximated by the transverse operator of the parabolic wave equation. This approach has been applied to derive two-way parabolic equations that account for single scattering.[2–4] In this paper, we derive a two-way parabolic equation that accounts for multiple scattering.

A range-dependent medium is approximated by a set of range-independent regions. In each region, the field is composed of a sum of outgoing and incoming fields. The outgoing fields at the beginnings of the regions and the incoming fields at the ends of the regions are obtained by iteration. Parabolic equation operators are used to enforce conditions across the vertical interfaces between regions and propagate the fields through the regions. This approach is similar to two-way coupled modes,[5] but the fields are not decomposed into modes. Parabolic equation solutions are presented for problems involving multiple scattering from waveguide features and compact objects. We describe the parabolic equation operators that are used in the two-way solution in Sec. II. We describe an iterative approach for solving the two-way problem in Sec. III. We present examples in Sec. IV.

[a]Electronic mail: jfl@aslan.nrl.navy.mil

II. PARABOLIC EQUATION OPERATORS

In this section, we discuss the basic parabolic equation formulas that are used in Sec. III to solve elliptic problems. We consider a two-dimensional problem and work in Cartesian coordinates, where the range x is the horizontal distance from a source and z is the depth below the surface. For range-independent problems, the complex pressure p satisfies

$$\frac{\partial^2 p}{\partial x^2}+\tilde{k}^2(1+X)p=2i\rho(z_0)\,\delta(x)\,\delta(z-z_0), \tag{1}$$

$$X\equiv\tilde{k}^{-2}\left(\rho\frac{\partial}{\partial z}\rho^{-1}\frac{\partial}{\partial z}+k^2-\tilde{k}^2\right), \tag{2}$$

where z_0 is the depth of a line source, $k=\omega/c$ is the wave number, ω is the circular frequency, c is the sound speed, ρ is the density, and $\tilde{k}$ is a representative wave number. The normal mode solution of Eq. (1) is

$$p(x,z)=\sum_n k_n^{-1}\phi_n(z_0)\,\phi_n(z)\exp(ik_n x), \tag{3}$$

$$\tilde{k}^2(1+X)\,\phi_n=k_n^2\phi_n, \tag{4}$$

where $\phi_n(z)$ and k_n^2 are the nth mode and eigenvalue.

Factoring the operator in Eq. (1) for $x>0$, we obtain

$$\left(\frac{\partial}{\partial x}+i\tilde{k}(1+X)^{1/2}\right)\left(\frac{\partial}{\partial x}-i\tilde{k}(1+X)^{1/2}\right)p=0. \tag{5}$$

The solution appearing in Eq. (3) satisfies the outgoing wave equation,

$$\frac{\partial p}{\partial x}=i\tilde{k}(1+X)^{1/2}p, \tag{6}$$

which corresponds to one of the factors in Eq. (5). Backscattered energy in a range-dependent medium satisfies the incoming wave equation,

 0001-4966/2002/112(2)/476/5/$19.00

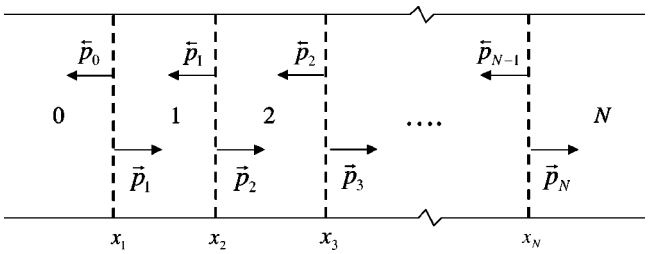


FIG. 1. Division of a range-dependent medium into a series of range-independent regions. Conditions for continuity of pressure and velocity are applied at the vertical interfaces between regions. The outgoing and incoming fields in each region are related by depth operators that are approximated by rational functions. The solution is obtained by an iterative sweep back and forth through the vertical interfaces.

$$\frac{\partial p}{\partial x} = -i\tilde{k}(1+X)^{1/2}p, \tag{7}$$

which corresponds to the other factor in Eq. (5).

It follows from Eq. (4) that

$$f(\tilde{k}^2(1+X))\,\phi_n = f(k_n^2)\,\phi_n \tag{8}$$

for an arbitrary analytic function f. From the modal expansion of the delta function,

$$\rho(z_0)\,\delta(z-z_0) = \sum_n \phi_n(z_0)\,\phi_n(z), \tag{9}$$

and Eqs. (3) and (8), we obtain

$$p(x,z) = \tilde{k}^{-1}(1+X)^{-1/2}\exp(i\tilde{k}x(1+X)^{1/2})\,\delta(z-z_0), \tag{10}$$

$$p(x+\Delta x,z) = \exp(i\tilde{k}\Delta x(1+X)^{1/2})p(x,z). \tag{11}$$

The self-starter[6] is an approach for generating initial conditions that is based on Eq. (10). The split-step Padé solution[7] is an approach for marching the solution through a range-independent region that is based on Eq. (11). Rational approximations for square roots, exponentials, and other functions have been used in the implementation of parabolic equation techniques.[8,9]

III. THE TWO-WAY PARABOLIC EQUATION

In this section, we derive a two-way parabolic equation that accounts for multiple scattering. A range-dependent medium is approximated by a sequence of $N+1$ range-independent regions that are separated by N vertical interfaces as shown in Fig. 1. The boundary between regions j and $j+1$ occurs at $x=x_{j+1}>0$. We assume that the medium is range-independent for $x<x_1$ and $x>x_N$. In the jth region, the field is decomposed into the outgoing and incoming fields. We define $\vec{p}_j$ to be the outgoing field at the beginning of region j and $\tilde{p}_j$ to be the incoming field at the end of region j. The incident field $\vec{p}_0$ is assumed to be known. From Eqs. (6), (7), and (11), we obtain the vertical interface conditions,

$$E_j\vec{p}_j + \tilde{p}_j = \vec{p}_{j+1} + E_{j+1}\tilde{p}_{j+1}, \tag{12}$$

$$D_j(E_j\vec{p}_j - \tilde{p}_j) = D_{j+1}(\vec{p}_{j+1} - E_{j+1}\tilde{p}_{j+1}), \tag{13}$$

$$D_j \equiv i\tilde{k}\rho_j^{-1}(1+X_j)^{1/2}, \tag{14}$$

$$E_j \equiv \exp(i\tilde{k}(x_{j+1}-x_j)(1+X_j)^{1/2}), \tag{15}$$

for conservation of pressure and x component of velocity at $x=x_{j+1}$. The unknowns are $\vec{p}_j$ for $1\leqslant j\leqslant N$ and $\tilde{p}_j$ for $0\leqslant j\leqslant N-1$. From Eq. (10), we obtain

$$E_0\vec{p}_0 = \tilde{k}^{-1}(1+X_0)^{-1/2}\exp(i\tilde{k}x_1(1+X_0)^{1/2})\,\delta(z-z_0) \tag{16}$$

for the case of a line source at $x=0$.

Since it is relatively difficult to solve Eqs. (12) and (13) directly, we derive an iterative approach. Rearranging Eq. (13) in two ways, we obtain

$$\vec{p}_{j+1} = E_{j+1}\tilde{p}_{j+1} + D_{j+1}^{-1}D_j(E_j\vec{p}_j - \tilde{p}_j), \tag{17}$$

$$\tilde{p}_j = E_j\vec{p}_j + D_j^{-1}D_{j+1}(E_{j+1}\tilde{p}_{j+1} - \vec{p}_{j+1}). \tag{18}$$

Substituting Eq. (12) into Eqs. (17) and (18), we obtain

$$2\vec{p}_{j+1} = 2E_j\vec{p}_j + (1 - D_{j+1}^{-1}D_j)(\tilde{p}_j - E_j\vec{p}_j), \tag{19}$$

$$2\tilde{p}_j = 2E_{j+1}\tilde{p}_{j+1} - (1 - D_j^{-1}D_{j+1})(E_{j+1}\tilde{p}_{j+1} - \vec{p}_{j+1}). \tag{20}$$

Rearranging Eqs. (19) and (20), we obtain the iteration formulas,

$$\vec{p}_{j+1} = \frac{\tau-2}{\tau}\vec{p}_{j+1} + \frac{2}{\tau}E_j\vec{p}_j + \frac{1}{\tau}(1 - D_{j+1}^{-1}D_j) \times (\tilde{p}_j - E_j\vec{p}_j), \tag{21}$$

$$\tilde{p}_j = \frac{\tau-2}{\tau}\tilde{p}_j + \frac{2}{\tau}E_{j+1}\tilde{p}_{j+1} - \frac{1}{\tau}(1 - D_j^{-1}D_{j+1}) \times (E_{j+1}\tilde{p}_{j+1} - \vec{p}_{j+1}), \tag{22}$$

where $\tau\geqslant 2$ is a convergence parameter. We have implemented Eqs. (21) and (22) with an approach that involves two sweeps through the interfaces for each iteration. On the outward sweeps, Eq. (21) is used to obtain improved estimates of the outgoing fields. On the inward sweeps, Eq. (22) is used to obtain improved estimates of the incoming fields. Increasing τ improves stability (at the expense of slower convergence) for problems involving relatively large sound speed and density contrasts. The depth operators D_j and E_j are implemented with rational approximations that are based on accuracy constraints for the propagating modes and stability constraints for the evanescent modes as described in Refs. 8 and 9.

IV. EXAMPLES

In this section, we illustrate the accuracy and capability of the multiple scattering parabolic equation solution. We consider two simple waveguide scattering problems and compare with exact and coupled mode solutions. We also consider a problem involving a compact scatterer in free space. For each of these 25 Hz examples, we take $\tau=4$ and use eight-term rational approximations.

Example A is a benchmark problem that has an exact solution. The waveguide has pressure release upper and lower boundaries. As Fig. 2 illustrates, the sound speed is

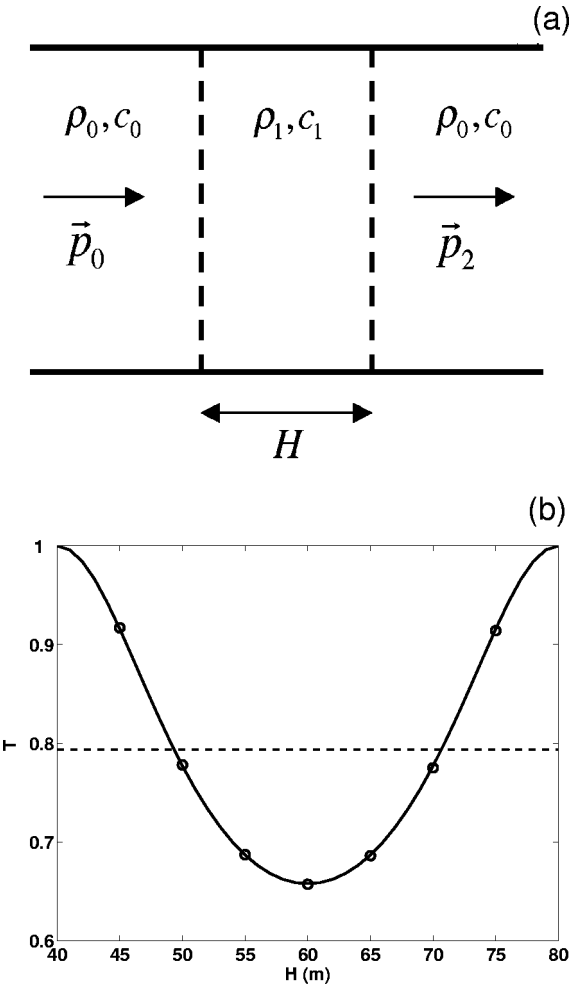


FIG. 2. Results for example A, which involves scattering in a waveguide composed of three homogeneous regions. (a) The geometry of the problem. The field is initialized with the first mode in the region to the left. The scattering region in the middle of is of width H. (b) Amplitude of the transmitted field to the right of the scattering region. Comparison of the exact solution (solid curve) with the multiple scattering (circles) and single scattering (dashed curve) parabolic equation solutions.

$c_0 = 1500$ m/s and the density is $\rho_0 = 1$ g/cm^3 everywhere except in a scattering region of width H, where the sound speed is $c_1 = 2000$ m/s and the density is $\rho_1 = 2$ g/cm^3. The field is initialized by the first mode, and the transmission coefficient through the inhomogeneity is computed as a function of H. Since the modes are identical in both regions, there is no coupling of energy between modes. The amplitude ratio of the transmitted field to the incident field is given by

$$T = \left| \frac{\vec{p}_2}{\vec{p}_0} \right| = \left| \frac{4\gamma_1\gamma_2}{(\gamma_1+\gamma_2)^2 - (\gamma_1-\gamma_2)^2 \exp(2ik_2H)} \right|, \qquad (23)$$

where $\gamma_j = k_j \rho_j^{-1}$ and $k_j = \omega c_j^{-1}$. Note that the maxima and minima of the amplitude ratio are spaced by quarter wavelengths in H. A comparison of solutions is shown in Fig. 2. The multiple scattering parabolic equation solution is in agreement with the exact solution. The amplitude of the single scattering parabolic equation solution is independent of H.

Example B compares parabolic equation and coupled mode solutions. This problem involves an isotropic source at

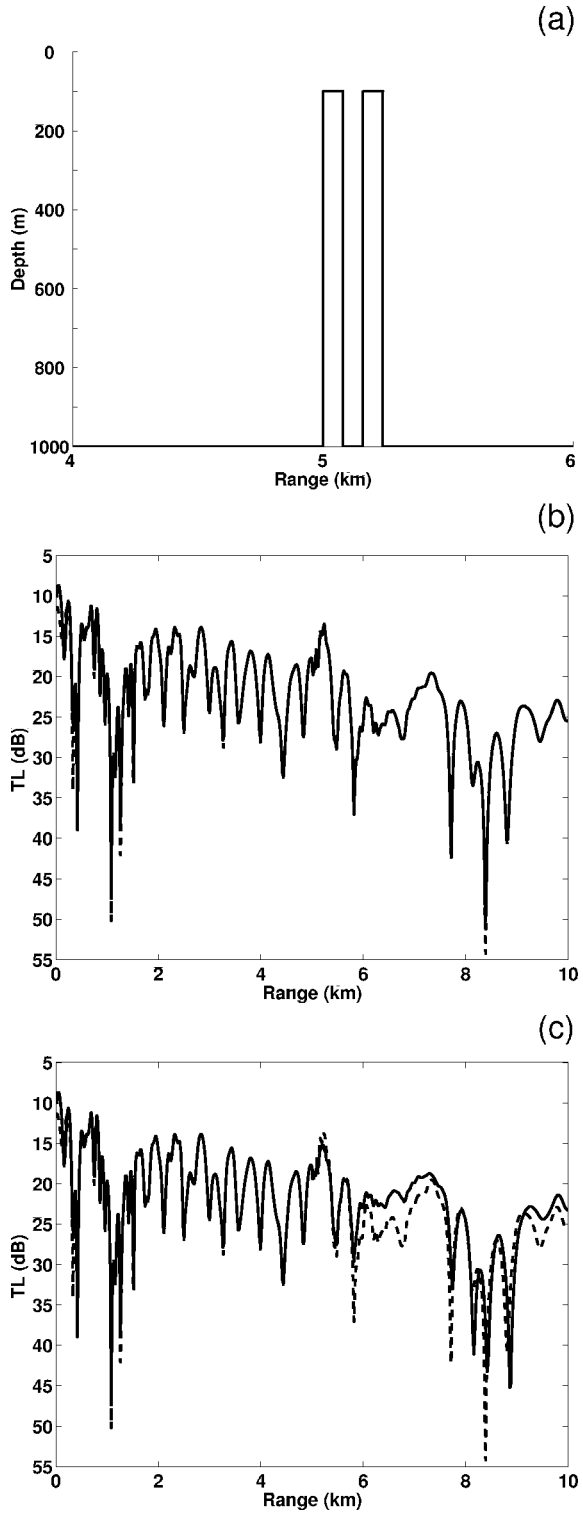


FIG. 3. Results for example B, which involves scattering in a waveguide with discontinuous changes in bathymetry between 100 and 1000 m. (a) The vertical interfaces are located at ranges of 5.0, 5.08, 5.16, 5.24 km from the source. (b) Comparison of the outgoing transmission loss at $z = 75$ m computed by coupled modes (dashed curve) and the multiple scattering parabolic equation (solid curve). (c) Comparison of the outgoing transmission loss at $z = 75$ m computed by the coupled modes (dashed curve) and the single scattering parabolic equation (solid curve).

$z = 500$ m in an ocean waveguide with four discontinuous changes in bathymetry between 100 and 1000 m as shown in Fig. 3. The sound speeds and densities are 1500 m/s and 1 g/cm^3 in the water and 1700 m/s and 2 g/cm^3 in the sedi-

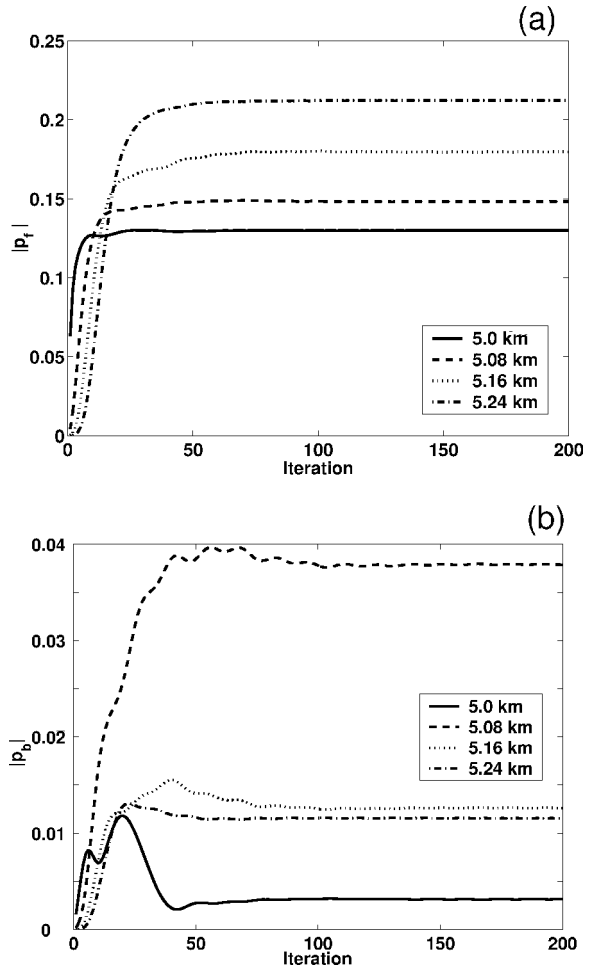


FIG. 4. Convergence of the multiple scattering parabolic equation solution at the vertical interfaces of example B. (a) Amplitude of the outgoing components at $z=75$ m. (b) Amplitude of the incoming component at $z=75$ m.

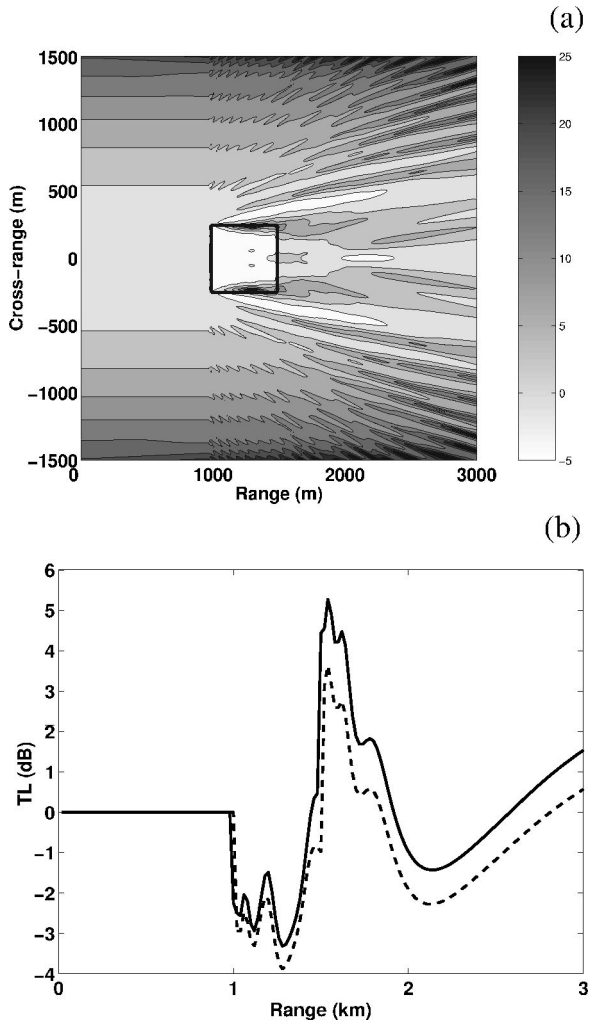


FIG. 5. Results for example C, which involves multiple scattering from a square object that is radiated by a Gaussian beam in a uniform background medium. (a) Intensity of the outgoing component. (b) Outgoing transmission loss at $z=0$ computed with the multiple scattering parabolic equation (solid curve) and the single scattering parabolic equation (dashed curve).

ment. The attenuation in the sediment is 0.1 dB/λ. Parabolic equation and coupled mode solutions appear in Fig. 3. The multiple scattering parabolic equation solution is nearly identical to the coupled mode solution. The single scattering parabolic equation solution breaks down for this problem. Convergence of the multiple scattering parabolic equation solution is shown in Fig. 4. Most of the outgoing and incoming components start to settle down after about 50 iterations. The slowest converging component settles down after about 100 iterations.

Example C involves scattering from a square inhomogeneity in a uniform background medium. The wave speeds and densities are 1500 m/s and 1 g/cm^3 in the background medium and 1700 m/s and 2 g/cm^3 in the 500 m square scatterer. The field is initialized with the Gaussian beam,

$$\vec{p}_0(z)=\exp\left(-\left(\frac{z}{1\ \mathrm{km}}\right)^2\right). \tag{24}$$

Absorbing layers are applied at the top and bottom of the computational grid to avoid artificial reflections. Multiple and single scattering parabolic equation solutions are shown in Fig. 5. Multiple scattering causes about a 1 dB shift in amplitude beyond the inhomogeneity.

V. CONCLUSION

The parabolic equation method has been generalized to solve problems in which multiple scattering is important. This method couples the incoming and outgoing field at a series of vertical interfaces by imposing continuity of pressure and horizontal particle velocity. The equations are solved with the parabolic equation method by iteratively sweeping the range-dependent region. The accuracy of the approach has been confirmed using exact and coupled mode solutions.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research.

[1] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, New York, 1994), pp. 343–412.
[2] M. D. Collins and R. B. Evans, "A two-way parabolic equation for acous-

tic backscattering in the ocean,'' J. Acoust. Soc. Am. **91**, 1357–1368 (1992).
[3] M. D. Collins, ''A two-way parabolic equation for elastic media,'' J. Acoust. Soc. Am. **93**, 1815–1825 (1993).
[4] M. J. Mills, M. D. Collins, and J. F. Lingevitch, ''Two-way parabolic equation techniques for diffraction and scattering problems,'' Wave Motion **31**, 173–180 (2000).
[5] R. B. Evans, ''A coupled mode solution for acoustic propagation in a waveguide with stepwise depth variations of a penetrable bottom,'' J. Acoust. Soc. Am. **74**, 188–195 (1983).
[6] M. D. Collins, ''The stabilized self-starter,'' J. Acoust. Soc. Am. **106**, 1724–1726 (1999).
[7] M. D. Collins, ''A split-step Padé solution for the parabolic equation method,'' J. Acoust. Soc. Am. **93**, 1736–1742 (1993).
[8] J. F. Lingevitch and M. D. Collins, ''Wave propagation in range-dependent poro-acoustic waveguides,'' J. Acoust. Soc. Am. **104**, 783–790 (1998).
[9] F. A. Milinazzo, C. A. Zala, and G. H. Brooke, ''Rational square-root approximations for parabolic equation algorithms,'' J. Acoust. Soc. Am. **101**, 760–766 (1997).

Surf-generated noise signatures: A comparison of plunging and spilling breakers

Steven L. Means and Richard M. Heitmeyer
Naval Research Laboratory, Code 7120, Washington, DC 20375

(Received 12 July 2000; revised 31 January 2002; accepted 6 May 2002)

Range-time-frequency distributions of surf-generated noise were measured within the surf zone during the SandyDuck'97 experiment at Duck, NC. A 24-phone, 138-m, bottom-mounted, linear array located along a line perpendicular to the shore at a depth of 1 to 3 m recorded the surf-generated noise. Concurrent video measurements of the location, size, and time-evolution of the individual breaking waves directly above the array were made from a nearby 43-m tower. Source level spectra are obtained by using a modified fast field program to account for water column and geoacoustic propagation from the distributed source region to an individual hydrophone. The length, location, and orientation of the leading edge of breakers are tracked in time from rectified video images. It is observed that the source levels from spilling breakers are lower (~5–10 dB) than those produced by plunging breakers that occurred during the same time period. Plunging breakers generated time-frequency signatures with a sharp onset while spilling breakers' signatures had a gradual low-frequency precursor. Range-time signatures of plunging breakers indicate a burst of acoustic energy while spilling breakers' signatures depict sound being generated over a longer time period with the source region moving with the breaking surface wave. [DOI: 10.1121/1.1491256]

PACS numbers: 43.30.Nb [SAC-B]

I. INTRODUCTION

This paper reports on differences between surf-generated noise signatures produced by the two predominately occurring breaker types that occur within the surf zone: spilling and plunging. In 1987, Papanicolaou and Raichlen[1] noted differences in wave tank measurements of noise signatures generated by the two breaker types, though there emphasis was mainly on wave and bubble cloud properties. Several papers have reported on field measurements of surf noise obtained either outside of the surf zone[2–5] or within a small region (<7 m cross-shore) of the surf zone.[6,7] The present paper reports measurements along a 138-m array deployed directly within the surf zone. This allows one to investigate properties of the surf-generated signatures over a substantial width of the surf zone.

The previous field measurements did not attempt to relate acoustic signature properties to quantitative video measurements of individual breaking waves nor characterize differences in breaker type. Although, one[6] did use video observations to qualitatively track the position of individual breakers. The present paper makes use of rectified video images to relate quantitative measurements of location, length, and orientation of the active region of breaking to the noise signatures generated by the two breaker types. Source level spectra are then obtained by accounting for propagation effects between the distributed source region and individual hydrophones within the array. These results are similar to the source level densities (SLDs) reported by Wilson[3] and Fabre,[4] except their results average over an along-shore portion of the coastline and thus do not apply to individual breakers.

The surf zone noise component of the SandyDuck experiment is described in Sec. II, which states the experimental layout and a description of the equipment used for acoustic and video measurements. The processing methods used in the analysis of the acoustic and video data are specified in Sec. III. Comparisons of received level signatures are made in Sec. IV. A definition of the source level per unit area is given in Sec. V. In Sec. VI we compare averaged source level spectra for plunging and spilling breakers. Section VII contains concluding remarks.

II. EXPERIMENTAL MEASUREMENTS

During September and October 1997, a multi-laboratory field experiment was conducted at the US Army Corps of Engineers Field Research Facility on the Outer Banks of North Carolina to study surf zone physics.[8] As part of that experiment, researchers from NRL conducted measurements of the space-time-frequency distribution of the noise generated by the individual breaking waves within the surf zone. The incident off-shore wave heights, measured in 8 m of water, varied from less than 0.5 m to over 3.5 m during the measurement time period. The extended duration of the experiment allowed noise measurements to be obtained during a wide variety of surf conditions.

A. Acoustic array

Acoustic measurements were made beneath the breaking waves on a 24-phone bottom-mounted array ranging from 60 to 198 m off the nominal shoreline along the 730 m alongshore. A diagram of the position of this array in the Field Research Facility (FRF) coordinate system is shown in Fig. 1. The hydrophones were spaced 6 m apart, as shown in Fig. 2, and were in water depths of 1–5 m. Since breaking surf is generated predominately by wave shoaling, the array was positioned to span an inner sandbar and a portion of an outer

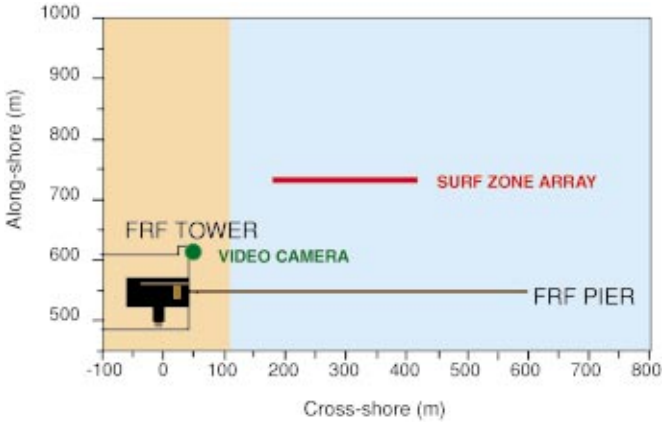


FIG. 1. Experimental layout for surf zone noise component of Sandy-Duck'97. The coordinate system shown is that used at the Army Corps of Engineers' Field Research Facility (FRF) in Duck, NC. The red line portrays the location of the 24-element, 138-m surf zone hydrophone array deployed along the 730 m along-shore. The video camera is located atop the FRF's 43-m tower.

sandbar. It is also evident in the figure that some of the hydrophones became buried during the experiment due to the sediment motion. The depth of burial was determined by taking the lowest bathymetry, measured daily, since the date of deployment at each hydrophone's cross-shore position along a nearby along-shore line. Although the covering of the hydrophones with sediment added difficulty in accounting from propagation effects from the source region to the receiver, it had the beneficial effect of preventing flow-induced pseudo-sound from corrupting the data for such shallow water measurements.

The hydrophones were designed for shallow water deployments with a two-pole high-pass filter at 40 Hz to prevent pressure fluctuations due to waves passing overhead from clipping their internal preamplifier. The hydrophone signals were brought ashore via double armored cable, pre-

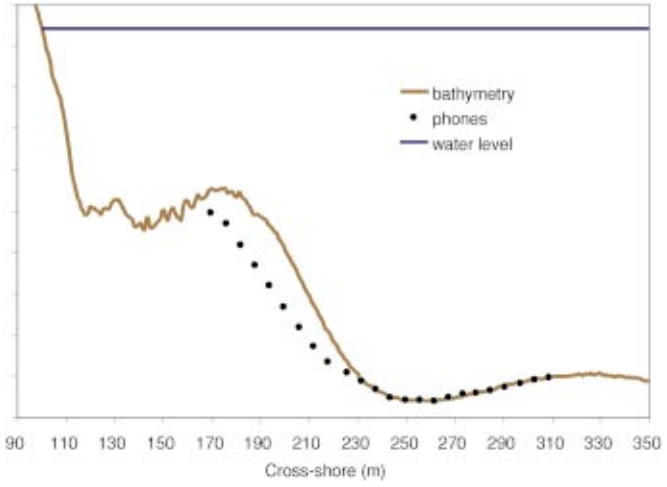


FIG. 2. Bathymetry measurement for 21 October 1997. The individual elements of the hydrophone were positioned to span an inner sandbar and a portion of an outer sandbar. As depicted in the figure, the depth of burial for individual hydrophones was estimated as the lowest bathymetry measured (daily) since the date of deployment. The tide level, illustrated in blue, was 0.694 m above the NGDV during the time period of measurement for the presently analyzed data.

whitened with a single-pole high-pass filter at 700 Hz, and digitized at a sampling rate of up to 50 kHz on an ICS System-1000 and recorded on digital tape. The data presented here were sampled at 12 kHz. In addition to the acoustic signals a time-code signal (IRIG-B) was also recorded simultaneously.

B. Video

Concurrent with the acoustic measurements, video recordings of individual breaking waves were made from a 43-m tower located nearby on the shore. A remote-controlled black-and-white camera, enclosed in a weathertight enclosure, captured images onto VHS video tape at a rate of 30 frames per second (fps). A time-code signal was also recorded on the audio track so that the video measurements could be synchronized with the acoustic measurements. The camera was oriented such that all of the 24 hydrophones of the surf zone array were within view as indicated in Fig. 3. Since photogrammetric transformations[9] from image coordinates to ground coordinates are to be performed, the exact position of the camera was surveyed with a laser transit. Ground control points were also needed for the photogrammetric transformations. These were obtained by observing the position of a Global Positioning System (GPS) antenna aboard the Army Corps of Engineers' Coastal Research Amphibious Buggy (CRAB)[10] during some of its daily survey operations, Fig. 4.

III. DATA ANALYSIS

The data chosen for analysis here were taken on 21 October 1997 during a time period with a 1.7 m significant wave height (four times the standard deviations of water surface displacement) and a tide level of 0.694 m above the 1929 National Geodetic Vertical Datum (NGVD). The surf was characterized by well-formed swell-induced breakers. The breaker type was predominately plunging, with an occasional spilling breaker.

A. Acoustic processing

Signals from the 24 hydrophones of the surf zone array were sampled at 12 kHz; 0.34-s time intervals were Hanning shaded and transformed to obtain range-time-frequency volumes of the surf-generated noise field. A simple detection algorithm selected surf noise signatures from the other noise sources. The selection algorithm consisted of three steps: (1) normalizing each frequency band by its 3-min time average; (2) setting a threshold on the summation over frequency for each time step; and (3) rejecting signatures of short duration. The selected received level signatures can then be compared for different breaker types or processed along with the video measurements to obtain source level spectra.

B. Video processing

The video and IRIG-B audio track recorded on VHS tape were played through a time-code reader to insert a time stamp onto the video images, then captured at 4 fps via a framegrabber in a Macintosh computer using the public domain NIH Image program (developed at the U.S. National

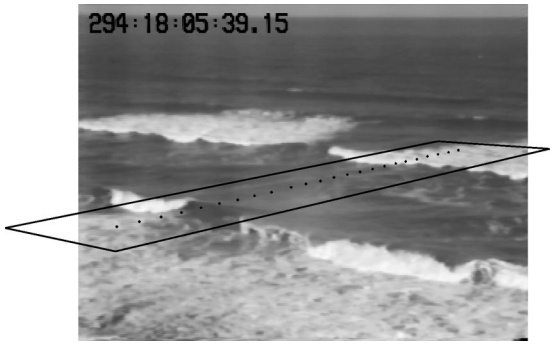


FIG. 3. A video snapshot of the surf zone during the time period of the presently analyzed data. The locations of the individual hydrophones of the surf zone array are indicated by the black circles. The portion of the image to be rectified for quantitative measurements is indicated as the circumscribed region. The time period was characterized by well-formed, swell-induced breakers, predominately of the plunging type.

Institutes of Health and available on the Internet at http://rsb.info.nih.gov/nih-image/). The breaker type of individual waves was then determined through observation.

Portions of the captured images, as outlined in Fig. 3, were imported into MATLAB for rectification to the FRF coordinate system. The photogrammetric transformation parameters needed for this rectification were obtained through a procedure similar to that described by Holland *et al.*[9] Once the images had been rectified, quantitative measurements of the size, location, and orientation of individual breaking waves, as seen in Fig. 5(a), could then be obtained.

It is commonly believed that sound is generated within the actively breaking region of the wave.[6] Therefore, one should confine measurements to that of active breaking. Since breaking surf leaves a trailing foam residue on the surface, which may remain for several seconds, it is expedient to subtract a preceding image from the current image to obtain an image representation of the active breaking region, as seen in Fig. 5(b). NIH Image has the automated capability of measuring the location, orientation, and major and minor axes length (of an ellipse of equal area) of the active breaking regions as a function of time. These measurements can then be used with the acoustic measurements in determining the source level of the observed distributed source regions.

FIG. 4. Video image of CRAB during one of its daily bathymetry measurement excursions. The position of the GPS antenna mounted aboard the CRAB was used as a ground control point (GCP) during the photogrammetric transformations. The position of the antenna within successive images (indicated by black dots) served as multiple GCPs necessary for accurate transformations.

IV. RECEIVED LEVEL COMPARISONS

A 3-minute time period containing ten distinct breaking surf occurrences, over the hydrophone at 181.8 m cross-shore, was processed to determine differences in the noise signatures of plunging and spilling breakers. It was determined through video observations that the first (~7 s), seventh (~89 s), and eighth (~145 s) signatures were generated by spilling breakers. The remaining seven were generated by plunging breakers. Figure 6 shows the range-time signatures at 300 Hz for the full range (138 m) of the surf zone array. The dark horizontal bands are due to hydrophones which had ceased to function due to the harsh surf zone environment. The signatures of the largest breakers begin at the seaward (upper) end of the array and propagate shoreward leaving a trace whose slopes yield the propagation velocity of the incoming waves. However, most of the breakers cease to break within the region between the inner and outer sandbars. Measurements of slope of the signatures give a velocity of roughly 5 m/s, which is consistent with video measurements of the wave propagation velocities observed in the surf zone.

Figure 7 shows an expanded view of the range-time signature for a plunging breaker at the shoreward end (170–212 m cross-shore) of the surf zone array. A strong pulse at 181.8 m is evident in the signature. It was determined from video observations that the overturning wavecrest began plunging

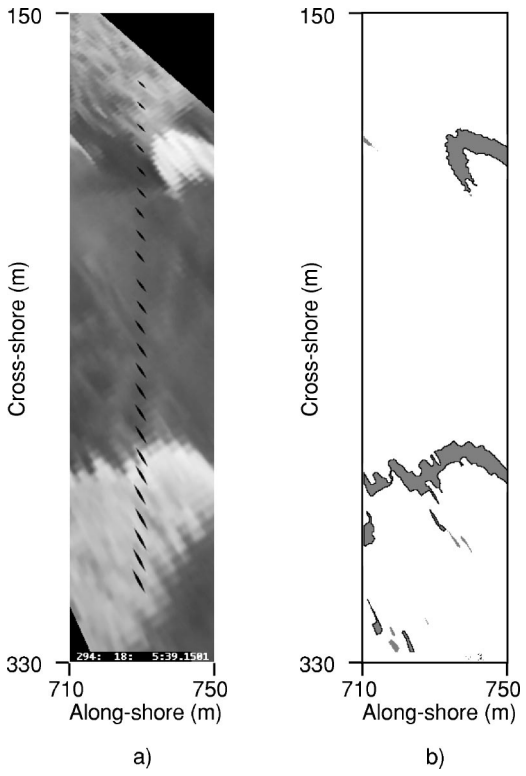


FIG. 5. The rectified image on the left (a) shows an example of two breaking waves propagating towards the beach (top of the image). The black ovals along the vertical centerline are the rectified dots (see Fig. 3) showing the position of the individual hydrophones. The image on the right (b) shows the active breaking region of the same two breaking waves. They were obtained by subtracting two successive images to yield a representation of the actively breaking region. This image is processed within NIH image to obtain measurements of the region's location, size, and orientation for later use in synthesizing with acoustic measurements.

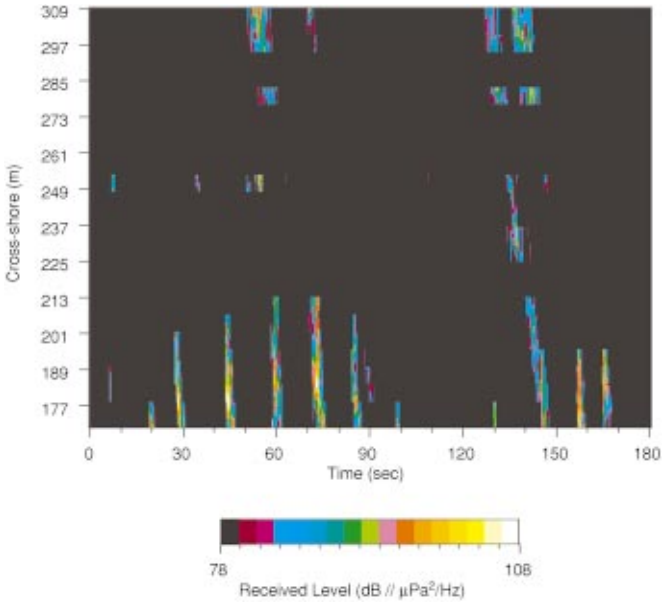


FIG. 6. Range-time signatures of the received levels measured along the full range of the surf zone array at 300 Hz with a 3-Hz bandwidth. Ten events are seen to occur over the third hydrophone (181.78 m) during the time period. The first (~7 s), seventh (~89 s), and eighth (~145 s) signatures were generated by spilling breakers. The remaining were generated by plunging breakers. The dark horizontal bands are due to hydrophones that ceased to operate in the harsh surf zone environment. The downward slope of the signatures corresponds to the phase velocity of the incoming surface gravity waves (~5 m/s).

into the water's surface directly above the third hydrophone (at 181.8 m cross-shore). The portion of the signature seaward of this cross-shore position is generated by acoustic propagation of the energy producing the vertical structure of this signature. Acoustic velocities yield a trace of nearly infinite slope at the shown time resolution. This vertical structure can be observed in the other plunging breakers' signatures in the shoreward portion of Fig. 6 as well.

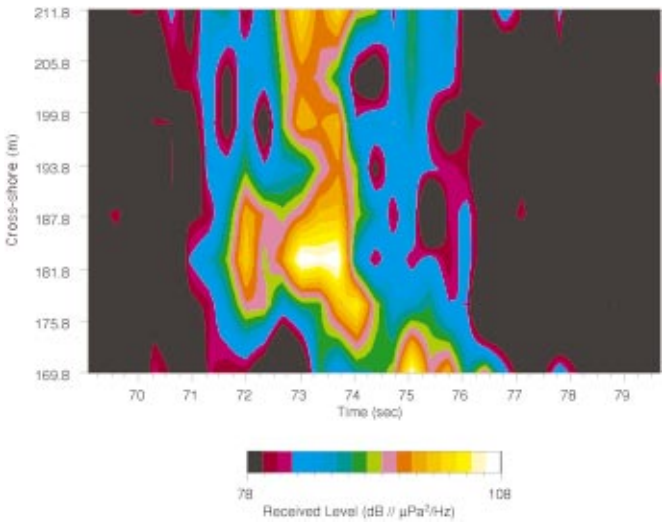


FIG. 7. An expanded view of the shoreward component of a range-time signature generated by a plunging breaker at 300 Hz with a 3-Hz bandwidth. The vertical structure of the signature is due to sound being generated directly over the third hydrophone (181.8 m cross-shore) with seaward propagation of the energy at acoustic velocities, which yield a trace of nearly infinite slope at this time resolution.

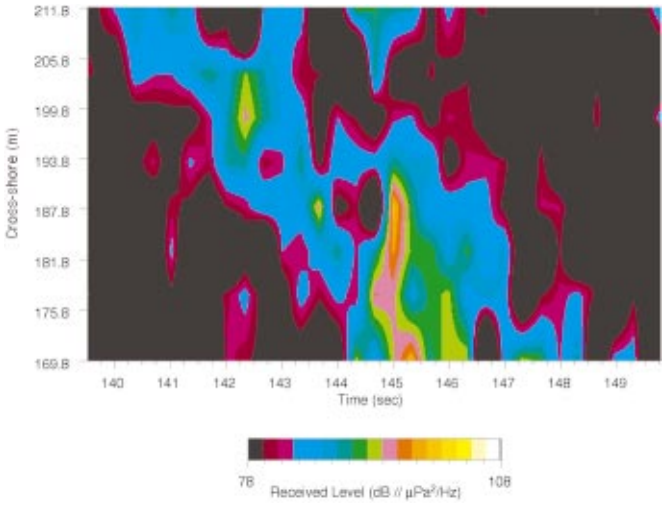


FIG. 8. An expanded view of the shoreward component of a range-time signature generated by a spilling breaker at 300 Hz with a 3-Hz bandwidth. In comparison to the plunging breaker's signature, the sound is produced over a longer time period. The source region progresses with the propagation velocity of the surface gravity wave, which yields a signature with a downward slope from top to bottom in time. The acoustic energy level appears to remain relatively steady for the duration of the passage over the shoreward portion of the array.

The less steep slope of the spilling breaker's signature, Fig. 8, is due to the sound being generated as the source region moved over the array with the propagating surface wave. The energy level appears to remain nearly steady for the duration of the passage over this portion of the array. Although this breaker began breaking seaward of the array, the sudden onset of breaking exhibited in the plunging breaker's signature was not observed in the other spilling breakers whose onset occurred over the array. It was visually observed that the spilling breakers during this time period generally began breaking on the inner bar seaward of the cross-shore

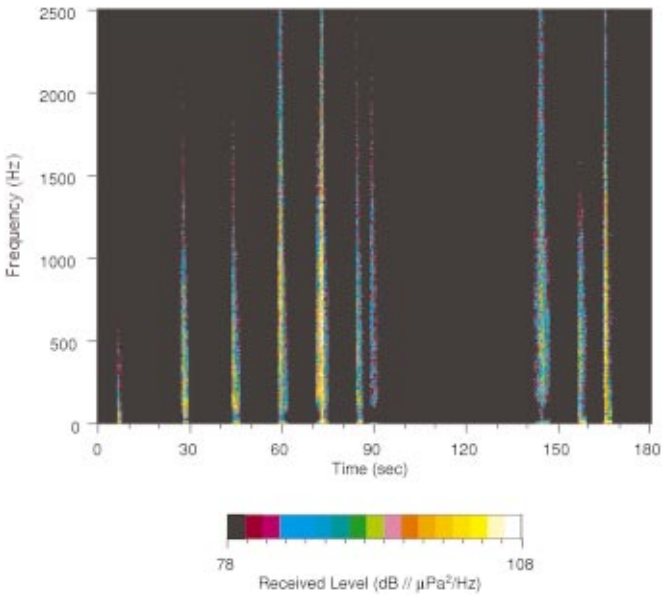


FIG. 9. Time-frequency signatures for the received level measured on the third surf zone hydrophone at 181.8 m cross-shore during a 3-min period. The signatures are broad band in nature with only the largest events having significant energy levels above 2500 Hz.

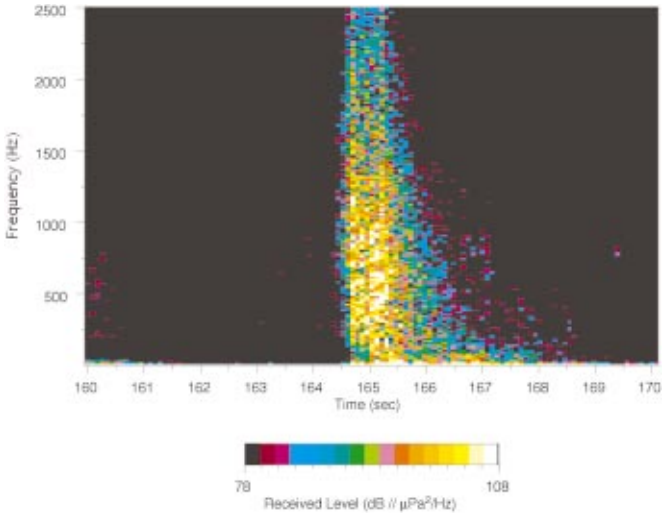


FIG. 10. Expanded view of the tenth time-frequency signature (in previous figure) generated by a large plunging breaker. The time resolution was increased, from the previous figure, to 0.085 s. A sharp onset is characteristic of plunging breakers and is due to the impact of the overturning jet of water plunging in to the water surface. The low-frequency tail is observed in many of the signatures independent of breaker type.

position at which plunging breakers began. It was also noted that the plunging breakers quickly evolved into ''borelike'' propagation as mentioned in Deane's description of the evolution of breaking events.[6] This phenomenon reveals itself in the less steep portion of the plunging breaker's signature near the shoreward end of the array in Fig. 7. The spilling breakers appeared to retain their ''spilling'' characteristics longer before eventually evolving into the ''borelike'' propagation as well.

Characteristic differences between the two breaker-type signatures may also be seen in time-frequency signatures. Figure 9 shows a time-frequency surface of the noise signals received at the hydrophone at 181.8 m cross-shore (third from shore), which was within the region of breaking occurring over the inner sandbar. The hydrophone was buried

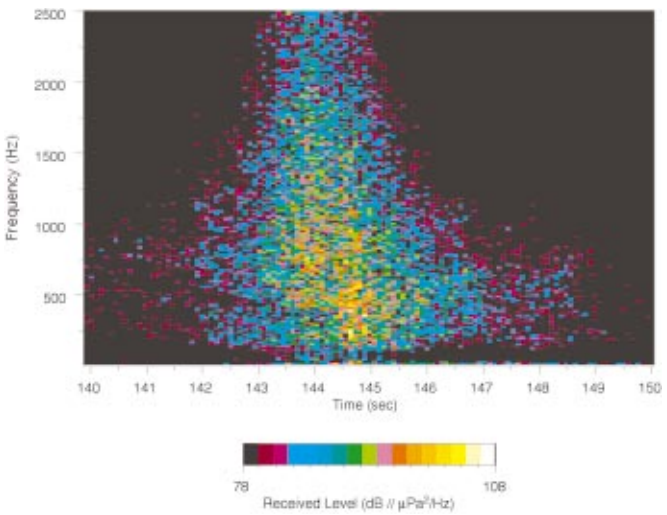


FIG. 11. Expanded view of a time-frequency signature generated by a large spilling breaker, eighth signature in Fig. 9. The time resolution is 0.085 s, as in the previous figure. A low-frequency precursor and tail are evident in the spilling breaker's signature.

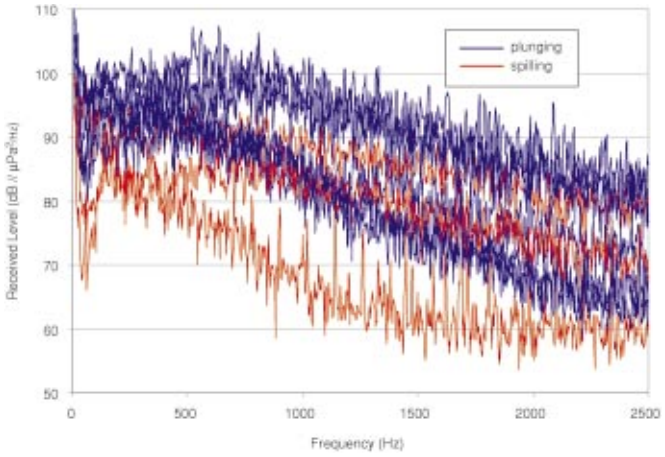


FIG. 12. Time-averaged received level spectra recorded on the third surf zone hydrophone (181.8 m cross-shore). The three red curves indicate spectra generated by spilling breakers. The remaining seven blue curves are spectra generated by plunging breakers. The spectra for plunging breakers appear to separate into two groups above 750 Hz, with larger events being roughly 10–15 dB higher in level.

roughly 0.6 m within the fine sand bottom, as depicted in Fig. 2. The signatures are broad band with energy above 2500 Hz for the larger events. However, higher energy levels are predominately at frequencies below 1 kHz. The duration of the signatures received as the breaker propagates overhead is on the order of several seconds.

The differences between the two breaker-type signatures may be seen in comparing the enlargements of two signatures in Fig. 10 (tenth signature) and Fig. 11 (eighth signature). The acoustic time series of these two signatures were processed to obtain higher time resolution (0.085 s). The signature in Fig. 10 was generated by a large plunging breaker. The sharp onset of the signature is characteristic of the impulsive nature of plunging breakers. The low-frequency tail, evident in this signature, is observed in most of the signatures independent of breaker type.

The spilling breaker's signature, Fig. 11, has both a low-frequency precursor and tail. At present it is unclear if this is a source or propagation phenomenon. The large bubble cloud

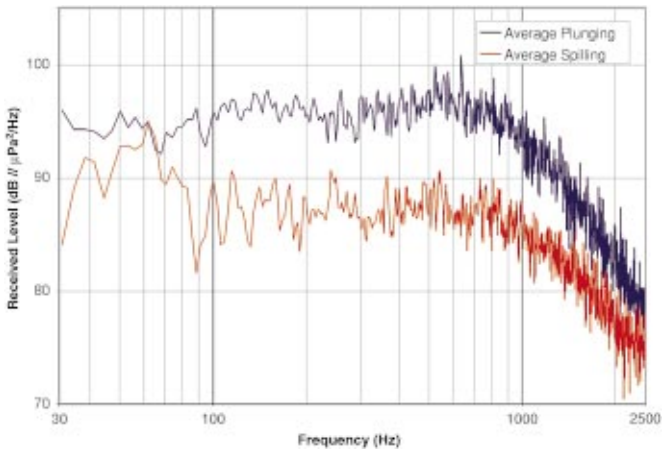


FIG. 13. Power-averaged received level spectra for both the plunging and spilling breakers which occurred during the 3-min interval. Each spectrum is generated by averaging over each individual, seven plunging and three spilling, breakers' spectra time-averaged over the duration of the signature.

entrained due to breaking could explain the attenuation of higher frequencies in the received signal after the breaker has passed overhead. In the absence of ambient bubble measurements, it is unclear whether bubbles may explain the absence of high frequencies in the precursor received ahead of the breaker's passage.

Time-averaged received-level spectra for all ten signatures at 181.8 m cross-shore are displayed in Fig. 12. Each 0.34-s time step for the duration of each breaker's signature was averaged to obtain a characteristic spectra for each event. The three spilling waves' spectra are displayed in red along with the seven plunging wave spectra in blue. The plunging breakers' spectra above 750 Hz appear to fall into two groups. Those belonging to the largest events being 10–15 dB higher in received level than the more moderated events. The two largest spilling spectra (seventh and eighth) were considerably lower in level across the full frequency band than the large plunging breaker spectra. The spilling spectra contained even relatively lower levels at frequencies below 750 Hz. For comparisons with later results, the power-averaged spectra for all of the events of each breaker type, seven plunging and three spilling, are shown in the log-log plot of Fig. 13. Although distinct differences between the spectra can be seen, it is necessary to account for gross features of the propagation between the source regions and the hydrophone to make quantitative comparisons.

V. SOURCE LEVEL

In the shallow water of the surf zone the acoustic propagation environment is quite complex. The analysis described here is meant to account for the gross features of the acoustic propagation and defines an effective source level so that differences between the spectra of the two breaker types may be compared. Although it is well known that bubbles strongly affect the propagation characteristics of sound, the source level definition given here categorizes these effects as properties of the source, as they are not considered in the propagation environment. A sheetlike distribution of closely spaced (0.01 m) incoherent point sources over the first meter beneath the water surface is chosen, in the absence of bubble plume depth measurements, to represent the sound generation region of the breaker. The location, orientation, and length of the sheet as a function of time is obtained from the video analysis as described above. This definition is similar to that chosen by Deane[6] except for neglecting the thickness of the sheet which would have little effect on the results since one is ''averaging'' the propagation effects over a considerable length of the observed wavefronts (10–30 m).

TABLE I. Table of acoustic properties of sediment layers at the location of the surf-zone hydrophone at the 181.8-m cross-shore. The depth values are relative to the water–bottom interface. The properties vary linearly from the upper surface to the lower, unless stated as being constant.

Layer	Depth (m)	Speed (m/s)	Attenuation (dB/m/kHz)	Density (g/cm^3)
water column	−1.96	1518.1	0.00	1.03
	0.0	1518.5	0.00	1.03
1	0.0	1720.0	0.50	1.90
	1.1	1800.9	0.49	1.90
2	1.1	1860.3	0.45	1.96
	6.6	1910.9	0.34	1.97
3	6.6	1954.5	0.34	2.02
	9.5	1964.9	0.32	2.03
4	9.5	1778.1	0.45	1.84
	15.4	1790.8	0.41	1.85
5	15.4	1933.6	0.29	1.99
	24.5	1947.4	0.26	2.01
6	24.5	2050.0	0.18	2.09
	50.0	2081.6	0.18	2.12
7	100.0	2141.6	0.17	2.17
8	200.0	2251.6	0.16	2.20

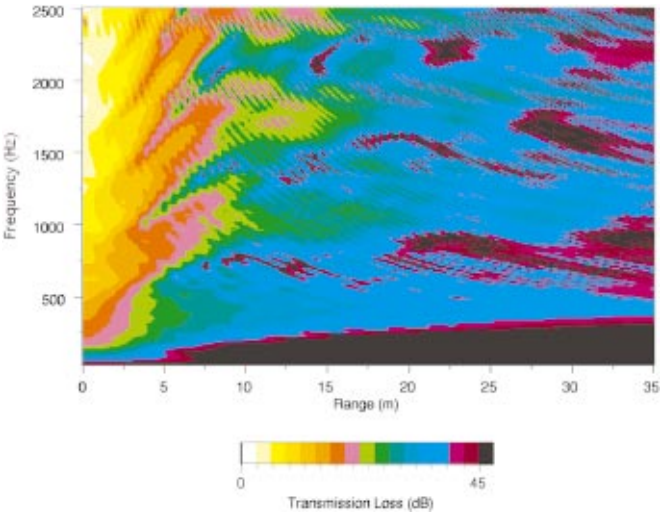


FIG. 14. Horizontal range-frequency calculations of the transmission loss in propagating from a point source, 0.15 m beneath the water's surface, to the third surf zone hydrophone (181.8 m cross-shore). Results from this calculation for various frequencies, depths, and ranges are used in estimating the source level per unit area spectra for breaking waves. For horizontal ranges over 10 m the transmission loss rises 12 dB per doubling of distance for frequencies below ~100 Hz and roughly 8 dB per doubling of distance at frequencies over 1500 Hz.

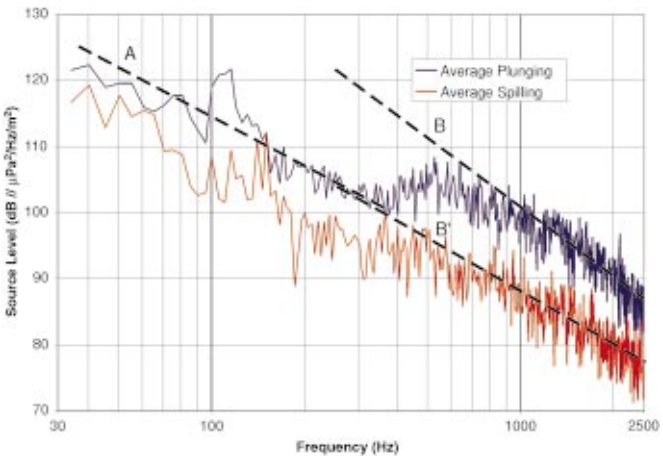


FIG. 15. Power-averaged source level spectra for both the plunging and spilling breakers which occurred during the 3-min interval. Each spectrum is generated by averaging over each individual, seven plunging and three spilling, breakers' spectra time-averaged over the duration in which a portion of the breaking region is within 5 m of the third surf zone hydrophone (181.8 m cross-shore). The slope (line A) of the spectra of both the plunging and spilling breakers drop off at 25 dB per decade below 300 Hz. At frequencies above 1 kHz the plunging waves spectra drop off at 35 dB per decade (line B) and spilling wave spectra drop off at 27 dB per decade (line B′).

The present analysis uses a modified version of Kutschale's range-independent fast field program[11] (FFP) to determine the transmission loss for the relatively short propagation distances, including geoacoustic effects of the bottom sediments, from the sheet distribution of point sources to individual hydrophones of the surf zone array. Although the FFP is capable of handling shear, it is neglected here. The FFP is modified by implementing the exact expression of the Hankel functions, as opposed to the long-range approximation usually implemented. A general introduction to fast-field programs may be found in the text by Jensen *et al.*[12] The analysis undertaken considers wave events only during the time period in which a portion of the breaking is within 5 m of the chosen hydrophone. Thus, the use of a range-independent propagation model is justified.

The geoacoustic properties of the sediments are derived, using methods developed by Hamilton,[13] from core samples as published by Fabre.[4] The properties at each hydrophone's position were obtained by a linear interpolation between core samples at the shoreline and 750 m offshore. Table I shows the acoustic properties for the various sediment layers used in determining the propagation loss to the surf-zone phone at 181.8 m cross-shore. Figure 14 shows a typical calculated transmission loss for a horizontal source–receiver distance up to 35 m at frequencies between 30 and 2500 Hz. At ranges over 10 m and frequencies below ~100 Hz the transmission loss rises roughly 12 dB per doubling of distance, which is what one would expect from a dipole. In the higher frequency bands (>1500 Hz) it rises roughly 8 dB per doubling of distance for ranges over 10 m.

Since the fast field computation yields transmission loss for propagation from a single point source, one may use transmission loss results for a point source at various ranges and depths in determining the source level per unit area for the distributed sheet source. This may be done by first considering the intensity received at a hydrophone from a continuous sheet source which can be expressed as the area integral

$$RI(\mathbf{r}) = \int \int \frac{SI(l,d)}{IL(\mathbf{r})} dl\, dd, \tag{1}$$

where $SI(l,d)$ is the source intensity per unit area and $IL(\mathbf{r})$ is the loss factor in propagating from the source to receiver (e.g., r^2 for spherical spreading). If the above is discretized to a sheet distribution of $N_l \times N_d$ closely spaced, incoherent point sources, the received intensity is then

$$RI(\mathbf{r}) = SI \sum_{j=1}^{N_d} \sum_{i=1}^{N_l} \frac{(\Delta l)^2}{IL_j(r_i)}, \tag{2}$$

where Δl is the separation, horizontal and vertical, between sources, r_i is the horizontal distance between source and receiver, and *SI* is assumed constant over the distribution. The loss factor for each point source along the sheet distribution is obtained from the modified fast field program. The source level, expressed in terms of the transmission loss for individual point sources composing the source distribution and the received level obtained at a given hydrophone, is given as

$$SL = RL - 20 \log \Delta l - 10 \log \sum_{j=1}^{N_d} \sum_{i=1}^{N_l} IL_j^{-1}(r_i). \tag{3}$$

The source level at each time step is then used to obtain a time-averaged source level spectrum over the duration that a portion of the wave is within 5 m of the hydrophone.

VI. SOURCE LEVEL COMPARISONS

The source level spectra for all ten signatures were obtained by synthesizing the video and acoustic measurement as described above. A power-averaged spectrum for each breaker type is displayed in Fig. 15. The general character of both spectra are similar in form. At frequencies below 300 Hz both spectra drop off at 25 dB per decade as indicated by line A for the plunging breakers. The average plunging spectrum drops off at 35 dB per decade at frequencies over 1 kHz, as indicated by line B and the spilling spectrum drops off at 27 dB per decade (line B′). The high-frequency roll-off of these spectra begin at a lower frequency (~1 kHz) than those of the wave tank measurements of Papanicolaou and Raichlen[1] (~5 kHz). This difference could be attributed to differences in the water's bubble content (salt versus fresh), wave characteristics, or effects of the wave tank itself.

The form of the plunging breaker spectrum is similar to that presented by Deane[6] in that a peak occurs in the spectra within the frequency range between the two extrema. No apparent peak occurs in the spilling spectrum. The peak for the plunging spectrum occurs between 500 and 600 Hz, which is different than the peak occurring at 1 kHz reported by Deane. Since this portion of the spectrum is governed by the oscillation of individual bubbles as stated by Deane, this difference can be attributed to the maximum radii of bubbles which are entrained during the breaking process. By the same reasoning, the different slopes at frequencies above 1 kHz between the spilling and plunging spectrum shown here and those reported by Deane (50 dB/decade) can be attributed to differences in the radii distribution of the bubbles entrained by the waves under different conditions.

The power-averaged source spectrum of the plunging breakers is roughly 5–10 dB higher than those of the large spilling breakers over most of the frequency range. A possible explanation for this difference may be obtained in reviewing the differences observed in the range-time signatures of the two breaker types. As stated previously, the plunging breakers generally began breaking near the third hydrophone (181.8 m cross-shore), generating a strong burst of acoustic energy then quickly evolved into the ''borelike'' propagation, which presumably generates lower acoustic source levels. The spilling breakers generate lower acoustic source levels, relative to actively plunging breakers, over a longer time period (and thus longer cross-shore range) while in their ''spilling'' mode before eventually evolving into the ''borelike'' propagation. Since the source level analysis was conducted while the waves were near the phone at which the plunging breaker began breaking, it is reasonable to expect the source level to be higher for the plunging waves. An analysis to compare the source spectra during the evolution to ''borelike'' propagation for the initially plunging and spill-

ing breakers was not conducted as these evolutions occurred shoreward of the deployed array. This level difference was observed by the authors in other data sets from different time periods with various wave conditions. However, an automated means, currently under development, of synthesizing the video and acoustic measurements is needed to process a large number of events in various wave conditions to establish this trend on a statistical basis.

VII. SUMMARY

In this paper we have reported measurements of surf-generated noise taken along a 138-m array within the surf zone. These measurements allow one to identify various characteristics of the noise signatures within range-time-frequency volumes. Distinguishing characteristics between signatures of plunging and spilling breakers were noted in both time-frequency and range-time distributions. Plunging breakers generated time-frequency signatures with a sharp onset due to the impulsive nature of the initial impact of the overturning wavecrest plunging into the water surface. Spilling breakers generated time-frequency signatures with a more gradual low-frequency precursor. Range-time signatures of plunging breakers show a strong burst of acoustic energy which subsequently propagates away at acoustic velocities. Spilling breakers' range-time signatures depict sound being generated over a longer time period and cross-shore range with the source region moving with the breaking surface wave.

A procedure was developed to use quantitative video measurements of breaker location, length, and orientation with acoustic measurements to obtain estimates of source level spectra. To that end, a modified fast field program was used to account for water-column and sediment effects on propagation from the distributed source regions to hydrophones which had buried within the bottom. The resulting source level spectra showed differences between plunging and spilling breakers in level and slope at high frequencies. The difference in slope was attributed to the different bubble radii distribution entrained during the different breaking events.

ACKNOWLEDGMENTS

The authors are grateful to Dr. K. T. Holland for many helpful discussions regarding photogrammetric transformation of video images. This work has been supported by The Office of Naval Research base funding at The Naval Research Laboratory.

[1] P. Papanicolaou and F. Raichlen, "Wave and bubble characteristics in the surf zone," in *Sea Surface Sound*, edited by B. R. Kerman (Kluwer Academic, Dordrecht, Netherlands, 1988) pp. 97–109.
[2] O. B. Wilson, S. N. Wolf, and F. Ingenito, "Measurements of acoustic ambient noise in shallow water due to breaking surf," J. Acoust. Soc. Am. **78**, 190–195 (1985).
[3] O. B. Wilson, M. S. Stewart, J. H. Wilson, and R. H. Bourke, "Noise Source Level Density Due to Surf-Part I: Monterey Bay, Ca," IEEE J. Ocean. Eng. **22**, 425–433 (1997).
[4] J. P. Fabre and J. H. Wilson, "Noise Source Level Density Due to Surf-Part II: Duck, NC," IEEE J. Ocean. Eng. **22**, 434–444 (1997).
[5] G. B. Deane, "Long time-base observation of surf noise," J. Acoust. Soc. Am. **107**, 758–770 (2000).
[6] G. B. Deane, "Sound generation and air entrainment by breaking waves in the surf zone," J. Acoust. Soc. Am. **102**, 2671–2689 (1997).
[7] G. B. Deane, "Acoustic hot-spots and breaking wave noise in the surf zone," J. Acoust. Soc. Am. **105**, 3151–3167 (1999).
[8] http://www.frf.usace.army.mil/SandyDuck/Overview/OverviewSandyDuck.stm
[9] K. T. Holland, R. A. Holman, T. C. Lippman, J. Stanley, and N. Plant, "Practical Use of Video Imagery in Nearshore Oceanographic Field Studies," IEEE J. Ocean. Eng. **22**, 81–92 (1997).
[10] http://www.frf.usace.army.mil/vehicles2.stm
[11] H. W. Kutschale, "Rapid Computation by Wave Theory of Propagation Loss in the Arctic Ocean," Lamont-Doherty Geological Observatory of Columbia University, Technical Report No. 8, CU-8-73 (1973).
[12] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP, Woodbury, NY, 1994), pp. 203–270.
[13] E. L. Hamilton, "Geoacoustic modeling of the seafloor," J. Acoust. Soc. Am. **68**, 1313–1340 (1980).

Multibeam volume acoustic backscatter imagery and reverberation measurements in the northeastern Gulf of Mexico[a)]

Timothy C. Gallaudet and Christian P. de Moustier
Marine Physical Laboratory, Scripps Institution of Oceanography, Mail Code 0205, 8602 La Jolla Shores Drive, La Jolla, California 92037-0205

(Received 4 June 2001; revised 16 November 2001; accepted 4 February 2002)

Multibeam volume acoustic backscatter imagery and reverberation measurements are derived from data collected in 200-m-deep waters in the northeastern Gulf of Mexico, with the Toroidal Volume Search Sonar (TVSS), a 68-kHz cylindrical sonar operated by the U.S. Navy's Coastal System Station. The TVSS's 360-degree vertical imaging plane allows simultaneous identification of multiple volume scattering sources and their discrimination from backscatter at the sea surface or the seafloor. This imaging capability is used to construct a three-dimensional representation of a pelagic fish school near the bottom. Scattering layers imaged in the mixed layer and upper thermocline are attributed to assemblages of epipelagic zooplankton. The fine scale patchiness of these scatterers is assessed with the two-dimensional variance spectra of vertical volume scattering strength images in the upper and middle water column. Mean volume reverberation levels exhibit a vertical directionality which is attributed to the volume scattering layers. Boundary echo sidelobe interference and reverberation is shown to be the major limitation in obtaining bioacoustic data with the TVSS. Because net tow and trawl samples were not collected with the acoustic data, the analysis presented is based upon comparison to previous biologic surveys in the northeastern Gulf of Mexico and reference to the bioacoustic literature.
[DOI: 10.1121/1.1490597]

PACS numbers: 43.30.Ft, 43.30.Sf, 43.30.Vh [DLB]

I. INTRODUCTION

Demographic pressures, overharvesting of marine fish stocks, and pollution are threatening marine ecosystems worldwide and provide incentives to focus fisheries management and plankton research efforts on conservation measures, thus requiring comprehensive knowledge of population dynamics and habitat variability.[2] Such knowledge may be gained through ocean volume acoustic backscatter measurements that have been used to study populations of fish and plankton since the deep scattering layer was identified by Eyring *et al.*[3]

Sonar echoes from volume scatterers are usually quantified using target strength (TS), which is the ratio of the intensity scattered by an object at a reference distance of 1 m, to the incident intensity, and volume scattering strength (S_V), which is the target strength of a unit volume.[4] For volume scattering from n_V scatterers per unit volume each with a mean differential scattering cross section $\langle \sigma_{\mathrm{bs}} \rangle$, S_V and TS are related by

$$\mathrm{TS} = 10 \log_{10} \langle \sigma_{\mathrm{bs}} \rangle = S_V - 10 \log_{10}(n_V) \quad (\mathrm{dB}). \tag{1}$$

Typically, measurements of S_V with a high frequency sonar are used to estimate abundance and/or biomass by estimating $\langle \sigma_{\mathrm{bs}} \rangle$ from trawl or net tow samples with an appropriate scattering model, and using Eq. (1) to compute n_V. Such efforts are difficult because marine ecosystems contain multi-species, multi-size assemblages of organisms.

Another problem affecting these types of measurements is the spatial variability of zooplankton and fish, which may lead to large errors in abundance and biomass estimates obtained with single beam sonars. Single beam, vertically oriented echo-sounders have seen the most use in bioacoustic applications,[5] but horizontally directed sonars have been used to characterize fish school structure, shape, and movement,[6–9] as well as zooplankton distributions.[10,11] Multibeam echo-sounders have been used in only a few bioacoustic studies to observe the swimming behavior of individual zooplankters,[12] to demonstrate that vertical and lateral vessel avoidance by fish negatively bias abundance estimates derived from vertical echo-sounding sonars,[13] and to provide more precise mapping and abundance of pelagic fish stocks in near-surface schools than can be obtained by vertical echo-sounding.[14] The use of multibeam sonars in place of single-beam sonars for future bioacoustic surveys is warranted by their increased coverage, which can improve acoustic estimates of biomass and abundance and better characterize spatial distributions of organisms.

In this context, we present volume acoustic backscatter and reverberation measurements derived from data collected by the U.S. Navy's Toroidal Volume Search Sonar (TVSS), a 68-kHz multibeam sonar capable of 360-degree imaging in a vertical plane perpendicular to its axis. We take advantage of

[a)]Portions of this paper were presented in talks at the Fourth Annual Symposium on Technology and the Mine Problem[1], 13–16 March 2000, Monterey, CA, the First Oceanic Imaging Conference, 2000, 3–5 May 2000, in Newport, RI, and Acoustical Society of America Meetings in Seattle, WA, 20–26 June 1998, Berlin, Germany, 15–19 March 1999, and Newport Beach, CA, 3–8 December 2000.

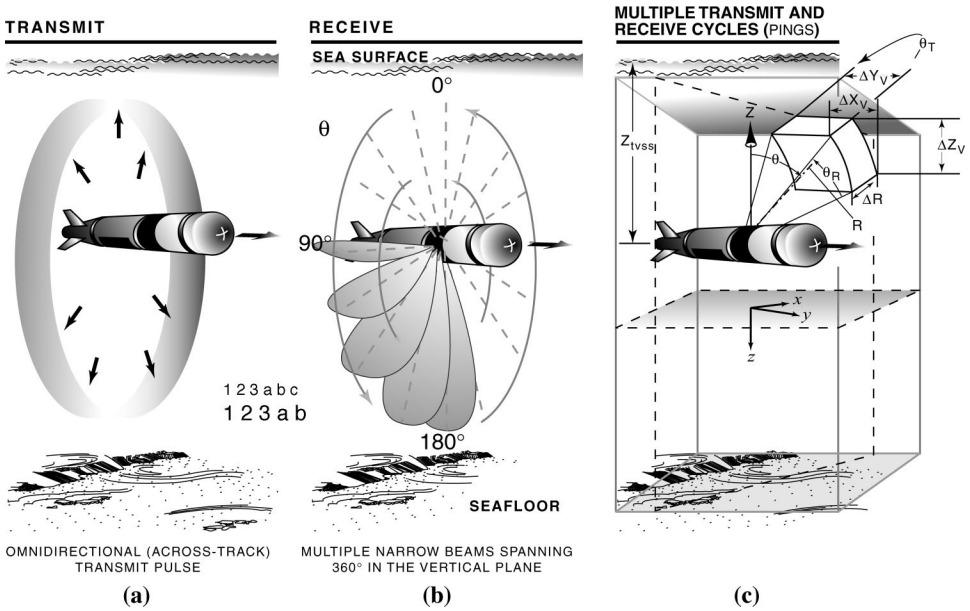


FIG. 1. The TVSS is designed for mine-hunting while deployed on an autonomous or unmanned underwater vehicle (AUV/UUV). In this study, a towed vehicle was used. (a) After transmission of a ''toroidal'' pulse, the sonar extracts the returned signal in directions spanning 360 degrees about the TVSS's axis (b). This geometry is used to construct undersea imagery for multiple horizontal and vertical planes using data collected over successive transmit-receive cycles (pings) (c). Only vertical imagery is presented in this study, where *x*, *y*, and *z* correspond to across- and along-track directions relative to the towfish and depth relative to the sea surface. The resolution of the TVSS data is defined by the dimensions of the ensonified volume in each direction (ΔX_V, ΔY_V, ΔZ_V), which depend upon the range from the TVSS (R), the transmitted beamwidth (θ_T), the receive beamwidth (θ_R), and the beam angle (θ).

the unique synoptic three-dimensional perspective afforded by this sonar system to explore and characterize the spatial and temporal structure of pelagic fish schools and the patchiness of epipelagic zooplankton found in the northeastern Gulf of Mexico. The data were collected during engineering tests of the TVSS conducted by the U.S. Navy's Coastal System Station, Panama City, FL, in a 2-nm^2 area centered at 29°30′N, 86°30′W. This is a relatively flat area on the continental shelf, with water depths ranging from 190 to 200 m, situated roughly 65 nm southwest of Panama City, and 20 nm southeast of De Soto Canyon.

The TVSS includes separate cylindrical projector and hydrophone arrays, with the same 0.53-m diameter, mounted coaxially on a cylindrical tow body. The projector array has 32 elements equally spaced 11.25 degrees apart around the cylinder and designed to produce a ''toroidal'' beampattern that is meant to be omni-directional in the plane perpendicular to the cylinder's axis (usually across-track) and 3.7 degrees wide at −3 dB in any plane containing the cylinder's axis (usually along-track). The hydrophone array consists of 120 elements equally spaced every 3 degrees around the cylinder. In the work presented here, beamforming of the hydrophone array yielded 120 receive beams, each 4.95 degrees wide at −3 dB and spaced 3 degrees apart to cover the full 360 degrees around the array in the plane perpendicular to the array's axis.

Using this multibeam geometry, we have adapted existing oceanic imaging techniques to construct acoustic backscatter imagery of horizontal and vertical planes in the ocean volume (Fig. 1). Only vertical imagery is presented in this study to characterize the spatial distributions of bioacoustic scatterers and to partially discriminate between desired bioacoustic signals and volume or boundary reverberation. This is not possible with conventional single- or dual-beam echo-sounders because of the temporal lags introduced by the successive transects required to cover a volume of ocean comparable to that sampled by the multibeam sonar on one track. However, multibeam echo-sounders have important limitations due to the beamforming process and our intent is to demonstrate some of the capabilities and limitations of multibeam sonars in bioacoustic applications.

We begin by describing the TVSS data and signal processing methods that help to interpret the acoustic backscatter images. Because net tow or trawl samples were not collected in parallel with the acoustic data, our interpretations of these images rely on comparison and reference to previous biologic surveys in the northeastern Gulf of Mexico, and on the general bioacoustic literature.

II. TVSS DATA

A. Data collection

The acoustic data presented here were collected on 9 November 1994 between 1026AM and 1131AM local time. The wind speed and sea state recorded at 0658AM were 6 knots (3 m/s) and 1.5, respectively. A CTD cast, taken at 0658AM approximately 100 m north of the location for run 1 (Fig. 2), revealed the presence of a 24.8 °C isothermal mixed layer extending to a depth of 49 m, a thermocline between 49 and 150 m depth, and a nearly isothermal layer above the bottom with a temperature of 15.6 °C. The surface salinity was 35.1 PSU, and the surface sound speed was 1534 m/s. Historical data indicate a relatively weak circulation in the region during fall months,[15] which, with the light winds, suggests that the surface currents were either weak or absent.

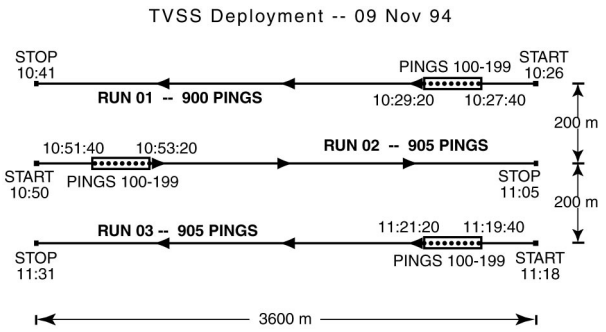


FIG. 2. TVSS track lines for three consecutive runs. The data presented in this paper are processed from pings 100–199 in each of the three runs. In the figure, north is upward.

The TVSS was towed at a depth of 78 m, approximately 735 m astern a ship moving at a nearly constant speed of 4.1 m/s. Three runs of 100 consecutive pings of acoustic backscatter data, from a 200 μs CW pulse of 68 kHz transmitted at 1 Hz recorded during the tests, are analyzed here. Towfish roll, pitch, heading, speed, and depth were sampled at 1 Hz (once per ping). Further information regarding the TVSS, the data collection, and the signal processing is described in Ref. 16.

B. Environmental context

Because we lacked the net tow and trawl data required to verify that biologic sources are responsible for scattering features seen in our data, we rely on historical and concurrent data for evidence. The abundance and distribution of pelagic fish in the Gulf of Mexico (GOM) were investigated by the US National Marine Fisheries Service (NMFS) during 15 cruises in the spring and fall seasons between 1988 and 1996.[17] Trawling was performed only during daylight hours near the bottom, with the highest concentration of sampling stations centered approximately 10 nm northwest of the TVSS experiment site. For several species, the highest abundances were sampled at four sites within 1 nm of the TVSS experiment site. Thus, the NMFS survey encompassed the year, season, time of day, geographic location, and depth corresponding to the collection of the TVSS acoustic data.

Zooplankton species known to scatter sound near 68 kHz, such as euphausiids, have been observed in the northeastern and central GOM as well. Hopkins' observations in a region about 120 nm south of the TVSS experiment showed that zooplankton biomass was concentrated in the upper 50 m of the water column.[18] Zimmerman and Biggs observed acoustic scattering due to zooplankton in warm- and cold-core eddies in the eastern central GOM during June 1995 using a 153-kHz acoustic Doppler current profiler.[19] Also in the eastern central GOM, Hopkins *et al.* examined the landward distributions of zooplankton between May and June 1977, and found many oceanic species distributed across the Florida shelf.[20] Ortner *et al.* investigated the vertical distributions of zooplankton during January and February 1981, and found zooplankton distributions to be closely tied to mixed layer depth.[21] Ortner *et al.* reported observations of euphausiids and decapod shrimp in a region about 50 nm southeast of the TVSS experiment site.[22]

C. Processing methods

The processing scheme that we have developed for the TVSS acoustic data is designed for conformal arrays and includes quadrature sampling, resampled amplitude shading, element-pattern compensation, and broadside beamforming on phase-compensated, overlapping subarrays with asymmetric projected element spacings.[23] This procedure permits split aperture processing of the beamformed output, which is performed because the processed data also were used to study acoustic backscatter from the ocean boundaries, and the phase zero-crossing of the output phasor is the most accurate means to detect the arrival time of boundary reflections on the maximum response axis of the beam.[24,25]

For each ping, the split aperture process is used to form a phasor $P(\theta,t)$ for a given beam direction θ and time sample t every 3 degrees for directions spanning the 360 degrees around the TVSS axis. The magnitude of $P(\theta,t)$ has units of volts^2, and is converted to squared echo amplitude according to

$$10\log_{10}(\mathrm{EL}^2(\theta,t)) = 10\log_{10}(\tfrac{1}{2}|P(\theta,t)|) - \mathrm{RVR} - \mathrm{FG} - \mathrm{DI} - \mathrm{TVG}, \quad (2)$$

where RVR$=-168$ dB $re{:}1\ V_{\mathrm{rms}}/\mu$Pa is the receive voltage response of each hydrophone, FG$=29$ dB is the preamplifier fixed gain, DI$=13$ dB is the array gain associated with the beamforming and split aperture processing, and TVG is the system time-varying gain. The left side of Eq. (2) is equivalent to the plane wave reverberation level (RL) in the active sonar equation,[4] which we use to compute volume scattering strength:

$$S_V = \mathrm{RL} - \mathrm{SL} + 2\mathrm{TL} - 10\log_{10} V \quad \text{with} \quad \mathrm{TL} = 20\log_{10} R + \alpha R, \quad (3)$$

where R is the range from the TVSS determined from t and θ using constant gradient ray-tracing techniques, SL $=216.8$ dB $re{:}1\ \mu$Pa@1 m is the calibrated TVSS source level, $\alpha=0.024$ dB/m, and V is the volume ensonified by the transmitted pulse within the receive beam.

Volume scattering strength images were constructed by first computing S_V from Eqs. (2) and (3) for each beamformed sample in each ping, yielding 120 backscattering strength times series. These may be displayed together as a vertical slice of volume acoustic scattering strength in polar coordinates of angle versus time (Fig. 3). In this representation, echoes from the sea surface and seafloor appear as the high backscatter, horizontal features. Scattering from resonant microbubbles in the towship's wake and from bubble clouds formed by breaking ship waves are responsible for the high backscattering strength features near the sea surface. Other scattering structures are apparent upon adjusting the dynamic range of the display, and these will be discussed in Sec. III.

Volume scattering strength images in the plane parallel to the towfish's direction of travel were formed by extracting S_V data recorded over successive pings. Echo-integrated vertical volume backscattering strength images on either side of

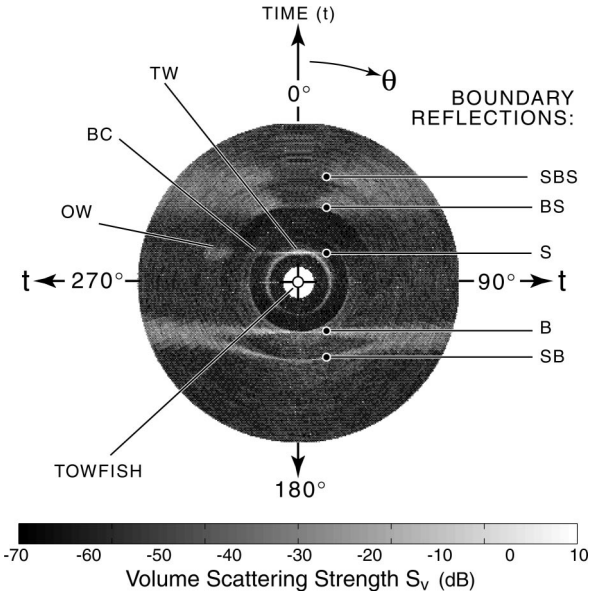


FIG. 3. One TVSS ping displayed as a vertical slice of acoustic volume scattering strength (S_V) in polar coordinates of angle versus two-way travel time (t). Labels refer to the following features: TW—township's wake; OW—township's wake generated during previous run; BC—bubble cloud generated by a breaking bow wave from the township; B—seafloor echo; S—sea surface echo; SB—surface-bottom (multiple) echo; BS—bottom-surface (multiple) echo; SBS—surface-bottom-surface (multiple) echo. The maximum two-way travel time t represented by the data in this figure is 0.54 s.

the towfish were constructed with a time gate δt corresponding to a 1.4-m depth interval, and an along-track averaging interval of three pings.

D. TVSS beam geometry

The sampling and resolution characteristics of the backscattering strength images are determined by the acoustic geometry of each TVSS ping. The angular sample spacing between the maximum response axes of adjacent beams in Fig. 3 is $\theta_s = 3$ degrees, whereas the quadrature sampled time increment within each beam is $\tau_s = 160\ \mu$s, which corresponds to a 12-cm sampling interval assuming a sound speed in seawater $c = 1500$ m/s. With the TVSS pulse length $\tau_p = 200\ \mu$s, the bandwidth $W = 0.88/\tau_p = 4.4$ kHz yields a nominal range resolution $\Delta R = c/2W = 17$ cm.

The volumetric resolution in each ping is determined by the spatial dimensions of the volume (V) ensonified by the TVSS transmit pulse within each receive beam (Fig. 1). We compute V as the ellipsoidal shell section formed from the intersection of the pulse, the transmit beam pattern, and the receive beam pattern at each sampling point:

$$V = \tfrac{2}{3}\,\theta_T \sin(\theta_R/2)[R'^3 - R^3] \quad (\mathrm{m}^3), \tag{4}$$

where R is the range from the center of the TVSS in meters, $R' = R + \Delta R$, the -3-dB receive beamwidth (θ_R) is 4.95 degrees, and the -3-dB transmit beamwidth (θ_T) is 3.7 degrees. The spatial dimensions of V in the across-track, along-track, and vertical dimensions (Δx_V, Δy_V, Δz_V, Fig. 1) vary with range. The along-track dimension is the same for all beam angles θ:

$$\Delta y_V = 2R' \sin(\theta_T/2). \tag{5}$$

However, the across-track and vertical dimensions of V vary with range and θ_a, which is defined in relation to θ in Fig. 1 as $\theta_a = \theta$ for $\theta = 0$ to 90 degrees; $\theta_a = |180 - \theta|$ for $\theta = 91$ to 270 degrees; and $\theta_a = (360 - \theta)$ for $\theta = 271$ to 359 degrees. Thus, for volume cells in beams directed towards zenith and nadir, ($\theta_a = 0$ degrees)

$$\begin{aligned} \Delta x_V &= 2R' \sin(\theta_R/2) \quad (\mathrm{m}), \\ \Delta z_V &= \Delta R + R(1 - \cos(\theta_R/2)) \quad (\mathrm{m}), \end{aligned} \tag{6a}$$

and for volume cells in horizontal beams, ($\theta_a = 90$ degrees)

$$\begin{aligned} \Delta x_V &= \Delta R + R(1 - \cos(\theta_R/2)) \quad (\mathrm{m}), \\ \Delta z_V &= 2R \sin(\theta_R/2) \quad (\mathrm{m}), \end{aligned} \tag{6b}$$

where Δx_V exceeds the range resolution at $\theta_a = 90$ degrees, and Δz_V exceeds ΔR at $\theta_a = 0$ degrees because of the curvature of the wavefront. For angles between the horizontal and vertical, the x and z dimensions of V can be expressed as

$$\begin{aligned} \Delta x_V &= 2R \sin(\theta_R/2)\cos\theta_a + \Delta R \sin(\theta_a + \theta_R/2), \\ \Delta z_V &= 2R \sin(\theta_R/2)\sin\theta_a + \Delta R \cos(\theta_a - \theta_R/2) \\ &\quad \text{for } 0 < \theta_a < 90 - \theta_R/2. \end{aligned} \tag{6c}$$

Samples in adjacent beams overlap because $\theta_s < \theta_R/2$. The overlapping volume increases with range:

$$V_{\mathrm{ol}} = \tfrac{2}{3}\theta_T \sin((\theta_R - \theta_s)/2)[R'^3 - R^3] \quad (\mathrm{m}^3) \quad (\text{adjacent beams}), \tag{7}$$

whereas the percent overlapping volume is

$$\mathrm{VP}_{\mathrm{ol}} = [\sin((\theta_R - \theta_s)/2)/\sin(\theta_R/2)] \times 100\% \quad (\text{adjacent beams}), \tag{8}$$

yielding a range-independent volume overlapping percentage of 39.4% between adjacent TVSS beams within the same ping. Volumes ensonified on successive pings in the same beam angle overlapped and increased with range beyond 62 m due to the towfish's speed of $V_{\mathrm{tvss}} = 4.1$ m/s and -3-dB transmit beamwidth $\theta_T = 3.7$ degrees.

The backscatter images in this study were constructed using the sidescanning techniques described in Ref. 24, where samples of acoustic backscatter are extracted for the time and angle pairs that correspond to the desired horizontal or vertical plane (Fig. 1). Between discrete beam angles, samples are interpolated in time increments corresponding to the quadrature sampling interval τ_s. Thus, the sample spacings in the images depend upon τ_s, as well as R, V_{tvss}, and θ_a. The along-track sample spacing is the ping separation distance $\delta y_V = \Delta y_{\mathrm{ping}}$, whereas across-track and vertical sample spacings are

$$\begin{aligned} \delta x_V &= c\,\tau_s/2 \sin\theta_a, \quad (\mathrm{m}) \quad \text{for } \theta_a > \theta_R, \\ \delta z_V &= c\,\tau_s/2 \cos\theta_a, \quad (\mathrm{m}) \quad \text{for } \theta_a < 90° - \theta_R. \end{aligned} \tag{9}$$

A consequence of the sidescanning procedure is that the number of samples per beam angle increases as θ_a decreases.

E. System-related sources of error

A large source of error in our measurements was the angular variation in the transmitted beampattern. A maximum deviation of -9 dB from the calibrated source level existed when hydrostatic tests were performed on the TVSS a year before the data collection. We were unable to correct for these variations because the actual beampattern may have changed slightly during the year between the hydrostatic testing and the TVSS deployment.[24]

The beampatterns of the 120 individual hydrophones introduced another source of error, because they were not identical, and their RVRs, FGs, and TVGs varied as much as ± 2 dB from average values. To simplify the beamforming, we used an average of the calibrated hydrophone sensitivity values of -168 dB $re{:}1V_{\mathrm{rms}}/\mu$Pa, and we approximated the beampatterns of individual hydrophones with a cardioid-shaped magnitude response that closely matched that provided by the manufacturer. In the absence of hydrophone phase response data, we approximated the phase pattern for each element with a function proportional to the $\sin^2$ of the angle between the maximum response axes of the element and the subarray containing it. We used these response patterns because they successfully removed pointing errors which were observed in previous beamforming efforts with the TVSS.[23]

Computer simulations, which used the TVSS transmit pattern and towfish attitude data for each run, indicated that the TVSS array beampatterns produced a maximum bias error of -7 dB and a maximum random error of ± 3.3 dB in the TVSS data. After echo-integrating and averaging over successive pings, these values were reduced to approximately -5 dB for the bias error and ± 1 dB for the random error.

Another important limitation of the data presented here is that boundary echoes at normal incidence ($\theta = 0$ or 180 degrees, Fig. 3), received through the sidelobes of beams directed towards the volume, contaminated the volume backscatter received in their mainlobes. This problem is illustrated in the returns for a single ping (Fig. 3) and for a 97-ping average [Fig. 4(a)], where boundary echoes appear as high backscatter circular features tangent to the sea surface and seafloor.

The linear bands extending diagonally from the bottom into the volume in Fig. 4(a) also result from bottom sidelobe returns. We verified through computer simulations that their unique structure is a consequence of the uniform 3-degree spacing between the receive beams, and of the nonuniform sidelobe spacing within each receive beampattern. The simulation results appear in Fig. 4(b) as dashed lines that match closely the underlying linear bands and represent the times and angles of each volume beam with a sidelobe directed at the bottom during the time of the bottom echo arrival. Minor differences probably result from uncertainties in the towfish's attitude.[24]

This sidelobe interference is somewhat enhanced by our choice of a resampled Dolph–Chebyshev amplitude shading window designed to produce a nearly uniform sidelobe level between -28 and -30 dB for all the receive beampatterns.[23] Lower sidelobe levels are achievable, but the corresponding increase in mainlobe beam width would degrade the spatial resolution in the images. Due to the strong boundary echo sidelobe returns in the TVSS data, sidelobe cancellation methods are necessary to use data in the upward looking beams beyond slant ranges equal to the towfish's depth ($\sim$78 m), and in the downward looking beams beyond slant ranges equal to the towfish's altitude ($\sim$115 m).

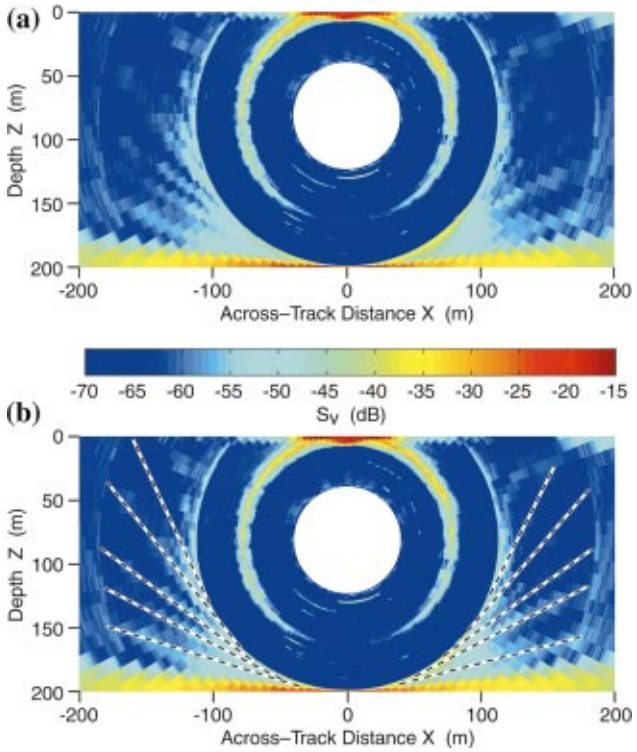


FIG. 4. (a) and (b) Average volume acoustic scattering strength of 97 pings collected during run 2. (b) Computer simulations determined the time-angle pairs (dashed lines) for which a sidelobe of a beam pointing in the ocean volume was directed towards the bottom at the time of arrival of each bottom echo.

One additional limitation of the data presented here is the void region with a radius of $\sim$40 m around the TVSS (Figs. 3 and 4) caused by a fixed 53-ms blanking delay between transmission and the start of digitization. This was done because the TVSS deployment was conducted primarily to evaluate the sonar's mine-detection capability between 50 and 750 m, so the void region helped to reduce the large amount of data collected by the sonar. This feature of the TVSS is not a problem for future bioacoustic applications because the time interval between transmit and receive is adjustable, so that the void region may be reduced.

III. IMAGERY RESULTS AND INTERPRETATION

A. Fish schools

Along-track images [Figs. 5(a) and (c)] were formed by using a vertical echo-sounding procedure to extract backscatter data in the down-looking beams at each along-track sampling point. Across-track sections [Figs. 5(b) and (d)] were formed by using the same procedure as that used to create Fig. 3. Based on the nearly concurrent NMFS trawl data, the prominent backscatter features centered at ($y = 285$ m, $z = 190$ m) in Fig. 5(a) and ($y = 340$ m, $z = 200$ m), in Fig.

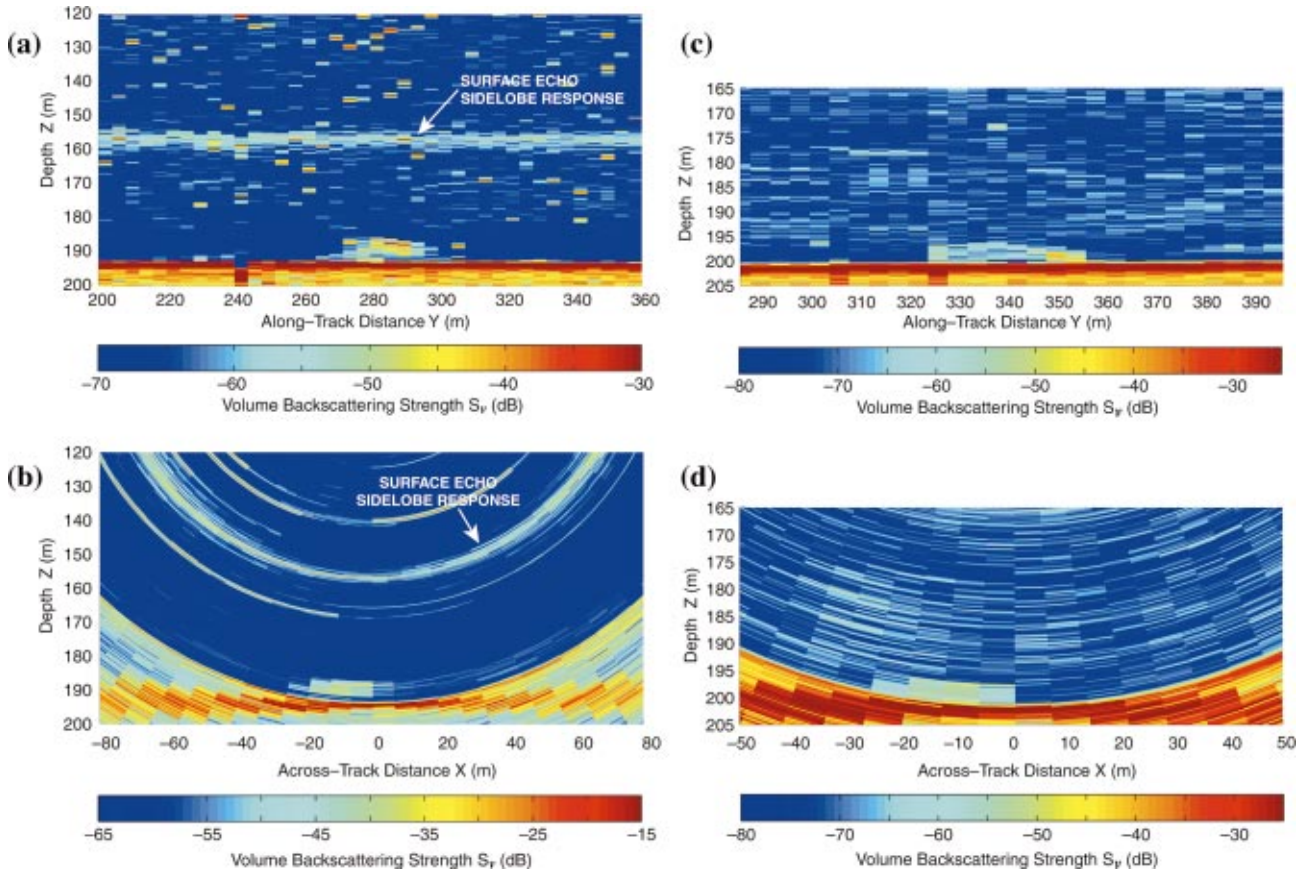


FIG. 5. Vertical volume acoustic backscattering strength (S_V) images of two near bottom fish schools in run 1 [(a), (b)], and run 2 [(c), (d)]. The seafloor is the high backscatter, horizontal feature at $z=193$ m in (a) and (b), and 202 m in (c) and (d). The horizontal feature with moderately high backscatter at 158-m depth in (a) is the sidelobe response of the sea surface echo, seen also in the individual (Fig. 3) and averaged (Fig. 4) ping data. The randomly distributed samples in (a) where $S_V > -40$ dB in the water column are signals that were saturated in the TVSS data acquisition system electronics. These appear as the thin arcs in Fig. 5(b) above and below the sea surface echo sidelobe return.

5(c) are attributed to schools of small pelagic fish. Together, the orthogonal image pairs in Fig. 5 characterize the size and scattering characteristics of the two fish schools (Table I). Although the images are limited by surface echo sidelobe interference and saturated samples, these artifacts were distinguished from biologic scatters by their arclike across-track structure [Figs. 5(b) and (d)]. Had we only constructed the along-track sections [Figs. 5(a) and (c)], these samples might have been incorrectly attributed to the acoustic backscatter from individual fish.

TABLE I. Fish school data corresponding to Figs. 5 and 6. The seafloor backscattering strength S_B was computed and analyzed in Ref. 24. The target strength (TS) values are for ensonified volumes, and not individual fish.

	Run 1	Run 2
Maximum height above seafloor (m)	7.1	4.7
Maximum along-track dimension (m)	42	32
Maximum across-track dimension (m)	31	26
Maximum TS (dB)	−24.3	−28.9
Mean TS (dB)	−45.5	−47.3
Maximum S_V (dB)	−31.9	−40.1
Mean S_V (dB)	−56.1	−59.8
Mean S_B (dB) at normal incidence	−8.0	−9.7
Standard deviation of S_B (dB) at normal incidence	2.3	2.0
Normal incidence S_B (dB) below fish school	−8.1	−9.9

We constructed a 3D shape representation of the school in Figs. 5(a) and (b) by processing multiple along-track (x,z) sections over successive pings. Bottom detection processing was used first to discriminate between acoustic backscatter from the seafloor and from the school.[24] A threshold of 5 dB over the ambient scattering strength level was applied to detect samples in the school, and their corresponding (x,y,z) positions were determined using constant gradient ray tracing. The result [Fig. 6(a)] resembles a ''stack'' of elongated ''tubes'' extending across-track, with a few apparently suspended above the ''stack.'' These features are a consequence of the TVSS sampling and resolution characteristics that are fundamental limitations for any sonar (cf. Ref. 7). The resolution of each sample is defined by the spatial dimensions of the ensonified volume [Fig. 1(c)], which are $\Delta x_V = 9.9$ m, $\Delta y_V = 7.4$ m, and $\Delta z_V = 0.28$ m at the school's center. Therefore, the elongated features in Fig. 6(a) are most likely returns from individual fish above the center of the school, because ensonified volumes in adjacent beams overlap by 39.4%, and beams adjacent to those directed towards large scatterers or boundaries exhibit a significant sidelobe response. Sidelobe response from the fish school appears in Fig. 6(b) as light blue samples at depths of 180–185 m, 40 m across-track on either side of center. These samples are about 5 dB above the ambient level and immediately precede the arc corresponding to the bottom echo sidelobe response.

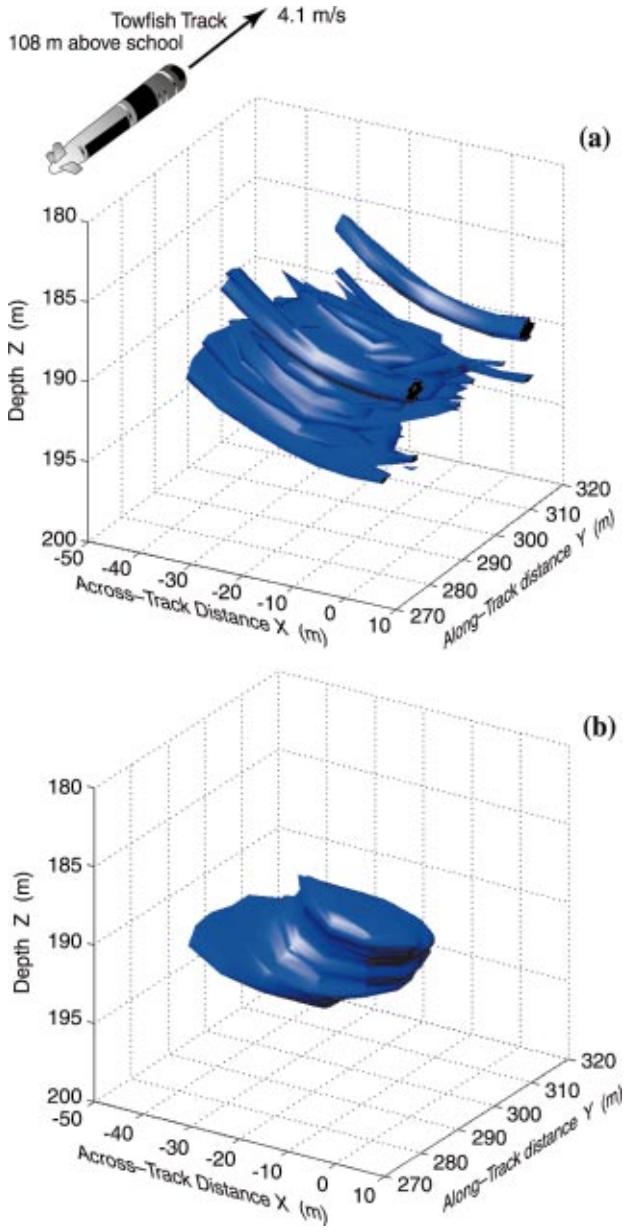


FIG. 6. (a) Three-dimensional representation of the fish school detected in run 1 [Figs. 5(a) and (b)]. The thin, elongated features are volume cells in adjacent, overlapping beams in which fish were detected. (b) Volumetric representation of the school after filtering the samples in and around the school with a $3\times3\times3$ moving average filter and thresholding the output 5 dB above the ambient scattering strength values.

To compensate for these artifacts and produce a shape characterization potentially more representative of the actual fish school, a $3\times3\times3$-sample moving average filter was applied to the scattering strength samples in and around the school. Filtered samples with scattering strengths greater than 5 dB above the ambient scattering strength were retained, yielding an oblong volumetric shape that extends slightly diagonally to the across-track direction [Fig. 6(b)]. The shape's maximum length is three times its maximum width and over six times its maximum height (Table I). In this representation, the school appears to be concentrated near the bottom, which was generally flat with a mean depth of 193 m in this run. Although the thresholds and filter dimensions were subjectively chosen and the representation does not account for relative motion between the school and the TVSS, the school shape in Fig. 6(b) is generally similar to those observed for other schools of small pelagic fish near the bottom during daytime (cf. Ref. 26).

With our limited data, we have no way of identifying the species of fish responsible for the enhanced volume acoustic backscatter seen in Fig. 5. However, three species representing 73% of the total catch between 150 and 200 m in the NMFS trawls could be considered candidates: the round herring (*Etrumeus teres*), the rough scad (*Trachurus lathami*), and the Gulf butterfish (*Peprilus burti*). The latter is the least likely candidate because it does not have a swimbladder for sizes greater than 75 mm.[27] The rough scad is a good candidate because its maximum observed length was 75.8 cm, and there was a tendency for the largest fish of most species to be found deeper than 150-m depth. Nonetheless, we favor the round herring (*E. teres*) as the most likely candidate because it was the most abundant species caught in the NMFS trawls near the TVSS experiment site and over all stations between 150- and 200-m depths, with an average catch rate of 2341/h for the entire GOM. Similar inferences were made by Nero *et al.*[26] in their detailed acoustic study of schooling fish about 3 nm southwest of the TVSS experiment site with a 38-kHz echo-sounder. They concluded that *E. teres* was responsible for the features in their acoustic backscatter data because of the tendency of this species to form compact schools near the bottom during the day, and ascend to the middle of the water column and disperse at night.

The acoustic characteristics of the schools in the TVSS data are generally comparable to those expected from small pelagic fish. For example, *E. teres* lengths recorded during the NMFS survey averaged 13.2 cm ($L/\lambda=6$ at 68 kHz) and had a maximum of 51.2 cm, corresponding respectively to mean and maximum target strengths of -42 and -31 dB according to Love's model,[28] or -49.5 and -37.5 dB using Foote's[29] empirical relation for clupeoids (herring and sprat) ($\mathrm{TS}=20\log_{10}L-71.9$ dB). The mean and maximum target strengths of ensonified volumes within the two schools (Table I) span these model predictions, suggesting that small pelagic fish within individual volume cells are a possible source for the observed data. However, this is only conjecture because physical capture was not performed during the acoustic data collection.

The two- and three-dimensional (2D/3D) characterizations of near-bottom fish schools derived from the TVSS data demonstrate the potential advantages and limitations of multibeam sonars in fisheries acoustics. Two-dimensional characterizations of schooling,[7] diel migration,[30] seasonal migration,[6] and feeding[26] have been useful in understanding fish behavior. But, as far as we know, the economic and scientific advantages of multibeam sonars for 3D characterization of fish schools have only been suggested and not documented.[31] Because fish distributions are heterogeneous, further 3D analyses may provide important information about the structure and composition of aggregations not available in 2D studies.

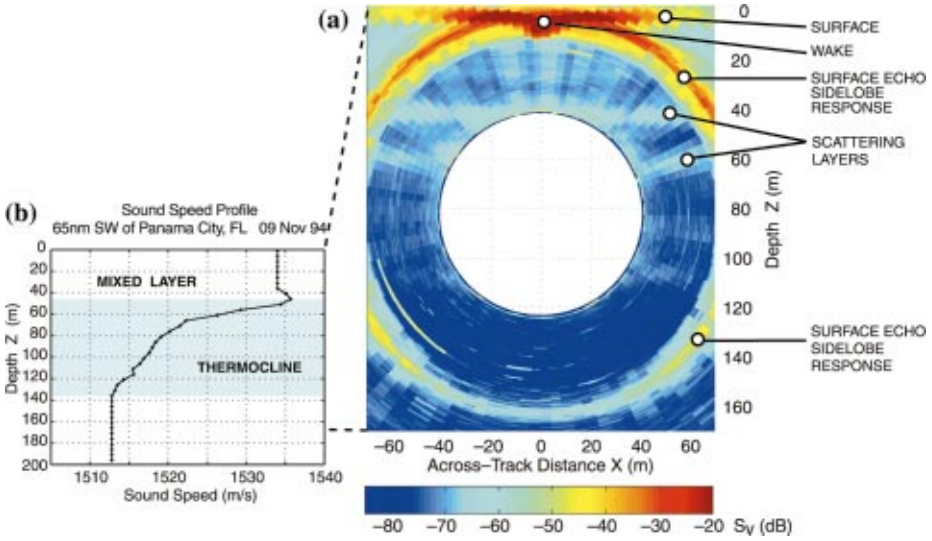


FIG. 7. (a) This 97-ping average for run 2 reveals scattering layers between 40- and 80-m depth. These depths comprised the base of the mixed layer and upper thermocline, which was well described by the sound speed profile (b) obtained 100 m north of the location for run 1.

B. Volume scattering layers

We constructed two types of volume backscattering strength images from the TVSS data collected in the middle and upper water column, and both revealed distinct scattering layers in the upper thermocline and at the base of the mixed layer. The first type of these images [Fig. 7(a)] was formed by averaging 97 pings along the length of run 2 (Fig. 2). The figure also shows the sidelobe response of the surface echo for beams directed away from zenith, and the high backscatter near zenith due to resonant microbubbles in the towship's wake.[25]

The second type of image consisted of vertical slices of volume acoustic backscattering strength formed along-track on each side of the TVSS by echo-integrating between 30- and 130-m depth, and then averaging in three-ping intervals along-track (Fig. 10). In the absence of net and trawl data, historical data suggest that epipelagic zooplankton, and to a lesser extent, micronekton, are the sources for the scattering layers, particularly the previously cited studies which showed that zooplankton distributions in the GOM are centered near the mixed layer depth and upper 50 m of the water column.[18,21]

Outside the Gulf of Mexico, the results of other acoustic studies using single-frequency sonars near 68 kHz generally support our hypothesis. Pluddemann and Pinkel[30] observed the diel migration of scattering layers in the eastern North Pacific with a 67-kHz Doppler sonar, and used Johnson's[32] fluid sphere model to reason that the majority of scattering in their data was due to organisms with equivalent spherical radii 0.1 to 0.4 cm, and lengths 0.5 to 2.0 cm, which includes micronekton and large zooplankton such as decapod shrimp and euphausiids. However, their reasoning assumed that the animals were well described by the fluid sphere model because they also lacked coincident *in situ* data.

The 70-kHz results of Stanton *et al.*[33] are more insightful because they collected mid-water trawl samples and compared them to their acoustic data. The peak scattering strengths in their data were 1–5 dB lower than those observed during the TVSS experiment (Table II) and were attributed to fish or large numbers of arthropods, including euphausiids and shrimp. Their scattering strength predictions were based on Love's[34] model for fish and Johnson's[32] model for arthropods and were within 5 dB of those measured acoustically. Because their scattering strengths were comparable to those measured in the scattering layers by the TVSS, it is likely that the TVSS data also were influenced by zooplankton and micronekton.

One of the largest sources of error in biomass estimates of oceanic zooplankton is their fine-scale patchiness (i.e., spatial variability on scale of 1 m to 1 km), so we examined this aspect of the scattering layers in the TVSS data. Most studies of zooplankton patchiness have relied on one-dimensional variance spectra or patch-finding methods, where ''patch'' is defined by some arbitrary criteria.[35,36] Wiebe's[37] definition of a patch as ''a concentration of individuals exceeding the central value in the data set'' implies that patch sizes vary with the length scales covered by the data set, the ''window'' of samples over which the patch is determined, and the threshold concentration beyond which a patch is defined. Nero and Magnuson[36] used two-dimensional acoustic transects of the Gulf Stream to illustrate the dependencies of patch size and internal patch characteristics on threshold values and window size. With knowledge of the water mass boundaries and characteristics in their data set, they were able to determine the thresholds and window sizes which produced patches that best represented the finescale features of interest. The limited coverage and *in situ* data in our experiment prevented us from using their approach without a large degree of subjectivity, so we characterized the patchiness of the scattering layer data by estimating the 2D variance spectra from the volume scattering coefficients corresponding to the six S_V images in Fig. 8 (see the Appendix).

The average 2D variance spectrum (Fig. 9) reveals the dominant scales of variability in the TVSS scattering layer data. Vertical variability is distributed over a relatively wide portion of the available range of spatial frequencies, with corresponding length scales of 8–33 m. These values are close to the scattering layer thicknesses in Figs. 7 and 8.

TABLE II. Data computed from the mean profiles of echo-integrated volume scattering strength in Fig. 8, where $x<0$ m was left of the towfish's track. The target strength (TS) values are for ensonified volumes, and not individual scatterers.

	Run 1		Run 2		Run 3	
	$x=-47$ m	$x=47$ m	$x=-47$ m	$x=47$ m	$x=-47$ m	$x=47$ m
Layer 1						
depth (m)	⋯	⋯	30	31	33	35
thickness (m)	⋯	⋯	9	11	14	18
maximum S_V (dB)	⋯	⋯	−67.5	−64.1	−63.2	−64.1
depth of maximum S_V (m)	⋯	⋯	29	35.5	30	30
maximum TS (dB)	⋯	⋯	−62.3	−58.5	−57.5	−58.0
depth of maximum TS (m)	⋯	⋯	29	35.5	29	30
Layer 2						
depth (m)	40	41	47	49	58	60
thickness (m)	28	25	21	16	29	22
maximum S_V (dB)	−61.5	−61.5	−64.1	−63.5	−61.5	−62.0
depth of maximum S_V (m)	49	33	50.5	53	55	60
maximum TS (dB)	−62.1	−56.3	−59.5	−59.5	−57.4	−58.1
depth of maximum TS (dB)	39	33	50.5	53	50.5	60
Layer 3						
depth (m)	70	76	70	74	⋯	⋯
thickness (m)	27	22	23	22	⋯	⋯
maximum S_V (dB)	−59.5	−59.0	−64.4	−66.2	⋯	⋯
depth of maximum S_V (m)	69	81	69	75	⋯	⋯
maximum TS (dB)	−56.5	−56.5	−61.5	−64.3	⋯	⋯
depth of maximum TS (m)	69	81	69	75.5	⋯	⋯

Horizontal variability is confined to a relatively narrow portion of the range of available spatial frequencies, with corresponding length scales of 30–100 m. These values are consistent with other upper water column observations of zooplankton patch lengths (cf. Refs. 36 and 37). The two dominant peaks in the region near the horizontal spatial frequency (μ) of 0.02 m^{-1} and between the vertical spatial frequencies (ν) of 0.03 and 0.09 m^{-1} may indicate a coupling between the vertical and horizontal variability at the dominant horizontal scale of 50 m. Greenlaw and Pearcy[38] used a 20-kHz sonar to suggest that such a phenomenon affected the distributions of mesopelagic micronekton off Oregon, but they based their hypothesis on separate one-dimensional vertical and horizontal variance spectra. The two-dimensional spectrum in Fig. 9 provides evidence for this type of coupling in the TVSS scattering layer data.

The average variance spectra for each run (Fig. 10) show that the patterns in the individual images are similar, but the magnitudes vary with mean scattering strengths. The spectral levels from imagery on the left side of the TVSS always exceed their counterparts on the right side because a hydrophone on the right side of the TVSS failed during the experiment, yielding a decreased gain on the right side of the array. The spectral levels for run 2 were half an order of magnitude lower than those for runs 1 and 3, suggesting that a larger scale of variability might exist which could not be resolved by the spectra in each run. These differences are most likely due to patches longer than 1 km because the three parallel runs were roughly 3.6 km long and spaced 200 m apart (Fig. 2), and coarse scale $[O(1-100\ \text{km})]$ horizontal patchiness of zooplankton is common.[36,39]

Despite the limitations of boundary echo sidelobe interference, we contend that the multibeam geometry of the TVSS provides a bioacoustic remote sensing capability superior to that of the conventional single beam echo-sounder. Because distributions of marine organisms vary in four dimensions (x, y, z, t), they can be sampled more completely by a multibeam sonar system that provides quasi-synoptic coverage through simultaneous horizontal and vertical soundings.

Multibeam systems such as the TVSS also are perfectly suited for the type of three-dimensional visualizations of biological scattering fields presented in Green *et al.*[40] Valuable qualitative and quantitative bioacoustic assessments, which have been mostly studied with single-beam systems, could be obtained from images like those in Fig. 8 with their 2D-point-kriging/3D-gridding techniques. TVSS data collected over the same track at various intervals (hours, days, weeks) could be used to characterize the 4D spatiotemporal dynamics of diel migration, interaction with dynamic features, and seasonal migration. In the future, we envision merging the multibeam geometry with a multifrequency or broadband capability to provide information regarding the 4D dynamics of species interactions and community structure that is presently unobtainable.

C. Mean volume reverberation

High-frequency acoustical scattering from zooplankton is important in non-bioacoustic applications because it can be a significant component of the total volume reverberation level. To quantify this for the TVSS experiment site, we used

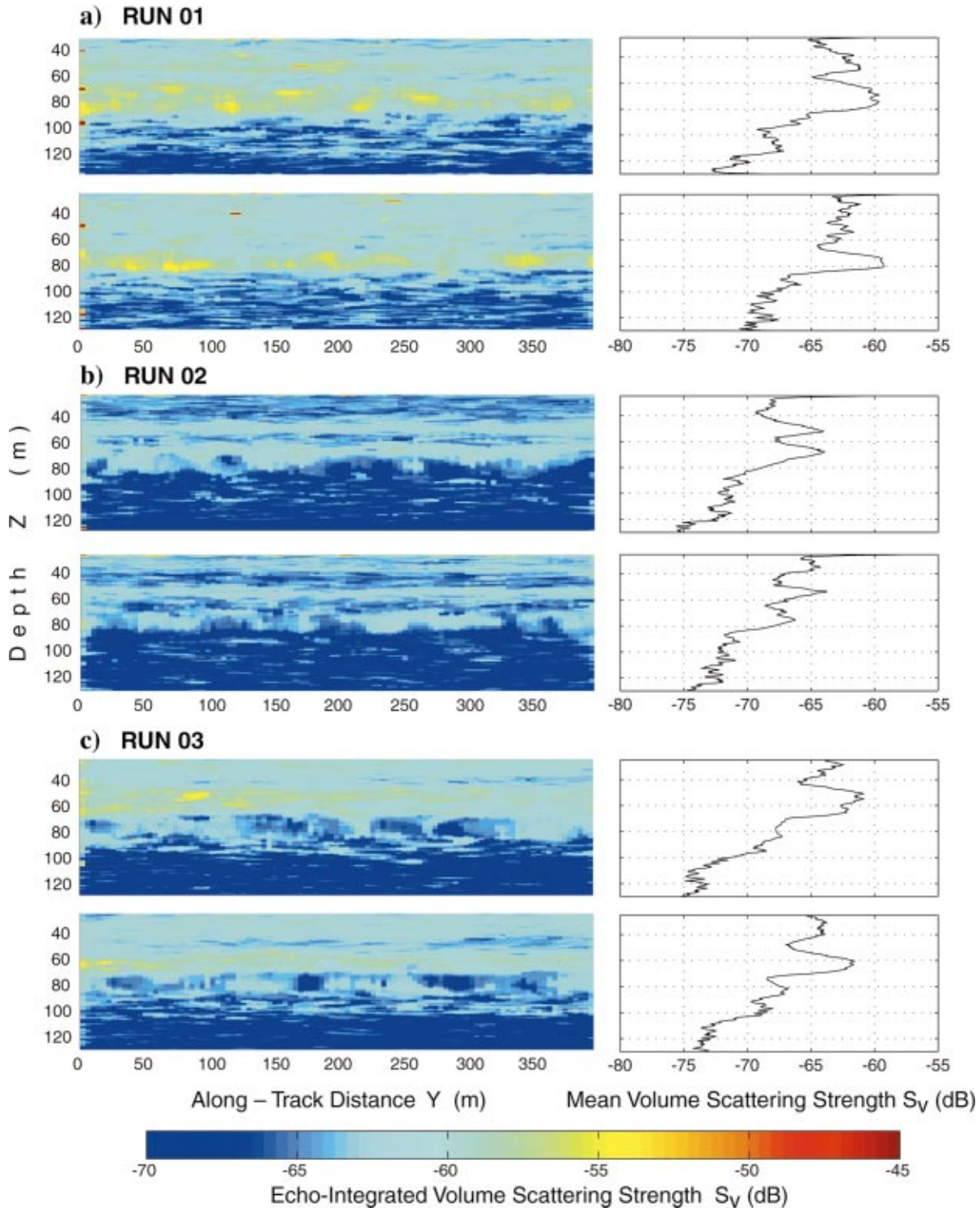


FIG. 8. (a)–(c) Along-track vertical sections of echo-integrated volume acoustic backscattering strength (S_V) from data 47 m across-track to the left and right of the towfish (respectively top and bottom image in each run). The corresponding mean ($\langle S_V \rangle$) profiles averaged over the number of usable pings (N_p) in each run are displayed to the right. Pings in which an excessive number of samples were saturated were removed from the images and not used in the averages to prevent upward biases in the S_V values. $N_p = 98$, 97, and 99 for runs 1, 2, and 3, respectively.

Eq. (2) to compute reverberation levels (RL) in each beam angle and averaged them over all pings and samples prior to the first echo from the township's wake. Pings which suffered from saturation at recording time were not used in computing the averages, so the number of pings used to average RL was less than 100 in each run. The average reverberation levels ($\overline{\mathrm{RL}}$) for the three TVSS runs (Fig. 11) exhibit a similar angular dependence, with reverberation levels over 10 dB higher in the beams directed towards the sea surface than those directed towards the bottom (Table III). Comparison between the image in Fig. 8 and the plots in Fig. 11 indicates that the scattering layers between 40- and 80-m

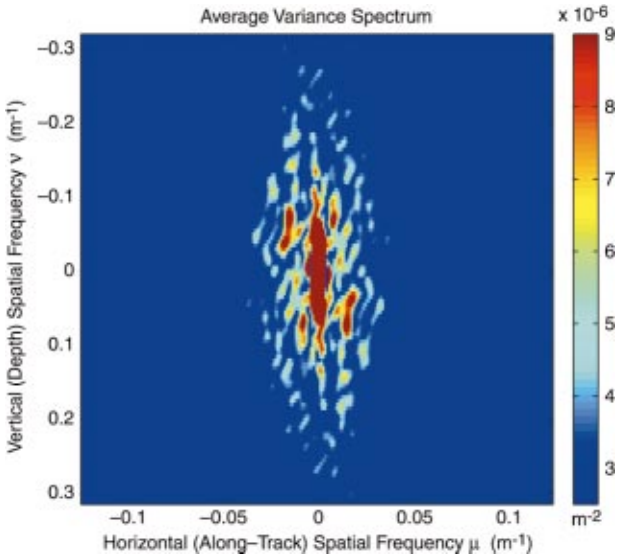


FIG. 9. The average of the six two-dimensional variance spectra of volume scattering coefficients corresponding to the images in Figs. 8(a)–(c). The Appendix describes the calculations, which included a two-dimensional Hanning window to reduce sidelobe leakage. The maximum and minimum length scales resolved are 2.8 and 50 m in the vertical, and 8 and 205 m in the horizontal.

depth, directly above the TVSS are the primary cause for the vertical directionality of the mean volume reverberation levels, as well as the relative peaks near 290 and 70 degrees in each run. The smaller peaks and nulls in $\overline{\mathrm{RL}}$ correspond to peaks and nulls in the transmit beampattern.

Figure 11 directly quantifies the angular-dependent noise floor for non-bioacoustic applications of these TVSS data. The minimum value in the downlooking beams corresponds to a minimum S_V measurable by the TVSS of -75 dB. The plots in Fig. 11 also emphasize that detection near the surface will be more difficult than near the bottom. Although we can expect similar results in many shallow water environments, the shape of $\overline{\mathrm{RL}}$ will vary with changes in scattering layer depths and thicknesses, as well as with the sonar's depth, beampattern, source level, and receiver characteristics. This information is as important in Naval applications (target detection) as in acoustic studies of the sea surface, seafloor, air-sea interaction, and mixing processes. A potential application of the TVSS in a passive mode would be to image the ambient noise field of the sea surface to determine the spatial characteristics of breaking waves and the time variability of the sea surface wave spectrum.[41]

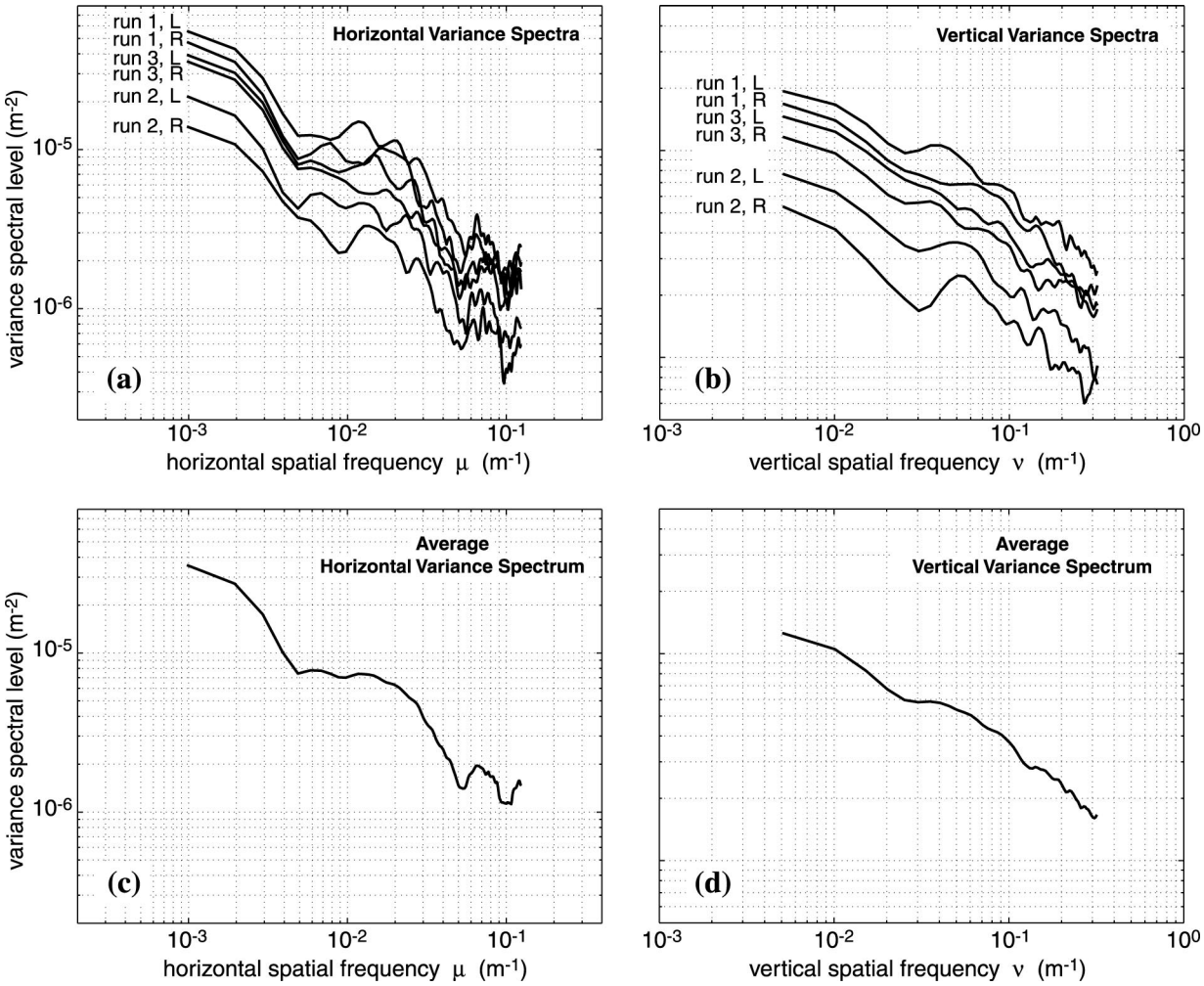


FIG. 10. Log-log plots of (a) horizontal variance spectra averaged over all vertical spatial frequencies in the two-dimensional spectra computed from each image in Fig. 8; L and R refer to images formed 47 m to the left and right side of the TVSS, respectively; (b) vertical variance spectra averaged over all horizontal spatial frequencies; (c) average horizontal, and (d) vertical variance spectrum computed from (a) and (b), respectively.

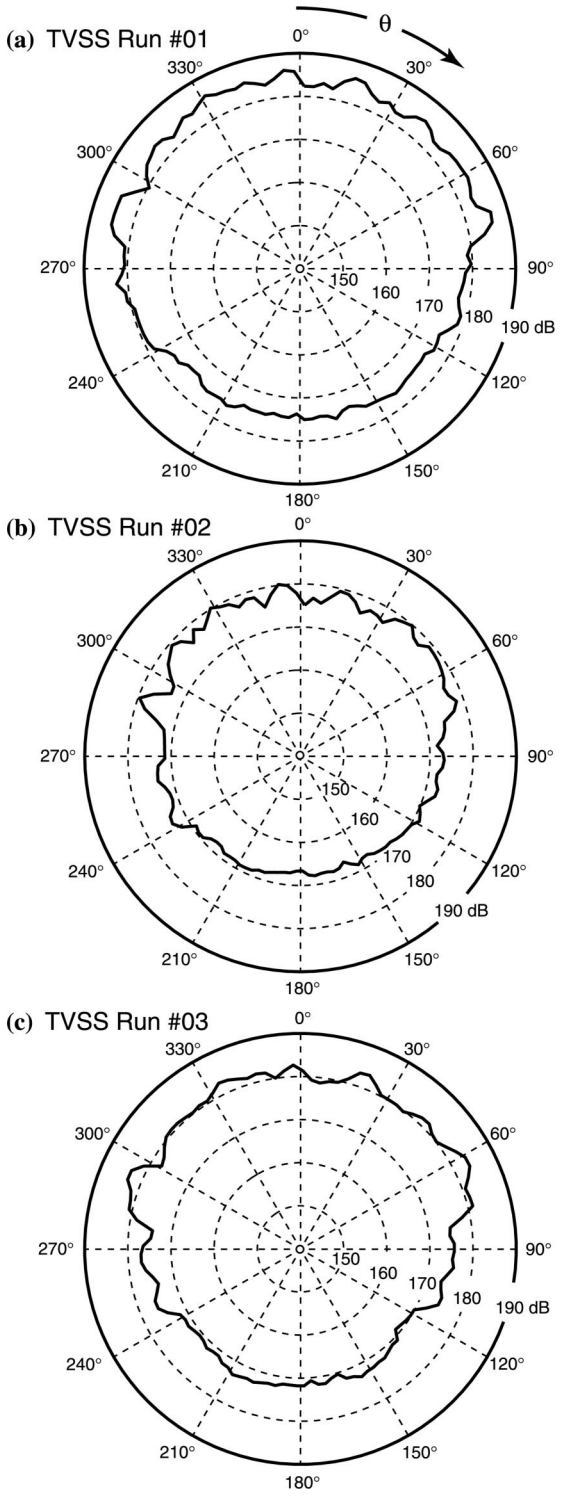


FIG. 11. N_p-ping average of volume reverberation levels ($\overline{\mathrm{RL}}(\theta)$) (dB) for runs 1–3 [(a)–(c)]. $\theta=0$ and 180 degrees are respectively the towfish's zenith and nadir, facing the towfish's direction of travel.

TABLE III. Average volume reverberation characteristics from Fig. 11, and computed over N_p pings. $\overline{\mathrm{RL}}_{\max}$ is the maximum over all angles θ at $\theta_{\max}$, and $\overline{\mathrm{RL}}_{\min}$ is the minimum at $\theta_{\min}$.

	Run 1	Run 2	Run 3
N_p	98	97	99
$\overline{\mathrm{RL}}_{\max}$ (dB)	187.3	180.2	184.9
$\theta_{\max}$	355°	352°	64°
$\overline{\mathrm{RL}}_{\min}$ (dB)	174.6	166.8	169.1
$\theta_{\min}$	181°	157°	130°
$\overline{\mathrm{RL}}_{\max}-\overline{\mathrm{RL}}_{\min}$ (dB)	12.7	13.4	15.8

D. Environmental sources of uncertainty

The most significant environmental source of uncertainty was the boundary echo sidelobe interference illustrated in Figs. 3 and 4. The scattering layer data (Table II) were not affected because we limited the slant range coverage of Fig. 8 to distances shorter than the first echo from the towship's wake. Although the fish school S_V and TS data (Table I) were obtained after the wake and surface echoes, analysis of the scattering strength time series in beams directed towards the fish schools showed that the upward bias in the fish school TS and S_V due to surface reverberation was no greater than 0.5 dB. This was fortunate, because the fish school echoes would have been significantly biased if they were closer to the TVSS or if the TVSS were towed at a greater depth, particularly if they were received at the same time as the surface echo. This illustrates that, in any shallow water multibeam application, the extent to which the data are limited by boundary echo sidelobe interference depends on the experimental geometry. Although this problem was enhanced in our results because the TVSS operated with 360 degree ensonification in the vertical plane perpendicular to the array axis, it could be reduced if the TVSS ensonified in discrete angular sectors, each electronically scanned over several pings to cover the 360 degrees about the array's axis.

There was some uncertainty regarding the sound speed used in this study because only one CTD cast was obtained about 100 m north of the location for run 1 (Fig. 2), approximately 4.5 h prior to the acoustic data collection. Any spatial variability in the local sound speed environment would have produced distortions in the processed imagery, hence errors in the scattering layer depth and thickness estimates (Table II), and fish school shape representation (Fig. 6). To evaluate the presence and effects of variability in the local sound speed environment, we processed seafloor bathymetry and sea surface relief maps using the sound speed profile in Fig. 7(b) and constant gradient ray-tracing methods, then compared them to known environmental data.[24,25] If the sound speed profile were not representative of the local sound speed environment, uncompensated ray bending would produce errors in the echo arrival angles, causing the sea surface relief maps and bathymetry to ''curl'' upward or downward in a symmetric manner about the track's centerline. Because the processed bathymetry was consistent with that previously obtained during bathymetric surveys performed by the Naval Oceanographic Office, and the relatively flat sea surface relief maps were consistent with the calm conditions observed during the experiment, we deemed sound speed errors negligible.

Conditions near the seafloor that may adversely affect the acoustic backscatter from near-bottom fish include seafloor relief, vegetation, and suspended sediment, but none was seen to influence our results. As shown in Fig. 5, the

seafloor in the region was relatively flat, so acoustic shadowing was not a potential source of uncertainty. Similarly, aquatic vegetation could not have been a source of ambiguity for the fish schools in Fig. 5 because the seagrasses capable of strongly scattering sound in the northeastern GOM are limited to depths shallower than 30 m.[42] Finally, acoustic backscatter from suspended sediment was extremely unlikely in the TVSS data because the lifting velocities of the fine-grained sediments in the TVSS experiment region are much greater than the typical bottom currents expected over the Florida shelf in the winter,[15] and because the scattering strengths expected from sediments at the site would have been much less than those in Table I (cf. Ref. 39).

In both the fish and zooplankton S_V and TS measurements, uncertainties due to multiple scattering and extinction also were deemed negligible. These effects can occur in fish schools with densities of 50/m^3 and greater,[33] and may occur in dense swarms of macrozooplankton.[43] The maximum S_V values in the TVSS scattering layer data (Figs. 7 and 8; Table II) were not large enough to suggest that the density of scatterers was sufficiently high to produce multiple scattering or extinction. To evaluate the effects of extinction through the fish schools, we compared the mean normal incidence bottom backscattering strength over runs 1 and 2 to the instantaneous bottom backscattering strength values below the fish schools. Values below the school were less than a standard deviation from the mean values, and comparable to theoretical predictions for the silt and sand in the region,[24] demonstrating that extinction was insignificant (Table I). In a similar manner, multiple scattering was ruled out by adjusting the dynamic range in Fig. 5 and analyzing the backscatter imagery under the schools (not shown). Multiple scattering returns would be evident as a ''tail'' below each school and ''under'' the bottom (e.g., Fig. 4 in Ref. 44), but such features were not found in the images.

Scattering and attenuation from resonant microbubbles are more common sources of uncertainty near the sea surface. They can be formed by breaking waves, propeller cavitation, and even zooplankton and fish.[5,25] As with boundary reverberation, they can produce upward biases in TS and S_V estimates when they are undetected. At the TVSS's acoustic frequency of 68 kHz, resonant scattering comes from bubbles with radii equal to 48 μm, with a single bubble target strength of -66 dB. For down-looking sonars near the surface, attenuation through the bubble layer will decrease the backscattered energy. These effects were negligible in this study because we used the acoustic backscatter imagery to delineate regions where acceptable data could be taken. The vertical volume images in Fig. 7 and Ref. 25 were used to define the maximum bubble depth of the wake that we set as the maximum slant range for both the scattering layer imagery and mean volume reverberation calculations.

We also used horizontal volume imagery to distinguish near-surface bubble clouds from schooling fish. Bubble clouds were observed at 3-m depth 50–100 m to either side of the ship's track, with mean volume scattering strengths of -35 dB, suggesting that they might be due to dense schools of large fish. Comparison between the near-surface horizontal imagery and other acoustic observations of ship wakes showed them to be bubble clouds generated by the towship's breaking bow waves, so they were not investigated with the other fish school data.

Ambient noise was a final source of uncertainty suggested by the similarity between the vertical directionality in mean volume reverberation levels (Fig. 11) and that observed for high-frequency ambient noise.[45] However, for biologic, surface, and ship-generated ambient noise sources, the contributions were insignificant. Biologic sources known to generate sound around 68 kHz include several species of dolphin and porpoise, and snapping shrimp[46] that are found in the Gulf of Mexico. However, acoustic backscatter from large scatterers like dolphin and porpoise were not observed in any of the volume and near-surface imagery, and snapping shrimp are not likely in water depths deeper than 60 m.[46] Because the sea state was only 1.5, surface-generated noise contributions were probably less than 30 dB *re* 1μ Pa. Propeller cavitation, as evidenced by the dense bubble clouds in the towship's wake (Fig. 7), would have dominated the towship-generated noise sources, but a review of cavitation noise data at frequencies near 68 kHz for ships with characteristics similar to the TVSS towship's[47] (e.g., Fig. 10.15 in Ref. 4) suggests that the cavitation noise level during the TVSS experiment would have been 90–100 dB *re*: 1 μPa @1 m, and even less at the TVSS due to spherical spreading and absorption losses. Had the TVSS operated at a lower frequency, and closer to the towship, the cavitation noise levels may have been significant.

IV. SUMMARY AND CONCLUSIONS

In this study, we have used the data collected by the TVSS to demonstrate the advantages and limitations of multibeam sonars in bioacoustic applications. TVSS imagery showed that the most significant limitation was boundary echo sidelobe interference, which prevented zooplankton remote sensing in the volume near the range of the first boundary echo, and fish detection near the range of the first bottom echo. The 3D representation of a near-bottom fish school demonstrated the fundamental limitation of sonar resolution inherent in both multibeam and single beam characterizations of scattering fields. Angular variation in the transmit array beampattern also contributed to uncertainties in the TVSS data.

Despite these limitations, the TVSS still provided more coverage than that possible with a single beam sonar, and we used this advantage to characterize the 3D acoustic structure of near-bottom schooling fish and zooplankton scattering layers in a shallow water region of the northeastern Gulf of Mexico. Supporting previous studies in the region, the TVSS scattering layer imagery indicated that the vertical distribution of zooplankton is closely associated with the mixed layer depth. The TVSS geometry also provided the unique capability to characterize the vertical angular dependence of volume reverberation, shown to be affected by bioacoustic scattering layers lying above the sonar. In the future, modifications of the transmit array beampattern or ping repetition cycle, and the application of sidelobe cancellation techniques, could reduce the effects of boundary reverberation

and increase the potential for toroidal multibeam sonars in bioacoustic remote sensing applications.

ACKNOWLEDGMENTS

This work was funded by the Office of Naval Research under ONR-NRL Contract No. N00014-96-1-G913. The authors would like to thank CAPT Tim Schnoor, USN (ONR), Sam Tooma, and Maria Kalcic (NRL) for their support; Candy Robertson and Lisa Tubridy (CSS) for information on the TVSS; and Pat Jordan (MPL) for administrative support. Chris Gledhill (NMFS) and Jason Link (NMFS) are thanked for providing the NMFS Small Pelagics Report.[17] Thanks are due to Jo Griffith (MPL) for help with the figures.

APPENDIX: TWO-DIMENSIONAL VARIANCE SPECTRUM

The two-dimensional variance spectrum in Fig. 9 is calculated from the backscattering coefficients (s_V) corresponding to each image in Fig. 8, where $S_V = 10\log_{10}(s_V)$. The s_V data for each image were first resampled vertically to 1.4-m depth intervals. Because the sidescanning procedure used to form the images resulted in nonuniform sample spacing which decreases away from normal incidence, the majority of the samples have spacing smaller than 1.4 m, so they were subsampled. A few samples near normal incidence ($z = 78$ m) are spaced 2.5 m apart, and were interpolated to 1.4-m spacing. The average horizontal sample spacing in each image is 4.1 m. The resulting $M \times N$ image, denoted $s_V(m,n)$, is demeaned according to

$$s'_V(m,n) = s_V(m,n) - (1/MN)\sum_{m=1}^{M}\sum_{n=1}^{N} s_V(m,n) \quad (\mathrm{m}^{-1}). \tag{A1}$$

The two-dimensional discrete Fourier transform of $s'_V(m,n)$ is computed as[48]

$$F_V(p,q) = (1/MN)\sum_{m=1}^{M}\sum_{n=1}^{N} s'_V(m,n) \times e^{-i2\pi(p(m-1)/M + q(n-1)/N)}, \tag{A2}$$

where p and q are vertical and horizontal spatial frequency indices, respectively. The variance spectrum is then

$$V_V(p,q) = (1/MN)|F_V(p,q)|^2 \quad (\mathrm{m}^{-2}). \tag{A3}$$

Thus, the variance spectrum is simply the two-dimensional power spectrum of the demeaned image. Figure 9 is the result of averaging the six spectra corresponding to Fig. 8.

The relationship between the indices p and q and their corresponding spatial frequencies μ and ν is

$$\nu = (1/dz)(p - M/2)/M, \quad \mu = (1/dy)(q - N/2)/N, \tag{A4}$$

where dz and dy are the vertical and horizontal sampling intervals, respectively. The maximum and minimum positive resolvable spatial frequencies are given by

$$\nu_{\max} = 1/(2dz), \quad \mu_{\max} = 1/(2dy), \quad \nu_{\min} = 2/(Mdz), \quad \mu_{\min} = 2(Ndy). \tag{A5}$$

If the total depth and along-track distance spanned by each image in Fig. 8 are denoted Z and Y, then the minimum resolvable spatial frequencies are also given by

$$\nu_{\min} = 2/Z, \quad \mu_{\min} = 2/Y. \tag{A6}$$

Thus, for Figs. 8 and 9, $dz = 1.4$ m, $dy = 4.1$ m, $Z = 100$ m, and $Y = 410$ m, so that

$$\nu_{\max} = 0.35\ \mathrm{m}^{-1}, \quad \mu_{\max} = 0.12\ \mathrm{m}^{-1}, \quad \nu_{\min} = 0.02\ \mathrm{m}^{-1}, \quad \mu_{\min} = 0.005\ \mathrm{m}^{-1}. \tag{A7}$$

The total sample variance in the backscattering coefficients $s_V(m,n)$ corresponding to each image in Fig. 8 is given by[49]

$$1/(MN)\sum_{m=1}^{M}\sum_{n=1}^{N} |s'_V(m,n)|^2, \tag{A8}$$

which can be related to the variance spectrum by applying Parseval's theorem generalized for the two-dimensional Fourier transform to Eq. (A2):

$$(1/MN)^2\sum_{p=1}^{M}\sum_{q=1}^{N} |F_V(p,q)|^2 = (1/MN)\sum_{m=1}^{M}\sum_{n=1}^{N} |s'_V(m,n)|^2. \tag{A9}$$

Substituting (A3) into (A9) yields

$$(1/MN)\sum_{p=1}^{M}\sum_{q=1}^{N} V_V(p,q) = (1/MN)\sum_{m=1}^{M}\sum_{n=1}^{N} |s'_V(m,n)|^2, \tag{A10}$$

which shows that the average value of the variance spectrum is equal to the total sample variance in the original data.

[1] T. C. Gallaudet, C. P. de Moustier, and M. Kalcic, "Imaging the ocean boundaries and volume with the Toroidal Volume Search Sonar (TVSS)," in *Proceedings of the Fourth Annual Symposium on the Mine Problem*, 13–16 March 2000, Monterey, CA.

[2] G. Rose, "Acoustics in fisheries in the 21st century," J. Acoust. Soc. Am. **108**, 2457 (2000).

[3] C. F. Eyring, R. J. Christensen, and R. W. Raitt, "Reverberation in the sea," J. Acoust. Soc. Am. **20**, 462–475 (1948).

[4] R. J. Urick, *Principles of Underwater Sound*, 3rd ed. (Peninsula, Los Altos, CA, 1983).

[5] H. Medwin and C. S. Clay, *Fundamentals of Acoustical Oceanography* (Academic, San Diego, 1998).

[6] B. Pedersen and M. Trevorrow, "Continuous monitoring of fish in a shallow channel with a fixed horizontal sonar," J. Acoust. Soc. Am. **105**, 3126–3135 (1999).

[7] O. A. Misund, A. Aglen, and E. Fronaes, "Mapping the shape, size, and density of fish-schools by echo-integration and a high resolution sonar," ICES J. Mar. Sci. **52**, 11–20 (1995).

[8] T. W. Steig and S. V. Johnston, "Monitoring fish movement patterns in a reservoir using horizontally scanning split-beam techniques," ICES J. Mar. Sci. **53**, 435–441 (1996).

[9] D. M. Farmer, M. V. Trevorrow, and B. Pedersen, "Intermediate range fish detection with a 12 kHz sidescan sonar," J. Acoust. Soc. Am. **105**, 2481–2490 (1999).

[10] P. Greenblatt, "Distributions of volume scattering observed with an 87.5 kHz sonar," J. Acoust. Soc. Am. **71**, 879–885 (1982).

[11] D. V. Holliday and R. E. Pieper, "Volume scattering strengths and zooplankton distributions at acoustic frequencies between 0.5 and 3 MHz," J. Acoust. Soc. Am. **67**, 135–146 (1980).

[12] D. McGehee and J. S. Jaffe, "Three-dimensional swimming behavior of individual zooplankters: observations using the acoustical imaging system FishTV," ICES J. Mar. Sci. **53**, 363–369 (1996).

[13] M. Soria, P. Freon, and F. Gerlotto, "Analysis of vessel influence on spatial behaviour of fish schools using multi-beam sonar and conse-

quences for biomass estimates by echo-sounder,'' ICES J. Mar. Sci. **53**, 453–458 (1996).
[14] O. Misund, A. Aglen, J. Hamre, E. Ona, I. Rottingen, D. Skagen, and J. Valdemarsen, ''Improved mapping of schooling fish near the surface: comparison of abundance estimates obtained by sonar and echo integration,'' ICES J. Mar. Sci. **53**, 383–388 (1996).
[15] G. A. Maul, ''The annual cycle of the Gulf Loop Current Part I: Observations during a one-year time series,'' J. Mar. Res. **35**, 29–47 (1977).
[16] T. C. Gallaudet, ''Shallow water acoustic backscatter and reverberation measurements using a 68 kHz cylindrical array,'' Ph.D. dissertation, University of California, San Diego, 2001.
[17] J. S. Link, T. A. Henwood, and C. T. Gledhill, *Small Pelagics in the Gulf of Mexico: A Description of the Abundance Geographic Distribution, Size, and Depth Distribution of Major Pelagic Species* (National Marine Fisheries Service, Southeast Fisheries Science Center, Pascagoula, MS, 1997).
[18] T. L. Hopkins, ''The vertical distribution of zooplankton in the eastern Gulf of Mexico,'' Deep-Sea Res., Part A **29**, 1069–1083 (1982).
[19] R. A. Zimmerman and D. C. Biggs, ''Patterns of distribution of sound scattering zooplankton in warm- and cold-core eddies in the Gulf of Mexico, from a narrowband acoustic Doppler current profiler survey,'' J. Geophys. Res. **104**, 5251–5262 (1999).
[20] T. L. Hopkins, D. M. Milliken, L. M. Bell, E. J. McMichael, J. J. Heffernan, and R. V. Cano, ''The landward distribution of oceanic plankton and micronekton, over the west Florida continental shelf as related to their vertical distribution,'' J. Plankton Res. **3**, 645–658 (1981).
[21] P. B. Ortner, R. L. Ferguson, S. R. Piotrowicz, L. Chesal, G. Berberian, and A. V. Palumbo, ''Biological consequences of hydrographic and atmospheric advection within the Gulf Loop Intrusion,'' Deep-Sea Res., Part A **31**, 1101–1120 (1984).
[22] P. B. Ortner, L. C. Hill, and S. R. Cummings, ''Zooplankton community structure and copepod species composition in the northern Gulf of Mexico,'' Cont. Shelf Res. **9**, 387–402 (1989).
[23] T. C. Gallaudet and C. P. de Moustier, ''On optimal amplitude shading for arrays of irregularly spaced or non-coplanar elements,'' IEEE J. Ocean. Eng. **25**, 553–567 (2000).
[24] T. C. Gallaudet and C. P. de Moustier, ''Using environmental information to estimate and correct for errors in bathymetry and seafloor acoustic imagery,'' IEEE J. Ocean Eng. submitted (2002).
[25] T. C. Gallaudet and C. P. de Moustier, ''Sea surface and volume backscattering strength measurements in the microbubble field of a ship's wake,'' J. Acoust. Soc. Am. submitted (2001).
[26] R. W. Nero, C. H. Thompson, J. R. Dubberley, and R. H. Love, ''Herring hydroglyphics in littoral waters of the northern Gulf of Mexico,'' J. Acoust. Soc. Am. **108**, 2489 (2000).
[27] R. C. Herron, T. D. Leming, and J. Li, ''Satellite-detected fronts and butterfish aggregations in the northeastern Gulf of Mexico,'' Cont. Shelf Res. **9**, 569–588 (1989).
[28] R. Love, ''Target strength of an individual fish at any aspect,'' J. Acoust. Soc. Am. **62**, 1397–1403 (1977).
[29] K. G. Foote, ''Fish target strengths for use in echo integrator surveys,'' J. Acoust. Soc. Am. **82**, 981–987 (1987).
[30] A. J. Pluddemann and R. Pinkel, ''Characterization of the patterns of diel migration using a Doppler sonar,'' Deep-Sea Res., Part A **36**, 509–530 (1989).
[31] L. Lecornu, V. Burdin, C. Scalabrin, and C. Hamitouche, ''Fish school analysis from multibeam sonar image processing,'' *Proc. IEEE Oceans '98*, Vol. 1, pp. 587–591 (1998).
[32] R. K. Johnson, ''Sound scattering from a fluid sphere revisited,'' J. Acoust. Soc. Am. **61**, 375–377 (1977).
[33] T. K. Stanton, R. D. Nash, R. L. Eastwood, and R. W. Nero, ''A field examination of acoustical scattering from marine organisms at 70 kHz,'' IEEE J. Ocean. Eng. **12**, 339–348 (1987).
[34] R. H. Love, ''Dorsal-aspect target strength of an individual fish,'' J. Acoust. Soc. Am. **49**, 816–823 (1971).
[35] D. L. Mackas and C. M. Boyd, ''Spectral analysis of zooplankton spatial heterogeneity,'' Science **204**, 62–64 (1979).
[36] R. W. Nero and J. J. Magnuson, ''Characterization of patches along transects using high resolution 70 kHz integrated acoustic data,'' Can. J. Fish. Aquat. Sci. **46**, 2056–2064 (1989).
[37] P. H. Wiebe, ''Small scale distribution in oceanic zooplankton,'' Limnol. Oceanogr. **15**, 205–217 (1970).
[38] C. F. Greenlaw and W. G. Pearcy, ''Acoustical patchiness of mesopelagic micronekton,'' J. Mar. Res. **43**, 163–178 (1985).
[39] P. H. Wiebe, T. K. Stanton, M. C. Benfield, D. G. Mountain, and C. H. Greene, ''High-frequency acoustic volume backscattering in the Georges Bank coastal region and its interpretation using scattering models,'' IEEE J. Ocean. Eng. **22**, 445–464 (1997).
[40] C. H. Greene, P. H. Wiebe, C. Pelkie, M. C. Benfield, and J. M. Popp, ''Three-dimensional acoustic visualization of zooplankton patchiness,'' Deep-Sea Res., Part II **45**, 1201–1217 (1998).
[41] P. A. Crowther and A. Hansla, ''The lifetimes, velocities, and probable origin of sonic and ultrasonic noise sources on the sea surface,'' in *Natural and Physical Sources of Underwater Sound*, edited by B. R. Kerman (Kluwer, Boston, 1993), pp. 379–392.
[42] S. Z. El-Sayed, W. M. Sackett, L. M. Jeffrey, A. D. Fredricks, R. P. Saunders, P. S. Conger, G. A. Fryxell, K. A. Steidinger, and S. A. Earle, ''Chemistry, primary productivity, and benthic algae of the Gulf of Mexico,'' in *Serial Atlas of the Marine Environment Folio 22* (American Geographical Society, New York, 1972).
[43] K. G. Foote, ''Correcting acoustic measurements of scatterer density for extinction,'' J. Acoust. Soc. Am. **88**, 1543–1546 (1990).
[44] T. K. Stanton, ''Effects of second-order scattering on high resolution sonars,'' J. Acoust. Soc. Am. **76**, 861–866 (1984).
[45] P. A. Crowther, H. J. S. Griffiths, and A. Hansla, ''Dependence of sea surface noise in narrow beams on windspeed and vertical angle,'' in *Natural and Physical Sources of Underwater Sound*, edited by B. R. Kerman (Kluwer, Boston, 1993), pp. 31–44.
[46] Applied Physics Laboratory, *APL-UW High Frequency Ocean Environmental Acoustic Models Handbook*, Technical Report APL-UW TR 9407 (Univ. of Washington, Seattle, 1994), Part III.
[47] F. R. Young, Cavitation (McGraw Hill, 1989).
[48] R. N. Bracewell, *The Fourier Transform and Its Applications*, 2nd ed. (McGraw Hill, New York, 1986).
[49] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*, 2nd ed. (Wiley, New York, 1986).

Thermodynamic efficiency of thermoacoustic mixture separation

D. A. Geller and G. W. Swift
Condensed Matter and Thermal Physics Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 1 March 2002; revised 15 May 2002; accepted 21 May 2002)

The acoustic power loss in the thermoacoustic mixture-separation process is derived, including the contributions due to a nonzero gradient in concentration. The significance of the gradient-dependent term is discussed. The limiting thermodynamic efficiency of the separation is calculated. Under reasonable circumstances, the efficiency approaches $10^{-2}n_H n_L(\Delta m/m_{\mathrm{avg}})^2$, where n_H and n_L are the mole fractions of the two components of the mixture, and $\Delta m/m_{\mathrm{avg}}$ is the fractional difference between the molar masses of the two components. This efficiency is of the same order of magnitude as that of some other, more conventional separation methods. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1494446]

PACS numbers: 43.35.Ud, 43.35.Ty [RR]

I. INTRODUCTION

For 20 years, the field of thermoacoustics has been loosely oriented around the problems of energy conversion in engines and refrigerators.[1] With the discovery of thermoacoustic mixture separation,[2–4] though, another energy-related aspect of thermoacoustics has emerged.

Sound waves in a binary gas mixture in a duct can cause separation of the mixture because the quasiadiabatic temperature oscillation in the bulk of the gas alternately drives each component of the mixture toward the isothermal wall of the duct through thermal diffusion. Then, whichever component has migrated into the boundary layer is trapped there by viscosity during the ensuing motion of the gas. This cycle of thermal diffusion and viscous motion causes the time-averaged net flow of the mixture components in opposite directions along the duct. Thermoacoustic mixture separation is thus intrinsically thermodynamically irreversible, because it is based on two diffusive processes. Ordinary mass diffusion, which is present in all other statistical separation processes as well, is an additional source of irreversibility.

Research and industry have myriad needs to separate mixtures. Whether thermoacoustics can play a meaningful role in this arena depends in part on the thermodynamic efficiency of the thermoacoustic mixture-separation process, which we consider in this paper. First, we derive the full expression for the acoustic power loss to first order in the concentration gradient. We show that the deviation of this loss from the ordinary thermal and viscous dissipative losses is generally very small. The implications of this result for the functioning of thermoacoustic engines and refrigerators are briefly discussed. Next, we derive expressions for the thermodynamic efficiency of thermoacoustic separation in various circumstances. Relative to the fundamental limit imposed by the first and second laws of thermodynamics, the efficiency is only 0.0023 for a 50–50 He–Ar mixture. It is even lower when the components of the mixture are more nearly equal in mass or less equal in concentration. Nevertheless, the efficiency of thermoacoustic mixture separation is actually of the same order of magnitude as that of other widely used irreversible separation processes. Hence, thermoacoustic mixture separation is *not* hopelessly inefficient. Other features, such as simplicity, reliability, low fabrication cost, ability to operate at ambient temperature and pressure, and independence of gravity might make this technique useful for some applications.

II. CALCULATION OF THE DISSIPATION

In order to assess the efficiency of thermoacoustic mixture separation, we first must calculate the rate of acoustic power loss per unit length in the duct. The acoustic loss may not be strictly dissipative, because some of the acoustic power is converted into the free energy of separating the gases. Nevertheless, we expect the dissipation to be substantial, because the separation originates from viscosity and thermal diffusion in the boundary layer, both of which are dissipative processes. To obtain the time-averaged acoustic power loss per unit length, $d\dot{E}_2/dx$, we start from the general formula

$$\frac{d\dot{E}_2}{dx} = -A\frac{d\langle\overline{pu}\rangle}{dx} = -\frac{1}{2}A\Re\left[\langle\tilde{u}_1\rangle\frac{dp_1}{dx} + \tilde{p}_1\frac{d\langle u_1\rangle}{dx}\right], \quad (1)$$

where p_1 and u_1 are the complex amplitudes of oscillating pressure and velocity, respectively, and A is the cross-sectional area of the duct. In our notation,[2,4] $\langle g\rangle$ denotes the average of a quantity g over the cross section of the duct, $\Re[g]$ denotes the real part, $\overline{g}$ denotes the average over time, and $\tilde{g}$ denotes the complex conjugate. The gradient dp_1/dx can be eliminated through the equation for spatially averaged velocity, i.e., the solution of the momentum equation in the acoustic approximation with $u_1 = 0$ at the walls of the duct

$$\frac{dp_1}{dx} = -\frac{\iota\omega\rho_m\langle u_1\rangle}{1-f_\nu}, \quad (2)$$

where ω is the angular frequency, ρ_m is the mean density, and $1-f_\nu$ depends on the profile of velocity on a cross section of the duct and therefore depends on the duct geometry. The gradient of velocity is eliminated using the spatial aver-

 0001-4966/2002/112(2)/504/7/$19.00

age of the equation of continuity to first order in differential quantities

$$\rho_m \frac{d\langle u_1\rangle}{dx} = -\iota\omega\langle\rho_1\rangle - \frac{d\rho_m}{dx}\langle u_1\rangle. \tag{3}$$

Note that we have kept a term in the gradient of the mean density, because for a binary mixture undergoing separation this is not generally zero.

Towards expressing Eq. (1) in terms of $|\langle u_1\rangle|$ and $|p_1|$ alone, without their derivatives, we write the expression for a differential change in the density, following from the ideal gas law

$$\frac{d\rho}{\rho} = \frac{dp}{p} - \frac{dT}{T} + \frac{m_H - m_L}{m_{\text{avg}}} dn_H, \tag{4}$$

where T is the temperature, n_H is the mole fraction of the heavy component, m_H and m_L are the molar masses of the heavy and light components, respectively, and $m_{\text{avg}} = n_H m_H + (1-n_H)m_L$ is the molar average mass. This identity can be used to expand both the first-order oscillating component $\langle\rho_1\rangle$ and the gradient $d\rho_m/dx$ in Eq. (3). As with our earlier derivation of the saturation value of the concentration gradient,[4] we continue to restrict the problem to the case of $dT_m/dx = 0$, and we also assume $dp_m/dx = 0$. Then, Eq. (3) becomes

$$\frac{d\langle u_1\rangle}{dx} = \iota\omega\frac{\langle T_1\rangle}{T_m} - \iota\omega\frac{\gamma}{\rho_m a^2}p_1 - \frac{m_H - m_L}{m_{\text{avg}}}\left(\iota\omega\langle n_1\rangle + \frac{dn_H}{dx}\langle u_1\rangle\right), \tag{5}$$

where a is the sound speed in the mixture, γ is the ratio of specific heats, and we have used the identity $\gamma p_m = \rho_m a^2$. To simplify the final term, we use the single-component continuity equation,[5] which in the present notation is

$$\rho_m\left(\iota\omega\langle n_1\rangle + \frac{dn_H}{dx}\langle u_1\rangle\right) = -\frac{m_{\text{avg}}}{m_H}\langle\boldsymbol{\nabla}\cdot\mathbf{i}_1\rangle \tag{6}$$

to first order, where $\mathbf{i}$ is the diffusive mass-flux density vector. On the right side, $\langle di_{1,x}/dx\rangle$ is negligibly small, and $\langle\boldsymbol{\nabla}_r\cdot\mathbf{i}_1\rangle$ (where r represents the coordinates perpendicular to x) is zero by virtue of the divergence theorem and the boundary condition of zero mass diffusion into the duct wall. Hence, Eq. (5) becomes simply

$$\rho_m\frac{d\langle u_1\rangle}{dx} = \iota\omega\frac{\rho_m}{T_m}\langle T_1\rangle - \iota\omega\frac{\gamma}{a^2}p_1, \tag{7}$$

which is identical to the spatially averaged first-order equation of continuity for a single-component, homogeneous fluid.

For both the boundary-layer limit and for a cylindrical tube of arbitrary diameter, we have shown[4] explicitly that $T_1 = (p_1/\rho_m c_p)(1-h)$, where h is a dimensionless function of the transverse coordinate normal to the wall and c_p is the isobaric specific heat. The first-order equation of continuity can then be rewritten, using the identity $\gamma - 1 = a^2/Tc_p$, as

$$\rho_m a^2\frac{d\langle u_1\rangle}{dx} = -\iota\omega[1 + (\gamma-1)\langle h\rangle]p_1. \tag{8}$$

Substituting Eqs. (2) and (8) into (1), the dissipation is

$$\frac{d\dot{E}_2}{dx} = -\frac{\omega A}{2}\Im\left[\frac{\rho_m|\langle u_1\rangle|^2}{1-f_\nu}\right] - \frac{\omega A}{2}\Im\left[\frac{(\gamma-1)\langle h\rangle|p_1|^2}{\rho_m a^2}\right], \tag{9}$$

where $\Im[g]$ denotes the imaginary part of g.

Consider the two terms in brackets separately. The first term is

$$\frac{\omega A}{2}\frac{\rho_m|\langle u_1\rangle|^2}{|1-f_\nu|^2}\Im[-f_\nu], \tag{10}$$

which is just the ordinary viscous damping term, as noted in Ref. 2. The second term must describe the other diffusive losses (both thermal and mass diffusion) plus the useful work done in separating the components of the mixture. Using Eq. (30) from Ref. 4 for T_1, and with $f_i = \langle h_i\rangle$, we have

$$-\frac{\omega A}{2}\frac{(\gamma-1)|p_1|^2}{\rho_m a^2}\Im[\langle h\rangle] = -\frac{\omega A}{2}\frac{(\gamma-1)|p_1|^2}{\rho_m a^2}\Im[Bf_\nu + Cf_{\kappa D} + (1-B-C)f_{D\kappa}], \tag{11}$$

where the dimensionless coefficients B and C are defined as in Ref. 4 for either the boundary-layer limit or for the case of an arbitrary-diameter tube.

Equations (10) and (11) are still general expressions for the two terms in Eq. (9). To make further progress, we now proceed explicitly in the boundary-layer limit. In that case $f_i = (1-\iota)\delta_i/2r_h$, where r_h is the hydraulic radius.[6] Expression (10) becomes

$$\frac{\omega A}{4r_h}\rho_m|\langle u_1\rangle|^2\delta_\nu. \tag{12}$$

From Ref. 4 we recall

$$B = \frac{\iota e^{-\iota\theta}\varepsilon}{1-f_\nu}\frac{\sigma}{(1-\sigma)(1-\sigma L) - \varepsilon\sigma}\Gamma_c, \tag{13}$$

$$C = C_{\text{S\&S}}\left[1 - B\left(1 + \frac{\sigma-1}{\sqrt{\sigma}}\frac{\delta_\kappa}{\sqrt{L}\delta_{\kappa D} - \delta_{D\kappa}}\right)\right], \tag{14}$$

$$C_{\text{S\&S}} = \frac{\sqrt{L}\delta_{\kappa D} - \delta_{D\kappa}}{(1+\sqrt{L})(\delta_{\kappa D} - \delta_{D\kappa})}, \tag{15}$$

where

$$\Gamma_c = \frac{dn_H/dx}{(dn_H/dx)_{\text{sat}}}, \tag{16}$$

$$\left(\frac{dn_H}{dx}\right)_{\text{sat}} \equiv \frac{\gamma-1}{\gamma}k_T\frac{|p_1|}{p_m}\frac{\omega}{|\langle u_1\rangle|}. \tag{17}$$

In these definitions, k_T is the thermal diffusion ratio, $\varepsilon = (\gamma-1)k_T^2/\gamma n_H(1-n_H)$, σ is the Prandtl number, $L = (\delta_\kappa/\delta_D)^2$, the δ_i are penetration depths all defined in Ref. 2 or 4, and θ is the phase angle by which p_1 leads $\langle u_1\rangle$ in time. Insertion of these coefficients in Eq. (11) yields

$$\frac{\omega A}{4r_h}\frac{(\gamma-1)|p_1|^2}{\rho_m a^2}\frac{\sqrt{L}}{1+\sqrt{L}}(\delta_{\kappa D}+\delta_{D\kappa})\times\left[1-\Gamma_c\frac{\varepsilon\sigma(\sin\theta-\cos\theta)}{(1-\sigma)(1-\sigma L)-\varepsilon\sigma}\times\left(1-\frac{1+\sigma\sqrt{L}}{\sqrt{\sigma L}}\frac{\delta_\kappa}{\delta_{\kappa D}+\delta_{D\kappa}}\right)\right] \tag{18}$$

to lowest order in δ/r_h. This expression for the second term of Eq. (9) is similar in form to the expression for the rate of change of acoustic power in a duct due to a longitudinal temperature gradient but ignoring viscosity.[1] Multiplying the prefactor through the square brackets, the first, θ-independent term is just the thermal damping term found previously in Ref. 2, which could have been obtained more rapidly here by setting $\Gamma_c=0$ in Eq. (11) and comparing with Eq. (57) in Ref. 2. The other term depends on both θ and on the concentration gradient through Γ_c. It is proportional to ε, so it is generally small.

The full expression for acoustic power loss can thus be expressed as the sum of three terms in the boundary-layer limit: a viscous term $d\dot{E}_\nu/dx$, a thermal- and diffusion-loss term $d\dot{E}_{\kappa,D}/dx$, and a term due to the concentration gradient, $d\dot{E}_\nabla/dx$. That is,

$$\frac{d\dot{E}_2}{dx}=\frac{d\dot{E}_\nu}{dx}+\frac{d\dot{E}_{\kappa,D}}{dx}+\frac{d\dot{E}_\nabla}{dx}, \tag{19}$$

where

$$\frac{d\dot{E}_\nu}{dx}=\frac{\omega A}{4r_h}\rho_m|\langle u_1\rangle|^2\delta_\nu, \tag{20}$$

$$\frac{d\dot{E}_{\kappa,D}}{dx}=\frac{\omega A}{4r_h}\frac{(\gamma-1)}{\rho_m a^2}|p_1|^2\frac{\sqrt{L}}{1+\sqrt{L}}(\delta_{\kappa D}+\delta_{D\kappa}), \tag{21}$$

$$\frac{d\dot{E}_\nabla}{dx}=\frac{\delta_\kappa}{4r_h}\frac{\gamma-1}{\gamma}\frac{k_T}{n_H(1-n_H)}\frac{dn_H}{dx}\frac{|p_1||U_1|(\cos\theta-\sin\theta)}{(1-\sigma)(1-\sigma L)-\varepsilon\sigma}\times\frac{\sigma\sqrt{L}}{1+\sqrt{L}}\left[\frac{\delta_{\kappa D}}{\delta_\kappa}+\frac{\delta_{D\kappa}}{\delta_\kappa}-\frac{1+\sigma\sqrt{L}}{\sqrt{\sigma L}}\right], \tag{22}$$

with $U_1=A|\langle u_1\rangle|$ the volume flow rate. The new term, $d\dot{E}_\nabla/dx$, must include the reversible work stored in the concentration gradient, but Eq. (18) suggests that $d\dot{E}_\nabla/dx$ be viewed as a correction to the thermal- and mass-diffusion loss accounting for the presence of this gradient. Note that in the limit $k_T\to 0$, we find $d\dot{E}_\nabla/dx\to 0$, and $d\dot{E}_{\kappa,D}/dx$ approaches the well-known boundary-layer thermal-loss term for acoustics alone.

As described in Ref. 4, Watson[7] has also considered mixtures subjected to oscillating flow in ducts, but neglecting thermal diffusion and setting $p_1=0$. Watson took the acoustic dissipation per unit length in a duct to be

$$\frac{d\dot{E}_{\text{Watson}}}{dx}=\Re\left\{\left\langle\tilde{u}_1\frac{dp_1}{dx}\right\rangle\right\}, \tag{23}$$

neglecting the $\tilde{p}_1\,du_1$ portion of the differential acoustic power. As a result, Watson's expressions for the power dissipation are equal only to our expression (10) for the viscous loss. Specifically, Watson's expression (107) for $\omega\to\infty$, which is equivalent to our boundary-layer limit, is exactly equal to our Eq. (20). By discarding the differential $\tilde{p}_1\,du_1$, Watson neglected both the thermal loss and the gradient-dependent loss. Strictly speaking, even if one attempted to keep $p_1=0$ in this situation, the concentration oscillations would drive temperature oscillations (via ε), which in turn could drive density and pressure oscillations, contributing to the gradient-dependent loss. However, we will see in the next section that neglect of this effect is not serious.

III. CONCENTRATION-GRADIENT-DEPENDENT LOSSES

To quantify the relative importance of the new, gradient-dependent term $d\dot{E}_\nabla/dx$, we define the ratio

$$b=\frac{(d\dot{E}_\nabla/dx)}{(d\dot{E}_\nu/dx+d\dot{E}_{\kappa,D}/dx)}, \tag{24}$$

which equals the fractional deviation of the total acoustic loss from the acoustic loss calculated in Ref. 2. This ratio is a function of k_T, T_m, p_m, ρ_m, n_H, ω, r_h, the δ_i, θ, γ, dn_H/dx, $|p_1|$, and $|\langle u_1\rangle|$. However, it is not difficult to show that b is only weakly dependent on most of these parameters, and we can obtain a rough upper limit on this ratio for cases of practical interest.

The ratio b is proportional to the concentration gradient. For any closed system, the highest gradient will occur when the thermoacoustic separation process saturates at $\dot{N}_H=0$. Using Eqs. (41), (44), and (56) from Ref. 4, we can write this limiting value as

$$\left(\frac{dn_H}{dx}\right)_{\text{lim}}=\frac{(F_{\text{trav}}\cos\theta+F_{\text{stand}}\sin\theta)(\gamma-1)k_T|p_1||\langle u_1\rangle|/\gamma p_m}{4D_{12}r_h/\delta_\kappa-F_{\text{grad}}|\langle u_1\rangle|^2/\omega}, \tag{25}$$

where D_{12} is the coefficient of mutual diffusion for the two components of the gas. Although it is experimentally possible to create much higher concentration gradients than this, such gradients must decay rapidly unless streams of unmixed gases are continuously injected into the system. In the context of using thermoacoustics to separate gases which are initially well mixed, we assume that the maximum concentration gradient is this limiting value.

Using Eq. (25) in Eq. (24), we can maximize b in terms of the acoustic amplitudes $|p_1|$ and $|\langle u_1\rangle|$. It is most convenient to replace $|p_1|$ by $z|\langle u_1\rangle|$, where z is the magnitude of the specific acoustic impedance. In that case one finds that b increases asymptotically toward the value

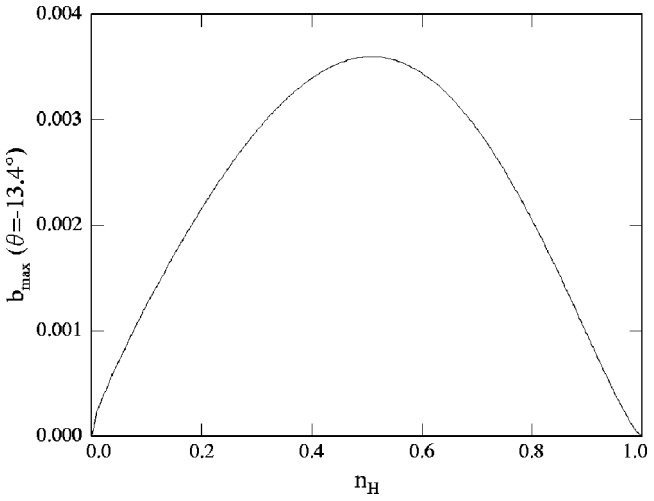


FIG. 1. The maximum ratio of gradient-dependent loss to thermal and viscous losses, for He–Ar mixtures in the boundary-layer limit with $\theta=-13.4°$.

$$b_{\max}=\frac{(F_{\text{trav}}\cos\theta+F_{\text{stand}}\sin\theta)(\sin\theta-\cos\theta)\varepsilon\sigma}{F_{\text{grad}}[(1-\sigma)(1-\sigma L)-\varepsilon\sigma]}\times\left[1-\frac{1+\sigma\sqrt{L}}{\sqrt{\sigma L}}\frac{\delta_\kappa}{\delta_{\kappa D}+\delta_{D\kappa}}\right], \tag{26}$$

as $z\to\infty$ and as $|\langle u_1\rangle|\to\infty$. "Infinite" $|\langle u_1\rangle|$ cannot be approached in reality because the oscillating flow would become turbulent, likely destroying the separation effect. Therefore, we take this limit to imply velocities just below the onset of turbulence. For the He–Ar mixtures studied in Refs. 2–4, $b_{\max}$ is maximized when $\theta\approx-13.4°$. By analogy with Fig. 5 from Ref. 2, we show $b_{\max}$ at $\theta=-13.4°$ as a function of n_H in Fig. 1. Because He–Ar mixtures have among the highest thermal diffusion ratios found in nature, it is unlikely that a $b_{\max}$ much greater than 0.004 will be encountered in any mixture separation.[8,9] Since $d\dot{E}_{\kappa,D}/dx$ differs little[2] from the ordinary thermal loss in a homogeneous gas, we conclude in general that, even with $dn_H/dx\neq0$, the added dissipation caused by thermoacoustic separation is much smaller than the ordinary thermal and viscous losses which would be calculated assuming a pure gas.

It is fortunate that $b_{\max}$ is small. If the acoustic dissipation arising from the concentration gradient had turned out to be large, then we would have to reformulate much of thermoacoustics: Rott's equations[10] assume a simple, nonseparable gas, and hence do not strictly apply to the propagation of sound in mixtures. Because $d\dot{E}_\nabla/dx$ is generally much smaller than the other acoustic losses, though, and because $d\dot{E}_{\kappa,D}/dx$ typically differs from the pure-gas thermal dissipation by less than 1%,[2] Rott's equations accurately describe the local properties of most thermoacoustic engines and refrigerators in practice. There are at least two important exceptions. First, the mixture-separation phenomenon can easily create a spatial distribution of the gas composition, and hence a significant x dependence of the gas properties that must be used in Rott's equations. This fact is illustrated by the resonator coupling experiment[2] in which the separation effect was discovered. Second, dramatic changes to the equations of thermoacoustics result when one component of the mixture experiences evaporation and liquefaction in the acoustic process.[11]

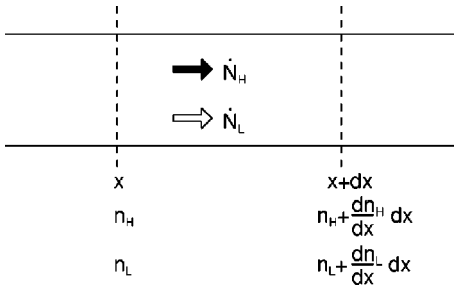


FIG. 2. Sign conventions used to derive the increase in Gibbs free energy associated with mixture separation along a duct. For the process considered here, we have $\dot{N}_L=-\dot{N}_H$, $n_L=1-n_H$, and $dn_L/dx=-dn_H/dx$.

Finally, because dn_H/dx can have either sign, the concentration gradient seems to offer the possibility of creating or amplifying sound, much like the temperature gradient across the stack or regenerator in a thermoacoustic engine, and it is interesting to consider whether such mixing-driven acoustic power production could be experimentally demonstrated. Unfortunately, $d\dot{E}_\nabla/dx$ is overwhelmed by the ordinary viscous and thermal dissipation unless the concentration gradient is extremely abrupt. We have not discovered a way to demonstrate this effect.

IV. THERMODYNAMIC EFFICIENCY

We can define the efficiency of thermoacoustic separation by the ratio of the rate at which useful separation work is done to the total acoustic power consumed in performing the separation

$$\eta=\frac{d\dot{G}/dx}{d\dot{E}_{\text{tot}}/dx}, \tag{27}$$

where $d\dot{G}/dx$ is the rate at which the Gibbs free energy of the mixture increases, per unit length. This definition of efficiency is thermodynamically meaningful, because $\dot{G}$ is the rate at which exergy (i.e., useful work) is built up in the separating mixture[12] and $\dot{E}_{\text{tot}}$ is the rate at which exergy is consumed by the wave[13] if the duct walls are in contact with a thermal reservoir at a temperature at which heat has no value.

We can derive a simple expression for $d\dot{G}/dx$ by starting with the well-known[14] concentration contribution to the free energy of a *homogeneous* binary mixture of N total moles

$$G=NR_{\text{univ}}T_m(n_H\log n_H+n_L\log n_L). \tag{28}$$

As illustrated in Fig. 2, the length dx of the thermoacoustic separation duct moves $\dot{N}_H\,dt$ moles of the heavy component from a location where the concentration is n_H to one where the concentration is $n_H+(dn_H/dx)dx$, and similarly moves $\dot{N}_L\,dt$ moles of the light component from n_L to $n_L+(dn_L/dx)dx$. The change in G for this process is

$$d^2G = R_{\text{univ}} T_m \Big[\dot{N}_H\, dt \log\Big(n_H + \frac{dn_H}{dx} dx \Big) - \dot{N}_H\, dt \log n_H + \dot{N}_L\, dt \log\Big(n_L + \frac{dn_L}{dx} dx \Big) - \dot{N}_L\, dt \log n_L \Big]. \tag{29}$$

Combining the logarithms and using the identities given in the caption of Fig. 2 quickly yields

$$\frac{d\dot{G}}{dx} = \frac{R_{\text{univ}} T_m}{n_H(1-n_H)} \frac{dn_H}{dx} \dot{N}_H, \tag{30}$$

where $\dot{N}_H$ is the mole flux of the heavy component. Using Eqs. (41), (44), and (56) from Ref. 4 for the mole flux through second order, this may be rewritten as

$$\frac{d\dot{G}}{dx} = \frac{1}{n_H(1-n_H)} \frac{dn_H}{dx} \Big[\frac{\delta_\kappa}{4r_h} \frac{\gamma-1}{\gamma} k_T |p_1| \times |U_1| (F_{\text{trav}} \cos\theta + F_{\text{stand}} \sin\theta) + \frac{\delta_\kappa}{4r_h} \frac{p_m}{\omega A} F_{\text{grad}} |U_1|^2 \frac{dn_H}{dx} - p_m A D_{12} \frac{dn_H}{dx} \Big]. \tag{31}$$

Although $\dot{G}$ and $\dot{E}_\nabla$ are both powers associated with separating the mixture along a concentration gradient, they differ dramatically. Whereas Eq. (22) is linear in the concentration gradient, Eq. (31) is quadratic in dn_H/dx. The term linear in dn_H/dx is quite similar to $d\dot{E}_\nabla/dx$, but it differs in its phase dependence and its lack of the combinations of L, σ, $\delta_{D\kappa}$, and $\delta_{\kappa D}$ appearing in Eq. (22). Moreover, while $d\dot{G}/dx \to 0$ as the separation approaches saturation, $d\dot{E}_\nabla/dx$ does not generally vanish, and if one first establishes a concentration gradient and then turns off the acoustics, $d\dot{E}_\nabla/dx = 0$ while $d\dot{G}/dx$ gives a nonzero energy loss as the concentration gradient decays by ordinary diffusion. Hence, $d\dot{E}_\nabla/dx$ must describe not only energy transferred from the acoustic wave into separation of the gases but also more subtle features of the dissipation which are lost in the heuristic description of separation given in Ref. 2. For example, while Eq. (21) describes losses due to diffusion of heat and of mass at the boundary layer and driven by the oscillating temperature T_1 when $dn_H/dx = 0$, this loss must be adjusted higher or lower by Eq. (22) in the presence of a longitudinal concentration gradient. This adjustment depends not only on the gradient but on the phasing of the acoustic fields, because the enhancement or depletion of one species outside the boundary layer will lead to either more or less thermal diffusion radially, depending on whether the motion is in phase or out of phase with respect to the temperature oscillations.

In the denominator of Eq. (27) we could use Eq. (19). However, Sec. III above and Sec. V of Ref. 2 show that, to a high degree of accuracy, we can ignore all the contributions to acoustic dissipation arising from thermal diffusion and simply use the ordinary expression for the acoustic dissipation

$$\frac{d\dot{E}_{\text{tot}}}{dx} \approx \frac{\omega A}{4r_h} \rho_m |\langle u_1 \rangle|^2 \delta_\nu + \frac{\omega A}{4r_h} \frac{\gamma-1}{\rho_m a^2} |p_1|^2 \delta_\kappa, \tag{32}$$

yielding some immediate simplification.

For comparison with other separation methods, it is necessary to find the conditions that maximize η. In particular, one must determine at what concentration gradient the maximum η occurs. Since neither of the terms in (32) depends on dn_H/dx, the maximum efficiency must occur when $d\dot{G}/dx$ is itself a maximum. Equation (31) shows that this rate is quadratic in dn_H/dx, and it is straightforward to show that the only maximum occurs when

$$\frac{dn_H}{dx} = \frac{1}{2} \Big(\frac{dn_H}{dx} \Big)_{\text{lim}}, \tag{33}$$

where $(dn_H/dx)_{\text{lim}}$ is the value of dn_H/dx for which $\dot{N}_H = 0$, obtained in Eq. (25). (Because effects which remix the gases, such as ordinary diffusion, enter the expression for the mole flux $\dot{N}_H$ at first order in dn_N/dx for many other thermophysical separation processes as well, the maximum rate at which useful separative work can be done often occurs when the concentration gradient is at half of its saturation value.[15]) Inserting this gradient into Eq. (31), and using Eq. (32), we obtain the efficiency at the best dn_H/dx

$$\eta(\text{best } dn_H/dx) = \frac{(\gamma-1)^2 k_T^2 (z')^2 (F_{\text{trav}} \cos\theta + F_{\text{stand}} \sin\theta)^2}{4\gamma n_H(1-n_H)[(\gamma-1)(z')^2 + \sqrt{\sigma}](-F_{\text{grad}} + 4D_{12}\omega r_h/\delta_\kappa |\langle u_1 \rangle|^2)}, \tag{34}$$

where $z' = |p_1|/\rho_m a |\langle u_1 \rangle|$. This efficiency is maximized asymptotically as $z' \to \infty$ and as $|\langle u_1 \rangle| \to \infty$, where $|\langle u_1 \rangle|$ again must be kept below the onset of turbulence. The only other adjustable parameter is the wave's phasing θ. To maximize the efficiency with respect to θ, $\tan\theta = F_{\text{stand}}/F_{\text{trav}}$, so that the limiting efficiency for thermoacoustic separation is finally

$$\eta_{\text{best}} = \frac{\varepsilon}{4} \frac{F_{\text{trav}}^2 + F_{\text{stand}}^2}{-F_{\text{grad}}}. \tag{35}$$

In the context of Refs. 2–4, $\eta_{\text{best}} = 0.0023$ for a 50–50 He–Ar mixture.

Taking $\varepsilon \ll 1$ leads to great simplification in the three

F's in Eq. (35), so that in this limit the best thermoacoustic efficiency becomes

$$\eta_{\mathrm{ta,best}} \approx \frac{\varepsilon}{4} \frac{2(1+\sqrt{\sigma L})}{\sqrt{L}(1+\sqrt{L})^2(1+\sigma)}. \tag{36}$$

We will compare this expression to the efficiencies of other mixture-separation methods in the next section.

In practice, it is easy to achieve values of dn_H/dx, z', $|\langle u_1 \rangle|$, and θ that bring η close to $\eta_{\mathrm{ta,best}}$. For example, some of the measurements reported in Ref. 4 had θ at its optimum value, $z' \approx 4$, and $|\langle u_1 \rangle|$ large enough to make $4D_{12}\omega r_h/\delta_\kappa |\langle u_1 \rangle|^2(-F_{\mathrm{grad}}) \approx 0.02$. At some time during the approach to the saturation data reported in Ref. 4, Eq. (33) was satisfied as well. Under these circumstances, $\eta/\eta_{\mathrm{ta,best}} \approx 0.9$.

V. COMPARISON TO TRADITIONAL METHODS

Mixture-separation methods can be broadly divided into two categories, those that are intrinsically reversible and those that are intrinsically irreversible. The use of perfect semipermeable membranes is an example of an intrinsically reversible separation method; for example, helium and hydrogen can be separated from much heavier gases using such membranes, with a thermodynamic efficiency near unity. Distillation separation methods[16,17] (e.g., air-separation plants) are also impressively efficient, often exceeding $\eta = 0.5$ in industrial practice. Relying as it does on intrinsically irreversible processes—diffusion of heat, mass, and momentum—thermoacoustic mixture separation cannot approach these high efficiencies.

However, many important mixtures, including those of most isotopes and of most isomers, must be separated by intrinsically irreversible methods, because the more efficient methods are inapplicable. The intrinsically irreversible methods include traditional thermal diffusion, gaseous diffusion, and mass diffusion.[18] To get an initial impression of how the efficiency of thermoacoustic mixture separation compares with those of these other methods, we will focus on isotope separation and on the first two intrinsically irreversible methods listed above.

The very process of thermal diffusion was only discovered between 1911 and 1917,[19] and attempts to separate isotopes with it followed quickly, with Clusius and Dickel[20] separating neon isotopes in 1938. Their apparatus comprised two vertical concentric tubes, the inner one heated and the gas mixture between the two. Thermal diffusion draws the lighter isotope toward the inner tube, and the heavier toward the outer tube, so that gravity-driven convection separates them vertically.

While thermal diffusion was discovered as a consequence of the kinetic theory of gases, the origins of gaseous diffusion are much older, tracing back to the experiments of Graham in the mid-19th century on the effusion of gases through porous materials. Although gaseous diffusion was used in the discoveries of the noble gases at the end of the 19th century, it was not until 1920 that the process was used for an isotope separation, in neon.[21] In a gaseous-diffusion system, the gas mixture is supplied at a constant pressure to one side of a porous barrier which is held at lower pressure on the other side. If the pore size is small enough, gas moves through the pores in Knudsen flow, and the components of the mixture escape to the low-pressure side at rates inversely proportional to the square root of their molecular masses.

Onsager[22] derived fundamental bounds on the efficiencies of separations based on thermal diffusion and gaseous diffusion. Our derivation of Eq. (30) is similar to his, and he further argued that the exergy spent in thermal-diffusion separation must be at least as large as the exergy lost in thermal conduction through the gas, and the exergy spent in gaseous-diffusion separation must be at least as large as the exergy dissipated in the free expansion. In our notation, his results are

$$\eta_{\mathrm{td,best}} = \frac{\varepsilon}{4}\frac{1}{L}, \tag{37}$$

$$\eta_{\mathrm{gd,best}} = \frac{n_H(1-n_H)}{4}\left(\frac{\Delta D}{D}\right)^2, \tag{38}$$

where D is the average gas diffusion constant, ΔD is the difference between the Ds for the two components of the gas, and the subscripts td and gd signify thermal diffusion and gaseous diffusion, respectively. Note the obvious similarity between Eqs. (35) and (37). For a 50–50 He–Ar mixture, these equations show that $\eta_{\mathrm{best}} = 0.57\,\eta_{\mathrm{td,best}}$.

To make further progress in approximately comparing the efficiencies of thermoacoustic, thermal-diffusion, and gaseous-diffusion separation, we will make some estimates using the hard-sphere kinetic-theory values[16,23] of the various parameters, assuming that the molar mass difference $\Delta m = m_H - m_L$ is much smaller than m_{avg}

$$\gamma = \tfrac{5}{3}, \quad \sigma = \tfrac{2}{3}, \quad L = \tfrac{5}{4}, \tag{39}$$

$$k_T = n_H(1-n_H)\frac{105}{236}\frac{\Delta m}{m_{\mathrm{avg}}}, \tag{40}$$

$$\frac{\Delta D}{D} = \frac{\Delta m}{2m_{\mathrm{avg}}}. \tag{41}$$

With these values, Eqs. (36), (37), and (38) become

$$\eta_{\mathrm{ta,best}} = 0.009 n_H(1-n_H)(\Delta m/m_{\mathrm{avg}})^2, \tag{42}$$

$$\eta_{\mathrm{td,best}} = 0.016 n_H(1-n_H)(\Delta m/m_{\mathrm{avg}})^2, \tag{43}$$

$$\eta_{\mathrm{gd,best}} = 0.063 n_H(1-n_H)(\Delta m/m_{\mathrm{avg}})^2. \tag{44}$$

Remarkably, these are all of the same order of magnitude, and share the same dependence on concentration and mass difference.

As an example of an economically important, intrinsically irreversible separation, consider the enrichment of uranium.[24] Most commercial power reactors use uranium that has been enriched by gaseous diffusion of UF_6 to about 3% U^{235}; the natural abundance is 0.7%. Evaluation of Eq. (44) for this situation yields $\eta_{\mathrm{gd,best}} \approx 10^{-7}$. The Tricastin (France) uranium enrichment facility actually operates at an efficiency of 10^{-8}. This example merely demonstrates that an extremely inefficient separation process can nevertheless

be important. In considering possible applications of thermoacoustic mixture separation, the intrinsic thermodynamic efficiency is not prohibitively low, so issues such as cost and reliability might be decisive.

ACKNOWLEDGMENTS

This work was supported by the Division of Materials Sciences in the U.S. Department of Energy's Office of Science.

[1] G. W. Swift, ''Thermoacoustic engines,'' J. Acoust. Soc. Am. **84**, 1145–1180 (1988).
[2] G. W. Swift and P. S. Spoor, ''Thermal diffusion and mixture separation in the acoustic boundary layer,'' J. Acoust. Soc. Am. **106**, 1794–1800 (1999); **107**, 2299(E) (2000); **109**, 1261(E) (2001).
[3] P. S. Spoor and G. W. Swift, ''Thermoacoustic separation of a He–Ar mixture,'' Phys. Rev. Lett. **85**, 1646–1649 (2000).
[4] D. A. Geller and G. W. Swift, ''Saturation of thermoacoustic mixture separation,'' J. Acoust. Soc. Am. **111**, 1675–1684 (2002).
[5] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon, New York, 1982), Eq. (57.3).
[6] The hydraulic radius r_h of a duct is defined as the ratio of the cross-sectional area to the perimeter. For a right circular cylinder, the hydraulic radius is equal to one half of the cylinder radius.
[7] E. J. Watson, ''Diffusion in oscillating pipe flow,'' J. Fluid Mech. **133**, 233–244 (1983).
[8] Note that in the above expression, the δ_i enter only in ratios. The mass diffusion coefficient D_{12}, the kinematic viscosity ν, and the thermal diffusivity κ are all proportional to $1/\rho$ and therefore to $1/p$. Furthermore, each of these parameters has approximately the same weak power-law dependence on temperature; in the simplest mean-free-path theory for a dilute gas of rigid spheres, this power law is $\sqrt{T}$. Thus, ratios of the δ_i, such as σ and L, are essentially independent of pressure and roughly independent of temperature in the regime that the gases in the mixture can be considered ideal. See Ref. 23.
[9] In general, k_T can be a function of both temperature and pressure. The temperature dependence typically exhibits a maximum value (at fixed pressure), and for the room-temperature He–Ar mixtures of our previous experiments, k_T is 0.8 of its maximum value at atmospheric pressure. The experimentally measured k_T may either increase or decrease with pressure, but this effect is due to the nonidealities in the equations of state of the gases. See K. E. Grew and T. L. Ibbs, *Thermal Diffusion in Gases* (Cambridge University Press, Cambridge, 1952), Chaps. 4–5. Nonideal gas behavior is beyond the scope of our treatment.
[10] N. Rott, ''Damped and thermally driven oscillations in wide and narrow tubes,'' Z. Angew. Math. Phys. **20**, 230–243 (1969), and N. Rott, ''Thermally driven oscillations. III. Second-order heat flux,'' *ibid.* **26**, 43–49 (1975).
[11] R. Raspet, C. J. Hickey, and J. M. Sabatier, ''The effect of evaporation-condensation on sound propagation in cylindrical tubes using the low reduced frequency approximation,'' J. Acoust. Soc. Am. **105**, 65–73 (1999); W. V. Slaton and R. Raspet, ''Wet-walled thermoacoustics,'' *ibid.* **110**, 2677 (2001); William V. Slaton, ''Inert gas–vapor mixtures in thermoacoustics,'' Ph.D. thesis, University of Mississippi, 2001.
[12] A. Bejan, *Advanced Engineering Thermodynamics*, 2nd ed. (Wiley, New York, 1997).
[13] G. W. Swift, *Thermoacoustics: A Unifying Perspective for Some Engines and Refrigerators*, Chap. 6 (Acoustical Society of America, 2002)
[14] This can be found in many introductory thermodynamics texts. For example, see Eqs. (4.128) and (4.139) in Ref. 12.
[15] H. London, *Separation of Isotopes* (Newnes, London, 1961).
[16] S. Villani, *Isotope Separation* (American Nuclear Society, La Grange Park, IL, 1976).
[17] *Encyclopedia of Separation Technology*, edited by D. M. Ruthven (Wiley, New York, 1997), Vol. 1.
[18] Gaseous diffusion describes effusion, or the molecular flow of a gas through narrow pores into vacuum. In mass diffusion, the mixture is separated by the difference in the components' rates of diffusion through a carrier gas (preferably one which is easily stripped from the product). See Ref. 16 for details of these methods.
[19] For an account of the theoretical discovery by Enskog and Chapman separately, and the experimental confirmation by Chapman and Dootson, see the review by R. Clark Jones, and W. H. Furry, ''The separation of isotopes by thermal diffusion,'' Rev. Mod. Phys. **18**, 151–224 (1946).
[20] K. Clusius and G. Dickel, ''Neues Verfahren zur Gasentmischung und Isotopentrennung,'' Naturwissenschaften **26**, 546(L) (1938).
[21] D. Massignon, in *Topics in Applied Physics: Uranium Enrichment*, edited by S. Villani (Springer, New York, 1979), Vol. 35, pp. 55–56.
[22] L. Onsager, ''Separation of gas (isotope) mixtures by irreversible processes,'' Phys. Rev. **55**, 1137(A) (1939).
[23] J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (Wiley, New York, 1954).
[24] Anthony V. Nero, *Guidebook to Nuclear Reactors* (University of California Press, Berkeley, 1979); see also the Uranium Information Centre at www.uic.com.au.

Study on the prestressed sandwich piezoelectric ceramic ultrasonic transducer of torsional-flexural composite vibrational mode

Lin Shuyu
Applied Acoustics Institute, Shaanxi Normal University, Xian, Shaanxi 710062, People's Republic of China

(Received 16 August 2001; accepted for publication 14 May 2002)

Based on the classical torsional and flexural vibrational theory of a slender rod, the prestressed sandwich torsional-flexural composite mode piezoelectric ceramic ultrasonic transducer is studied. This type of transducer consists of the slender metal rods and the longitudinally and tangentially polarized piezoelectric ceramic rings. The resonance frequency equations for the torsional and flexural vibrations in the transducers are derived. The simultaneous resonance of the torsional and flexural vibrations in the transducer is acquired by correcting the length of the metal slender rods resulting from the piezoelectric ceramic elements. The experimental results show that the measured resonance frequencies of the transducers are in good agreement with the computed ones, and the measured resonance frequencies of the torsional and the flexural vibrations in the composite transducers are also in good agreement with each other. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1492819]

PACS numbers: 43.38.Ar, 43.38.Fx [SLE]

I. INTRODUCTION

Traditional ultrasonic technologies such as ultrasonic cleaning, ultrasonic plastic and metal welding, ultrasonic machining, and ultrasonic fatigue testing have found widespread applications in the fields of industry, electronics, and material science. In these applications, the prestressed sandwich transducers in longitudinal vibrational mode are widely used.[1–5] The reason is that the design theory for the longitudinal sandwich transducer is well developed and the longitudinal transducers have high electro-acoustical efficiency and large power capacity. In recent years, some new ultrasonic technologies such as ultrasonic motor and ultrasonic surgery have also received increased attention. Corresponding to this development, ultrasonic transducers and the vibrating systems have become important research subjects recently. Apart from the traditional single vibrational mode ultrasonic transducer, such as the longitudinal, the torsional, and the flexural transducers, some compound transducers such as longitudinal-torsional and longitudinal-flexural transducers have found more and more applications.[6–15]

For the creation of torsional, flexural, longitudinal-flexural, and longitudinal-torsional compound vibration, two methods can be used. One is by means of the vibrational mode conversion of longitudinal into torsional or flexural vibration[16] and the other is by using the longitudinally or tangentially polarized piezoelectric ceramics.[17] In the first method, the conversion of longitudinal into torsional or flexural vibration is by means of the sandwich longitudinal transducers that are attached to a transmission cylinder or a radiating plate. In this case, the total input electric power can be large and the system can be used in high-power ultrasonic applications, such as in ultrasonic metal welding, ultrasonic plastic welding, ultrasonic drying, ultrasonic levitation, and ultrasonic machining. However, since the volume of this kind of vibrating system is large and the efficiency of conversion of longitudinal into torsional or flexural vibration is low, the system has not been widely used. Another method to create the torsional and the longitudinal-torsional vibration by means of conversion of longitudinal into torsional vibration is by using a drilling tool with spiral slots that is attached to the output end of a longitudinal sandwich transducer. In this case, the vibrating system can create longitudinal-torsional compound vibration. Since the geometrical shape of the transmission cylinder with spiral slots is complex, the computation and design of this kind of transducer is difficult, and the vibrational characteristics cannot be studied analytically. Even if the numerical methods such as the finite element method are used, the analysis is still complex and cumbersome.

The design theory for the creation of torsional, flexural, longitudinal-flexural, and longitudinal-torsional compound vibrations by means of the tangentially and the longitudinally polarized piezoelectric ceramics has been established.[6,11,13,17] In this case, the torsional sandwich transducer is composed of the tangentially polarized piezoelectric ceramic tubes and the front and back metal cylinders, the longitudinal-torsional compound transducer is composed of the longitudinally and tangentially polarized piezoelectric ceramic tubes and the front and back metal cylinders, while the sandwich flexural transducer or the longitudinal-flexural composite transducer is composed of the longitudinally polarized piezoelectric ceramic elements. Although the sandwich torsional and the sandwich flexural transducers are similar to the sandwich longitudinal transducer in appearance, the concrete manufacturing of the torsional and flexural transducer is complex, especially for the processing of the tangentially polarized piezoelectric ceramic ring. Since the polarization is along the circumferential direction, large sectors are difficult to polarize and electrical breakdown may happen. In this case, an entire polarization is difficult to achieve. To overcome these difficulties, the piezoelectric ce-

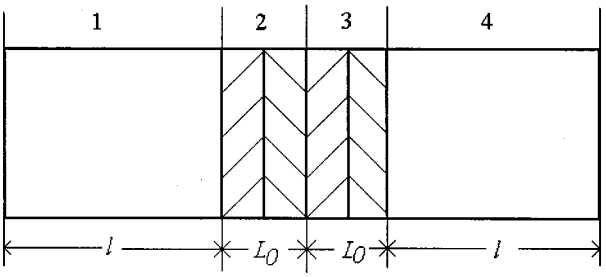


FIG. 1. A schematic diagram of the torsional-flexural composite mode ultrasonic transducer.

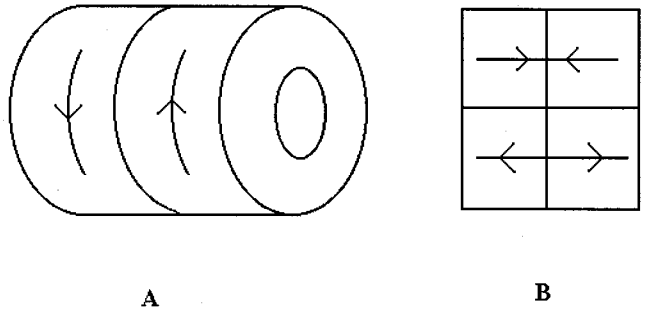


FIG. 2. The tangentially and longitudinally polarized piezoelectric ceramic elements in the torsional-flexural composite mode transducer.

ramic sectors must be small and the lateral dimension (such as the diameter of the segmented ceramic ring and the wall thickness of ring) is limited. For the prestressed sandwich flexural transducer, the longitudinally polarized piezoelectric ceramic rings must be cut into two equal half-circles, and their polarization direction must be carefully arranged. On the other hand, for the longitudinal-torsional and longitudinal-flexural compound transducers, since the sound speeds of longitudinal, flexural, and torsional vibrations are different, the longitudinal, flexural, and torsional vibrations are difficult to resonate at the same frequency. This makes the effective electrical excitation of the longitudinal-torsional or longitudinal-flexural compound system complex. Although the simultaneous resonance of longitudinal and torsional vibrations can be achieved by properly choosing the shape and dimension of the transducer,[13] the computation is complex.

In this paper, the prestressed torsional-flexural compound ultrasonic transducer is analyzed. The resonance frequency equations for the torsional and the flexural vibrations in the compound transducer are derived. Since the sound speeds for the torsional vibration and the flexural vibration are different and the sound speed of the flexural vibration depends on the frequency, it is difficult for the torsional vibration and the flexural vibration in the compound transducer to resonate at the same resonance frequency. In order to overcome this problem and simplify the electrical excitation of the torsional-flexural compound transducer, two procedures are taken in the following analysis. One is by choosing the cross-sectional radius of the transducer, and the other is by correcting the length of the metal cylinder resulting from the piezoelectric ceramic elements in the compound transducer.

II. RESONANCE FREQUENCY EQUATIONS OF THE TORSIONAL-FLEXURAL COMPOSITE MODE ULTRASONIC TRANSDUCER

The schematic diagram of the torsional-flexural composite mode ultrasonic transducer is shown in Fig. 1. In the figure, parts 1, 2, 3, and 4 represent the back metal cylinder, the tangentially polarized piezoelectric ceramic elements, the longitudinally polarized piezoelectric ceramic elements, and the front metal cylinder, respectively. The back and the front metal masses are solid cylinders; their radius is a. The piezoelectric ceramic elements are rings, their outer and inner radius are R_0 and r. For simplicity, the transducer has a symmetrical structure in geometry. In this case, the back and front cylinders have the same material parameters, length, and radius; the tangentially and longitudinally polarized piezoelectric rings also have the same material parameters and geometrical dimensions. The length of the back and the front metal cylinder is l. The length of the tangentially and the longitudinally polarized piezoelectric ceramic elements is L_0. There are four pieces of ceramic rings in all in the compound transducer. Two pieces of rings that are polarized tangentially create torsional vibration; the other two pieces that are polarized longitudinally and are cut into four equal half-circles create flexural vibration. It should be pointed out that the two pieces of piezoelectric ceramic rings polarized in longitudinal direction for creating flexural vibration in the composite mode transducer are cut into four equal half-circles and their polarization direction must be arranged correctly. The tangentially and longitudinally polarized piezoelectric ceramic elements are shown in Fig. 2. In Fig. 2, the arrow denotes the polarization direction. Figure 2(a) represents the tangentially polarized piezoelectric ceramics, and Fig. 2(b) represents the longitudinally polarized piezoelectric ceramics. In order to simplify the analysis, it is assumed that the cross-sectional radius of the transducer should be much less than the wavelength of the torsional and the flexural vibration in the transducer, and the theory for the flexural vibration and the torsional vibration of a slender rod can be used.

For the torsional-flexural compound ultrasonic transducer shown in Fig. 1, solving the torsional and the flexural wave equations and the piezoelectric constitutive equation for the piezoelectric ceramic elements and the metal back and front cylinder can derive its resonance frequency equations in principle. When the displacement distribution functions for the torsional and the flexural vibrations are derived, the resonance frequency equations for the torsional and flexural vibrations can be obtained by using the boundary conditions of the transducer. From the resonance frequency equations for the torsional and the flexural vibrations in the compound transducer, the resonance frequencies or the geometrical dimensions can be obtained and by choosing the dimensions of the transducer properly, the simultaneous resonance of the torsional and the flexural vibrations in the transducer can be realized. However, since the solutions for the flexural vibration in the transducer is very complex, the results from the above-mentioned method are not suitable for the engineering design of the torsional-flexural compound transducer.

In the following analysis, an approximate method is used to derive the resonance frequency equation of the torsional-flexural compound transducer. According to this method, two procedures are taken. First, the torsional-flexural compound transducer is regarded as a slender metal rod with a constant radius, using the traditional torsional and flexural vibrational theory, the resonance frequency equations for the torsional and the flexural vibrations can be derived, and the simultaneous resonance can be obtained. Second, using the boundary conditions of the displacements and forces at the boundaries between the metal cylinder and the piezoelectric ceramic elements, the correction lengths of the piezoelectric ceramic elements in the transducer can be obtained. On the other hand, when the piezoelectric ceramic elements are located at the displacement node or antinode, the displacement distribution functions in the piezoelectric ceramic elements can be regarded as functions of sine or cosine. For the flexural vibration, the flexural displacement and the shearing force are continuous at the boundary between the metal cylinder and the piezoelectric ceramic elements, while for the torsional vibration, the torsional angular displacement and the torsional moment are also continuous at the boundary between the metal cylinder and the piezoelectric ceramic elements. In the following analysis, for simplicity, it is assumed that the torsional-flexural compound transducer has a symmetrical structure; its geometrical center axis that is perpendicular to the length is the symmetrical axis. It divides the transducer into two equal halves.

A. Theoretical consideration for a slender metal rod in which the torsional vibration and the flexural vibration resonate at the same resonance frequency

According to the traditional torsional and flexural vibrational theory of a slender rod, the resonance frequency equations for a slender metal rod of length L with free boundary conditions can be obtained as

$$f_t = mc_t/2L, \quad m = 1,2,3,..., \tag{1}$$

$$f_w = \pi c_l R(2n+1)^2/8L^2, \quad n = 1,2,3,..., \tag{2}$$

where f_t and f_w are the resonance frequencies of the torsional vibration and the flexural vibration in a slender metal rod, m and n are two positive integers, which represent the vibrational orders of the torsional and the flexural vibration, $c_t = (G/\rho)^{1/2}$, $c_l = (E/\rho)^{1/2}$, and c_t and c_l are the sound speeds of the torsional vibration and the longitudinal vibration in a slender rod with constant cross section. E and G are Young's modulus and the shearing modulus of the material; L is the length of the metal slender rod; R is the gyration radius of the cross section of the rod, its definition is $SR^2 = \int\int r^2\, dS$; r is the cross sectional radius of the rod; and S is the cross-sectional area of the rod. For a slender rod whose radius is a, we have $R = a/2$ and $S = \pi a^2$. Let $f_t = f_w$. From Eqs. (1) and (2), the following equation can be obtained:

$$L = \frac{\pi a}{8}\sqrt{2(1+\sigma)}\cdot\frac{(2n+1)^2}{m}. \tag{3}$$

In Eq. (3), σ is the Poisson's ratio of the material, $G = E/[2(1+\sigma)]$. From Eqs. (1)–(3), it can be seen that when the length and radius of the metal slender rod satisfy Eq. (3), the torsional vibration of order m and the flexural vibration of order n will resonate at the same resonance frequency, and this resonance frequency can be expressed as

$$f = f_t = f_w = \frac{2}{\pi a(1+\sigma)}\left(\frac{E}{\rho}\right)^{1/2}\frac{m^2}{(2n+1)^2}. \tag{4}$$

From Eqs. (3) and (4), when the material parameters are given and the vibrational orders are specified, the resonance length and the resonance frequency of a slender metal rod can be obtained. In this case, by exciting the metal slender rod properly, the torsional vibration and the flexural vibration can be created and made resonate at the same resonance frequency.

B. Length correction resulting from the piezoelectric ceramic elements in the compound transducer

In the above analysis, the torsional vibration of order m and the flexural vibration of order n are designed to resonate at the same resonance frequency in a slender metal rod. In a prestressed sandwich torsional-flexural composite mode transducer, for different vibrational orders, the vibrational displacement distribution in the piezoelectric ceramic elements is different; the correction length resulting from the piezoelectric ceramic elements is different accordingly. For different torsional and flexural vibrational orders in a composite mode transducer, the length correction will be analyzed separately in the following sections.

1. Vibrational orders m and n are odd numbers

In this case, for the torsional vibration in the compound transducer, the piezoelectric ceramic elements are at the torsional angular displacement node. The torsional angular displacement φ and the torsional moment M can be expressed as

$$\varphi = A\sin(k_t x), \tag{5}$$

$$M = IG\frac{d\varphi}{dx} = IGAk_t\cos(k_t x). \tag{6}$$

Here A is a constant that can be determined by the boundary condition; $k_t = \omega/c_t$ is the torsional wave number; and $I = \int\int r^2\, ds$ is the polar moment of the cross section. For a solid rod with radius a, $I = (\pi/2)a^4$. For a ring whose outer and inner radius are R_0 and r, $I = (\pi/2)[R_0^4 - r^4]$. From Eqs. (5) and (6), the torsional angular displacement and the torsional moment at the interface between the metal cylinder and the ceramics in the piezoelectric ceramic element in the transducer whose length is L_0 are

$$\varphi_0 = A_0\sin(k_{t0}L_0), \tag{7}$$

$$M_0 = I_0G_0A_0k_{t0}\cos(k_{t0}L_0). \tag{8}$$

If the correction length resulting from the piezoelectric ceramic elements of length L_0 is L_t, which is the distance from the geometrical center of the transducer, the angular displacement and the torsional moment at the position L_t from the geometrical center in the metal cylinder are

$$\varphi_1 = A_1\sin(k_{t1}L_t), \tag{9}$$

$$M_1 = I_1 G_1 A_1 k_{t1} \cos(k_{t1} L_t). \tag{10}$$

In order not to disturb the vibrational characteristics of the transducer arising from the length correction of the piezoelectric ceramic elements, the angular displacement and the torsional moment before and after length correction should be equal, i.e., $\varphi_0 = \varphi_1$, $M_0 = M_1$. Therefore, we have the following equations:

$$A_0 \sin(k_{t0} L_0) = A_1 \sin(k_{t1} L_t), \tag{11}$$

$$I_0 G_0 A_0 k_{t0} \cos(k_{t0} L_0) = I_1 G_1 A_1 k_{t1} \cos(k_{t1} L_t). \tag{12}$$

From Eqs. (11) and (12), we have

$$\tan(k_{t0} L_0)/\rho_0 c_{t0} I_0 = \tan(k_{t1} L_t)/\rho_1 c_{t1} I_1. \tag{13}$$

In the above equations, $c_{t0} = (G_0/\rho_0)^{1/2}$ is the sound speed of torsional vibration in the tangentially polarized piezoelectric ceramic elements, $G_0 = 1/s_{55}^E$, s_{55}^E is the elastic compliance constant, $k_{t0} = \omega/c_{t0}$, $k_{t1} = \omega/c_{t1}$, and $c_{t1} = (G_1/\rho_1)^{1/2}$ is the sound speed of torsional vibration in the slender metal rod with constant cross section. From Eq. (13), it can be seen that when the geometrical dimensions of the piezoelectric ceramic elements are given, the correction length L_t resulting from the piezoelectric ceramic elements in torsional vibration can be determined.

For the flexural vibration in the compound transducer, when the vibrational order n is an odd number, the flexural vibrational displacement at the geometrical center of the transducer is antinode of displacement distribution. In this case, the flexural vibrational displacement y and the shearing force F near the geometrical center can be expressed as the following approximate forms:

$$y = B \cos(k_f x), \tag{14}$$

$$F = ESR^2 \frac{d^3 y}{dx^3} = BESR^2 k_f^3 \sin(k_f x). \tag{15}$$

Here B is a constant, $k_f = [\omega/(c_l R)]^{1/2}$. For the piezoelectric ceramic elements in flexural vibration whose length is L_0, the flexural displacement and the shearing force at the boundary in contact with the metal cylinder are

$$y_0 = B_0 \cos(k_{f0} L_0), \tag{16}$$

$$F_0 = B_0 E_0 S_0 R_p^2 k_{f0}^3 \sin(k_{f0} L_0). \tag{17}$$

In Eqs. (16) and (17), $k_{f0} = [\omega/(c_{l0} R_p)]^{1/2}$, $c_{l0} = [1/(s_{33}^E \rho_0)]^{1/2}$, s_{33}^E is the elastic compliance constant of the piezoelectric ceramics, $R_p^2 = (R_0^2 + r^2)/4$, R_p is the gyration radius of a ring, and $S_0 = \pi(R_0^2 - r^2)$. If the correction length resulting from the piezoelectric ceramic elements of length L_0 in flexural vibration is L_f, the flexural displacement and the shearing force at the position L_f from the geometrical center of the transducer are

$$y_1 = B_1 \cos(k_{f1} L_f), \tag{18}$$

$$F_1 = B_1 E_1 S_1 R_1^2 k_{f1}^3 \sin(k_{f1} L_f). \tag{19}$$

Here, $k_{f1} = [\omega/(c_{l1} R_1)]^{1/2}$, $c_{l1} = (E_1/\rho_1)^{1/2}$, $R_1 = a/2$ is the gyration radius of a solid cross section, and $S_1 = \pi a^2$. According to the similar analysis, in order not to disturb the flexural vibrational characteristics of the transducer, the flexural displacement and the shearing force must be continuous. From the above analysis, we have

$$B_0 \cos(k_{f0} L_0) = B_1 \cos(k_{f1} L_f), \tag{20}$$

$$B_0 E_0 S_0 R_p^2 k_{f0}^3 \sin(k_{f0} L_0) = B_1 E_1 S_1 R_1^2 k_{f1}^3 \sin(k_{f1} L_f). \tag{21}$$

Using Eqs. (20) and (21), we have

$$E_0 S_0 R_p^2 k_{f0}^3 \tan(k_{f0} L_0) = E_1 S_1 R_1^2 k_{f1}^3 \tan(k_{f1} L_f). \tag{22}$$

From Eq. (22), it can be seen that when the geometrical dimensions of the piezoelectric ceramic elements of length L_0 are given, the correction length L_f resulting from the piezoelectric ceramic elements in flexural vibration of order n can be determined.

2. Vibrational orders m and n are even numbers

In this case, for the torsional vibration of order m in the torsional-flexural composite mode transducer, the torsional angular displacement at the geometrical center of the composite transducer is antinode. The angular displacement and the torsional moment can be expressed as

$$\varphi = A \cos(k_t x), \tag{23}$$

$$M = IG \frac{d\varphi}{dx} = -AIGk_t \sin(k_t x). \tag{24}$$

Using the similar procedures, the metal cylinder correction length L_t resulting from the piezoelectric ceramic element in torsional vibration in the transducer can be obtained using the following expression,

$$I_0 G_0 k_{t0} \tan(k_{t0} L_0) = I_1 G_1 k_{t1} \tan(k_{t1} L_t). \tag{25}$$

In Eq. (25), the definitions of the parameters are the same as those in the above analysis. For the flexural vibration of order n in the composite transducer, the flexural vibrational displacement at the geometrical center of the transducer is node. In this case, the flexural vibrational displacement ant shearing force near the geometrical center of the transducer can be expressed as

$$y = B \sin(k_f x), \tag{26}$$

$$F = -BESR^2 k_f^3 \cos(k_f x). \tag{27}$$

Using the similar procedures, the metal cylinder correction length L_f resulting from the piezoelectric ceramic elements of length L_0 near the center of the composite mode transducer can be obtained using the following equation,

$$\tan(k_{f0} L_0)/E_0 S_0 R_p^2 k_{f0}^3 = \tan(k_{f1} L_f)/E_1 S_1 R_1^2 k_{f1}^3. \tag{28}$$

From Eq. (28), when the material parameters and the resonance frequency are given, the metal cylinder correction length resulting from the piezoelectric ceramic element near the geometrical center can be computed.

3. Vibrational order m is even number, while n is odd number

In this case, for the torsional vibration in the composite mode transducer, the torsional angular displacement at the geometrical center of the transducer is displacement antin-

ode, while for the flexural vibration in the composite mode transducer, the flexural displacement at the geometrical center of the transducer is also antinode of displacement. Using the similar procedures, the metal cylinder correction length L_t resulting from the piezoelectric ceramic elements near the geometrical center of the transducer in torsional vibration and the metal cylinder correction length L_f resulting from the piezoelectric ceramic elements near the geometrical center in flexural vibration can be obtained according to the following two equations,

$$I_0 G_0 k_{t0} \tan(k_{t0} L_0) = I_1 G_1 k_{t1} \tan(k_{t1} L_t), \tag{29}$$

$$E_0 S_0 R_p^2 k_{f0}^3 \tan(k_{f0} L_0) = E_1 S_1 R_1^2 k_{f1}^3 \tan(k_{f1} L_f). \tag{30}$$

4. Vibrational order m is odd number, while vibrational order n is even number

In this case, for the torsional vibration in the composite mode transducer, the torsional angular displacement at the geometrical center of the transducer is displacement node, while for the flexural vibration in the composite mode transducer, the flexural displacement at the geometrical center of the transducer is also displacement node. Using the similar procedures, the metal cylinder correction length L_t resulting from the piezoelectric ceramic elements near the geometrical center of the transducer in torsional vibration and the metal cylinder correction length L_f resulting from the piezoelectric ceramic elements near the geometrical center in flexural vibration can be obtained according to the following two equations,

$$\tan(k_{t0} L_0)/\rho_0 c_{t0} I_0 = \tan(k_{t1} L_t)/\rho_1 c_{t1} I_1, \tag{31}$$

$$\tan(k_{f0} L_0)/E_0 S_0 R_p^2 k_{f0}^3 = \tan(k_{f1} L_f)/E_1 S_1 R_1^2 k_{f1}^3. \tag{32}$$

From Eqs. (31) and (32), when the material parameters and the resonance frequency of the composite mode transducer are given, the metal cylinder correction length resulting from the piezoelectric ceramic elements in the transducer can be obtained analytically.

In the above analysis, four types of combinations of the torsional and the flexural vibrational order in the torsional-flexural composite mode transducer are analyzed. For every combination, the metal cylinder correction length L_t and L_f are obtained analytically. It can be seen that for every combination of the torsional and the flexural vibrational orders, the metal cylinder correction length L_t resulting from the piezoelectric ceramic elements in torsional vibration and L_f resulting from the piezoelectric ceramic elements in flexural vibration are not the same. Therefore, in the design of the torsional-flexural composite mode transducer, a compromise about the correction length L_t and L_f must be considered. If the length of the piezoelectric ceramic elements in the composite mode transducer is short compared with the total length of the composite mode transducer, and considering the fact that the resonance frequency of the transducer in flexural vibration is inversely proportional to the square of its length, while the resonance frequency of the transducer in torsional vibration is inversely proportional to the length, an approximate expression about the metal cylinder correction length L_c in total resulting from the piezoelectric ceramic elements in torsional and flexural vibration can be given:

$$L_c = 2\left(L_f - \frac{L_f L_t}{L_f + L_t}\right) = \frac{2L_f^2}{L_f + L_t}. \tag{33}$$

It can be seen from Eq. (33) that when the metal cylinder correction length L_f resulting from the piezoelectric ceramic elements in flexural vibration is smaller than the correction length L_t resulting from the piezoelectric ceramic elements in torsional vibration, the total correction length L_c is smaller than L_f, while when the correction length L_f is larger than the correction length L_t, the total correction length L_c is larger than the correction length L_f. According to the relationship between the resonance frequency and the length of the transducer in torsional and flexural vibrations, the total approximate correction length L_c can guarantee the simultaneous resonance of the torsional and the flexural vibration in the composite mode transducer basically. This can be further illustrated in the following experiments.

Based on the above analysis, after the length correction of the piezoelectric ceramic elements, the length of the metal cylinder l in the composite mode prestressed transducer with a symmetrical structure can be obtained according to the following equation,

$$l = L/2 - L_c. \tag{34}$$

In this case, the total length L_A of the prestressed composite mode transducer is

$$L_A = 2(L/2 - L_c + L_0) = L - 2L_c + 2L_0. \tag{35}$$

Summing up the above analyses, the procedures to design the prestressed torsional-flexural composite mode transducer can be outlined as follows. (a) Choose and specify the material parameters, the geometrical dimensions of the piezoelectric ceramic elements and the resonance frequency of the composite mode transducer. (b) Determine the orders of the torsional and the flexural vibrations in the transducer, i.e., determine the values of *m* and *n*. (c) From Eqs. (3) and (4), compute the cross-sectional radius and the length L of the metal slender cylinder. (d) Using Eqs. (13) and (22), (25) and (28), (29) and (30), and (31) and (32), compute the metal cylinder correction length L_t and L_f. (e) Using Eq. (33), compute the approximate correction length of the piezoelectric ceramic elements in the transducer. (f) Using Eq. (34), determine the length l of the metal cylinder in the transducer.

III. EXPERIMENTS

The objective of this paper is to analyze the vibrational characteristics of the pre-stressed torsional-flexural composite mode ultrasonic transducer and to realize the simultaneous resonance of the torsional and the flexural vibrations, so that the electrical excitation of the composite mode transducer can be simplified. To verify the results derived in the above analysis, some prestressed torsional-flexural composite mode transducers were designed and manufactured. The material of the front and the back metal cylinders is stainless steel, its material parameters are $\rho = 7800$ kg/m^3, $c_l = 5000$ m/s, and $\sigma = 0.28$. The piezoelectric ceramic ele-

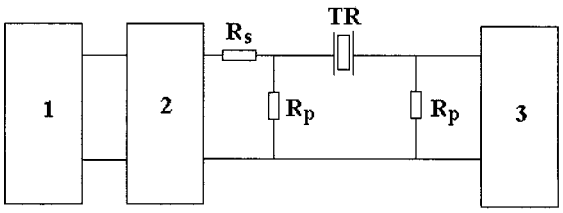


FIG. 3. Experimental transmission line circuit for measuring the resonance frequency.

ments are four pieces of circular rings, in which two pieces are cut into four equal half-circles that are polarized longitudinally, while the other two ceramic rings are polarized tangentially. The ceramic material is an equivalent of PZT-4. The ceramic material parameters are $\rho_0 = 7500\,\mathrm{kg/m^3}$ and $c_{t0} = 1850\,\mathrm{m/s}$; the geometrical dimensions of the four-piece ceramic elements are the same, the outer and inner radiis are $R_0 = 12.5\,\mathrm{mm}$ and $r = 4\,\mathrm{mm}$, the thickness of every ceramic ring has three specifications, they are 4, 3, and 2 mm. The resonance frequencies of the torsional and the flexural vibrations in the transducers were measured using the transmission line method as shown in Fig. 3. The measured results of the resonance frequencies were tabulated in Table I. In the table, L is the length of a slender rod whose resonance frequency is equal to that of the transducer; L_0 is the half-length of the piezoelectric ceramic elements in the transducer; l is the length of the metal cylinder in the transducer; L_t and L_f are the corrected lengths of the piezoelectric ceramic elements in torsional and flexural vibration; L_c is an approximate correction length of the piezoelectric ceramic elements, its relationship with L_t and L_f is described by Eq. (33); f is the designed resonance frequency of the composite mode transducer, and f_{tm} and f_{wm} are the measured resonance frequencies of the torsional vibration and the flexural vibration in the transducer. It can be seen that f_{tm} and f_{wm} are basically the same to each other, and the measured resonance frequency is also in good agreement with the designed frequency. Therefore, the objective to realize the simultaneous resonance of the torsional and the flexural vibrations in the composite mode transducer is accomplished. As for the frequency measurement error, the following factors should be taken into account. First, the standard material parameters used in the calculation of the resonance frequency are different from the actual values. Second, the process to polarize the piezoelectric ceramic rings tangentially is complex; the polarization direction is not along the ideal circumferential direction. Third, in Eq. (33), the metal cylinder correction length is approximate; it will cause some frequency error. Finally, in the experiment, the torsional and the flexural vibrations are excited separately; the coupling between the torsional and the flexural vibrations in the composite mode transducer is neglected. In practical applications, however, the torsional and the flexural vibrations are excited using the same electric generator; the coupling between the two vibrations should be considered. In later work, the interaction between the different vibrational modes in composite mode transducer, the effect of this interaction on the resonance frequency, and other vibrational characteristics will be further studied.

IV. CONCLUSIONS

In this paper, the torsional-flexural composite mode transducer is studied, the resonance frequency equation is derived, and the expression for the correction length resulting from the piezoelectric ceramic elements is obtained. To sum up the above analysis, the following conclusions can be drawn.

(1) Based on the torsional and flexural vibrational theory of a slender rod, the simultaneous resonance of the torsional vibration and the flexural vibration in a slender rod in torsional-flexural composite mode vibration is realized.
(2) For different combinations of the vibrational orders of the torsional and the flexural vibrations in the prestressed composite mode transducer, the expression for the correction length resulting from the piezoelectric ceramic elements is obtained. By means of this length correction, the simultaneous resonance can be realized.
(3) Since the sine or cosine functions are assumed for the angular and the flexural displacement distribution near the geometrical center of the transducer, the total length of the piezoelectric ceramic elements in the transducer cannot be large. On the other hand, the piezoelectric ceramic elements should be placed at the displacement node or anti-node, so that the practical displacement distribution in the piezoelectric ceramic elements can satisfy the assumed since or cosine function distribution.
(4) In this paper, the total correction length of the piezoelectric ceramic elements is an approximate value that can be determined by the correction length resulting from the piezoelectric ceramic elements in torsional and flexural vibrations, respectively. This approximation can cause the measured resonance frequency error compared with the designed resonance frequency. However, the measured resonance frequency of the torsional vibration is in good agreement with that of flexural vibration.

TABLE I. Theoretical and experimental results of the resonance frequencies of the torsional-flexural composite mode transducers.

L_0 (mm)	L (mm)	l (mm)	L_t (mm)	L_f (mm)	L_c (mm)	m	n	f (Hz)	f_{tm} (Hz)	f_{wm} (Hz)
4	70.3	33.7	11.3	3.5	1.7	1	1	22 105	20 947	20 787
6	98.2	36.8	6.3	10.0	12.3	2	2	31 831	33 214	33 375
8	196.3	87.4	23.2	14.2	10.8	1	2	7958	7643	7589

(5) In the analysis, the torsional and flexural vibrational theory of a slender rod is assumed. The lateral dimension of the transducer should be much less than its length. This is to say, the radius is much less than the length.
(6) The prestressed sandwich torsional-flexural composite mode transducer is assumed to be a symmetrical structure; and the piezoelectric ceramic elements are located near the geometrical center of the transducer.

[1] E. A. Neppiras, ''The pre-stressed piezoelectric sandwich transducer,'' Ultrasonics International 1973 Conference Proceedings, p. 295.
[2] L. Shuyu and Z. Fucheng, ''Study of vibrational characteristics for piezoelectric sandwich ultrasonic transducers,'' Ultrasonics **32**, 39–42 (1994).
[3] K. F. Graff, ''Macrosonics in industry, 5. Ultrasonic machining,'' Ultrasonics **13**(3), 103–109 (1975).
[4] B. Dubus and J. C. Debus, ''Analysis of mechanical limitations of high power piezoelectric transducers using finite element modeling,'' Ultrasonics **29**(3), 201–207 (1991).
[5] A. Bangviwat, H. K. Ponnekanti, and R. D. Finch, ''Optimizing the performance of piezoelectric drivers that use stepped horns,'' J. Acoust. Soc. Am. **90**, 1223–1229 (1991).
[6] L. Shuyu, ''Study of the sandwich piezoelectric ultrasonic torsional transducer,'' Ultrasonics **32**(6), 461–465 (1994).
[7] J. A. Gallego-Juarez, G. Rodriguez-Corral, and L. Gaete-Garreton, ''An ultrasonic transducer for high power applications in gases,'' Ultrasonics **16**, 267–271 (1978).
[8] H. Nagashima and S. Ueha, ''Study of longitudinal-torsional composite vibrator system,'' J. Acoust. Soc. Jpn. **45**(2), 94–100 (1989).
[9] K. Nakamura, M. Kurosawa, and S. Ueha, ''Design of a hybrid transducer type ultrasonic motor,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **UFFC-40**, 395–401 (1993).
[10] S. Ueha, H. Nagashima, and M. Masuda, ''Longitudinal-torsional composite transducer and its applications,'' Jpn. J. Appl. Phys., Suppl. **26** Supplement 26-2, 188–190 (1987).
[11] Y. Watanabe, Y. Tsuda, and E. Mori, ''A longitudinal-flexural complex-mode ultrasonic high-power transducer system with one-dimensional construction,'' Jpn. J. Appl. Phys., Part 1 **32**(5B), 2430–2434 (1993).
[12] O. Ohnishi, O. Myohga, and T. Ucnikawa, ''Piezoelectric ultrasonic motor using longitudinal-torsional composite vibration of a cylinder resonator,'' Ultrasonics Symposium (1989), pp. 739–743.
[13] L. Shuyu, ''Sandwich piezoelectric ultrasonic transducers of longitudinal-torsional compound vibrational modes,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **UFFC-44**, 1189–1197 (1997).
[14] K. Nakamura, M. Kurosawa, and S. Ueha, ''Characteristics of a hybrid transducer-type ultrasonic motor,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **UFFC-38**, 188–193 (1991).
[15] Y. Toikawa, K. Adachi, M. Aoyagi, T. Sagae, and T. Takamo, ''Some constructions and characteristics of rod-type piezoelectric ultrasonic motors using longitudinal and torsional vibrations,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **UFFC-39**, 600–608 (1992).
[16] L. D. Rozenberg, *Sources of High-Intensity Ultrasound* (Plenum, New York, 1969), Vol. 2.
[17] L. Shuyu, ''Torsional vibration of coaxially segmented, tangentially polarized piezoelectric ceramic tubes,'' J. Acoust. Soc. Am. **99**, 3476–3480 (1996).

Applications of the direct Trefftz boundary element method to the free-vibration problem of a membrane

Jiang Ren Chang[a)]
Department of System Engineering and Naval Architecture, National Taiwan Ocean University, 2 Pei-Ning Road, Keelung, Taiwan 202, Republic of China

Ru Feng Liu, Weichung Yeih, and Shyh Rong Kuo
Department of Harbor and River Engineering, National Taiwan Ocean University, 2 Pei-Ning Road, Keelung, Taiwan 202, Republic of China

(Received 28 November 2001; revised 24 May 2002; accepted 28 May 2002)

In this paper, the direct Trefftz method is applied to solve the free-vibration problem of a membrane. In the direct Trefftz method, there exists no spurious eigenvalue. However, an ill-posed nature of numerical instability encountered in the direct Trefftz method requires some treatments. The Tikhonov's regularization method and generalized singular-value decomposition method are used to deal with such an ill-posed problem. Numerical results show the validity of the current approach. [DOI: 10.1121/1.1494992]

PACS numbers: 43.40.Dx [LLT]

I. INTRODUCTION

The eigenproblem is encountered very often in both engineering practice and academic research. In the design stage of a structure system, it is well known that the engineer should avoid having the structural eigenfrequencies coincide with the driving force frequency.[1] In the linear theory of vibration analysis, it is known that the eigenvalues and corresponding eigenfunctions are used to represent arbitrary functions, which means that they construct the operator spectrum.[2] It is not surprising then that the eigenproblem analysis becomes vitally important in tackling the wonderful world of vibration.

For an arbitrarily shaped domain, the numerical method sometimes is required in analysis because the analytical solution might not be available. To date, the finite element method (FEM) and the boundary element method (BEM) have been attractive to both the academic and engineering fields because of their respective merits. Eigenproblem analysis using the boundary element method has been studied for a long time. The complex-valued singular type auxiliary functions have been adopted.[3] To avoid complex-valued computation, De Mey[4] proposed two alternatives: the real-part formulation and regular formulation. The real-part formulation basically adopts the real-part function of the complex-valued auxiliary function (the fundamental solution) as the auxiliary function. The multiple reciprocity method (MR/BEM),[5] which treats the Helmholtz equation as a Poisson's equation, has been developed for eigenproblem analysis.[6–9] Both the real-part formulation and MR/BEM result in the spurious eigenvalues reported by many researchers.[6–8] Yeih *et al.*[9] proved that the real-part formulation and MR/BEM are equivalent mathematically, and the spurious eigenvalues encountered in both formulations stem from lacking constraint equations contributed by the imaginary part of the complex-valued fundamental solution. Another approach proposed by De Mey is the regular formulation. This method adopts a nonsingular auxiliary function to construct constraint equations. Kim and Kang[10] used the wave-type base functions, one regular formulation in our opinion, to analyze the free vibration of membranes. In their paper, the wave-type base functions, which are periodic along each element and propagate into the domain of interest, were selected to construct the needed equations. They pointed out that some incorrect answers might appear, and they explained these phenomena as due to the incompleteness of the basis functions. Later, Kang *et al.*[11,12] proposed another regular formulation using the so-called nondimensional dynamic influence function. Simply speaking, their method took the response at any point inside the domain of interest as a linear combination of many nonsingular point sources located at the selected boundary nodes. They claimed that their method worked very well, and no numerical instability behaviors were reported. Recently, Chen *et al.*[13] used the circular domain and the property of circulants to theoretically examine the possibility of using the imaginary dual BEM as a solution for the Helmholtz eigenproblems. They reported that spurious eigensolutions also appear in the imaginary dual BEM; however, no numerical examples were illustrated in their paper. Kuo *et al.*[14] pointed out that the ill-posed behavior should exist in the regular BEM formulation. They also proposed a combination of the Tikhonov's regularization method and generalized singular-value decomposition to treat such an ill-posed formulation. The regular formulations Kuo *et al.* proposed were the imaginary-part dual BEM and the plane-wave method. Nevertheless, in Kuo's paper their methods failed when a multiply connected domain problem was treated.

Another nonsingular boundary-type approach is the Trefftz method, which has been widely used to deal with many types of problems[15–17] and gained considerable popularity in the BEM community.[18] The boundary-type Trefftz method basically employs a complete set of solutions satisfying the governing equation as the beginning step. To derive

a)Electronic mail: cjr@sena.ntou.edu.tw

the boundary integral equation, either the reciprocity law (which is similar to that used in the conventional BEM) or the weight residual method can be used. A main benefit of the Trefftz method is that it does not involve singular integrals because of the properties of its solution basis functions (T-complete functions); thus, it can be categorized into the regular boundary element method. Since it can avoid the difficulties with integration over singularities in the traditional BEM and often obtains more accurate results, various formulations of the Trefftz method have been developed and further applied to the engineering problems. Two important review articles about the Trefftz method[19] and its existing formulations[20] associated with comparisons with available boundary-type solution procedures can be found. In general, the formulations of the Trefftz method can be classified into the indirect and direct ones. For the indirect Trefftz formulation, the solutions of the problem are approximated by the superposition of the T-complete functions satisfying the governing equation, while in the direct one, the T-complete functions are taken as the weight function and the integral equations are derived from the governing equations. The mathematical bases of them are fairly different.[20] Although the Trefftz method has been successfully used to solve many problems, for the eigenproblem using the Helmholtz equation few attempts[21,22] have been found in the literature, to the authors' best knowledge. The reason may come from the ill-posed behavior nature of a regular formulation as Kuo *et al.*[14] have indicated, and it leads to the inaccuracy of the numerical results.[23] Most of the researchers have been studying the indirect Trefftz formulations. As a counterpart of the indirect Trefftz method, the direct Trefftz method is relatively new from its developing history,[20] and for some problems it performs in a superior way.[24] Besides, in direct Trefftz method there exist no spurious eigenvalues for the eigenproblem analysis, and it can deal with the multiply connected domain problem within its own formulation.

Based on the advantages over the traditional BEM, in this paper we will construct the direct Trefftz formulation to solve the free-vibration problem of a membrane. We prove that the direct Trefftz method has no spurious eigenvalues but has an ill-posed nature of numerical instability. The Tikhonov's regularization method[25] and generalized singular-value decomposition[26] are used to resolve such a problem. The direct Trefftz method can yield a solution for a multiply connected domain. Numerical results are provided to show the validity of our proposed approach.

II. DERIVATION OF DIRECT TREFFTZ FORMULATION

Consider a two-dimensional finite membrane Ω enclosed by the boundary Γ, the governing equation for the free-vibration problem is written as the Helmholtz equation, i.e.,

$$(\nabla^2+k^2)u(\mathbf{x})=0, \quad \mathbf{x}\in\Omega, \tag{1}$$

where ∇^2 is the Laplacian operator, k is the wave number, $u(\mathbf{x})$ is the physical quantity at $\mathbf{x}$.

The direct Trefftz formulation is derived as follows. Let a field $W(\mathbf{x})$ satisfying the Helmholtz equation, i.e.,

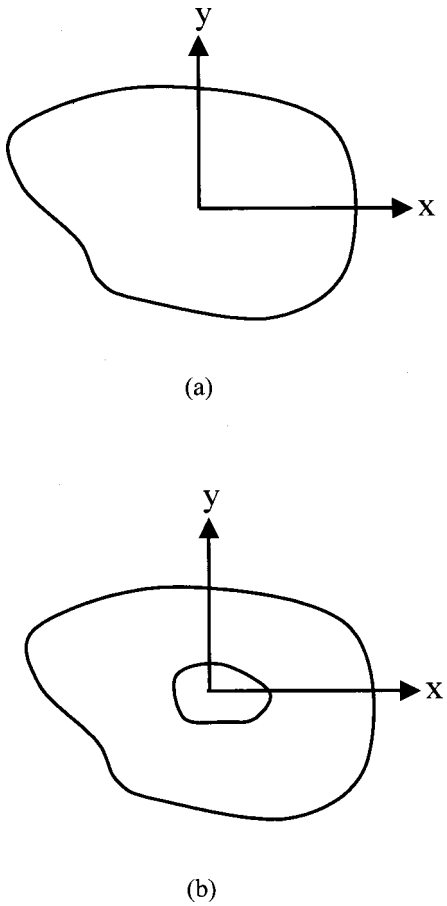


FIG. 1. (a) A simply connected domain. (b) A multiply connected domain of genus 1.

$$(\nabla^2+k^2)W(\mathbf{x})=0, \quad \mathbf{x}\in\Omega; \tag{2}$$

then, by the reciprocity theorem one can have

$$\int_\Gamma W(\mathbf{x})\frac{\partial u(\mathbf{x})}{\partial n}d\Gamma(\mathbf{x})=\int_\Gamma u(\mathbf{x})\frac{\partial W(\mathbf{x})}{\partial n}d\Gamma(\mathbf{x}), \tag{3}$$

where n denotes the out-normal direction at boundary point $\mathbf{x}$. The choice of $W(\mathbf{x})$ depends on the problem itself. A complete set of $W(\mathbf{x})$, written as $\{W_i(\mathbf{x})\}$, is chosen to give enough bases to represent all the physical quantities. This complete set is called the T-complete function set. In the mathematical language, the T-complete function set provides complete function bases to represent any physical field. For example, a simply connected domain shown in Fig. 1(a) and having the origin located inside the domain of interest, it is convenient to have the T-complete set as

$$\{J_0(kr), J_m(kr)\cos(m\theta), J_m(kr)\sin(m\theta)\}$$

for $m=1,2,3...$, in which J_m is the first kind of Bessel function of mth order, r is the distance from the origin to a domain point, and θ is the angle between the x axis and the radial vector from the origin to that domain point. For a multiply connected domain of genus 1 (i.e., a domain with one hole) and locating the origin inside the hole as shown in Fig. 1(b), the T-complete set is

$$\{J_0(kr), Y_0(kr), J_m(kr)\cos(m\theta), J_m(kr)\sin(m\theta),$$

$$Y_m(kr)\cos(m\theta), Y_m(kr)\sin(m\theta)\} \quad \text{for} \quad m=1,2,3,...,$$

where Y_m is the second kind of Bessel function of mth order.

For a boundary value problem, $\alpha_1 u+\beta_1 t=0$, where $t(\mathbf{x})\equiv[\partial u(\mathbf{x})]/\partial n$, one can assign

$$u=\beta_1\psi, \quad t=-\alpha_1\psi, \tag{4}$$

then substituting them into Eq. (3) produces

$$\int_\Gamma\left[\alpha_1 W(\mathbf{x})+\beta_1\frac{\partial W(\mathbf{x})}{\partial n}\right]\psi(\mathbf{x})d\Gamma(\mathbf{x})=0. \tag{5}$$

Changing the base functions, $W_i(\mathbf{x})$, and adopting constant element implementation for boundary unknowns ψ, one can have the following linear algebraic equation:

$$\{\alpha_1[\tilde{\mathbf{U}}]+\beta_1[\tilde{\mathbf{T}}]\}[\psi]=0, \tag{6}$$

where the components of the matrices are represented as

$$\tilde{U}_{ij}\equiv\int_{\Gamma_j} W_i(\mathbf{x})d\Gamma(\mathbf{x}), \tag{7a}$$

$$\tilde{T}_{ij}\equiv\int_{\Gamma_j}\frac{\partial W_i(\mathbf{x})}{\partial n}d\Gamma(\mathbf{x}), \tag{7b}$$

in which Γ_j is the jth element on the boundary and $W_i(\mathbf{x})$ is the ith base function.

There is something worth mentioning here; that is, the direct Trefftz method will not have spurious eigensolutions. To prove this, we need to take a look at Kuo's work.[14] Consider the original problem having boundary condition $\alpha_1 u+\beta_1 t=0$ on the boundary; the corresponding influencing matrix $\mathbf{A}_1$ is

$$\mathbf{A}_1=\alpha_1\tilde{\mathbf{U}}+\beta_1\tilde{\mathbf{T}}. \tag{8a}$$

Let us pick another complementary problem with boundary condition $\alpha_1 u+\beta_1 t=0$ on the boundary and

$$\det\begin{Vmatrix}\alpha_1 & \beta_1\\ \alpha_2 & \beta_2\end{Vmatrix}\neq 0;$$

the influencing matrix $\mathbf{A}_2$ is

$$\mathbf{A}_2=\alpha_2\tilde{\mathbf{U}}+\beta_2\tilde{\mathbf{T}}. \tag{8b}$$

These two systems cannot have the same eigensolution. That is, at a specific wave number k, it is impossible to have the same nontrivial boundary eigensolution $\psi(\mathbf{x})$ for both systems. This theorem is proven in Kuo's paper,[14] and we adopt their results as follows for the readers' convenience.

A. Lemma 1

Given that the governing equation is a Helmholtz equation, $(\nabla^2+k^2)u(\mathbf{x})=0$, for a domain Ω enclosed by the boundary Γ, and that the overspecified homogeneous boundary conditions are $u(\mathbf{x})=0$ and $t(\mathbf{x})=0$ for $\mathbf{x}$ on the subboundary $\Gamma_1\subset\Gamma$, there exists a unique solution, $u(\mathbf{x})=0$ for $\mathbf{x}\in\Omega+\Gamma$.

B. Definition

Two sets of boundary conditions, $\alpha_1(\mathbf{x})u(\mathbf{x})+\beta_1(\mathbf{x})t(\mathbf{x})=0$ and $\alpha_2(\mathbf{x})u(\mathbf{x})+\beta_2(\mathbf{x})t(\mathbf{x})=0$, where $\alpha_1(\mathbf{x})$, $\alpha_2(\mathbf{x})$, $\beta_1(\mathbf{x})$, and $\beta_2(\mathbf{x})$ are given functions, are said to be homogeneous, linearly independent boundary conditions if and only if

$$\det\begin{Vmatrix}\alpha_1(\mathbf{x}) & \beta_1(\mathbf{x})\\ \alpha_2(\mathbf{x}) & \beta_2(\mathbf{x})\end{Vmatrix}\neq 0$$

for any $\mathbf{x}$ on the boundary.

C. Theorem 1

For the Helmholtz equation, given two systems having homogeneous, linearly independent boundary conditions on part of the boundary denoted as Γ_1, it is impossible for both systems to have the same eigensolution.

Theorem 1 supports the conclusion we mentioned above. Theorem 1 also hints that if there exists an ''eigensolution'' to make two systems have homogeneous, linearly independent boundary conditions degenerated at the same time, it must be the spurious eigensolution. Following this, now let us give the proof.

D. Theorem 2

For the Helmholtz equation, given a boundary condition as $\alpha_1 u+\beta_1 t=0$, the direct Trefftz formulation $\mathbf{A}_1(k)\psi=0$ cannot have a spurious eigensolution.

E. Proof

Let us pick another system with a boundary condition as $\alpha_2 u+\beta_2 t=0$ and

$$\det\begin{Vmatrix}\alpha_1 & \beta_1\\ \alpha_2 & \beta_2\end{Vmatrix}\neq 0;$$

its corresponding eigenproblem is written as $\mathbf{A}_2(k)\psi=0$. Further, we assume that there exists a specific wave number k_c such that a nontrivial solution, $\psi(\mathbf{x})$, can satisfy $\mathbf{A}_1(k_c)\psi=0$ and $\mathbf{A}_2(k_c)\psi=0$ simultaneously. This means that $\psi(\mathbf{x})$ is a spurious eigensolution by Theorem 1. Suppose there are n constant elements on the boundary for both problems. Then, it can be said that the following linear system:

$$\begin{bmatrix}\mathbf{A}_1\\ \mathbf{A}_2\end{bmatrix}_{2n\times n}=[\psi]_{n\times 1}\begin{bmatrix}\alpha_1\tilde{\mathbf{U}}+\beta_1\tilde{\mathbf{T}}\\ \alpha_2\tilde{\mathbf{U}}+\beta_2\tilde{\mathbf{T}}\end{bmatrix}[\psi]_{n\times 1}$$

$$=\begin{bmatrix}\alpha_1\mathbf{I} & \beta_1\mathbf{I}\\ \alpha_2\mathbf{I} & \beta_2\mathbf{I}\end{bmatrix}_{2n\times n}\begin{bmatrix}\tilde{\mathbf{U}}\\ \tilde{\mathbf{T}}\end{bmatrix}_{2n\times n}[\psi]_{n\times 1}=0, \tag{9}$$

must be linear dependent where $\mathbf{I}$ is an n by n identity matrix. It can then be said that this is possible if and only if

$$\text{rank}\left(\begin{bmatrix}\tilde{\mathbf{U}}\\ \tilde{\mathbf{T}}\end{bmatrix}\right)<n. \tag{10}$$

If n is very large and equal or unequal length element is adopted, from Eqs. (7a) and (7b), we can say that Eq. (10) is equivalent to

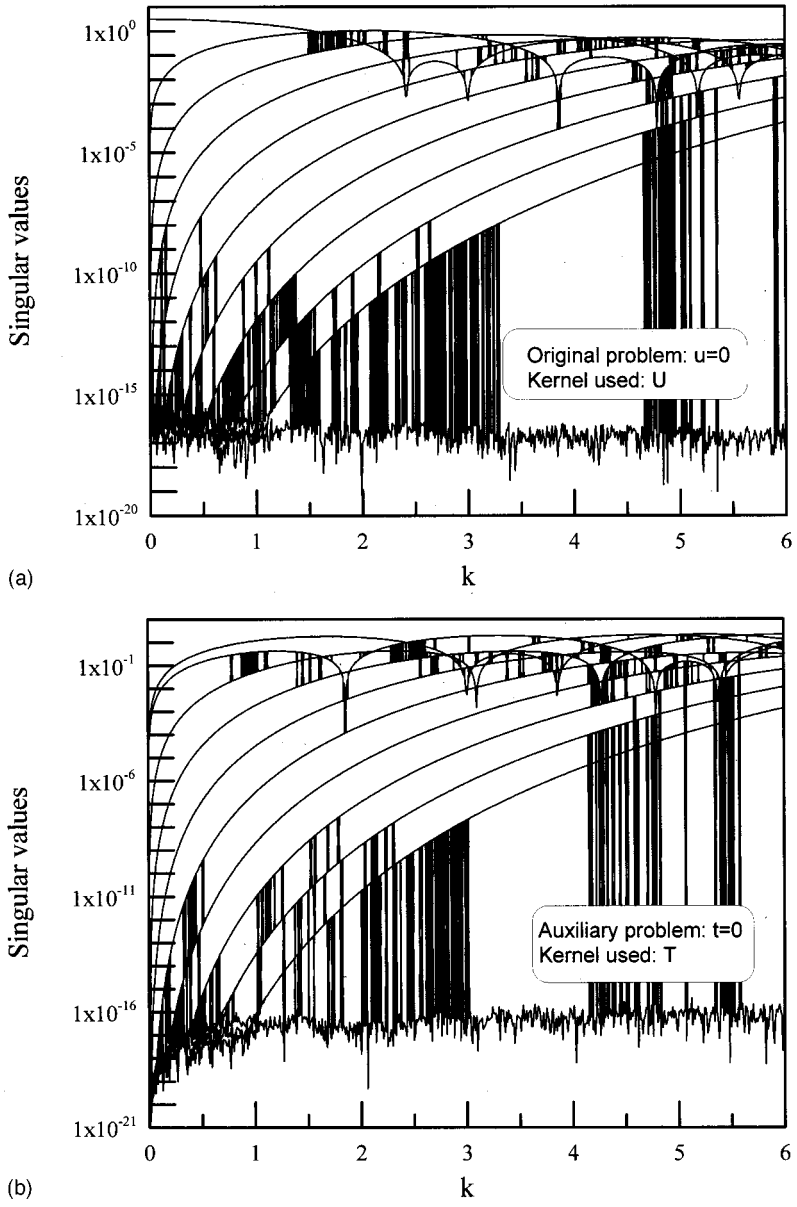


FIG. 2. (a) Numerical contamination exists by only performing the SVD technique for the original system. (b) Numerical contamination exists by only performing the SVD technique for the auxiliary system.

$$\operatorname{rank}\left(\begin{bmatrix} L_j W_i(\mathbf{x}_j) \\ L_j \dfrac{\partial W_i(\mathbf{x}_j)}{\partial n_j} \end{bmatrix}\right) < n, \tag{11}$$

where L_j is the element length of the jth element. The above equation is impossible to be achieved due to the linearly independent behavior of the base functions $W_i(\mathbf{x})$. Actually,

$$\operatorname{rank}\left(\begin{bmatrix} L_j W_i(\mathbf{x}_j) \\ L_j \dfrac{\partial W_i(\mathbf{x}_j)}{\partial n_j} \end{bmatrix}\right) = n.$$

This then leads to a contradiction and completes the proof.

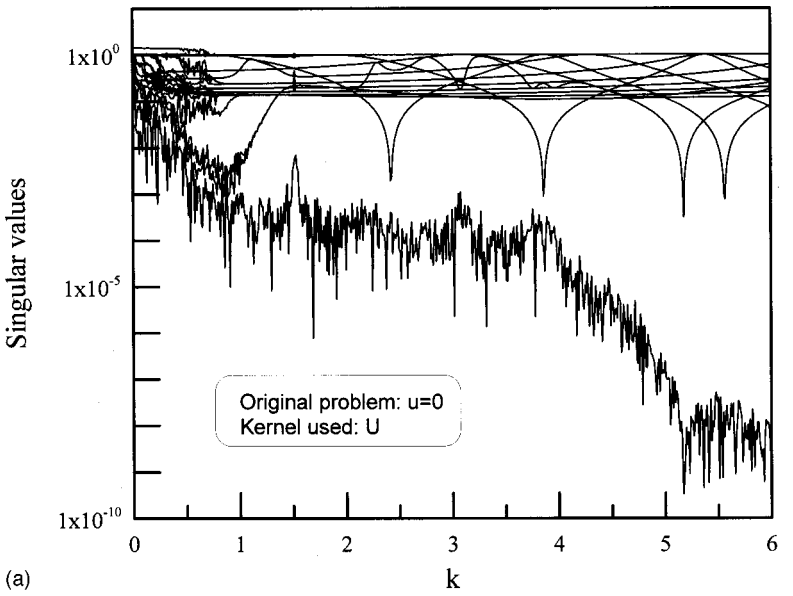
III. REGULARIZATION AND GENERALIZED SINGULAR-VALUE DECOMPOSITION METHODS TO DEAL WITH THE ILL-POSED PROBLEM

The Trefftz method adopts nonsingular base functions, and thus can be categorized into the regular BEM formulations.[14] However, the regular formulation leads to the ill-posed behaviors while the nodes (or elements) increase. Kuo *et al.*[14] explained the reason and proposed a method to fix it. Here, we simply introduce the method Kuo *et al.* suggested since we will use the same technique later on.

To treat the ill-posed behaviors, Kuo *et al.*[14] proposed using the Tikhonov's regularization method and generalized singular-value decomposition. Now, let us briefly introduce their idea. From Theorem 1, it can be seen that the spurious eigensolution will appear in two systems having homogeneous, linearly independent boundary conditions simultaneously. That is, we have a system as $[\mathbf{A}_1]_{n\times n}\psi_{n\times 1} = [\mathbf{A}_2]_{n\times n}\psi_{n\times 1} = \mathbf{0}$. Since both problems can have common spurious eigensolutions, we can intuitively decompose both matrices into the following form:

$$\mathbf{P}\mathbf{W}_1\mathbf{x} = \mathbf{P}\mathbf{W}_2\mathbf{x} = 0,$$

where $\mathbf{P}\mathbf{W}_1 = \mathbf{A}_1$ and $\mathbf{P}\mathbf{W}_2 = \mathbf{A}_2$. Then, spurious eigenvalues will result in the rank deficiency of matrix $\mathbf{P}$, and true eigenvalues will result in the rank deficiency of matrix $\mathbf{W}_1$ for the original problem. When the spurious eigenvalues are encoun-



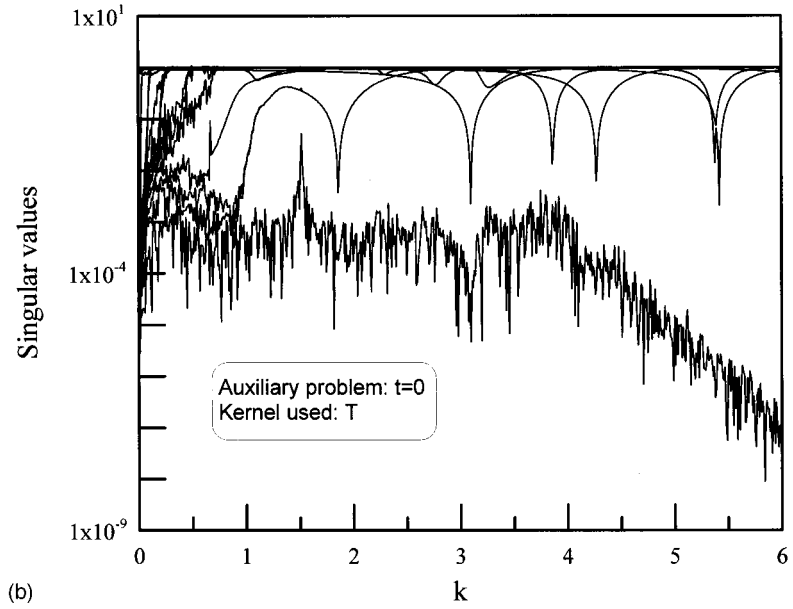


FIG. 3. (a) Numerical contamination still exists after using Tikhonov's regularization method and the SVD technique for the original system. (b) Numerical contamination still exists after using the Tikhonov's regularization method and the SVD technique for the auxiliary system.

tered, basically we want to extract them by finding matrix **P**. That is, to perform a numerical operation of L'Hospital rule on an indefinite form of 0/0. The above-mentioned technique can be achieved using the QR factorization, which is the first step of the generalized singular-value decomposition.

Remember that the serious problem we encounter is not spurious eigensolution but numerical instability of this algorithm. To treat this, we will add some small quantities into the matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ to make the numerically tiny singular values occurring in both matrices become "numerical spurious eigenvalues" such that the QR factorization can extract them. Let $\mathbf{A}_1$ and $\mathbf{A}_2$ have the following singular value decompositions:

$$\mathbf{A}_1 = \mathbf{P}\Sigma_1\mathbf{V}_1^*, \tag{12a}$$

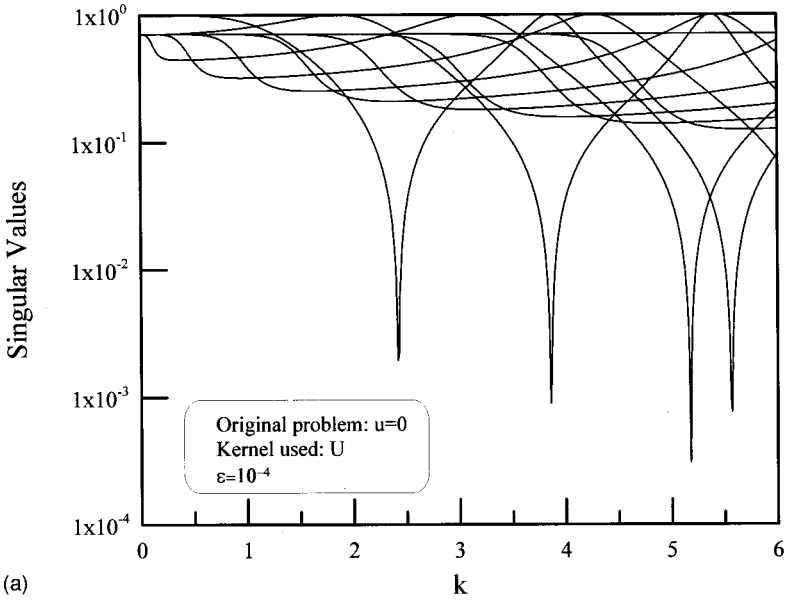
$$\mathbf{A}_2 = \mathbf{P}\Sigma_2\mathbf{V}_2^*, \tag{12b}$$

where $\mathbf{V}_i$ is the right unitary matrix of system i, the superscript "*" means take the transpose and complex conjugate of the matrix, and Σ_i is a singular value matrix of system i with singular values allocated in the diagonal line. When one of the singular values is numerically very small at a specific wave number, it can be said that the system has degenerated, i.e., that the wave number is an eigenvalue. However, when a nonsingular BEM is adopted, there exist many numerical tiny values in the singular values, which are not true zeros. This phenomenon becomes very severe when the number of elements increases and/or a direct eigenvalue search is used at a low wave number. Now, let us add two small quantities in the matrices to construct new influencing matrices as

$$\hat{\mathbf{A}}_1 = \mathbf{P}(\Sigma_1 + \varepsilon_1\mathbf{I})\mathbf{V}_1^*, \tag{13a}$$

$$\hat{\mathbf{A}}_2 = \mathbf{P}(\Sigma_2 + \varepsilon_2\mathbf{I})\mathbf{V}_2^*, \tag{13b}$$

where ε_i is the small value added to system i. The choice of ε_i is dependent on the problem itself; however, if they are larger than the unreasonable tiny values of singular values in the original two systems, but still small enough not to overcoat the true eigenvalue, one can then successfully extract



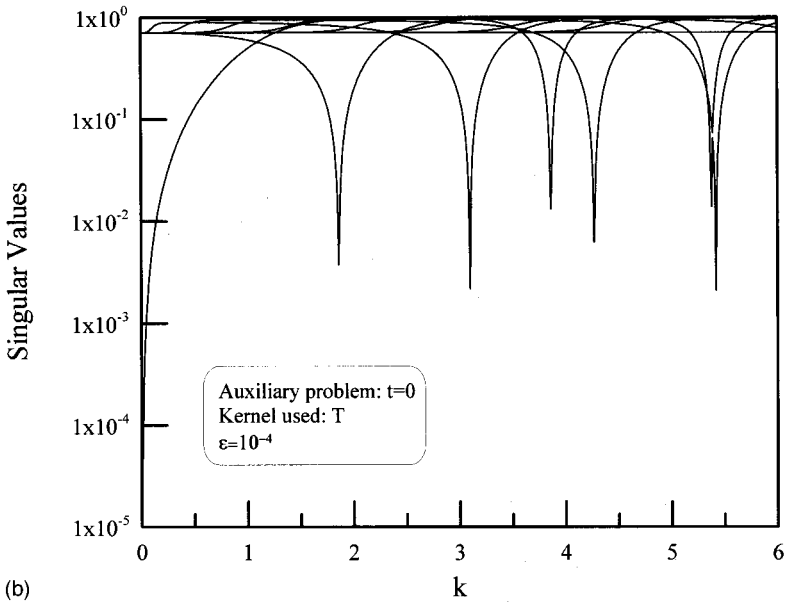


FIG. 4. (a) Numerical contamination is eliminated by using Tikhonov's regularization method, QR factorization, and the SVD technique for the original system. (b) Numerical contamination is eliminated by using Tikhonov's regularization method, QR factorization, and the SVD technique for the auxiliary system.

the contaminated tiny value. If one takes the QR factorization of $\hat{\mathbf{A}}_1$ and $\hat{\mathbf{A}}_2$, the unreasonable ones can be extracted. The idea can be seen in Fig. 2–Fig. 4. Before treatments shown in Fig. 2, at low wave numbers some singular values are very small for both systems. However, when the minimum singular value of system one occurs for "mode shape p" (or pth singular vector), the corresponding singular value of system two for the same pth singular vector may not be the smallest one in system two. After QR factorization, this singular value remains the smallest in system one such that we still cannot distinguish if it is an eigenvalue, as shown in Fig. 3. After treatment as shown in Fig. 4, the contaminated singular values for both systems are elevated. The QR factorization method extracts such singular values out and changes the order of the singular values. Adding such a small value (for instance, $\varepsilon = 10^{-4}$ in Fig. 4) in the singular value cannot change the facts of true degenerated singular value. That is, at the true eigenvalue, the singular value of system one should approach zero but its corresponding part in system two will not be close to zero. Using this method, we can successfully treat the ill-posed behaviors; numerical examples will be given in the next section.

IV. NUMERICAL EXAMPLES

A. Example 1

A circular domain with radius $R_o = 1.0$ and the Dirichlet boundary condition, $u = 0$, is given.

Fifty-one constant elements are used, and the Neumann condition problem, $t = 0$, is chosen as the auxiliary problem. Using the Tikhonov's regularization method and generalized singular-value decomposition, eigenvalues are found successfully, as shown in Fig. 5. In this figure, the value in the bracket is the analytical solution.

B. Example 2

A circular domain with radius $R_o = 1.0$ and the Neumann boundary condition, $t = 0$, is given.

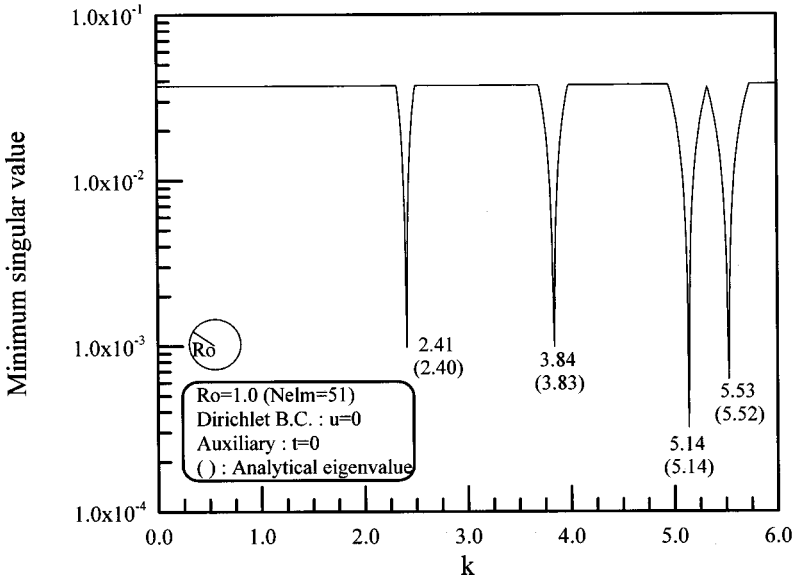


FIG. 5. Eigenvalues searching for the Dirichlet boundary condition of a unit circle by using the direct Trefftz method.

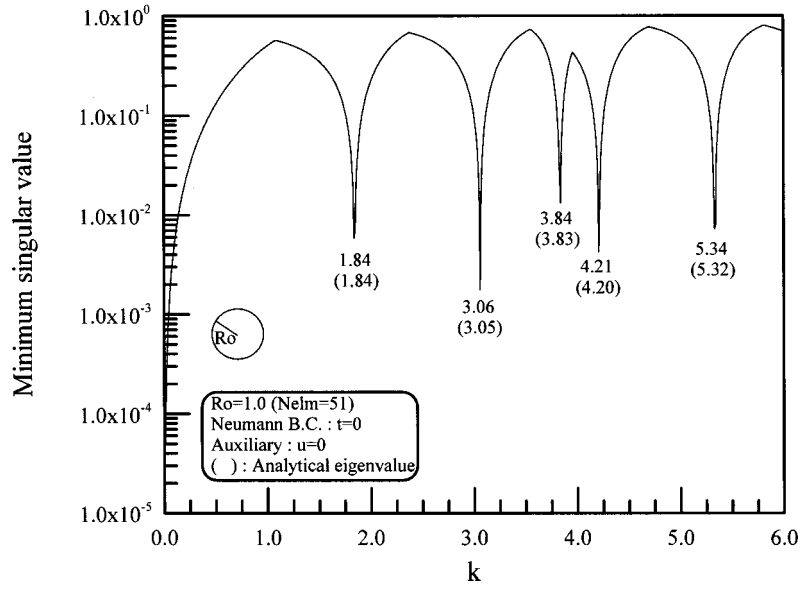


FIG. 6. Eigenvalues searching for the Neumann boundary condition of a unit circle by using the direct Trefftz method.

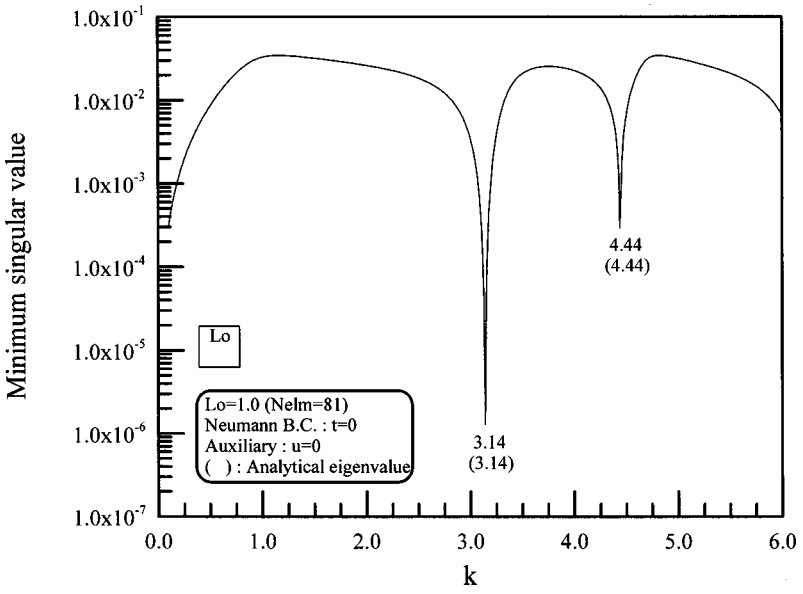


FIG. 7. Eigenvalues searching for the Neumann boundary condition of a square by using the direct Trefftz method with an auxiliary system, $u=0$.

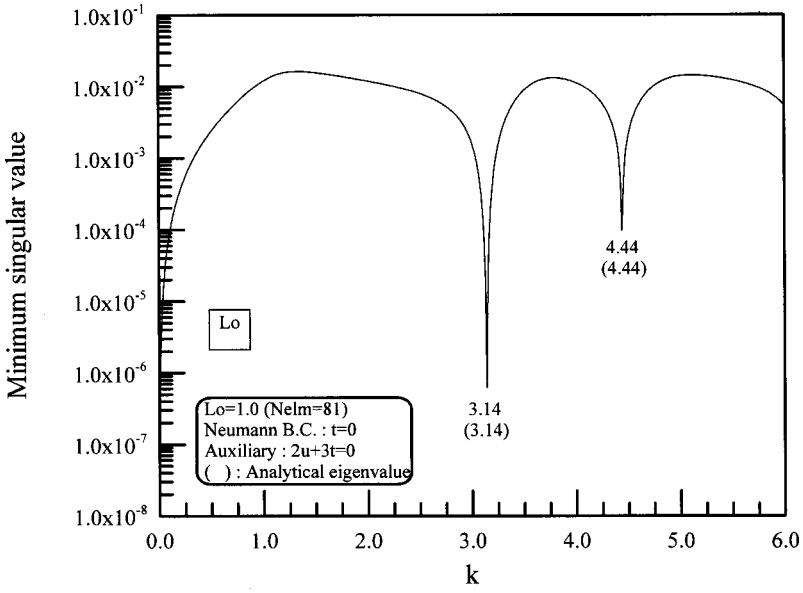


FIG. 8. Eigenvalues searching for the Neumann boundary condition of a square by using the direct Trefftz method with a different auxiliary system, $2u+3t=0$.

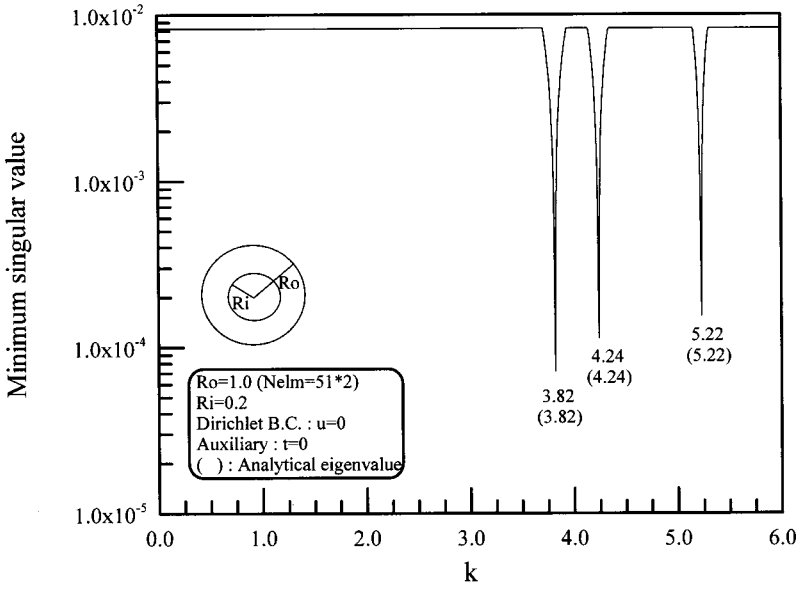


FIG. 9. Eigenvalues searching for the Dirichlet boundary condition of an annular region.

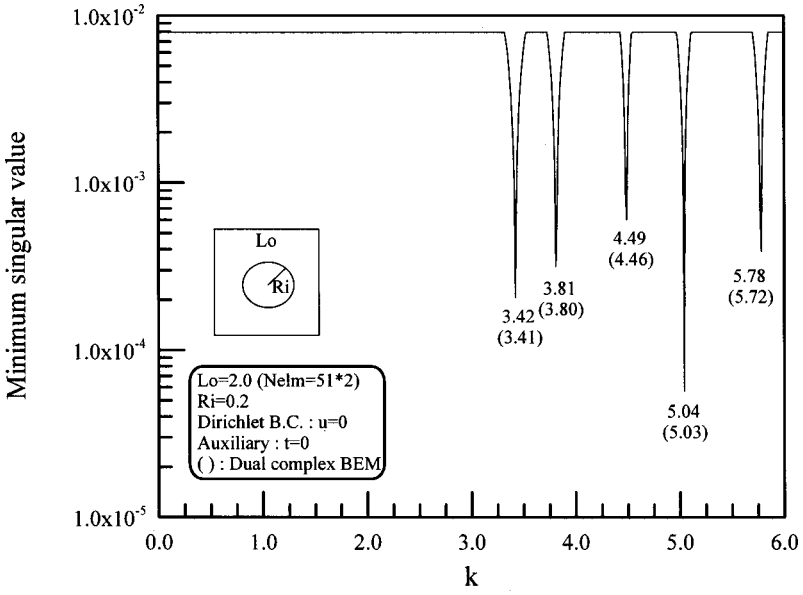


FIG. 10. Eigenvalues searching for the Dirichlet boundary condition of the multiply connected domain.

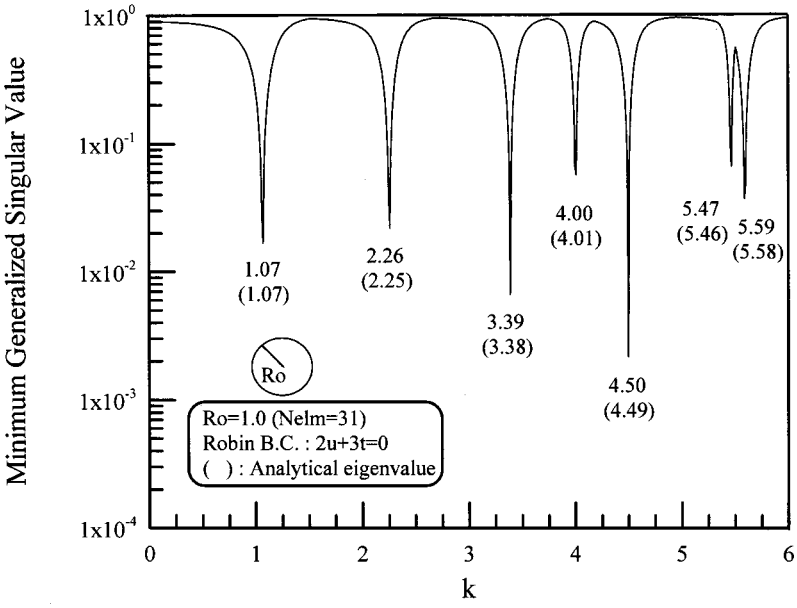


FIG. 11. Eigenvalues searching for the Robin-type boundary condition of a unit circle.

In this example, we can see that our method is valid for all kinds of boundary conditions. Again, 51 constant elements are used and the Dirichlet condition problem, $u=0$, is used as the auxiliary problem. Using the proposed method, eigenvalues are successfully found and are very close to the analytical values, as shown in Fig. 6.

C. Example 3

A square membrane with edge length $L_o=1.0$ and the Neumann boundary condition, $t=0$, is prescribed.

In this example, a domain without radial symmetry is illustrated. Eighty-one constant elements are used and the Dirichlet boundary problem, $u=0$, is chosen as the auxiliary problem. It can be found in Fig. 7 that the numerical results match the analytical solutions very well.

We have claimed that any problem having a linearly independent boundary condition to the original problem can be used as an auxiliary problem. In this example, we use another auxiliary problem $2u+3t=0$. The results are shown in Fig. 8, and our approach works as expected.

D. Example 4

An annular region with the outer radius $R_o=1.0$ and inner radius $R_i=0.2$, and a Dirichlet boundary condition, $u=0$, is prescribed on the boundary.

The domain is a multiply connected domain, which shows the superiority of the current approach over Kuo's method.[14] Their methods were proven to fail when a multiply connected domain is treated. However, the direct Trefftz method can easily overcome this problem by putting the origin inside the hole. In this example, 51 elements are used for the outer and inner boundaries. The auxiliary problem is the Neumann problem, $t=0$. As shown in Fig. 9, eigenvalues can be found successfully. The analytical values are obtained using the eigenequation[27]

$$[J_m(kR_o)Y_m(kR_i)-Y_m(kR_o)J_m(kR_i)]=0.$$

E. Example 5

A multiply connected domain with the outer boundary is a square with edge length $L_o=2.0$, and the inner boundary is a circle with a radius $R_i=0.2$. The origin of the circular hole is the geometric center of the whole domain. The boundary condition is the Dirichlet condition, $u=0$.

In this example, no analytical solution is available. We compared our results with those obtained from the complex-valued dual BEM. The auxiliary system is the Neumann problem, $t=0$. As shown in Fig. 10, numerical results obtained from the direct Trefftz method are close to those obtained from the complex-valued dual BEM. The reason why a complex-dual BEM is required is explained in Chang's dissertation.[27] He explained that solving an eigenvalue problem of a multiply connected domain by the complex-valued singular integral equation or the complex-valued hypersingular integral equation will result in an unreasonable numerical resonance. He named this kind of degeneracy of the direct BEM the pseudofictitious eigenvalue. To treat this unreasonable degeneracy, a combined use of singular and hypersingular integral equations was suggested. For more detail, readers can refer to Chang's dissertation.[27]

F. Example 6

A circular domain with radius $R_o=1.0$ and the Robin-type boundary condition, $2u+3t=0$, is given on the boundary.

The analytical values for this case can be obtained by using the true eigenequation as:[14] $2J_m(kR_o)+3J'_m(kR_o)=0$. In this case, 31 elements and 31 bases are used correspondingly. When the Tikhonov's regularization method and generalized singular-value decomposition are adopted, eigenvalues are found successfully and numerical results match analytical solutions very well, as shown in Fig. 11.

V. CONCLUSIONS

In this paper, the direct Trefftz method was used to solve the free-vibration problem of a membrane. It was found that the direct Trefftz method has no spurious eigenvalues, but leads to numerical instability when the number of elements increases and/or an eigenvalue search is conducted in the low wave number range. To treat this ill-posed behavior, we adopted the Tikhonov's regularization method and the generalized singular-value decomposition. The direct Trefftz method can easily treat a multiply connected domain of genus 1 by putting the origin inside the hole. Six numerical examples were provided to show the validity of the current approach, and good matches can be achieved in comparison with the analytical solutions or numerical results obtained from other methods.

ACKNOWLEDGMENTS

The first author would like to express his thank to the National Science Council, Taiwan, for their financial support under Contract Number: NSC-90-2611-E-019-018.

[1] R. W. Clough and J. Penzien, *Dynamics of Structures*, 2nd ed. (McGraw-Hill, New York, 1993).
[2] I. Stakgold, *Green's Functions and Boundary Value Problems*, 2nd ed. (Wiley, New York, 1998).
[3] N. Kamiya, E. Andoh, and K. Nogae, ''A new complex-valued formulation and eigenvalue analysis of the Helmholtz equation by boundary element method,'' Adv. Eng. Software **26**, 219–227 (1996).
[4] G. De Mey, ''A simplified integral equation method for the calculation of the eigenvalues of Helmholtz equation,'' Int. J. Numer. Methods Eng. **11**, 1340–1342 (1977).
[5] A. J. Nowak and C. A. Brebbia, ''The multiple reciprocity method—A new approach for transforming BEM domain integral to the boundary,'' Eng. Anal. Boundary Elem. **6**, 164–167 (1989).
[6] J. T. Chen and F. C. Wong, ''Analytical derivations for one-dimensional eigenproblems using dual BEM and MRM,'' Eng. Anal. Boundary Elem. **20**(1), 25–33 (1997).
[7] W. Yeih, J. T. Chen, and C. M. Chang, ''Applications of dual MRM for determining the natural frequencies and natural modes of an Euler–Bernoulli beam using the singular value decomposition method,'' Eng. Anal. Boundary Elem. **23**, 339–360 (1999).
[8] W. Yeih, J. R. Chang, C. M. Chang, and J. T. Chen, ''Application of dual BEM for determining the natural frequencies and natural modes of a rod using the singular value decomposition method,'' Adv. Eng. Software **30**(7), 459–468 (1999).
[9] W. Yeih, J. T. Chen, K. H. Chen, and F. C. Wong, ''A study on the multiple reciprocity method and complex-valued formulation for the Helmhotz equation,'' Adv. Eng. Software **29**(1), 1–6 (1998).
[10] Y. Y. Kim and J. H. Kang, ''Free vibration analysis of membranes using wave-type base functions,'' J. Acoust. Soc. Am. **99**, 2938–2946 (1996).
[11] S. W. Kang, J. M. Lee, and Y. J. Kang, ''Vibration analysis of arbitrarily shaped membranes using nondimensional dynamic influence function,'' J. Sound Vib. **221**(1), 117–132 (1999).
[12] S. K. Kang and J. M. Lee, ''Eigenmode analysis of arbitrarily shaped two-dimension cavities by the method of point-matching,'' J. Acoust. Soc. Am. **107**, 1153–1160 (2000).
[13] J. T. Chen, S. R. Kuo, and K. H. Chen, ''A nonsingular integral formulation for the Helmholtz eigenproblems of a circular domain,'' J. Chin. Inst. Eng. **22**(6), 729–739 (1999).
[14] S. R. Kuo, W. Yeih, and Y. C. Wu, ''Application of the generalized singular-value decomposition on the eigenproblem using the incomplete boundary element formulation,'' J. Sound Vib. **235**(5), 813–845 (2000).
[15] W. G. Jin, Y. K. Cheung, and O. C. Zienkiewicz, ''Applications of the Trefftz method in plane elasticity problems,'' Int. J. Numer. Methods Eng. **30**, 1147–1161 (1990).
[16] W. G. Jin, Y. K. Cheung, and O. C. Zienkiewicz, ''Trefftz method for Kirchhoff plate bending problems,'' Int. J. Numer. Methods Eng. **36**, 765–781 (1993).
[17] I. Harari, P. Barai, and P. E. Barai, ''Three-dimensional infinite elements based on a Trefftz formulation,'' J. Comput. Acoust. **9**(2), 381–394 (2001).
[18] N. Kamiya and E. Kita, ''Editorial—Trefftz method 70 years,'' Adv. Eng. Software **24**, 1 (1995).
[19] J. Jirousek and A. Wróblewski, ''T-elements: State of the art and future trends,'' Arch. Comput. Methods Eng. **3**(4), 323–434 (1996).
[20] E. Kita and N. Kamiya, ''Trefftz method: An overview,'' Adv. Eng. Software **24**, 3–12 (1995).
[21] Y. K. Cheung, W. G. Jin, and O. C. Zienkiewicz, ''Solution of Helmholtz equation by Trefftz method,'' Int. J. Numer. Methods Eng. **32**, 63–78 (1991).
[22] N. Kamiya and S. T. Wu, ''Generalized eigenvalue formulation of the Helmholtz equation by Trefftz method,'' Eng. Comput. **11**, 177–186 (1994).
[23] E. Kita, N. Kamiya, and T. Iio, ''Application of a direct Trefftz method with domain decomposition to 2D potential problems,'' Eng. Anal. Boundary Elem. **23**, 539–548 (1999).
[24] W. G. Jin and Y. K. Cheung, ''Trefftz direct method,'' Adv. Eng. Software **24**, 65–69 (1995).
[25] A. N. Tikhonov, ''Solution of incorrectly formulated problems and regularization method,'' Dokl. Akad. Nauk SSSR **151**(3), 501–504 (1963).
[26] G. H. Golub and C. F. Van Loan, *Matrix Computation*, 2nd ed. (Johns Hopkins University Press, Baltimore, 1989).
[27] J. R. Chang, ''Applications of domain partitioning on vibration problems using the dual BEM,'' Ph.D. dissertation of Department of Harbor and River Engineering, National Taiwan Ocean University, Keelung, Taiwan (1999).

Leaky helical flexural wave backscattering contributions from tilted cylindrical shells in water: Observations and modeling

Florian J. Blonigen and Philip L. Marston[a)]
Department of Physics, Washington State University, Pullman, Washington 99164-2814

(Received 26 January 2002; revised 12 May 2002; accepted 20 May 2002)

For tilt angles smaller than the meridional ray coupling condition previously investigated [S. F. Morse *et al.*, J. Acoust. Soc. Am. **103**, 785–794 (1998)], flexural helical waves on cylindrical shells can significantly enhance the backscattering. These contributions are compared and modeled here for an empty cylinder. Experiments using tone bursts were performed on a tilted stainless steel shell to investigate the contributions caused by flexural leaky Lamb waves above the coincidence frequency of the shell. In some of the measurements the tone bursts were of sufficient duration to superpose helical wave contributions of successive circumnavigations, along with the meridional contribution near the critical tilt, to arrive at a quasi-steady-state backscattering amplitude for the cylinder. These measurements are compared with an approximate numerical partial-wave series solution and a ray theory as a function of the tilt angle. The data for $ka=20$ follow the basic shape of the ray theory and the relevant features of the partial-wave model. They illustrate the importance of the interference of successive helical wave contributions. Measurements (also as a function of the tilt angle) using tone bursts that were sufficiently short to separate the earliest helical wave contribution from later contributions also support the ray theory. [DOI: 10.1121/1.1492822]

PACS numbers: 43.40.Fz, 43.30.Gv, 43.20.Fn [ANN]

I. INTRODUCTION

At frequencies above the coincidence frequency, it is found that leaky flexural waves make important contributions to the backscattering of sound by cylindrical shells in water.[1–4] Specifically, for moderately thick shells, the lowest-order antisymmetric (a_0) Lamb wave can make significant backscattering contributions out to large oblique tilt angles. Other studies typically have emphasized scattering by thin shells at lower frequencies (typically in the low- and mid-frequency regions) where torsional and compressional leaky waves make the primary contributions.[5–8] Those studies typically emphasize tilt angles relatively close to broadside illumination as a consequence of the shear and compressional wave cutoff angles. Torsional and compressional waves contribute to the scattering above the coincidence frequency in addition to the flexural waves considered here, however, they are weakly damped and are likely to be of less importance in high-frequency, high-resolution sonar imaging applications.[2,9,10] That is because when short tone bursts are used to illuminate the target, there can be insufficient time for the narrow resonances to build up to a steady response.

In this work, backscattering experiments were performed on a stainless steel cylindrical shell in water using incident tone bursts of various durations to investigate flexural wave contributions. The magnitude of the backscattering was compared against theoretical predictions of the quasi-steady-state response for plane wave incidence. Figure 1 shows the backscattering form function magnitude plotted as a function of the tilt angle γ at $ka=20$ for an empty shell. The solid points are the experimental data, which are to be compared with two theoretical curves. The solid curve is an approximate numerical partial wave series (PWS) solution for the backscattering[4] based on an extension of the method of Ref. 5. The dashed curve is given by a simpler ray model that was formulated in this work and is discussed in Sec. II. In the ray model, the large peak at $\gamma\approx 43$ degrees is identified as an a_0 flexural wave contribution due to *meridional* leaky rays on the shell, while the broad oscillations at smaller tilt angles are due to the interference of *helical* leaky ray arrivals with each other and with the meridional contribution. The emphasis of this study is on the helical contributions that are important in the range between 10 and 40 degrees in this example. The backscattering form function magnitude associated with these contributions is typically greater than unity, which may be considered large; and it is much larger than the backscattering magnitude of the rigid response alone at oblique tilt angles at this frequency.[11] The normalization used to express the dimensionless form function is such that for geometric reflection from a sphere having the same radius as the cylinder the form function is unimodular.[4]

Relevant ray paths for the ray model are shown in Fig. 2. Rays are launched at those points on the cylinder where the angle between the incident wave vector and the local shell normal is equal to the leaky wave coupling angle, $\theta_l = \sin^{-1}(c/c_l)$, for the lth class of leaky wave under consideration (the a_0 waves in this work). Here c_l is the phase velocity ($c_l > c$) of the leaky wave of type l and c is the speed of sound in the surrounding fluid. The energy associated with a leaky ray attenuates along its ray path due to radiation damping with a characteristic attenuation constant α, which

a)Author to whom correspondence should be addressed. Electronic mail: marston@wsu.edu

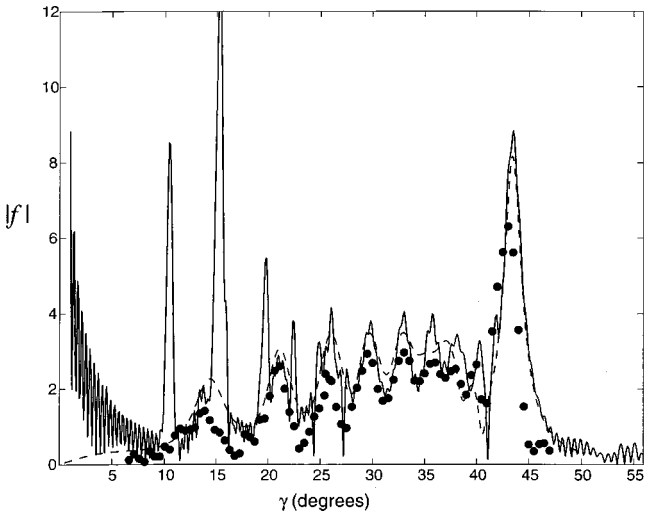


FIG. 1. Far-field backscattering form function magnitude as a function of tilt angle (measured from broadside incidence) and fixed frequency ($ka=20$) for an air-filled cylindrical shell target of finite length which is composed of stainless steel (SS304). The solid curve is an approximate numerical partial wave series solution described in Ref. 4. The dashed curve is a ray model described in this work which predicts the quasi-steady-state response arising from the a_0 leaky flexural waves on the shell. The solid points are experimental data. The large peak at $\gamma \approx 43$ degrees is the meridional leaky flexural wave enhancement. The broad oscillations at smaller tilt angles are the helical wave contributions. Narrow features of the series solution are discussed in Sec. IV.

may be regarded as the imaginary part of a complex leaky wave number $k_p = k_l + i\alpha$, where $k_l = k(c/c_l)$ is the real part of the leaky wave number and k is the wave number in water. The outgoing rays leave the cylinder at the same coupling angle relative to the shell normal at the detachment points. When the tilt angle of the cylinder, γ, measured from broadside incidence, is equal to the coupling angle θ_l, an incident ray A can launch a meridional leaky ray that runs strictly in the axial direction on the cylinder (ray B) and leaks radiation in the specular direction (ray C). Upon reflection from the cylinder's truncation, the meridional leaky ray reverses direction and radiates in the backscattering direction (ray D) as well as from subsequent locations on the meridian. For tilt angles less than the coupling angle, $\gamma < \theta_l$, an incident ray E launches a leaky ray on the shell at a nonzero azimuthal angle ϕ, measured from the meridian as shown in Fig. 2. The resulting leaky ray F travels along a helical path on the cylinder with helix angle, ψ, and radiates in the specular direction (ray G) at points corresponding to the azimuthal angle $-\phi$ along the leaky ray path, or in the backscattering direction if it has reflected off the cylinder truncation. By symmetry, helical rays may also be launched at azimuthal angle $-\phi$ which have the opposite helicity, or helix angle $-\psi$.

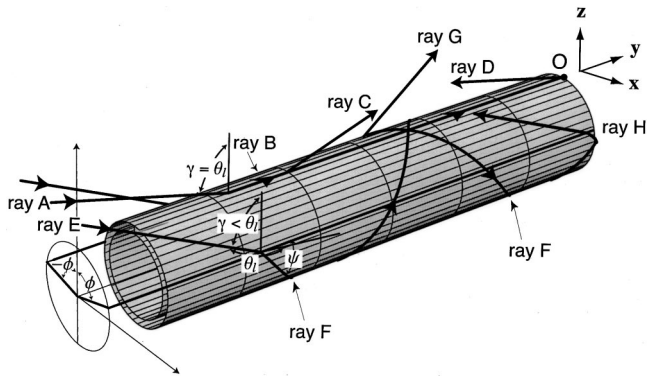


FIG. 2. Leaky ray paths on a cylinder. Leaky rays are launched at those points on the cylinder where the angle between the incident wave vector and the local shell normal is equal to the leaky wave coupling angle θ_l. When the tilt angle is equal to the coupling angle, $\gamma = \theta_l$, an incident ray (ray A) can launch a meridional leaky ray which runs along the meridian of the cylinder (ray B). This reradiates in the specular direction at the same coupling angle (ray C). Upon reflection from the cylinder's truncation at point O, the meridional ray reverses direction and radiates in the backscattering direction (ray D). For $\gamma < \theta_l$, an incident ray (ray E) may launch a helical leaky ray (ray F) at a point off the meridian at azimuthal angle, ϕ, measured from the meridian. It travels along a helical path with helix angle ψ. It reradiates in the specular direction at the azimuthal angle $-\phi$ (ray G), or in the backscattering direction (ray H) if it has reflected off the cylinder truncation.

It is evident in Fig. 1 that the meridional contribution gives the dominant feature in the backscattering form function over a wide range of tilt angle. This results from the fact that meridional contributions can be associated with short leaky ray paths on the shell. Helical rays, on the other hand, must traverse the backside of the cylinder at least once, so that they necessarily have longer ray paths on the shell than those of meridional rays. This leads to lesser contributions from the helical rays because of the radiation damping. Nevertheless, helical waves of the lowest-order antisymmetric a_0 Lamb mode make appreciable scattering contributions and are important for understanding the high-frequency response from shells at tilt angles smaller than the coupling angle $\gamma < \theta_l$, where the meridional contribution is diminished. While the present work emphasizes the helical contributions, we also observe interference between helical and meridional components in the backscattering. The meridional ray contribution was explained in experiments by Morse,[4] and the ray model for the meridional contribution was developed previously by Marston.[12] The present authors previously developed a ray model for the synthesis of helical and meridional specularly scattered contributions from an infinitely long empty cylindrical shell.[1] The ray theory is reformulated here for the case of a finite shell in Sec. II. These ray models are not based on the assumptions of thin-shell mechanics, typically used for thin shells at lower frequencies.[5,6] Section III discusses the experiments using short tone bursts and the comparison with theory for an air-filled shell. Section IV discusses the observations plotted in Fig. 1 which used longer tone bursts.

II. RAY THEORY FOR HELICAL FLEXURAL WAVE CONTRIBUTIONS FROM A FINITE CYLINDER

The ray theory presented here is an extension of a discussion in Sec. II of Ref. 1 in which a ray formulation was developed to predict the far-field scattering amplitudes in the specular direction due to helical rays on an infinitely long cylinder. In this section, we will be concerned with the backscattering contributions of helical waves on finite cylinders. The analogous ray prediction for the backscattering amplitudes due to meridional rays is derived in Ref. 12 and use will be made of this result in the final synthesis of the meridional and helical contributions.

In this ray formulation, the steady-state backscattering response from an incident plane wave will be considered. The plane wave is described by a wave vector **k**, which has magnitude $k=\omega/c$ and tilt angle γ with respect to broadside incidence. It is incident on a finite cylinder of outer radius a and length L. For a finite cylindrical shell, the target is also characterized by an inner radius b. Let the phase of the plane wave be referenced to the origin, given by point O in Fig. 2. Figure 2 also shows a set of three orthogonal unit vectors which define the x, y, and z coordinate axes of a Cartesian coordinate system. For future reference, the y-z plane will be referred to as the meridional plane, the plane defined by the wave vector and the cylinder's axis; the x-z plane will be referred to as the base plane, the plane perpendicular to the cylinder's axis; and the x-y plane will be referred to as the exit plane, the plane tangent to the cylinder on its illuminated side. This plane is the two-dimensional domain of integration for the propagation integrals that will be used to calculate backscattering amplitudes.

Section II of Ref. 1 gives an expression [Eq. (3) in that paper] for the pressure p_l of the outgoing wave radiated by the leaky wave at some point on the surface of the cylinder where an outgoing ray detaches from a helical path. This result was formulated using the method of Sec. V of Ref. 13. It is written here as follows:

$$p_l^{(0)} \approx \kappa (2/\pi k_p)^{1/2} p_o B e^{ik_p \sigma_o + i\pi/4} e^{-i\eta} e^{iky_o \sin\gamma} A_f \sigma_o^{-1/2}/\pi, \tag{1}$$

where k_p is the complex leaky wave number, p_o is the magnitude of the incident pressure, and $\kappa \approx -\alpha k_l \exp(i\varphi_{bl})$ is a leaky wave coupling parameter where the coupling background phase φ_{bl} is discussed in Appendix A of Ref. 13 and in Ref. 12. The phase factors depend on the geometry of the helical ray path, where σ_o is the path length of the helical ray from the launch point to the detachment point of interest and y_o is the y coordinate (measured along the cylinder's axis) of the helical ray launch point. The phase $\eta = ka\cos\theta_l - ka\cos\gamma$ is needed to properly reference the overall phase to the origin. The Fresnel area A_f is a measure of the area around the helical ray launch point which contributes to the final leaky wave pressure $p_l^{(0)}$ at the detachment point. The factor B does not appear in Eq. (3) of Ref. 1. It is a reflection coefficient for the reflection of helical rays off the cylinder truncation.

Figure 3 shows the plane of the unwrapped cylinder surface laid out flat in the exit plane and parametrized by the Cartesian coordinates x and y, where again the y axis is parallel to the cylinder's axis. Ray BA is a helical ray launched at point B with Cartesian coordinates (x_o, y_o). When the tilt angle is less than the coupling angle, $\gamma < \theta_l$, helical rays must be launched at an azimuthal angle ϕ measured from the meridian where ϕ satisfies the equation $\cos\phi = \cos\theta_l/\cos\gamma$. This yields $x_o = a\phi$ for the x coordinate of helical ray launching points. The angle ψ is the helix angle of the helical ray BA, where ψ satisfies the equation $\cos\psi = \sin\gamma/\sin\theta_l$. By symmetry arguments, leaky helical rays launched at $x=\phi a$ will reradiate in the backscattering direction at detachment points along $x=(2\pi-\phi)a$, such as point A shown in Fig. 3. However, this is only true for those rays which have reflected

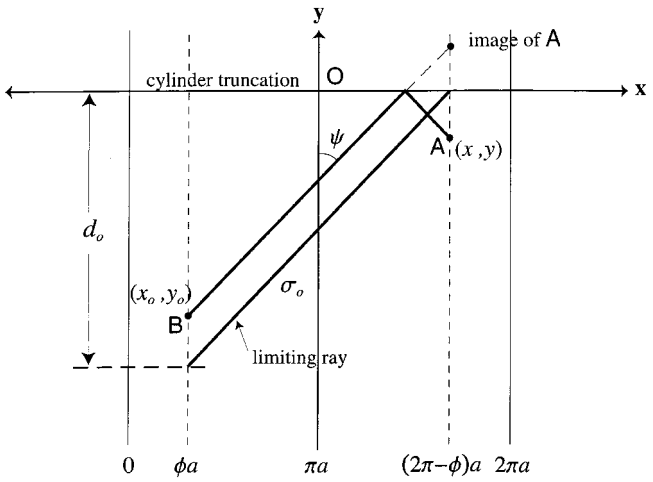


FIG. 3. Diagram of helical rays on the unwrapped plane surface of the cylinder. The y axis is parallel to the cylinder's axis. The lowest order leaky helical rays of helix angle ψ as shown are launched along the line $x=\phi a$ and travel once around the backside of the cylinder before reradiating in the backscattering direction at $x=(2\pi-\phi)a$. Only those rays launched within a distance d_o from the cylinder truncation reflect off the truncation and therefore reradiate in the backscattering direction at $x=(2\pi-\phi)a$ for helical rays of lowest order. For higher-order (nth-order) helical rays this distance is d_n. The images of the reradiation points reflected through the truncation line show that all rays of the same order have the same path length.

off the cylinder truncation. This limits the y coordinate of the launch points of relevant helical rays to lie within a distance d_o from the cylinder truncation at $y=0$. The value for d_o may be defined by the limiting ray which is launched at a distance d_o from the truncation and reaches the cylinder truncation at the reradiation point. This may be interpreted as the effective aperture length of the cylinder for this backscattering process. The path length of this ray is $\sigma_o = (2\pi - 2\phi)a/\sin\psi$. The effective aperture distance is then $d_o = \sigma_o \cos\psi$. Ray BA then has the same path length σ_o. This can be shown by extending the helical ray beyond the cylinder truncation to the image of point A. Thus, all helical rays launched within a distance d_o of the truncation have the same path length σ_o to their respective reradiation points. All such rays reflect once from the cylinder truncation and reradiate in the backscattering direction at $x=(2\pi-\phi)a$, where all the reradiation points also lie within a distance d_o from the truncation by symmetry. The y coordinates of the detachment points are then related to their corresponding launching points by the equation $y=-(d_o+y_o)$.

The superscript (0) of the symbol $p_l^{(0)}$ indicates that this is only the lowest-order contribution to the leaky wave pressure at $x=(2\pi-\phi)a$. This contribution is due to the ray that travels once around the backside of the cylinder from $x=\phi a$ to $x=(2\pi-\phi)a$, such as ray BA in Fig. 3. However, a sufficiently long cylinder has a family of helical rays that make contributions $p_l^{(n)}$ to the radiated leaky wave pressure at $x=(2\pi-\phi)a$ as discussed in Ref. 1. These rays correspond to higher-order helical rays which make n complete circumnavigations of the cylinder. They are launched at azimuthal angles $\phi - 2\pi n$ (where $n=0,1,2,...$), each of which satisfy the equation $\cos\phi = \cos\theta_l/\cos\gamma$ and propagate with the same helix angle ψ. The nth-order contribution $p_l^{(n)}$ can

be written in the form of Eq. (1) with σ_o replaced by σ_n where

$$\sigma_n = \sigma_o + 2\pi n a/\sin\psi \tag{2}$$

is the path length of the nth-order helical ray.

In addition, the relevant effective aperture length is different for each order of helical ray. Generalizing the result $d_o = \sigma_o \cos\psi$ to helical rays of any order n one obtains

$$d_n = \sigma_n \cos\psi. \tag{3}$$

It follows that the y coordinates for the detachment points are then given by equation $y = -(d_n + y_o)$. It is clear that for sufficiently large n, the aperture length d_n will become longer than the length of the cylinder and resulting helical ray paths will involve one or more reflections off both of the cylinder's truncations. Such rays have backscattering contributions that are highly attenuated (for the type of wave considered) due to their longer path lengths. These contributions involving multiple reverberations of waves across the length of the cylinder can be ignored assuming the cylinder is sufficiently long. The finite cylinder may then be treated a semi-infinite cylinder with only one truncation. In terms of the specific cylinder studied, only the first few orders of helical rays make appreciable contributions.

The far-field backscattered pressure $p_{\mathrm{bs}}^{(n)}$ due to the n-order helical ray contribution is calculated by evaluating a Rayleigh–Sommerfeld propagation integral in the exit plane,

$$p_{\mathrm{bs}}^{(n)} = (-ik\cos\gamma/2\pi r)\int\int_{-\infty}^{\infty} dx'\,dy'\,p^{(n)}(x',y')e^{ikr'}, \tag{4}$$

which is an integral over the pressure $p^{(n)}(x,y)$ in the exit plane due to that contribution. This pressure in the exit plane may be approximated using ray theory as will be explained subsequently. Equation (4) is a simplified version of the formula for far-field scattering by finite objects. In this expression, $r = (x^2+y^2+z^2)^{1/2}$ is the distance from the origin (point O in Fig. 2) to the backscattering observation point, and $r' = [(x-x')^2+(y-y')^2+z^2]^{1/2}$ is the distance from the observation point to a source point in the exit plane. From Fig. 2, we can infer that $y = -r\sin\gamma$, $x = 0$, $z = r\cos\gamma$. Inserting this into our expression for r' gives $r' \approx r + y'\sin\gamma$ for large r.

The outgoing radiated pressure in the exit plane $p^{(n)}(x,y)$ needs to be determined. Figure 4 shows a ray diagram of the helical ray scattering process as projected onto a base plane perpendicular to the cylinder's axis. The exit plane is perpendicular to the base plane, intersecting it along the gray line as shown in the figure. An incident ray of magnitude $k\cos\gamma$ is shown launching a helical ray on the cylinder at point A at azimuthal angle ϕ. The helical ray travels around the backside of the cylinder at least once before reradiating in the backscattering direction at azimuthal angle $-\phi$ (point B). The outgoing wavefront appears to diverge from a virtual focus at point F, such that its radius of curvature at point D in the exit plane is a, the radius of the cylinder. The x coordinate of point D is $-a\sin\phi$ as can be seen in the figure. Thus, the outgoing pressure *in the exit plane* at this value of x should be

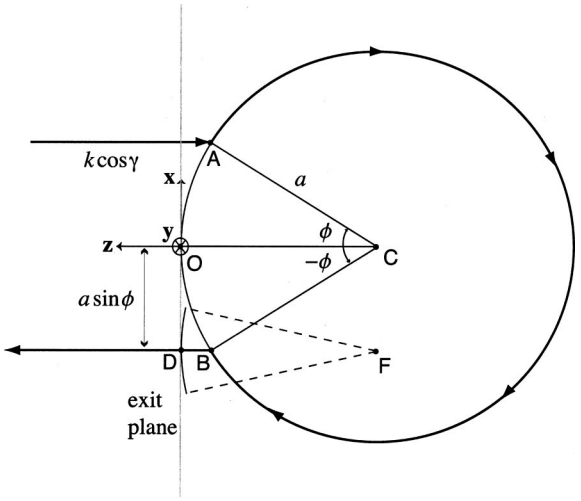


FIG. 4. Ray diagram for the projections of rays onto the base plane of the cylinder. The exit plane is perpendicular to the base plane, intersecting it along the gray line shown in the figure. An incident ray of projected magnitude $k\cos\gamma$ launches a helical ray on the cylinder at point A at azimuthal angle ϕ. The helical ray travels around the backside of the cylinder at least once before reradiating in the backscattering direction at azimuthal angle $-\phi$ (point B). The outgoing wavefront appears to diverge from a virtual focus at point F, such that its radius of curvature at point D in the exit plane is a, the radius of the cylinder.

$$p^{(n)}(x = -a\sin\phi, y) = |\cos\phi|^{1/2} p_l^{(n)}(y)\exp[ika\cos\gamma(1-\cos\phi)], \quad -d_n \leqslant y \leqslant 0, \tag{5}$$

where it is indicated explicitly that the outgoing radiated leaky wave pressure $p_l^{(n)}$ due to the nth-order helical ray depends on y, and the relevant region of y where $p^{(n)}(x,y)$ is nonzero is restricted to the effective aperture. The phase factor gives the phase change along the outgoing ray path in going from the detachment point to the point $(x = -a\sin\phi, y)$ in the exit plane. The factor $|\cos\phi|^{1/2}$ takes into account the wavefront spreading along the same piece of the ray path. The pressure in the exit plane along the x axis around $x = -a\sin\phi$ is obtained by multiplying Eq. (5) by an appropriate factor which takes into account the change in the phase as a function of x. Since the radius of curvature of the wavefront around $x = -a\sin\phi$ is a, this factor is $\exp[ik\cos\gamma(x + a\sin\phi)^2/2a]$.

Inserting Eq. (5) into Eq. (4) and integrating yields the desired result. The quantity $-(y + d_n)$ may be substituted for y_o in our expression for $p_l^{(n)}$. This yields a phase term for $p_l^{(n)}$ proportional to y which cancels out with another such term in the phase of the integrand of Eq. (4) when the approximation $r' \approx r + y'\sin\gamma$ is employed. The integration over y' in Eq. (4) then becomes trivial, simply producing a factor of d_n. The integral over x' may be performed using stationary phase methods. The result is the far-field backscattered pressure contribution due to the nth-order helical ray, $p_{\mathrm{bs}}^{(n)}$. For finite scatterers, it is usually convenient to express this pressure in terms of a dimensionless form function, f, defined by the following expression:

$$p_{\mathrm{bs}} = p_o(a/2r) f e^{ikr}. \tag{6}$$

Under this definition for the form function, f is unimodular for the scattering of sound rays by a fixed-rigid sphere having the same radius as that of the cylinder. The form function, f_n, due the nth-order helical wave backscattering contribution can be simplified to

$$f_n = -\left(\frac{8\alpha e^{i\varphi_{bl}}}{\sin\psi}\right) B d_n \exp(ik_p\sigma_n - 2i\eta - ikd_n \sin\gamma). \tag{7}$$

A factor of 2 is included in the final expression to account for an identical scattering contribution due to rays that circumnavigate the cylinder with the opposite helicity (helix angle $-\psi$). Note also that B should not depend on n, the order of helical rays, or the helicity, as all helical rays reflect off the cylinder truncation with the same angle of reflection, ψ, for any given tilt angle γ. In the ray model, it is assumed that $|B|=1$ because the frequency considered is far below the onset of next antisymmetric propagating Lamb wave. (See Sec. V.) The total leaky helical wave contribution to the far-field backscattering form function, f_{lh}, is given by a sum of such terms of the form shown in Eq. (7):

$$f_{lh} = \sum_n f_n. \tag{8}$$

The sum is over n, the helical ray ordering number. In computations, this sum is truncated when the condition $\exp(-\alpha\sigma_n) < 10^{-6}$ is satisfied. Only the first few terms in the sum will make appreciable contributions to f.

One may arrive at an expression for the total leaky ray backscattering form function, f_{lt}, by including the meridional ray contribution, f_{lm}. The analogous ray theoretic form function for the meridional wave contribution from finite cylinders was previously derived and is given below for the situation where the attenuation rate $\alpha \gg 1/L$. The steady-state backscattering contribution for a finite cylinder due to leaky waves of class l is given by

$$f_{lt} = f_{lh} + f_{lm}. \tag{9}$$

The meridional contribution is[12,14]

$$f_{lm} \approx -e^{i\pi/4} e^{i\varphi_{bl}} [2B(k_l/\alpha)^{1/2}] F_e, \tag{10}$$

$$F_e = (\mu')^{-1/2} F_\rho e^{\mu''} \int_{\mu''}^{\infty} \mathrm{erfc}(\sqrt{w})\, dw, \tag{11}$$

where $\mu' = (\cos\theta_l/\cos\gamma)[1 + i(k\sin\gamma - k_l)/\alpha]$, $F_\rho(\mu) = \{[\mu/F(\mu)] - \mu + \frac{1}{2}\}^{-1/2}$, $\mu = \alpha a \tan\theta_l$, $F(\mu) = (\pi\mu)^{1/2} \times e^{\mu}\,\mathrm{erfc}(\mu^{1/2})$, and $\mu'' = \mu\cos\theta_l/\cos\gamma$. The integral in Eq. (11) is numerically evaluated where erfc denotes the complementary error function. Since the phase φ_{bl} also appears in Eq. (7), it is not needed in the evaluation of $|f_{lt}|$.

III. EXPERIMENTS USING SHORT BACKSCATTERED TONE BURSTS: AIR-FILLED SHELL

Experiments were performed on a stainless-steel cylindrical shell of type SS304 which had a density of 7.57 g/cm^3, a bulk longitudinal speed of 5657 m/s, and a bulk shear speed of 3141 m/s. The cylindrical target had a radius of $a = 2.102$ cm and length $L = 24.505$ cm. This gives a length to radius ratio of $L/a = 11.66$ for the cylinder. This was large enough to justify ignoring length effects. The thickness of the shell, h, is given by the ratio $h/a = 0.1625$, a moderate thickness. This was the same thickness used in one of the computational tests of a related ray theory for scattering in the specular direction by an infinitely long evacuated steel cylindrical shell.[1] This paper discusses experiments done with air-filled targets. A plastic endcap was put on each end of the cylinder and rubber O-rings used as a seal to prevent water from leaking into the shell interior. The experiments were done in a large water-filled redwood tank, where the speed of sound in water was 1480 m/s. The target was placed about 285 cm away from the transducer which was used to both produce the incident sound and receive scattered signals. The experiment was performed in a transmit–receive (TR) mode which was controlled by a TR device. The transducer was a four panel 1-3 composite (Material Systems, Inc.), although only one of the panel transducers was used in the experiments.

TABLE I. Fourier coefficients of flexural wave parameters for $ka = 20$.

j	0	1	2	3	4
$k_{jl}a$	13.3167	0.4808	−0.0400	−0.0057	−0.0011
$\alpha_{jl}a$	0.215 37	0.011 91	0.002 52	0.000 27	0.000 10

It is the a_0 (antisymmetric) class of leaky Lamb waves which makes the dominant backscattering contributions for steel shells in the frequency range of interest in the experiments (about 200 kHz). The values of the a_0 helical leaky wave parameters were calculated using the method of Sec. IV of Ref. 1 for an infinitely long evacuated steel shell. The parameters depend on the helix angle ψ shown in Fig. 2 where the dependence is approximated by a truncated Fourier series

$$k_l(\psi) = k_{0l} + \sum_{j=1}^{4} k_{jl}\cos(2j\psi). \tag{12}$$

A similar expansion is used for the damping α, where the coefficients are denoted by α_{jl}. Table I lists the products of the coefficients with the shell radius when $ka = 20$. The resulting leaky wave numbers k_l and corresponding damping, α, are then substituted into Eq. (7). The a_0 meridional ray parameters are given by taking $\psi = 0$. These values are used in the meridional part of the total form function, f_{lm}. The interior is a good approximation to a vacuum for the purposes of this experiment.

From the analysis of the arrival times of backscattering contributions, it was decided to first study the discrete helical wave backscattering arrivals from the cylindrical shell target in the tilt angle range 10 degrees $< \gamma <$ 30 degrees. Using sufficiently short tone bursts, it was frequently possible to temporally separate the individual helical wave returns and also separate them from the edge diffraction contributions and the meridional contribution. A backscattering experiment using tone bursts of five cycles was performed at $ka = 20$ (frequency = 224 kHz). A waterfall plot of the backscattering records taken in the angular range 10 degrees $< \gamma <$ 30 degrees at intervals of 0.5 degrees is shown in Fig. 5. The horizontal axis is time in microseconds and measured from

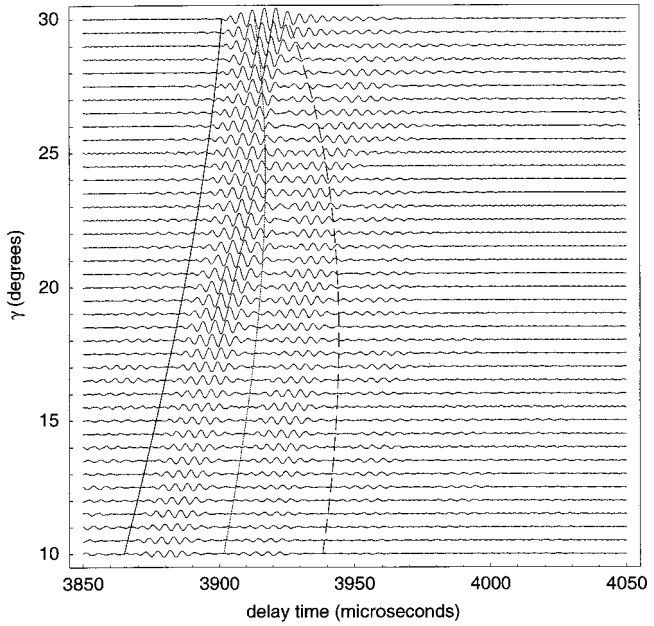


FIG. 5. Time records of backscattered five-cycle tone bursts from the cylindrical shell target cascaded as a function of the tilt angle at $ka=20$ (frequency 224 kHz). The delay time is referenced to the leading edge of the incident burst at the time it was first sent out from the transducer. Theoretical arrival times for the first three helical waves given by the solid, gray, and dashed curves coincide with the arrival of three discernible returns of about five cycles at sufficiently small tilt angle as seen in the records.

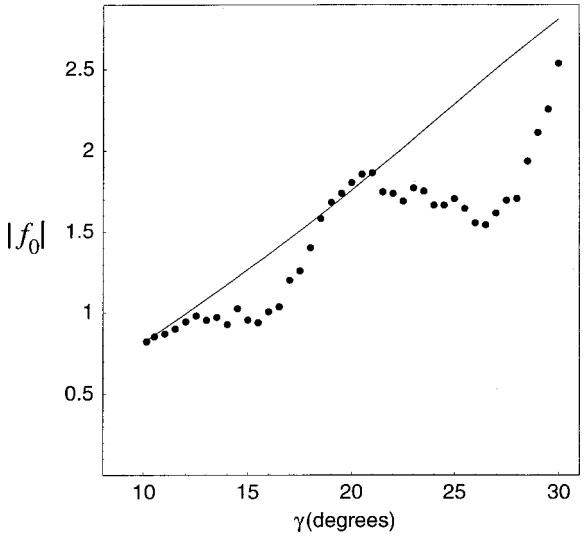


FIG. 6. The magnitude of the backscattering form function as a function of the tilt angle for the finite air-filled steel (SS304) cylindrical shell target due to the first helical wave arrival alone. The solid curve is the ray theory given by Eq. (7) for $n=0$. The solid points are the experimental results.

the leading edge of the incident burst at the time it was first sent out from the transducer. The theoretical curves for the helical wave arrival times are also plotted. These curves were computed by using the leaky wave phase velocity to determine the coupling locations and the group velocity to determine the helical ray contribution to the delay. The delay time is referenced to the time that the incident burst was sent from the transducer and the relationship between the timing and the geometry was determined by recording the echo for broad-side illumination. The solid curve, gray curve, and dashed curve correspond to the first, second, and third helical rays, respectively. The arrival of three tone bursts at these predicted arrival times may be identified as the first three helical wave returns in the backscattering records. The records show that only the first few helical wave returns are of significant amplitude. Each successive helical wave arrival is weaker in amplitude than the previous, as expected. The different helical arrivals are fairly well separated in time from each other using tone bursts of five cycles for tilt angles below about 25 degrees. For larger tilt angles, the helical waves begin to interfere with each other. A weaker burst preceding the helical arrival is associated with the edge diffraction return that has the same time delay as a meridional ray return, which will be referred to as the ''mr'' return.

The first helical wave arrival is largest in magnitude, so it was used in a comparison with the ray theory. It was observed that the five-cycle tone burst was sufficiently long for the first helical wave arrival to have reached a steady-state amplitude. The maximum in the amplitude of the burst was used to calculate an experimental form function magnitude for the first helical wave arrival. This was the maximum amplitude taken in a certain time window: from the theoretical first helical wave arrival time plus six cycles. The maximum amplitude is the maximum of the envelope of the received signal, calculated by constructing the analytic signal. One obtains similar results if one uses the maximum of the received signal itself in this time window instead of the maximum of the envelop. The backscattered specular reflections from a calibration sphere were used to convert the voltage amplitudes of the received signal to a form function magnitude as discussed in Appendix B of Ref. 14. The steady-state form function magnitude for the first helical wave return is $|f_n(ka,\gamma)|$ where $f_n(ka,\gamma)$ is given by Eq. (7), $n=0$ for the first helical wave arrival, and ka is set to 20. Note that the *second* helical wave is given by $n=1$ and so forth.

The theoretical (solid curve) and experimental (solid points) form function magnitudes are plotted in Fig. 6 as a function of tilt angle. Since only one helical ray arrival is included in the ray synthesis, the theoretical form function does not exhibit the regular interference structure evident in the computational tests of the ray theory in Ref. 1. The solid points roughly follow the theoretical curve but they deviate somewhat below the theoretical curve in a region centered around $\gamma=15$ degrees and beyond 20-degree tilt. Inspection of the time domain records (Fig. 5) suggests that the deviations occur because the tone burst signal of the first helical wave has not been adequately temporally isolated from other scattering contributions. For γ greater than about 20 degrees, and certainly for γ greater than 25 degrees, the first helical wave is interfering with subsequent helical waves. The ray theoretic form function of the first helical wave plotted in Fig. 6 from Eq. (7) with $n=0$ neglects the interference with other contributions. For γ around 15 degrees, the first helical wave is most likely interfering with an mr corner return. Specifically, the s_0 (symmetric compressional) leaky Lamb meridional wave may be interfering with the first helical wave return, as its arrival time should also be given by the mr corner timing. As can be seen in Fig. 5, there is a particu-

larly noticeable signal at around or just past $\gamma=15$ degrees as opposed to other angles at about the mr timing, which at this tilt occurs just a few cycles ahead of the first helical wave. This is probably the compressional meridional contribution superposed with the mr edge diffraction return. The importance of the measurements shown in Fig. 6 is that the $n=0$ helical contribution has an appreciable magnitude over a range of tilt angles.

As previously noted, the reductions in amplitude near γ of 15 and 25 degrees appear to be the result of interference by contributions not included in the prediction for the $n=0$ contribution. It is plausible that for some other values of ka, such localized interfering contributions may result in constructive interference instead of the destructive interference evident in Fig. 6.

IV. SUPERPOSITION OF HELICAL WAVE ARRIVALS: AIR-FILLED SHELL

The steady-state backscattering form function associated with the total contribution of leaky helical waves (the a_0 leaky waves) is given by Eq. (9), which consists of a superposition of all helical wave contributions of order n and the meridional contribution. It was of interest to test the ray theory for the total leaky wave contribution for a finite cylindrical shell against the results of experiment. To do this, longer tone bursts were used to superpose all of the relevant helical wave arrivals. The results are described in this section for the air-filled shell. The water sound speed in this experiment was 1483 m/s.

Only the first few helical wave arrivals were important. Higher-order helical waves are sufficiently damped out that they can be neglected in the experiment. It was decided that the first four helical wave contributions would be superposed, and they would have to be superposed over a duration of several cycles to ensure that a steady state signal is achieved. Consider, for example, the experiments performed at $ka=20$. By studying the arrival times of various scattering contributions, it was concluded that a tone burst length of about 150 μs would be sufficiently long. This is long enough to superpose the first four helical waves for a few cycles at about 5-degree tilt, the tilt where the helical waves are the most temporally separated in the experiments. Tilt angles less than 5 degrees were not investigated. A 34-cycle tone burst was used at $ka=20$, which corresponds to a frequency of about 224 kHz. At 34 cycles, this actually translates into a burst length of about 151 μs long.

A backscattering experiment using 34-cycle tone bursts was performed with the cylindrical shell target at $ka=20$ and at every 0.5 degrees in the tilt angle range 5 degrees$<\gamma<$47 degrees. The experiment was taken out to 47-degree tilt angle in order to capture the meridional leaky wave contribution as well. Figure 7 shows selected backscattering records from the experiment. A backscattered record is shown at every interval of 2.5 degrees between 5 and 42.5 degrees. As in Fig. 5, the horizontal axis is time in microseconds and measured from the time the incident burst was sent out from the transducer. Corresponding to each record, a horizontal line segment is shown which outlines the average envelope of the signal over a certain time window. The value

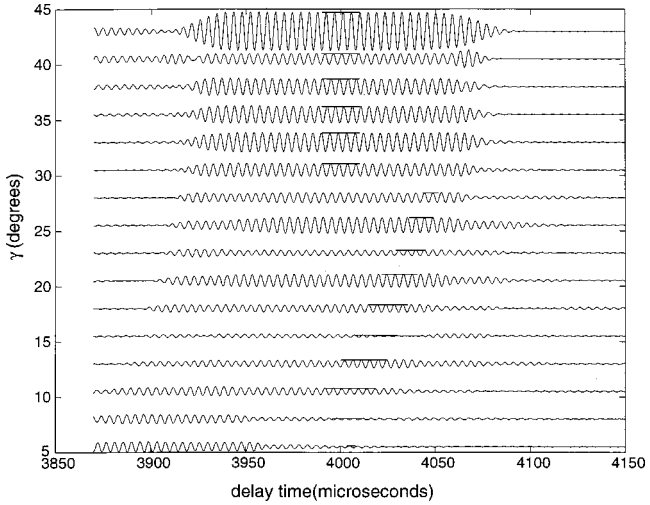


FIG. 7. Time records of backscattered 34-cycle tone bursts for the air-filled cylindrical shell target cascaded as a function of tilt angle at $ka=20$ (frequency 224 kHz). Only those records at 2.5-degree intervals starting at 5 degrees are shown. The delay time is referenced to the time that the incident burst was sent out from the transducer. The short horizontal lines outline the average envelope of the burst over a certain time window wherein the first few helical wave returns and, at larger tilt angles, the meridional wave return, were superposed. This measured signal average was then converted into an experimental form function magnitude.

of this average envelope in terms of voltage was converted to a steady-state form function magnitude according to the procedure discussed in Appendix B of Ref. 14. This was used as the experimental form function magnitude.

The position and length of the time window changes depending on the tilt angle. For tilt angles less than 30 degrees, the time window chosen was such that it spans the region over which the first four helical wave returns interfere. Thus, the amplitude of the signal in this region should approximate the steady-state leaky helical wave form function given by Eq. (8). The trailing edge of the window coincides with the trailing edge of the tone burst signal of the first helical wave. The leading edge of the window depends on the arrivals of signals due to the diffracted rays from the corners. Using this choice of window, we attempted to gate out the extraneous arrivals from the cylinder corners in the data analysis since these were not of interest in the experiment. Near the critical tilt angle, it is desirable to include that part of the tone burst signal around the mr arrival time. In this way, the leaky meridional wave signal is included in the experimental analysis and one can arrive at a true experimental backscattering form function due to the *total* a_0 leaky wave contribution. For $\gamma>30$ degrees, the time window is fixed at 20-μs duration, and starts at the delay time 3990 μs and ends at 4010 μs. It was estimated that during this time window, the meridional wave signal reaches a steady state, and any helical wave signals that are still important at these large tilt angles are also present within this time window.

Figure 1 shows plots of backscattering form function magnitudes as a function of the tilt angle. The experimental data is given by the solid points. The solid curve is an approximate numerical partial wave series (PWS) solution.[4] The gray curve gives the ray theoretic form function magnitude due to the a_0 leaky wave contributions, determined us-

ing Eq. (9). Results are shown for $ka=20$. There is a strong peak in the scattering at $\gamma=\theta_l$ ($\gamma=43$ degrees) corresponding to the meridional leaky wave contribution. The broad oscillations at smaller tilt angles are due to the interference of helical wave arrivals with each other and with the meridional contribution. The data displays this interference structure predicted by the ray theory and the partial-wave theory, as well the large peak at $\gamma=\theta_l$ due to the meridional ray. The importance of including several orders of helical rays when analyzing the backscattering is confirmed.

Note that the ray theory is only meant to model the a_0 leaky wave contributions. It is not meant to reproduce the large signal in the approximate PWS solution near $\gamma=0$ associated with broadside specular reflection. It is also not meant to reproduce the large but narrow peaks for $\gamma<30$ degrees which correspond to weakly damped torsional or compression waves as mentioned in the Introduction. Ignoring these scattering features, it appears that the ray theory compares well to the approximate PWS solution for the meridional peak and the helical wave peaks with $\gamma<35$ degrees. Admittedly, there is less agreement in the tilt angle range 30 degrees$<\gamma<$40 degrees. This discrepancy has not been resolved.

The important result is that the experimental data reproduce, at least somewhat quantitatively, the a_0 leaky Lamb wave features in the scattering form function as was designed. There is agreement between experiment and theory both in general magnitude and in the locations of the interference peaks. It should be pointed out once more that the experimental form function magnitudes are about unity or greater, even at large tilt angles, which means that the backscattering is significant. The oscillations in form function magnitude in the tilt angle range 30 degrees$<\gamma<$40 degrees are small compared with the theoretical curves. This is probably because the projected aperture length, $d_n \cos\gamma$, for all helical wave arrivals is too long for the phase and amplitude of the incident wave to be uniform. This is simply a consequence of the fact that the source is a finite distance R away. The Fresnel width $(R\lambda)^{1/2}$, where λ is the wavelength, only slightly exceeds $d_n \cos\gamma$ for some of the significant helical contributions. The experimental magnitudes also underestimate the theoretical curves at the meridional ray peak since the experimental setup was not optimized to take data for the meridional contribution.

As for the ray theory, the experimental data was not meant to reproduce the large, narrow peaks in the approximate PWS form function for $\gamma<30$ degrees. For example, the large peak at $\gamma\approx15$ degrees is most likely due to the meridional leaky wave peak corresponding to the compressional wave in the shell, as discussed in the previous section. It is not surprising that this feature does not appear in the experimental data since signals that arrived at the mr timing were purposely gated out at this tilt angle, where mr denotes the arrival time of meridional wave contributions. Furthermore, the torsional and compressional wave contributions, as mentioned before, have a very weak damping parameter, α, so that a tone burst much longer in duration is needed for those contributions to build up to a steady state signal. Even the 34-cycle tone bursts at $ka=20$ are probably insufficiently long for the shear and compressional contributions to reach a steady state amplitude that could reproduce the large, narrow peaks in the steady state form function that appear in the approximate PWS solution.

Measurements following a procedure similar to the one used for Fig. 1 were also taken for $ka=15$ at a frequency of 168.7 kHz. The location of the meridional ray peak is increased to $\gamma=48$ degrees with a measured peak magnitude of $|f|=6.4$. The location and magnitude of the peak are both in excellent agreement with ray theory. When γ is decreased below 48 degrees, ray theory predicts that the helical ray oscillations are broader than the corresponding oscillations in Fig. 1. The observations agree with this prediction and with the general magnitude and location of the predicted oscillations.

V. DISCUSSION

These experiments on steel cylindrical shells have shown that at high frequencies, a_0 leaky Lamb waves give the dominant backscattering contributions and that the ray methods developed here are useful for the modeling and interpretation of these scattering responses from shell structures. The focus in this work was on helical wave contributions. Though such contributions are not as large in magnitude as those of the previously studied meridional waves, they do yield form function magnitudes of unity or greater. A ray model for these backscattering contributions was developed and was in agreement with the experimental data as is evident in Fig. 1 at $ka=20$, showing that the ray theory is not restricted to very high frequencies. This frequency was well above the coincidence frequency for the shell, which is about ka of 7. Experiments were done at higher frequencies but are not shown here. The results were similar except that the magnitudes of the backscattering contributions were significantly diminished relative to the predictions for all tilt angles when ka was increased to 30. This is apparently a consequence of the breakdown of the assumption $|B|=1$ in the ray theory and in the partial-wave model, where B is the reflection coefficient for leaky helical rays at the cylinder truncation. Energy may be lost upon reflection due to mode conversion to higher-order modes of the same symmetry[15] (the threshold for the propagation of the a_1 mode occurs at $ka=41$), or some energy may be lost to the end caps. It is to be expected that even without the complication of end caps that the reflection becomes partial below the propagation threshold for the a_1 wave. That is a consequence of radiation into the water resulting from the evanescent elastic field near the end of the cylinder that is present for $ka<41$.[16]

The comparisons with ray theory displayed in Figs. 1 and 6 give support to the formulation developed in Refs. 12–14 for approximating elastic contributions to the scattering. It is noteworthy that computations.[17,18] of the scattering by buried spherical targets suggest that flexural waves remain relevant even for buried targets. Consequently, the present ray formulation may also find application to the scattering by buried targets.

The present investigation emphasizes features in the scattering by bluntly truncated shells that differ from the fea-

tures emphasized investigations carried out at relatively low frequencies.[5,6,8,19,20] Uberall *et al.*[19] predict that wave vector coupling conditions, along with a condition for reinforcement of the helical wave following repeated reflections from the truncations, constrain both ka and the tilt angle γ if there is to be strong acoustic coupling to weakly damped helical leaky modes. (See also Refs. 5 and 20.) The class of modes studied here, however, was sufficiently damped that reflections from the opposite ends of the cylinder were unimportant. Consequently, the condition that the frequency correspond to a natural frequency of the fluid-loaded cylinder is no longer present.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research.

[1] F. J. Blonigen and P. L. Marston, ''Leaky helical flexural wave scattering contributions from tilted cylindrical shells: Ray theory and wave-vector anisotropy,'' J. Acoust. Soc. Am. **110**, 1764–1769 (2001).

[2] G. Kaduchak, C. M. Wassmuth, and C. M. Loeffler, ''Elastic wave contributions in high-resolution acoustic images of fluid-filled, finite cylindrical shells in water,'' J. Acoust. Soc. Am. **100**, 64–71 (1996).

[3] S. F. Morse and P. L. Marston, ''Meridional ray contributions to scattering by tilted cylindrical shells above the coincidence frequency: Ray theory and computations,'' J. Acoust. Soc. Am. **106**, 2595–2600 (1999).

[4] S. F. Morse, P. L. Marston, and G. Kaduchak, ''High-frequency backscattering enhancements by thick finite cylindrical shells in water at oblique incidence: Experiments, interpretation, and calculations,'' J. Acoust. Soc. Am. **103**, 785–794 (1998).

[5] 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).

[6] L. B. Felsen, J. M. Ho, and I. T. Lu, ''Three dimensional Green's function for fluid- loaded thin elastic cylindrical shell: Alternative representations and ray acoustic forms,'' J. Acoust. Soc. Am. **87**, 554–569 (1990); A. N. Norris and D. A. Rebinsky, ''Acoustic coupling to membrane waves on elastic shells,'' *ibid.* **95**, 1809–1829 (1994); D. A. Rebinsky and A. N. Norris, ''Benchmarking an acoustic coupling theory for elastic shells of arbitrary shape,'' *ibid.* **98**, 2368–2371 (1995).

[7] F. Léon, F. Lecroq, D. Décultot, and G. Maze, ''Scattering of an obliquely incident acoustic wave by an infinite hollow cylindrical shell,'' J. Acoust. Soc. Am. **91**, 1388–1397 (1992).

[8] A. Tesei, W. L. J. Fox, A. Maguer, and A. Lovik, ''Target parameter estimation using resonance scattering analysis applied to air-filled, cylindrical shells in water,'' J. Acoust. Soc. Am. **108**, 2891–2900 (2000); L. Haumesser, A. Baillard, D. Décultot, and G. Maze, ''Behavior of first guided wave on finite cylindrical shells of various lengths: Experimental investigation,'' *ibid.* **109**, 583–590 (2001); N. Touraine, L. Haumesser, D. Décultot, G. Maze, A. Klauson, and J. Metsaveer, ''Analysis of the acoustic scattering at variable incidences from an extra thin cylindrical shell bounded by hemispherical endcaps,'' *ibid.* **108**, 2187–2196 (2000).

[9] V. Murino and A. Trucco, ''Three-dimensional image generation and processing in underwater acoustic vision,'' Proc. IEEE **88**, 1903–1946 (2000).

[10] P. T. Gough and D. W. Hawkins, ''Imaging algorithms for a strip-map synthetic aperture sonar: Minimizing the effects of aperture errors and aperture undersampling,'' IEEE J. Ocean. Eng. **22**, 27–39 (1997).

[11] T. K. Stanton, ''Sound scattering by cylinders of finite length. I. Fluid cylinders,'' J. Acoust. Soc. Am. **83**, 55–63 (1988).

[12] P. L. Marston, ''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); erratum *ibid.* **103**, 2236 (1998).

[13] P. L. Marston, ''Spatial approximation of leaky wave surface amplitudes for three-dimensional high-frequency scattering: Fresnel patches and application to edge-excited and regular helical waves on cylinders,'' J. Acoust. Soc. Am. **102**, 1628–1638 (1997).

[14] K. Gipson and P. L. Marston, ''Backscattering enhancements due to reflection of meridional leaky Rayleigh waves at the blunt truncation of a tilted solid cylinder in water: Observations and theory,'' J. Acoust. Soc. Am. **106**, 1673–1680 (1999).

[15] S. F. Morse and P. L. Marston, ''Degradation of meridional ray backscattering enhancements for tilted cylinders by mode conversion: wideband observations using a chirped PVDF sheet source,'' IEEE J. Ocean. Eng. **26**, 152–155 (2001).

[16] S. F. Morse, ''High frequency acoustic backscattering enhancements for finite cylindrical shells in water at oblique incidence,'' Ph.D. Dissertation, Washington State University, Department of Physics, 1998; S. F. Morse and P. L. Marston, ''Meridonal ray backscattering enhancements for empty truncated tilted cylindrical shells: Measurements, ray model and effects of a mode threshold'' (submitted for publication).

[17] R. Lim, K. L. Williams, and E. I. Thorsos, ''Acoustic scattering by a three-dimensional elastic object near a rough surface,'' J. Acoust. Soc. Am. **107**, 1246–1262 (2000).

[18] R. Lim, J. L. Lopes, R. H. Hackman, and D. G. Todoroff, ''Scattering by objects buried in underwater sediments: Theory and experiment,'' J. Acoust. Soc. Am. **93**, 1762–1783 (1993).

[19] H. Uberall, P. J. Moser, B. L. Merchant, A. Nagl, K. B. Yoo, S. H. Brown, J. W. Dickey, and J. M. D'Archangelo, ''Complex acoustic and electromagnetic resonance frequencies of prolate spheroids and related elongated objects and their physical interpretation,'' J. Appl. Phys. **58**, 2109–2124 (1985).

[20] F. Lecroq, G. Maze, D. Décultot, and J. Ripoche, ''Acoustic scattering from an air-filled cylindrical shell with welded flat plate endcaps: Experimental and theoretical study,'' J. Acoust. Soc. Am. **95**, 762–769 (1993).

The propagation of sound in narrow street canyons

K. K. Iu and K. M. Li[a)]
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

(Received 22 October 2001; revised 29 April 2002; accepted 20 May 2002)

This paper addresses an important problem of predicting sound propagation in narrow street canyons with width less than 10 m, which are commonly found in a built-up urban district. Major noise sources are, for example, air conditioners installed on building facades and powered mechanical equipment for repair and construction work. Interference effects due to multiple reflections from building facades and ground surfaces are important contributions in these complex environments. Although the studies of sound transmission in urban areas can be traced back to as early as the 1960s, the resulting mathematical and numerical models are still unable to predict sound fields accurately in city streets. This is understandable because sound propagation in city streets involves many intriguing phenomena such as reflections and scattering at the building facades, diffusion effects due to recessions and protrusions of building surfaces, geometric spreading, and atmospheric absorption. This paper describes the development of a numerical model for the prediction of sound fields in city streets. To simplify the problem, a typical city street is represented by two parallel reflecting walls and a flat impedance ground. The numerical model is based on a simple ray theory that takes account of multiple reflections from the building facades. The sound fields due to the point source and its images are summed coherently such that mutual interference effects between contributing rays can be included in the analysis. Indoor experiments are conducted in an anechoic chamber. Experimental data are compared with theoretical predictions to establish the validity and usefulness of this simple model. Outdoor experimental measurements have also been conducted to further validate the model.
[DOI: 10.1121/1.1492821]

PACS numbers: 43.50.Gf, 43.28.En, 43.50.Rq [MRS]

I. INTRODUCTION

The investigation of sound propagation outdoors has been the subject of extensive research since the 1950s. Piercy *et al.*[1] reviewed the development prior to 1977. Embleton[2] has summarized many intriguing features of outdoor sound by emphasizing field measurements and their physical interpretations. Most of these studies consider a relatively simple scenario—the source and receiver are located in the vicinity of an absorbing ground surface. However, due to the rapid urbanization of many countries and heightened environmental and other concerns in the past three decades, knowledge of the near-ground propagation of sound in complex urban environments is of great economic and social importance.

Indeed, theoretical studies of sound propagation in a city street were conducted in the 1970s by Schlatter,[3] Lee and Davies,[4] Lyon,[5] and Steenackers *et al.*[6] It was assumed in their analyses that the street was represented by a simple channel between two infinite-reflecting walls with image sources produced as a result of multiple reflections of sound. In their approach, attention was focused on the determination of a reverberant sound field in city streets. However, the possibilities of ground interference and further interference due to multiple reflections between two parallel walls were ignored in these early studies.

Oldham and Radwan[7] have proposed an alternative approach in which the sound levels in city streets are estimated by summing the sound intensities of direct and reflected sound. Kang[8] has compared the sound fields in street canyons with diffusely and geometrically reflecting boundaries. He has developed a radiosity-based theoretical model for diffusely reflecting boundaries. In his model, the sound propagation in the street can be simulated by energy exchange between the nodal points. The interference effects due to multiple reflections are expected to be less significant in the model. However, the radiosity approach is not included in the scope of our current study, but, rather, we prefer to study the interference effect of multiple reflections.

For geometrically reflecting boundaries, the energy response at the receiver is obtained by summing contributions from the source and all image sources. Some commercial software, such as Raynoise,[9] has implemented this idea to include the effect of wall/ground absorption in these models. However, the interference effects between the direct and reflected waves are again not included in their calculations. In addition, they have restricted their studies to the case where the width of the street canyon ranges from 10 to 40 m only. In some practical situations of a dense high-rise city, the width of many street alleys is often less than 10 m. This gives us a motivation to investigate the propagation of sound in this type of environment, which is referred as narrow street canyons in this paper.

We also note that Walerian *et al.*[10] have recently exploited a simple model for predicting the time-average sound level within an urban system. In their propagation model, a sound wave interacting with obstacles undergoes multiple

[a)]Author to whom correspondence should be addressed. Electronic mail: mmkmli@polyu.edu.hk

reflections from walls and diffractions from wedges. In the case of canyon streets, diffraction at the top edges of buildings has been omitted in their studies. Further, it is assumed that the total field is computed by summing the squared pressures of all image sources. As a result of using squared pressures in the summation, the interference effects are omitted in their analysis. Although Walerian *et al.* have pointed out that the interference effect is significant for a relatively narrow street canyon, no further investigations have been conducted.

In the late 1970s, Gensane and Santon[11] considered the wave nature of sound in their study for sound fields in bounded and arbitrarily shaped air spaces. Lemire and Nicolas[12] extended this approach by using a spherical wave reflection coefficient instead of a plane wave reflection coefficient to model sound wave propagation in a bounded space. In particular, they investigated the sound fields in a rectangular enclosure and the region bounded by two infinite parallel planes. Their predicted results agreed very well with the standard normal mode solutions, but no experimental validations were presented in their study. We also remark that Dance *et al.*[13] have pointed out the importance of interference effects in an enclosed space. By assuming perfectly reflecting boundaries and using an image-source method, they developed an interference model for calculating the total sound fields in an industrial space. However, their model was found to be more accurate near the vertical walls, floor, and ceiling, but less so at receivers located further from these reflecting surfaces. In view of these earlier studies, we endeavor to extend these models by considering theoretically and experimentally the effect of finite impedance on the overall sound fields in an enclosed space.

In this paper, we wish to investigate the propagation of sound in a narrow street such that the height of the buildings is much greater than the width of the street. It is expected that the interference effects caused by direct and reflected rays will play a significant role in the prediction of sound fields in such a street canyon. A brief review of other models is discussed also, but our principal aim is to establish a simple yet accurate model for predicting sound fields in narrow street canyons. The model is based on an analytic Green's function. We also conduct extensive experimental measurements both indoors and outdoors. Indoor experiments are conducted in an anechoic chamber and outdoor field measurements are carried out in a typical alley street. The theoretical predictions will be compared with the experimental measurements. The emphasis of this paper is the experimental verification of the proposed theoretical model.

II. PROPAGATION MODELS

A. Analytical formulations

In an earlier publication, Tang and Li[14] derived an expression for the sound field due to a harmonic source above an impedance ground in front of an absorbing facade. In this paper, we extend their work to consider the sound fields above an impedance ground between two parallel facade surfaces. This is the model we adopt to assess the sound propagation in a narrow city street. Essentially, the street is represented as two infinite vertical planes of normalized admittance, β_L and β_R, respectively. They are parallel, and located at the plane of $x=0$ for the left vertical wall and at the plane $x=W$ for the right vertical wall. An impedance ground of normalized admittance β_B, which is located between the vertical walls, is situated at the plane $z=0$. The impedance ground is assumed to be perpendicular to the vertical walls. The schematic diagram of the posed problem is shown in Fig. 1. We assume that a point source is placed at $(x_s,0,z_s)$, where the time-dependent factor, $e^{-i\omega t}$, is understood and suppressed throughout the analysis. As the sound field is symmetrical about the $y=0$ plane, we restrict our interest to the region where $x\in[0,W]$, $y\in[0,\infty)$, and $z\in[0,\infty)$. The sound field, $p(x,y,z)$, can be computed by solving the Helmholtz equation:

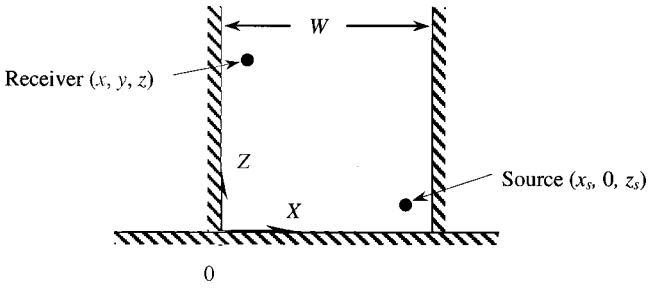


FIG. 1. Schematic diagram for sound propagation in a city street.

$$\nabla^2 p + k^2 p = -\delta(x-x_s)\,\delta(y)\,\delta(z-z_s), \tag{1}$$

where k is the wave number of the source. The governing equation is supplemented further by the boundary conditions of the parallel walls and the impedance ground as follows:

$$\frac{\partial p}{\partial x} + ik\beta_L p = 0 \quad \text{at } x=0, \tag{2a}$$

$$\frac{\partial p}{\partial x} - ik\beta_R p = 0 \quad \text{at } x=W, \tag{2b}$$

and

$$\frac{\partial p}{\partial z} + ik\beta_B p = 0 \quad \text{at } z=0. \tag{2c}$$

In the absence of the impedance ground, the solution can be expressed as a Fourier integral[15]

$$p = -\frac{1}{(2\pi)^3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{G_x \exp(ik_y y + ik_z|z-z_s|)}{(k^2-k_x^2-k_y^2-k_z^2)} \times dk_x\,dk_y\,dk_z, \tag{3}$$

taken over the space of all wavenumbers, $\mathbf{k}\equiv(k_x,k_y,k_z)$. The variable G_x in Eq. (3) may be regarded as the required Green's function in the wave number space given by

$$G_x=\begin{cases}\dfrac{[e^{-ik_x x_s}+V_L e^{ik_x x_s}][e^{ik_x x}+V_R e^{[ik_x(2W-x)]}]}{1-V_L V_R\exp(2ik_x W)} \\ \quad \text{for } x\geqslant x_s, \\ \dfrac{[e^{ik_x x_s}+V_L e^{ik_x x_s}][e^{-ik_x x}+V_R e^{[ik_x(2W-x)]}]}{1-V_L V_R\exp(2ik_x W)} \\ \quad \text{for } x<x_s,\end{cases} \tag{4}$$

where V_L and V_R are the reflection factors for the waves reflected by the left and right vertical walls, respectively. They can be determined according to

$$V_L=\frac{k_x-k\beta_L}{k_x+k\beta_L} \tag{5a}$$

and

$$V_R=\frac{k_x-k\beta_R}{k_x+k\beta_R}. \tag{5b}$$

Replacing the denominator by its binomial series, expanding the square brackets, and grouping all terms accordingly in Eq. (4), we obtain

$$G_x=\sum_{l=0}^{\infty}[e^{ik_x|x-x_s|}+V_L e^{ik_x(x+x_s)}+V_R e^{ik_x(2W-x-x_s)} + V_L V_R e^{ik_x(2W-x+x_s)}](V_L V_R e^{2ik_x W})^l. \tag{6}$$

First, let us consider the case when $l=0$ of the series in Eq. (6). The first term in the square bracket represents the direct wave. The second term in the square bracket is a result of the reflection from the left vertical wall. They are marked as S_{01} and S_{02} in Fig. 2. The last two terms are a pair of image sources, which are obtained by mirror reflection of the first two in the right vertical wall. They are marked as S_{03} and S_{04} in Fig. 2. The process can be repeated again and again (in which l ranges from 1 to ∞ in the series), leading to a row of an infinite number of image sources located at the height of z_s above the impedance ground. In the event that the waves are reflected by the impedance ground, we can apply the boundary condition (2c) in Eq. (3) to yield

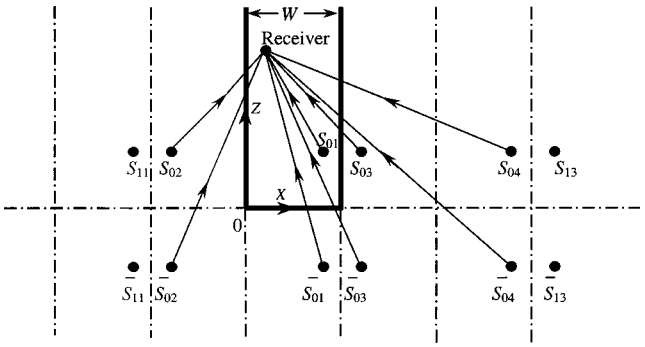


FIG. 2. Two rows of image sources are obtained by mirror reflection of the vertical walls and a reflecting ground. The source, S_{01}, is located at $(x_s,0,z_s)$ and the receiver at (x,y,z).

$$p=-\frac{1}{(2\pi)^3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{G_x e^{ik_y y}}{(k^2-k_x^2-k_y^2-k_z^2)} \times\{e^{ik_z|z-z_s|}+V_B e^{ik_z(z+z_s)}\}dk_x dk_y dk_z, \tag{7}$$

where V_B is the reflection factor of the waves from the ground, which can be expressed as

$$V_B=\frac{k_z-k\beta_B}{k_z+k\beta_B}. \tag{8}$$

The outer integral with respect to k_z can be evaluated by the method of contour integration[15] to give

$$p=\frac{i}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{G_x e^{ik_y y}}{2k_z^*} \times\{e^{ik_z^*|z-z_s|}+V_B^* e^{ik_z^*(z+z_s)}\}dk_x dk_y, \tag{9a}$$

where

$$k_z^*=+\sqrt{k^2-k_x^2-k_y^2} \tag{9b}$$

and

$$V_B^*=\frac{k_z^*-k\beta_B}{k_z^*+k\beta_B}. \tag{9c}$$

In this paper, the integral (9a) is used for the modeling of sound propagation in a street canyon. However, the exact evaluation of the integral is difficult if not impossible for all but the simplest situation. A well-known example is that all reflecting surfaces are acoustically hard where $\beta_B=\beta_L=\beta_R=0$. In this case, the reflection factors are all equal to 1, i.e., $V_L=V_R=V_B^*=1$. Substitution of Eq. (6) into (9a) leads to

$$p=\frac{i}{(2\pi)^2}\sum_{l=0}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{e^{ik_y y}e^{ik_z^*|z-z_s|}}{2k_z^*}[e^{ik_x|x-x_s|}+e^{ik_x(x+x_s)}+e^{ik_x(2W-x-x_s)}+e^{ik_x(2W-x+x_s)}](e^{2ik_x W})^l dk_x dk_y + \frac{i}{(2\pi)^2}\sum_{l=0}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{e^{ik_y y}e^{ik_z^*(z+z_s)}}{2k_z^*}[e^{ik_x|x-x_s|}+e^{ik_x(x+x_s)}+e^{ik_x(2W-x-x_s)}+e^{ik_x(2W-x+x_s)}](e^{2ik_x W})^l dk_x dk_y. \tag{10}$$

Obviously, the total sound field is composed of contributions from two rows of noise sources located at heights z_s and $-z_s$, respectively. Each of these integrals in Eq. (10) can be identified as a Sommerfeld integral with each source located at a different position. These integrals can be evaluated exactly (see, for example, Ref. 16), and the total sound field can be expressed as the sum of a series as follows:

$$p=\frac{1}{4\pi}\sum_{l=0}^{\infty}\left[\frac{e^{ikd_{l1}}}{d_{l1}}+\frac{e^{ikd_{l2}}}{d_{l2}}+\frac{e^{ikd_{l3}}}{d_{l3}}+\frac{e^{ikd_{l4}}}{d_{l4}}+\frac{e^{ik\bar{d}_{l1}}}{\bar{d}_{l1}}+\frac{e^{ik\bar{d}_{l2}}}{\bar{d}_{l2}}+\frac{e^{ik\bar{d}_{l3}}}{\bar{d}_{l3}}+\frac{e^{ik\bar{d}_{l4}}}{\bar{d}_{l4}}\right], \tag{11}$$

where the first four terms of the square bracket in the above series are the contributions due to the reflections from the vertical walls only. The path lengths, d_{l1}, d_{l2}, d_{l3}, and d_{l4} can be determined by simple geometrical considerations. The total sound field is also augmented by a set of image sources due to the presence of a reflecting ground. These are represented by the last four terms of the square bracket in Eq. (11). The corresponding path lengths are $\bar{d}_{l1}$, $\bar{d}_{l2}$, $\bar{d}_{l3}$, and $\bar{d}_{l4}$ that can also be determined straightforwardly.

The above theoretical formulation provides the basis for subsequent analyses. Two popular heuristic approximations, the incoherent and coherent models, will be described in the following sections.

B. A critical review of the incoherent model

Oldham and Radwan[7] have proposed a simple model for the prediction of sound propagated in a street canyon. The model is based on the construction of all image sources due to the reflections from the building facade on each side of the street. The total sound level is computed by summing the intensities of direct and all reflected sound levels incoherently. According to this model, the intensity of the sound field ($I\propto\overline{p^2}$) can be obtained from Eq. (11) by summing each contributing component incoherently to yield

$$I=I_{\mathrm{ref}}\,d_{\mathrm{ref}}^{\,2}\sum_{l=0}^{\infty}\left(\frac{1}{d_{l1}^2}+\frac{1}{d_{l2}^2}+\frac{1}{d_{l3}^2}+\frac{1}{d_{l4}^2}+\frac{1}{\bar{d}_{l1}^2}+\frac{1}{\bar{d}_{l2}^2}+\frac{1}{\bar{d}_{l3}^2}+\frac{1}{\bar{d}_{l4}^2}\right), \tag{12}$$

where I_{ref} is the reference sound intensity at a reference distance of d_{ref} from the source. An implicit assumption of this approach is that the mutual interference effects of the direct and all reflected waves are ignored. The assumption is justifiable because the existence of many possible rays is expected to smooth out the effect of interference. Benjegard[17] supported this assumption by arguing that typical distances between the source and the receiver in urban areas are shorter than in the case of major highways. As a result, the influence of any possible interference due to ground and facade reflections is insignificant for the propagation of sound. Chew[18] derived a model to predict sound fields for the case in which buildings flank both sides of an expressway. In his model, he followed the concept proposed by Davies,[19] which includes the effects of multiple reflections and diffuse scattering for noise propagated in street canyons. Reflection coefficients are used in their numerical method, leading to a better model for estimating the effects of ground and facade on the overall noise levels, but the direct wave and reflected waves are combined incoherently. In this paper, we refer to this type of prediction method as the incoherent model.

To allow for finite impedances of the facades and the ground surfaces, the concept of the absorption coefficient is sometimes introduced, such that a fraction of constant sound energy, α say, is absorbed for each reflection from the facades and impedance ground. In other words, the factor $(1-\alpha)$ is multiplied for each reflection from the reflecting boundaries. Hence, the total sound field can be estimated heuristically by

$$I=I_{\mathrm{ref}}\,d_{\mathrm{ref}}^{\,2}\sum_{l=0}^{\infty}\left(\frac{(1-\alpha_L)^l(1-\alpha_R)^l}{d_{l1}^2}+\frac{(1-\alpha_L)^{l+1}(1-\alpha_R)^l}{d_{l2}^2}+\frac{(1-\alpha_L)^l(1-\alpha_R)^{l+1}}{d_{l3}^2}+\frac{(1-\alpha_L)^{l+1}(1-\alpha_R)^{l+1}}{d_{l4}^2}+\frac{(1-\alpha_L)^l(1-\alpha_R)^l(1-\alpha_G)^{l+1}}{\bar{d}_{l1}^2}+\frac{(1-\alpha_L)^{l+1}(1-\alpha_R)^l(1-\alpha_G)^{l+1}}{\bar{d}_{l2}^2}+\frac{(1-\alpha_L)^l(1-\alpha_R)^{l+1}(1-\alpha_G)^{l+1}}{\bar{d}_{l3}^2}+\frac{(1-\alpha_L)^{l+1}(1-\alpha_R)^{l+1}(1-\alpha_G)^{l+1}}{\bar{d}_{l4}^2}\right), \tag{13}$$

where the subscripts L, R, and G represent the corresponding absorption coefficients for the left and right vertical walls and the impedance ground, respectively.

Building facades consist typically of brickwork or other masonry construction and glass windows. The absorption coefficient of brickwork varies from 0.02 at 125 Hz to 0.05 at 2 kHz. Windows, which act as panel absorbers, enhance the absorption at low frequencies. Delany[20] estimated the average absorption coefficient to be of the order of 0.1, taking into account the greater area of brickwork generally found in

traditional residential areas. In addition, Lee and Davies[4] studied the ground effect assuming the asphalt road surface to reflect 0.9 of the incident sound with no change in phase. Hence, it is reasonable for us to assume that all reflecting boundaries (including the vertical walls and the impedance ground) have the same absorption coefficients, i.e., $\alpha_L=\alpha_R=\alpha_G$. Indeed, Steenackers, Myncke, and Cops[6] determined the absorption coefficient of a typical urban street by measuring the decay curves of the reverberated sound in different streets and comparing them with calculated curves for different absorption coefficients. The absorption coefficient is determined by fitting the measured data with theoretical predictions. For streets of width ranging from 9 to 12 m, they found that the corresponding absorption coefficients vary between 0.1 and 0.2. In this paper, the absorption coefficient of 0.1 is used in our subsequent numerical analyses unless otherwise stated.

We also note that the direct sound pressure level is given, according to the incoherent model, by

$$L_D=L_{\mathrm{ref}}+20\lg(d_{\mathrm{ref}}/d_R), \tag{14}$$

where L_{ref} is the sound pressure level from the source at the reference distance and d_R is the distance between the source and receiver. The sound pressure level at the receiver due to a reflected ray ''hitting'' the boundaries n times is given by

$$L_{Rn}=L_{\mathrm{ref}}+20\lg(d_{\mathrm{ref}}/d_{Rn})+10n\lg(1-\alpha), \tag{15}$$

where n is the order of reflection, α is the absorption coefficient of the boundaries and d_{Rn} is the path length from the image sources to the receiver.

C. The coherent model

The incoherent model estimates the sound field by the energy approach in which the interference effects due to the source and its images are ignored. However, it becomes increasingly inadequate in some cases. A more accurate and intricate model is to allow the variation of the reflection factors, V_L, V_R, and V_B^*, with the angle of incidence in Eq. (9a). Before we proceed to give an analytical expression for the integrals in Eq. (9a), it is of interest to consider a special case where we have only an infinite facade, located at the left side, say, of the source. The theoretical and experimental studies of this case has been reported elsewhere.[14] In fact, this problem can be treated as a special case of two parallel walls where the width of the street becomes very large, $W\to\infty$. It is then straightforward to show $e^{ik_xW}\to 0$. We can simplify G_x from Eq. (4) to

$$G_x=e^{ik_x|x-x_s|}+V_Le^{ik_x(x+x_s)}. \tag{16}$$

Substituting it into Eq. (9a) leads to an analogous expression as developed in our earlier study[14] of sound fields above an impedance ground in front of an absorbing facade. An asymptotic solution has been derived in their study and it is given as follows:

$$\begin{aligned} p=&\frac{e^{ikd_{01}}}{4\pi d_{01}}+Q(d_{02},\theta_{02},\beta_L)\frac{e^{ikd_{02}}}{4\pi d_{02}}\\ &+Q(\bar{d}_{01},\bar{\theta}_{01},\beta_G)\frac{e^{ik\bar{d}_{01}}}{4\pi d_{01}}\\ &+Q(\bar{d}_{02},\bar{\theta}_{02},\beta_G)Q(\bar{d}_{02},\bar{\Theta}_{02},\beta_L)\frac{e^{ik\bar{d}_{02}}}{4\pi d_{02}}, \end{aligned} \tag{17}$$

where $Q(d,\theta,\beta)$ is the spherical wave reflection coefficient[21] that can be determined for a given separation of the image source and receiver d, the angle of incidence of the reflected wave θ, and the normalized admittance of the boundary β. It is determined according to

$$Q(d,\theta,\beta)=R_p+(1-R_p)F(w), \tag{18}$$

where

$$R_p=\frac{\cos\theta-\beta}{\cos\theta+\beta}, \tag{19}$$

$$F(w)=1+i\sqrt{\pi}we^{-w^2}\,\mathrm{erfc}(-iw), \tag{20}$$

and

$$w=+\sqrt{\tfrac{1}{2}ikd}(\cos\theta+\beta). \tag{21}$$

The angle of incidence of the reflected waves, θ_{02}, $\bar{\theta}_{01}$, $\bar{\theta}_{02}$, and $\bar{\Theta}_{02}$ can be found for a given source and receiver location as

$$\begin{aligned} &\cos\theta_{02}=\frac{x+x_s}{d_{02}},\quad \cos\bar{\theta}_{01}=\frac{z+z_s}{\bar{d}_{01}},\\ &\cos\bar{\theta}_{02}=\frac{z+z_s}{\bar{d}_{02}},\quad \text{and}\quad \cos\bar{\Theta}_{02}=\frac{x+x_s}{\bar{d}_{02}}. \end{aligned} \tag{22}$$

It is enlightening to show that our current theoretical formulation may be regarded as a generalization of the previous study.[14] Based on the analytical result of their approach, we can write down the heuristic solution for Eq. (9a) straightforwardly as

$$p=\frac{1}{4\pi}\sum_{l=0}^{\infty}\Bigg(Q^{l}(d_{l1},\theta_{l1},\beta_L)Q^{l}(d_{l1},\theta_{l1},\beta_R)\frac{e^{ikd_{l1}}}{d_{l1}}+Q^{l+1}(d_{l2},\theta_{l2},\beta_L)Q^{l}(d_{l2},\theta_{l2},\beta_R)\frac{e^{ikd_{l2}}}{d_{l2}}$$
$$+Q^{l}(d_{l3},\theta_{l3},\beta_L)Q^{l+1}(d_{l3},\theta_{l3},\beta_R)\frac{e^{ikd_{l3}}}{d_{l3}}+Q^{l+1}(d_{l4},\theta_{l4},\beta_L)Q^{l+1}(d_{l4},\theta_{l4},\beta_R)\frac{e^{ikd_{l4}}}{d_{l4}}$$
$$+Q^{l}(\bar{d}_{l1},\bar{\theta}_{l1},\beta_L)Q^{l}(\bar{d}_{l1},\bar{\theta}_{l1},\beta_R)Q(\bar{d}_{l1},\bar{\Theta}_{l1},\beta_G)\frac{e^{ik\bar{d}_{l1}}}{d_{l1}}+Q^{l+1}(\bar{d}_{l2},\bar{\theta}_{l2},\beta_L)Q^{l}(\bar{d}_{l2},\bar{\theta}_{l2},\beta_R)Q(\bar{d}_{l2},\bar{\Theta}_{l2},\beta_G)\frac{e^{ik\bar{d}_{l2}}}{d_{l2}}$$
$$+Q^{l}(\bar{d}_{l3},\bar{\theta}_{l3},\beta_L)Q^{l+1}(\bar{d}_{l3},\bar{\theta}_{l3},\beta_R)Q(\bar{d}_{l3},\bar{\Theta}_{l3},\beta_G)\frac{e^{ik\bar{d}_{l3}}}{d_{l3}}$$
$$+Q^{l+1}(\bar{d}_{l4},\bar{\theta}_{l4},\beta_L)Q^{l+1}(\bar{d}_{l4},\bar{\theta}_{l4},\beta_R)Q(\bar{d}_{l4},\bar{\Theta}_{l4},\beta_G)\frac{e^{ik\bar{d}_{l4}}}{d_{l4}}\Bigg), \tag{23}$$

where θ_{l1}, θ_{l2}, θ_{l3}, θ_{l4}, $\bar{\theta}_{l1}$, $\bar{\theta}_{l2}$, $\bar{\theta}_{l3}$, $\bar{\theta}_{l4}$, $\bar{\Theta}_{l1}$, $\bar{\Theta}_{l2}$, $\bar{\Theta}_{l3}$, and $\bar{\Theta}_{l4}$ are the angles of incidence of the reflected waves measured from the normal of the reflecting plane.

The validity of Eq. (23) can be confirmed by comparing it with precise model experiments conducted indoors in an anechoic chamber and outdoors in a narrow street. The details of these experimental measurements and numerical simulations will be described in the following section. Before we end this section, we remark that use of an impedance model for enclosing surfaces has somewhat complicated the analysis. However, its introduction is essential as it allows the consideration of more general situations. For instance, a hard ground is not accurate enough to model an asphalt road surface which is frequently used to reduce road traffic noise in urban areas.

III. NUMERICAL SIMULATIONS, EXPERIMENTAL RESULTS, AND DISCUSSIONS

A. Numerical simulations

To simplify the problems, we shall model the propagation of sound in city streets as an analogous situation to the determination of sound fields in a simple channel bounded by two parallel (vertical) infinite walls and a flat horizontal ground. Although it is straightforward to implement impedance boundary conditions in our numerical model, the analyses here are restricted to perfectly reflecting wall surfaces only. Both hard and impedance grounds will be studied. The width of the channel is W and the source is located at a height of z_s above the horizontal ground. To facilitate subsequent calculations, we assume that the origin coincides with the left corner of the street canyon (in the $y=0$ plane) as shown in Fig. 1. The source and receiver are located at $(x_s,0,z_s)$ and (x,y,z), respectively. We are interested in the case where $0\leqslant x,\ x_s\leqslant W$.

It is obvious from Fig. 2 that a series of image sources are produced by multiple reflections of the two parallel walls. Another set of image sources is also formed due to the presence of a reflecting ground. The total sound field can be estimated either by Eq. (13) or (23), depending on whether the incoherent or coherent models are required.

In the presentation of numerical and experimental results, we use excess attenuation (EA), which is defined as the ratio of the total field, P, to the direct field, P_d, as follows:

$$\mathrm{EA}=20\lg(P/P_d). \tag{24}$$

There are many rays connecting the source and receiver that contribute to give the total field. According to the incoherent model, the individual excess attenuation, EA_n due to each ray can be obtained from Eqs. (14) and (15) to yield

$$\mathrm{EA}_n=L_{Rn}-L_D=20\lg(d_R/d_{Rn})+10n\lg(1-\alpha), \tag{25}$$

where n is the order of rays. The excess attenuation, EA, due to the total field is obtained by combining each individual excess attenuation, EA_n logarithmically to give

$$\mathrm{EA}=10\lg\Bigg(\sum_{n=1}^{\infty}10^{\mathrm{EA}_n/10}\Bigg). \tag{26}$$

We wish to point out that only a few terms are normally required in Eq. (13), and hence Eq. (26), to ensure the convergence of the series. This is because of the large distances of the image sources from the receiver and the presence of $(1-\alpha)^n$ term in the series. Oldham and Radwan[22] suggested that the inclusion of higher orders beyond 8 (i.e., $n=8$) produced only a minor contribution to the total field, which can be ignored without significant errors. Our initial analysis supports their view. Indeed, the exact number of n depends on the distance between the source and the receiver, the separation of the boundaries, and the magnitude of the absorption coefficient. When the separation of the boundaries is reduced, it is found that up to 12th-order rays ($n=12$) are required to ensure the convergence of the series for the calculation of the total sound intensities. In Fig. 3, we display a plot of the excess attenuation, EA, versus the wall separation, W. The source and receiver are located at heights of 0.5

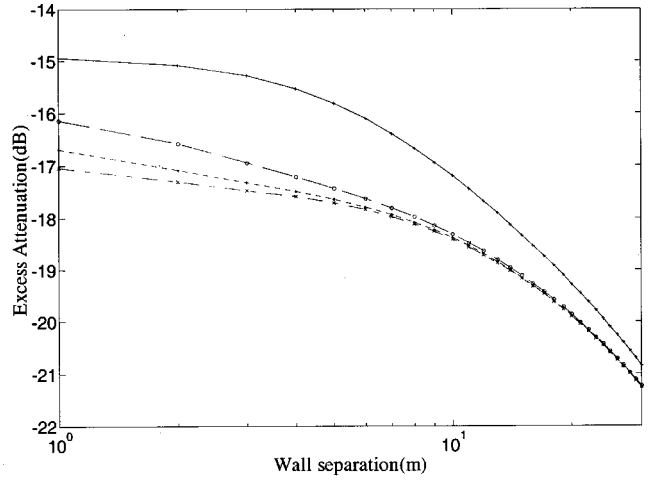


FIG. 3. The contribution of the higher-order rays on overall sound fields by plotting the predicted excess attenuation versus wall separation. The source is located at $(W/2,0,0.5)$ and receiver at $(W-1,20,1.5)$. ———, $n=4$; ---, $n=8$; ···, $n=12$; –·–, $n=16$.

and 1.5 m, respectively, above the ground, and the horizontal separation between them is 20 m. In numerical simulations, the source is placed equidistant from both vertical walls and the receiver is placed at $(W-1)$ m from the left vertical wall.

Different values of n (4, 8, 12, and 16, respectively) are used to calculate the sound intensities, which illustrate the relative significance of higher-order rays. The excess attenuation, which is defined as $L_{Rn}-L_D$ with a reference distance d_R of 1 m, is used in Fig. 3. It is obvious that there is a sizable contribution from the 8th higher-order ray and above $(n\geqslant 8)$ if W is less than about 4 m. Its contribution becomes insignificant if the separation is greater than 10 m.

Differences between noise levels with and without ground surfaces were found to be of the order of 2.5 dB within a typical urban distance of 200 m. In one of the incoherent models, further approximation is made: the ground effect is approximated by adding 3 dB to the predicted levels due to the wall reflections only.[7] Further, it is interesting to note that the excess attenuation, the total field in the incoherent model, is independent of the frequencies since the absorption coefficients of all reflecting surfaces are assumed to be constant at all frequencies. We reiterate here that the absorption coefficient of 0.1 is taken in our numerical simulations.

According to the coherent model, the total field, P, can be obtained from Eq. (23). It is worth noting that the spherical wave reflection coefficient is equal to 1 for a perfectly reflecting surface. In this case, Eq. (23) can be reduced to Eq. (11). No further simplification can be made if the total sound field is computed by summing all contributions from individual rays coherently. It is straightforward to use Eq. (23) to compute the sound pressure. Hence the excess attenuation can be determined by using Eq. (24) with the direct field given by

$$P_d=\frac{1}{4\pi}\left[\frac{e^{ikd_R}}{d_R}\right], \tag{27}$$

where d_R is the direct distance between the source and receiver. In laboratory measurements and theoretical predictions, d_R is taken to be 1 m for the reference sound level.

A close examination of Eq. (23) reveals that the sound field is calculated by summing an infinite series. Each term of the series is represented by a ''virtual'' source multiplied by a spherical wave reflection coefficient. As discussed previously, only a finite number of image sources contribute significantly to the total sound field. The contributions of these image sources, which become weaker in strength when they are located farther from the receiver, can be neglected. The rate of attenuation is even greater for boundaries of finite impedance since the spherical wave reflection coefficient is generally less than 1 for such cases. Hence, the strength of the reflected waves is further attenuated after a few reflections from the boundaries. In our experiments, as shown in the following sections, we find that the use of 8 to 12 sets of image sources in our theoretical predictions is normally adequate to give satisfactory results.

Figure 4 shows a comparison of the theoretical predictions for sound propagation in street canyons. The predicted excess attenuation is plotted against the width of the street canyon. The source frequencies of 125 Hz, 1 kHz, and 8 kHz are used in the coherent model for the prediction of sound fields. In Fig. 4(a), the source and receiver coordinates are set at $(W/2,0,0.65)$ and $(W/2,5,2.4)$, respectively. A somewhat different receiver location of $(W-1,10,5)$ is used in Fig. 4(b). Figures 4(a) and (b) show that coherent and incoherent models give rather similar results when the width of a street canyon is greater than about 10 m. However, when the width becomes narrower, deviations in EA as predicted by the coherent and incoherent models are more significant. This suggests that it is inappropriate to use the incoherent model for the prediction of noise propagation in a narrow street canyon.

B. Indoor laboratory measurements

In the present study, a one-tenth scale was used to model a street canyon and a data acquisition system was set up to conduct indoor measurements for sound propagation over different ground surfaces. A BSWA TECH $\frac{1}{2}$-in. condenser prepolarized microphone of type MK224 fitted with a BSWA TECH preamplifier of type MA201 was used as the receiver for indoor measurements. A Tannoy driver with a tube of internal diameter of 3 cm and length of 1.5 m was used as a point source. All indoor measurements were conducted in an anechoic chamber with internal dimensions of $6\times6\times4$ m^3. A PC-based maximum length sequence system analyzer (MLSSA)[23] was used both as the signal generator for the source and as the analyzer for subsequent data processing. The received signals were compared with the known output sequence such that the background noise effects were minimized. The time-domain data were converted to spectrum level data through the use of fast Fourier transform. Each spectrum level was then normalized by a ''standard'' prerecorded direct field measurement. All the steps described above were carried out with MLSSA software. The final output was the required excess attenuation spectrum. We note

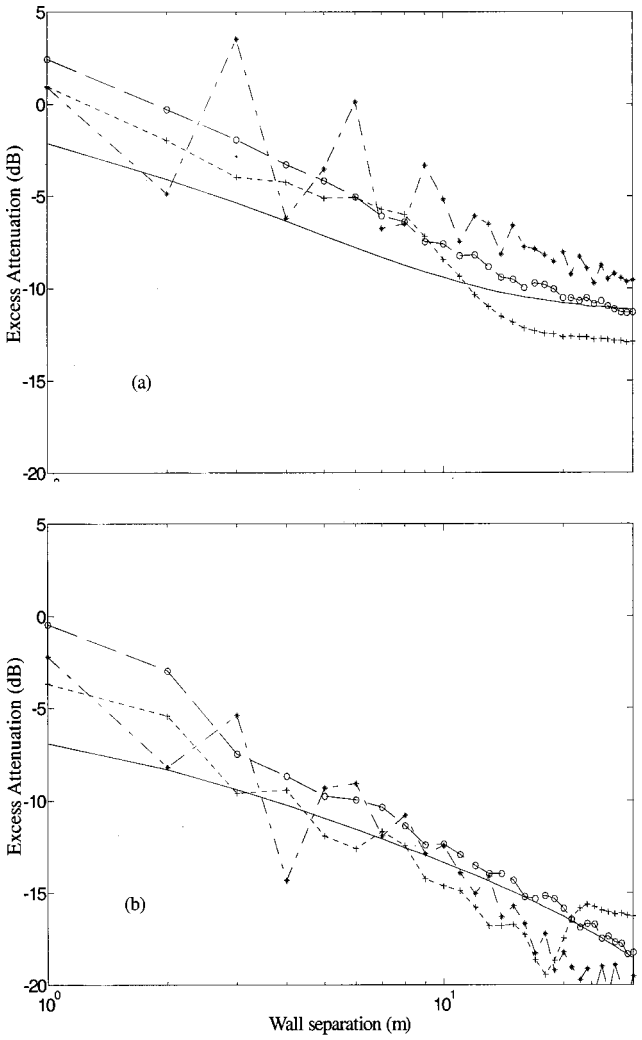


FIG. 4. Comparisons of the coherent and incoherent prediction on the excess attenuation versus wall separation. *−−−*, coherent prediction at 125 Hz; +----+, coherent prediction at 1000 Hz; ∘−−∘, coherent prediction at 8000 Hz; ———, incoherent prediction. (a) Both source and receiver are located at the centerline. The coordinates of the source and receiver are $(W/2,0,0.65)$ and $(W/2,5,2.4)$. (b) The source is located at the centerline and the receiver is offset from the centerline. The coordinates of the source and receiver are, respectively, $(W/2,0,0.65)$ and $(W-1,10,5)$.

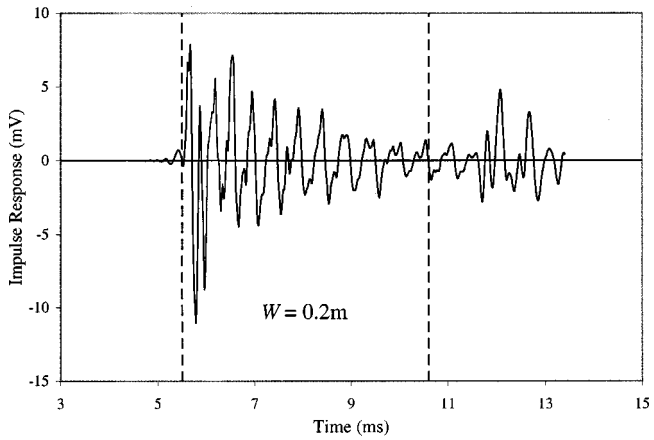


FIG. 5. The impulse response at the receiver for the direct ray up to the 12th-order ray; source at (0.1,0,0.065) and receiver at (0.1,0.8,0.065).

that the prerecorded direct field measurement is taken at 1 m from the source in the absence of any reflecting surfaces in the anechoic chamber.

In the current study, the maximum length sequence (MLS) technique is chosen to investigate sound propagation in a model street canyon as the technique improves the signal-to-noise ratio as compared to the conventional, stationary excitation technique. Heutchi *et al.*[24] have confirmed its usefulness in their study of outdoor sound propagation for distances of up to 200 m.

MLSSA operates in the time domain where the impulse response is measured. As we are only interested in reflection rays up to 12th order, the time-domain data chosen for processing are based on the time between the arrival of the direct and the 12th-order rays. Figure 5 shows a typical impulse response for sound propagation in a model street canyon for the source and receiver located at (0.1,0,0.065) and (0.1,0.8,0.065), respectively. For this particular source/receiver geometry, a time interval of between 5.5 and 10.6 ms is chosen for our subsequent analysis in the frequency domain, as the former is estimated to be the arrival time of the direct ray, while the latter is that for the 12th-order ray.

A varnished plywood board 18 mm in thickness was used to simulate an acoustically hard surface. A carpet, laid on the varnished plywood board, was used to simulate an impedance ground surface. Preliminary measurements were conducted for sound propagation over the plywood board. Both source and receiver were set at a height of 0.065 m above the surface at various receiver locations between 0.5 and 1.4 m from the source. Theoretical predictions of excess attenuation based on the model described in the previous section were in accord with the experimental results. A typical set of measurements, with a source/receiver separation of 0.5 m, is shown in Fig. 6(a).

Attenborough's two-parameter model[25] was used to describe the impedance of the ground surface. Measurement procedures were repeated as in the case of the hard surface described above. The two parameters, the effective flow resistivity (σ_e) and the effective rate of change of porosity with depth (α_e), were deduced from the measurements. The best-fit values for σ_e and α_e were 13.5 kPa s m^{-2} and 100 m^{-1}, respectively. Figure 6(b) shows a comparison of experimental results and theoretical predictions based on the best-fit parameters for the acoustical characterization of the absorbing ground—a carpet laid on a plywood board. We note that, in Fig. 6(b), the same source/receiver geometry was used as in Fig. 6(a).

After the acoustical characterization of the model ground, a one-tenth scale model of a street canyon was set up in the anechoic chamber. Since the maximum path length between the image source and the receiver was about 10 m (100 m at full scale) at frequency up to 10 kHz in our measurements, the effect of air absorption was not included in our analyses. The vertical walls and ground were made of 18-mm-thick plywood boards. The vertical walls, which were supported by a steel angle framework to ensure their orientation and stability, have a length and height of 2.4 and 1.8 m, respectively. The separation, W, between the vertical walls was set at 0.2, 0.5, and 0.8 m for experiments. In a set

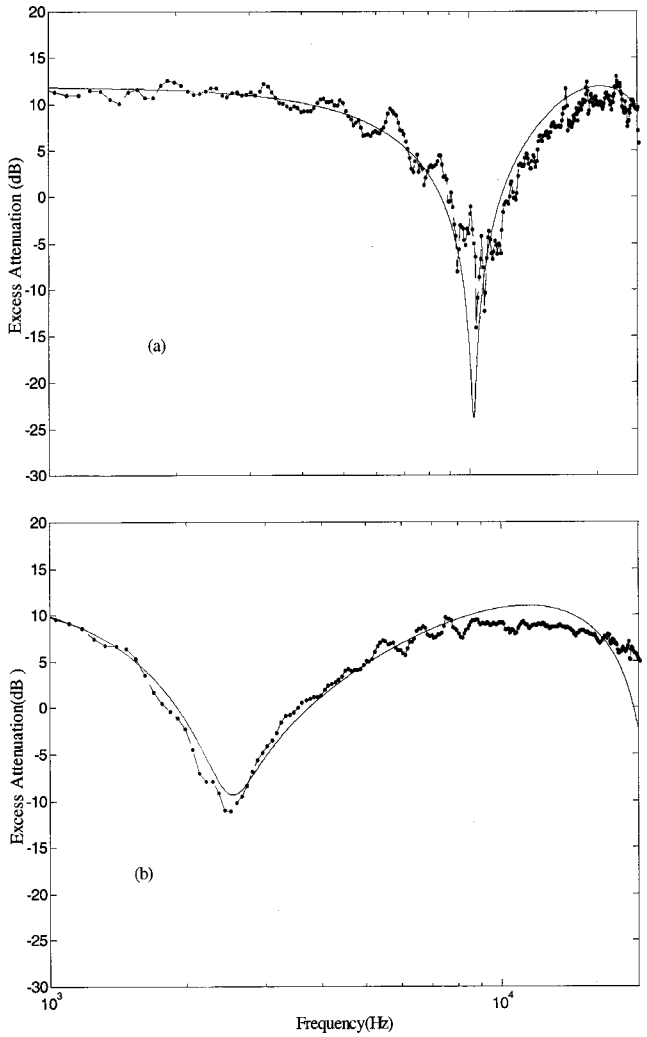


FIG. 6. Sound propagation over horizontal ground. Both source and receiver are 0.065 m above the ground and are separated from each other at 0.5 m. ——— theoretical prediction; ---, measurement. (a) Effect of hard ground reflection and (b) characterization of impedance ground.

of measurements, carpet was laid on the plywood board to simulate an absorbing ground surface in a street canyon.

At full scale, the model represents a canyon street section of 24 m in length with two 18-m-high flat vertical parallel building walls erected on either side of a street with widths of 2, 5, and 8 m. In the following paragraphs, we refer to all dimensions as the scaled distances unless otherwise stated. Although Busch *et al.*[26] have reported an improved technique for simultaneously selecting both an optimal scale factor and optimal model materials, no attempt is made in our laboratory measurements to select the most appropriate materials to model an outdoor porous ground surface. However, the use of carpet-covered ground will allow a validation of the coherent and incoherent models by comparing theoretical predictions with precise indoor measurements.

Indoor measurement of the excess attenuation was made at various receiver locations with the source located near the ground. This is to simulate the approximate location of ground-based sources such as engine noise emitted from light vehicles or other construction equipment for road repair

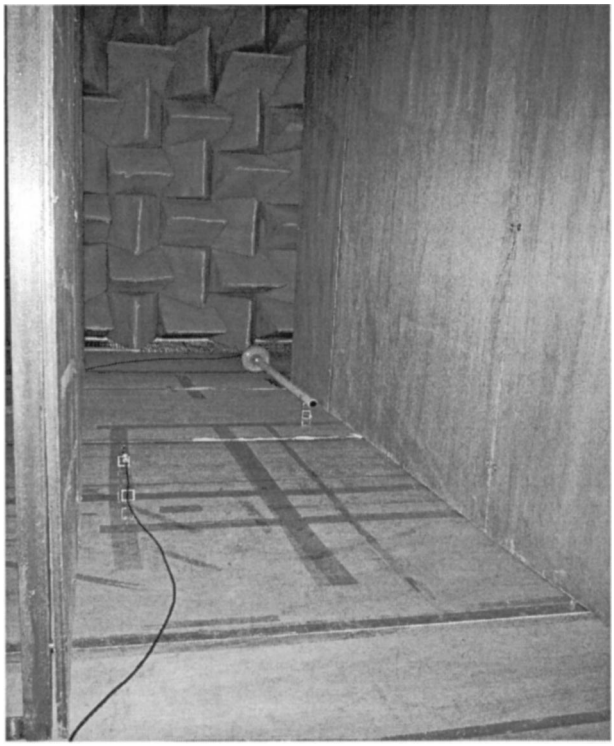


FIG. 7. A 1:10 scale model set up in an anechoic chamber for measurement of sound propagation in a street canyon.

work. Recently, Horoshenkov *et al.*[27] have used a 1:20 scale model to study sound propagation in a city street canyon. Noise sources, such as light and heavy vehicles, have been simulated by tubes from the source chambers elevated at 0.025 m (0.5 m at full scale) and 0.05 m (1.0 m at full scale) above the ground. In our measurements, the source was located at 0.065 m (0.65 m at full scale) above the ground and either at the centerline or on the side of the street at 0.1 m (1.0 m at full scale) from one of the vertical walls. Again, the chosen source height was used to simulate a realistic urban noise source such as the typical engine location of automobiles. For the 0.2-m-wide street, the receivers were set to locate both at the centerline of the street and on the side of the street at 0.02 m from one of the vertical walls. For the 0.5- and 0.8-m-wide streets, the receivers were also set at the centerline and on the side of the street but at 0.1 m from one of the vertical walls. The heights of the receivers were at 0.065, 0.12, 0.24 and 0.5 m from the ground. Due to the limitation of the available space in the anechoic chamber, the horizontal distances between the source and receivers were in the range of 0.2 to 1.0 m. A photograph showing the scale model set in the anechoic chamber is shown in Fig. 7.

To demonstrate the effect of mutual interference, the source was placed equidistant from the vertical walls and at a height of 0.065 m. Measurements were taken at a horizontal distance of 1 m in front of the source and at a height of 0.5 m above the ground. By reducing the width of the street, the interference effect became more significant. Figure 8(a) shows that distinct peaks occurred at 2000 and 6000 Hz (200 and 600 Hz at full scale) in the street of width 0.2 m (2 m at full scale). In this event, the measured excess attenuation caused by constructive interference is 10 dB higher than that

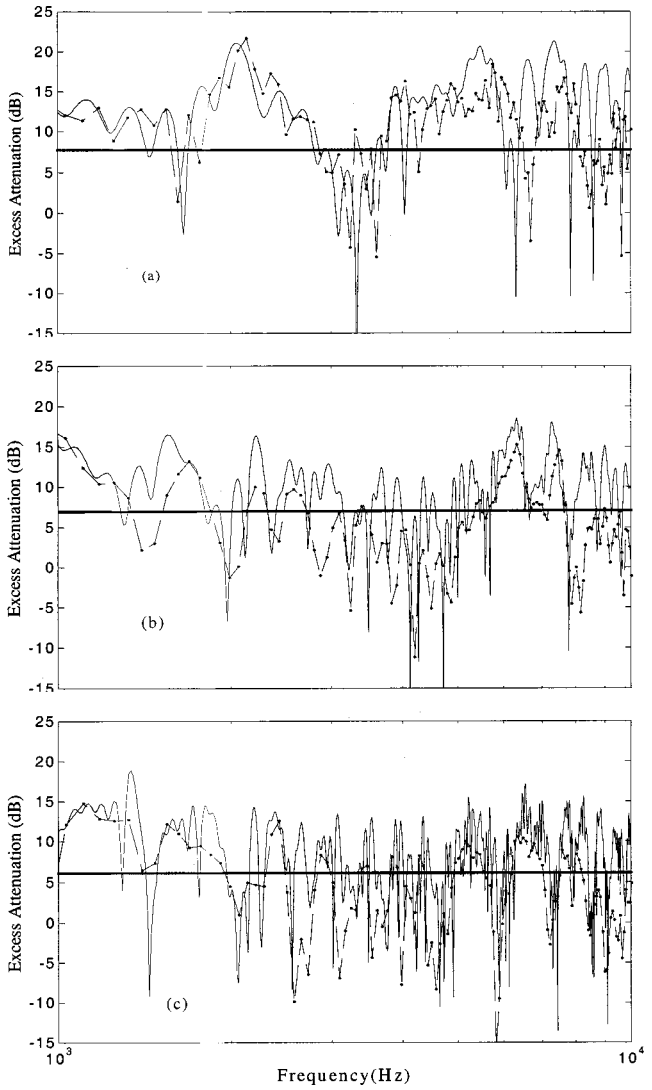


FIG. 8. Excess attenuation versus frequency for various street widths for hard ground. ———, coherent prediction; ——, incoherent prediction; ·––·, measurement. (a) Width=0.2 m; source at (0.1,0,0.065); receiver at (0.1,1,0.5); (b) width=0.5 m; source at (0.25,0,0.065); receiver at (0.25,1,0.5); and (c) width=0.8 m; source at (0.4,0,0.065); receiver at (0.4,1,0.5).

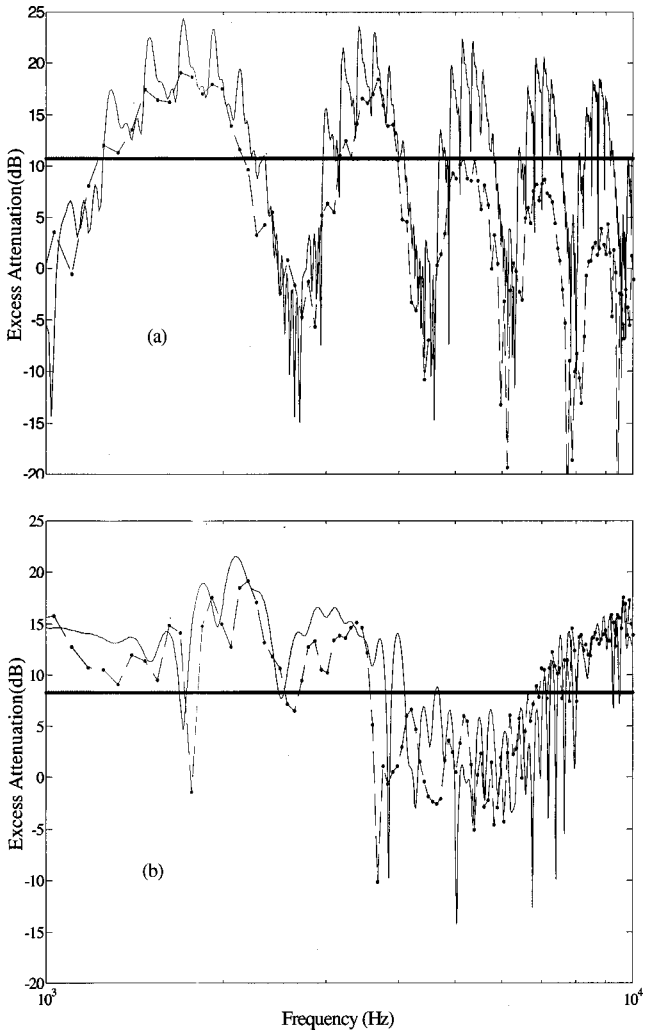


FIG. 9. Excess attenuation versus frequency for different ratios of the horizontal distance to street width. A hard ground is assumed in the prediction. ———, coherent prediction; ——, incoherent prediction; ·––·, measurement. (a) Ratio=0.25; width=0.8 m; source at (0.1,0,0.065); receiver at (0.7,0.2,0.065); and (b) ratio=5; width=0.2 m; source at (0.1,0,0.065); receiver at (0.18,1,0.24).

predicted by the incoherent model. In Figs. 8(b) and (c), peaks and dips are close to each other and such fluctuations will be smoothed out if the measurements are taken in octave bands. This shows that the interference effect becomes gradually less significant when the width of the street is increased to 0.5 and 0.8 m (5 and 8 m for a full-scale street).

The interference effect can also occur when the source is located at positions other than the center. In Fig. 9, the source is located 0.1 m from one of the vertical walls and the receiver is placed at 0.1 or 0.02 m from the opposite wall. To further demonstrate that the coherent model can be used to predict the interference effect in a wide range of horizontal distances to street width ratios, a change in the ratio from 0.25 to 5.0 is selected for the purpose of illustration. The theoretical and experimental results are shown in Figs. 9(a) and (b), respectively. In Fig. 9(a), the horizontal distance between the source and receiver is set at 0.2 m (the shortest distance used in the experiment) in the 0.8-m-wide street (the widest street used in the experiment). In Fig. 9(b), the horizontal distance is 1.0 m (the longest distance used in the experiment) in the 0.2-m-wide street (the narrowest street used in the experiment).

Elevated noise sources, such as air conditioners and cooling towers, installed on building facades are common in a complex urban environment. To simulate this situation, additional measurements of the excess attenuation were made for the street canyon with width of 0.5 m. The source was located at 0.05 m (0.5 m at full scale) from one of the vertical walls and the receivers were placed at 0.1 m from the opposite wall. While the source was set at 0.5 m (5 m at full scale) above the ground, the receivers were set at the heights of 0.15, 0.5 and 0.8 m. The horizontal distances between the source and receivers were set at 0.2, 0.5, 0.8 and 1 m.

The effect of mutual interference also occurs when the noise is emitted from the elevated source. Figure 10(b) illustrates that a distinct dip occurs at around 1700 Hz (170 Hz at full scale) when the receiver is located at the same height of the source. When the receiver is moved to either a lower

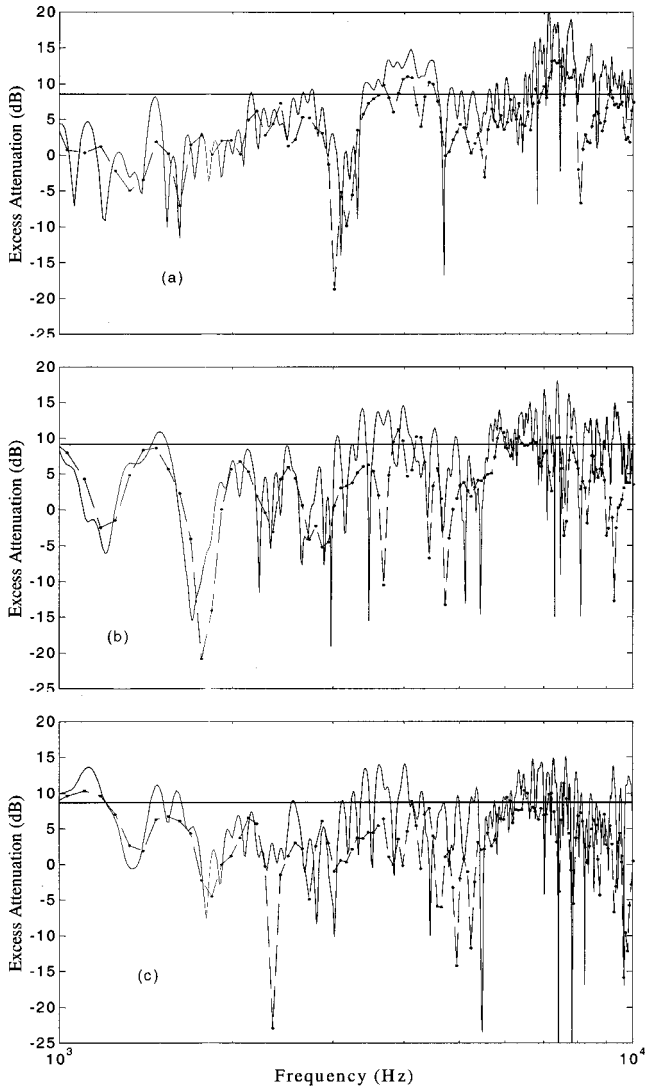


FIG. 10. Excess attenuation versus frequency for elevated source in a 0.5-m-wide street with hard ground. ———, coherent prediction; ——, incoherent prediction; ·——·, measurement. (a) Source at (0.05,0,0.5); receiver at (0.4,0.8,0.15); (b) source at (0.05,0,0.5); receiver at (0.4,0.8,0.5); and (c) source at (0.05,0,0.5); receiver at (0.4,0.8,0.8).

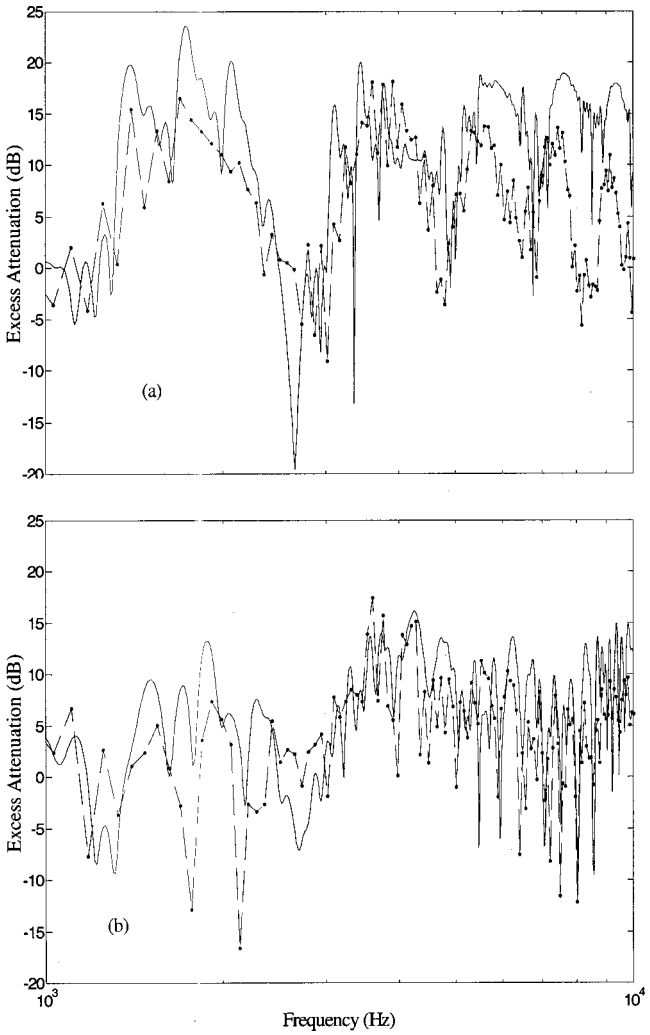


FIG. 11. Excess attenuation versus frequency for various receiver locations in the street with an impedance ground. ———, coherent prediction; ·——·, measurement. (a) Width=0.5 m; source at (0.1,0,0.065); receiver at (0.4,0.2,0.12); and (b) width=0.5 m; source at (0.1,0,0.065); receiver at (0.4,1,0.5).

level at 0.15 m above the ground in the scale model experiment in Fig. 10(a) or to a higher level at 0.8 m above the ground in Fig. 10(c), the effect of destructive interference becomes less significant. In Figs. 10(a)–(c), the source is separated from the receiver at a horizontal distance of 0.8 m. According to the incoherent model the predicted results will be the same in all the cases. However, it is worth pointing out that there are about 10 dB differences between the prediction results by the incoherent model and the measurements at the frequency range from 1000 to 2000 Hz (100 to 200 Hz at full scale).

The coherent model can also be used to predict the interference peaks and dips in the street with an absorptive ground. Experiments were conducted in the ''model'' street at the anechoic chamber. To compare the numerical and experimental results, the predicted excess attenuation together with the measured excess attenuation are plotted against the frequency in Figs. 11(a) and (b). In these comparisons, the width of the street was 0.5 m and the source was located 0.1 m from one of the vertical walls while the receiver was located 0.1 m from the opposite wall. Furthermore, the receiver was set at a horizontal distance of 0.2 m from the source and at 0.12 m above the ground in Fig. 11(a), and was moved to 1.0 m from the source and 0.5 m above the ground in Fig. 11(b). Again, good agreements between the theoretical predictions and experimental measurements are evident from the results as shown in the figures. The agreements are within 5 dB in most cases, especially at the frequency range from 3000 to 5000 Hz (300 to 500 Hz at full scale). However, the most important point to note is that the numerical model can predict well the general trend of the experimental data in the frequency spectrum.

Although it is increasingly common to use porous asphalt[28] as a noise abatement measure for reducing traffic noise, it may not be useful in some frequency ranges. Figure 12 illustrates that the sound levels are increased significantly at the frequency range of 3000 to 4000 Hz by the introduction of a finite impedance ground. Figure 12(a) shows the excess attenuation when measurements are conducted in a street of width 0.5 m with a plywood board representing hard

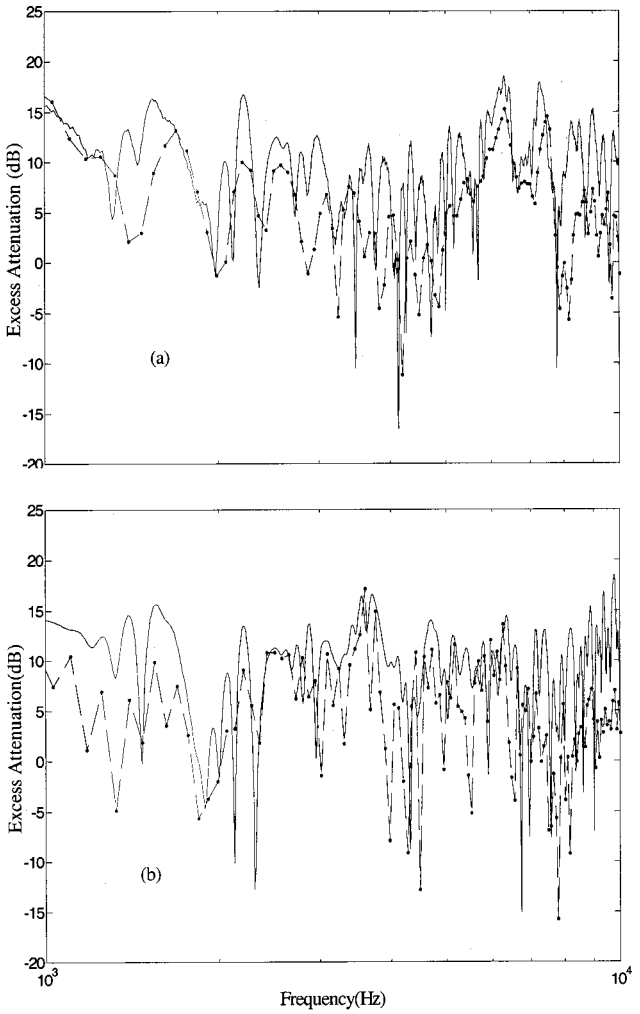


FIG. 12. Comparison of excess attenuation with the boundary conditions changing from (a) a hard ground to (b) an impedance ground. ———, coherent prediction; ·––·, measurement. In both cases, the width=0.5 m, the source is at (0.25,0,0.065), and receiver at (0.25,1,0.5).

ground. Figure 12(b) displays the results obtained in the same street with a carpet-covered board representing an absorbing ground. In both cases, the source and the receiver are placed equidistant from the vertical walls. The source is set at a height of 0.065 m. The receiver is located at a height of 0.5 m and at a horizontal distance of 1.0 m from the source.

In general, the theoretical predictions by the coherent model agree reasonably well with the indoor experimental data, which suggests that the interference effect should not be ignored, especially for narrow street canyons. Some discrepancies are found at frequencies between 1000 and 2000 Hz in the scale model experiments. This is probably due to the experimental limitations in the anechoic chamber and the diffraction effect at the edges of the scale model. We also note that some systematic discrepancies are observed in the high frequency region above 10 000 Hz (1000 Hz at full scale). In this region, experimental results are always lower than the predictions.

We also wish to point out that some indoor experimental results are presented in single frequencies for different wall separations. This allows clear and detailed comparisons of experimental data with the numerical models in a more precise manner. However, in the next section, we shall show outdoor data where experimental results will be presented in one-third octaves from the practical point of views.

FIG. 13. Experimental setup in a 1.55-m-wide street canyon for outdoor measurements.

C. Outdoor field measurements

We have demonstrated in the last section that the coherent model agrees well with indoor measurement data. To provide further evidence in a realistic urban environment, outdoor experimental measurements are required in order to establish the validity of the coherent model for the prediction of multi-path transmissions of sound in a street canyon. A side lane of 1.55 m in width was chosen. Two parallel building facades furnished with marble stone and ceramic tiles enclosed the side lane. The ground was made of concrete with a small drain channel formed at the edge of the side lane (see Fig. 13 for a photograph of the experimental test site).

A Roland type KC-300 amplifier and loudspeaker unit was used as the sound source. A broadband white noise, which was generated by an Ivie Noise Generator of type IE-20B, was used for all outdoor experiments. Measurements of the sound pressure levels in one-third octave bands were conducted using an Ono Sokki Precision Sound Level Meter type LA-5110 at various receiver locations, with the source situated at the centerline of the side lane and 0.5 m above the ground. The receiver was located either at the centerline of the street or at an offset position of 0.15 m from one of the building facades.

It is convenient to present the outdoor experimental data in a form of ''transmission loss'' spectrum for comparison with the theoretical predictions. The transmission loss (TL) is defined as the ratio of the total sound field, P is measured at the receiver position to the total sound field, and P_1 is measured at 1 m from the source at the same height. It can be expressed as

$$\mathrm{TL} = 20 \lg(P/P_1). \tag{28}$$

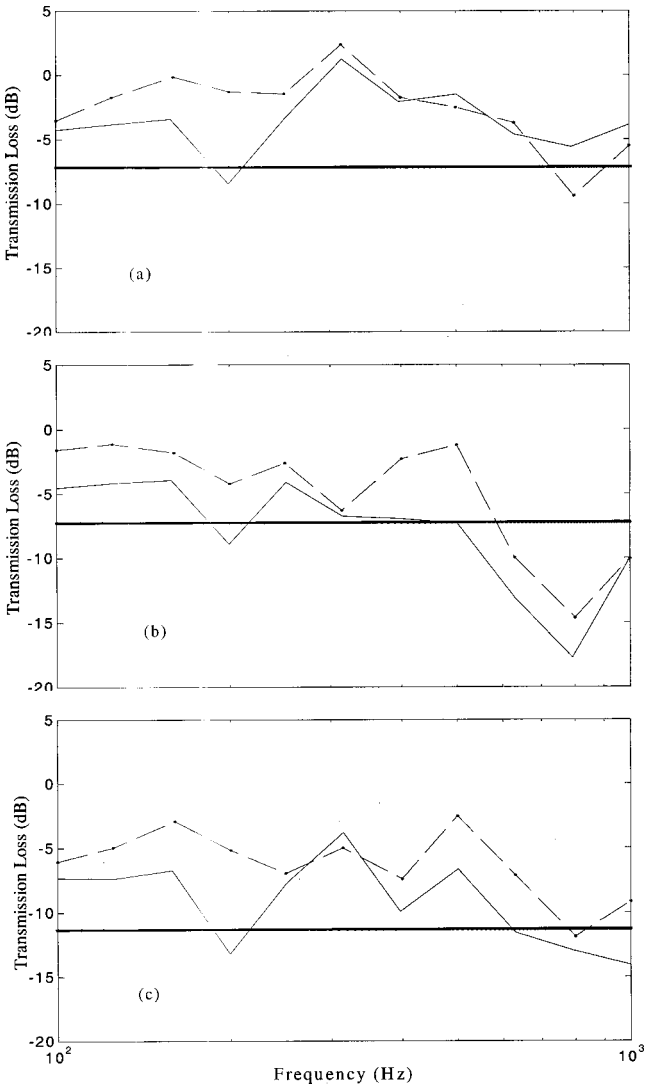


FIG. 14. Comparison of transmission loss spectra among the coherent model predictions, the incoherent model predictions, and outdoor measurements for a street canyon with a width of 1.55 m. ———, coherent prediction; ———, incoherent prediction; ·––·, measurement. (a) Source at (0.78,0,0.5); receiver at (0.78,3.88,0.5); (b) source at (0.78,0,0.5); receiver at (1.4,3.88,0.93); and (c) source at (0.78,0,0.5); receiver at (0.78,7.75,1.86).

Figure 14 (see the figure captions for the geometrical configurations of the source and receiver) shows comparisons of the theoretical predictions with the outdoor experimental results. We note that the theoretical predictions of sound propagation in the side lane in one-third octave bands are based on the coherent model as well as the incoherent model. Due to the limitation of the test site in our outdoor measurements, the horizontal distances between the source and the receiver are limited in the range from 1.0 to 7.75 m and all boundaries are modeled as perfectly reflecting planes. Nevertheless, the experimental results can provide us with useful information on the validity of the model in a more complex outdoor environment.

It is found from Figs. 14(a)–(c) that the coherent model can give a reasonable prediction of the general trend of the transmission loss spectra as compared with the experimental data. However, the incoherent model is unable to predict the trend because the interference effects from all contributing rays are ignored in this model. Inevitably, the incoherent model will not be sufficiently accurate when it is used to predict the propagation of sound in a narrow street canyon. In fact, we have conducted a series of measurements for other geometrical configurations as described earlier. In this paper, we only show three sets of representative data for illustration. All other measurements have led to a rather similar conclusion as described above. Hence, they are not shown here for the sake of brevity.

Finally, some significant discrepancies between outdoor measurements and theoretical predictions are found. In addition to the possible reasons explained in the last section, there are some possible factors that might affect the accuracy of outdoor measurements as follows,

(1) The surface roughness of the building facades and the drain channels on the ground could cause diffusion of sound;
(2) sound might reflect back from both ends of the side lane; and
(3) the atmospheric turbulence effect is especially important at high frequencies above 4000 Hz octave band. As a result, the predicted attenuation will be less reliable especially at high frequencies.

IV. CONCLUSIONS

Theoretical models of the sound field due to a stationary, point monopole source in a street canyon have been developed. The theory is an extension of the classical Weyl–van de Pol formula that includes the effect of multiple reflections from boundary surfaces. By identifying the corresponding image sources, the total field can be computed by summing all contributions coherently. Indoor experiments have been conducted in a one-tenth scale model facility. Locations of the source and the receiver are chosen to simulate those situations commonly occurring in city streets. It has been demonstrated that the theoretical predictions agree reasonably well with our experimental measurements. Further, outdoor experimental measurements have also been conducted to establish the validity of the proposed model for the prediction of multi-path transmissions of sound in a street canyon.

It is concluded that the interference effect has to be taken into account to predict the sound propagation in narrow street canyons of perfectly flat walls. Use of the incoherent model gives reasonable predictions only when the separation between the two vertical walls of a street canyon is greater than 10 m or in the case when there are structured walls leading to significant diffusions of sound energy.

ACKNOWLEDGMENTS

One of the authors (KKI) is supported by a Teaching Company Scheme jointly sponsored by the Industry Department of the HKSAR Government and NAP Acoustics (Far East) Limited. The research described in this paper was supported in part by the Hong Kong Polytechnic University and the Research Grants Council of the HKSAR Government.

[1] J. E. Piercy, T. F. W. Embleton, and L. C. Sutherland, ''Review of noise propagation in the atmosphere,'' J. Acoust. Soc. Am. **61**, 1403–1418 (1977).
[2] T. F. W. Embleton, ''Tutorial on sound propagation outdoors,'' J. Acoust. Soc. Am. **100**, 31–48 (1996).
[3] W. R. Schlatter, ''Sound power measurement in a semi-confined space,'' M.S. thesis, MIT, 1971.
[4] K. P. Lee and H. G. Davies, ''Monogram for estimating noise Propagation in urban areas,'' J. Acoust. Soc. Am. **57**, 1477–1480 (1975).
[5] R. H. Lyon, ''Stochastics and environmental noise,'' The 3rd US–Japan Joint Seminar in Applied Stochastics (1971).
[6] P. Steenackers, H. Myncke, and A. Cops, ''Reverberation in town streets,'' Acustica **40**, 115–119 (1978).
[7] D. J. Oldham and M. M. Radwan, ''Sound propagation in city streets,'' J. Build. Acoust., **1**, 65–88 (1994).
[8] J. Kang, ''Sound propagation in street canyons: Comparison between diffusely and geometrically reflecting boundaries,'' J. Acoust. Soc. Am. **107**, 1394–1404 (2000).
[9] RAYNOISE User's Manual, Version 2.1 LMS Numerical Technologies (2001).
[10] E. Walerian, R. Janczur, and M. Czechowicz, ''Sound level forecasting for city-centers. Part 1: Sound level due to a road within an urban canyon,'' Appl. Acoust. **62**, 359–380 (2001).
[11] M. Gensane and F. Santon, ''Prediction of sound fields in rooms of arbitrary shape: the validity of the image sources method,'' J. Sound Vib. **63**, 97–108 (1979).
[12] G. Lemire and J. Nicolas, ''Aerial propagation of spherical sound waves in bounded spaces,'' J. Acoust. Soc. Am. **86**, 1845–1853 (1989).
[13] S. M. Dance, J. P. Roberts, and B. M. Shield, ''Computer prediction of sound distribution in enclosed spaces using an interference pressure model,'' Appl. Acoust. **44**, 53–65 (1995).
[14] S. H. Tang and K. M. Li, ''The prediction of facade effects from a point source above an impedance ground,'' J. Acoust. Soc. Am. **110**, 278–288 (2001).
[15] J. Lighthill, ''Asymptotic behaviour of anisotropic wave systems stimulated by oscillated sources,'' in *Wave Asymptotics*, edited by P. A. Martin and G. R. Wickham (Cambridge U. P., Cambridge, 1992), Chap. 1. See also J. Lighthill, ''Emendations to a proof in the general three-dimensional theory of oscillating sources of waves,'' Proc. R. Soc. London, Ser. A **427**, 31–42 (1990).
[16] L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1980), p. 228.
[17] S. O. Benjegard, ''Traffic noise in urban areas,'' The 8th International Congress on Acoustics, London (1974), p. 114.
[18] C. H. Chew, ''Prediction of traffic noise from expressways—Part II: Buildings flanking both sides of expressway,'' Appl. Acoust. **32**, 61–72 (1991).
[19] H. G. Davies, ''Multiple-reflection diffuse-scattering model for noise propagation in streets,'' J. Acoust. Soc. Am. **64**, 517–521 (1978).
[20] M. E. Delany, ''Prediction of traffic noise levels,'' NPL Acoustic Report 56 (1972).
[21] K. Attenborough, ''Review of ground effects on outdoor sound propagation from continuous broadband sources,'' Appl. Acoust. **24**, 289–319 (1988).
[22] M. M. Radwan and D. J. Oldham, ''The prediction of noise from urban traffic under interrupted flow conditions,'' Appl. Acoust. **21**, 163–185 (1987).
[23] D. D. Rife and J. Van der Kooy, ''Transfer-function measurement with Maximum-Length Sequences,'' J. Audio Eng. Soc. **37**, 419–443 (1989).
[24] K. Heutschi and A. Rosenheck, ''Outdoor sound propagation measurements using an MLS Technique,'' Appl. Acoust. **51**, 13–32 (1997).
[25] K. Attenborough, ''Ground parameter information for propagation modelling,'' J. Acoust. Soc. Am. **92**, 418–427 (1992).
[26] T. A. Busch and M. R. Hodgson, ''Improved method for selecting scale factors and model materials for scale modelling of outdoor sound propagation,'' J. Sound Vib. **243**(1), 173–181 (2001).
[27] K. V. Horoshenkov, D. C. Hothersall, and S. E. Mercy, ''Scale modelling of sound propagation in a city street canyon,'' J. Sound Vib. **223**(1), 795–819 (1999).
[28] M. C. Berengier, M. R. Stinson, G. A. Daigle, and J. F. Hamet, ''Porous road pavements: Acoustical characterization and propagation effects,'' J. Acoust. Soc. Am. **101**, 155–162 (1997).

A multiple regression model for urban traffic noise in Hong Kong

W. M. To and Rodney C. W. Ip
Department of Mechanical Engineering, Hong Kong University of Science & Technology, Clear Water Bay, Hong Kong

Gabriel C. K. Lam and Chris T. H. Yau
Nature & Technologies (HK) Limited, Unit 2 & 3, 4/F, Wellborne Commercial Centre, 8 Java Road, North Point, Hong Kong

(Received 6 April 2001; revised 9 April 2002; accepted 21 May 2002)

This article describes the roadside traffic noise surveys conducted in heavily built-up urban areas in Hong Kong. Noise measurements were carried out along 18 major roads in 1999. The measurement data included L_{10}, L_{50}, L_{90}, L_{eq}, L_{max}, the number of light vehicles, the number of heavy vehicles, the total traffic flow, and the average speed of vehicles. Statistical analysis using the analysis of variance (ANOVA) and Tukey test ($p<0.05$) reveals that the total traffic flow and the number of heavy vehicles are the most significant factors of urban traffic noise. Multiple regression was used to derive a set of empirical formulas for predicting L_{10} noise level due to road traffic. The accuracy of these empirical formulas is quantified and compared to that of another widely used prediction model in Hong Kong—the Calculation of Road Traffic Noise. The applicability of the selected multiple regression model is validated by the noise measurements performed in the winter of 2000.
 [DOI: 10.1121/1.1494803]

PACS numbers: 43.50.Lj [MRS]

I. INTRODUCTION

Accurate noise prediction is important for environmental planning in a densely populated city such as Hong Kong. Over the past decades, various traffic noise prediction models based on the statistical analysis of noise measurements at different places[1–10] have been suggested. In 1967, Stephenson and Vulkan[1] presented a traffic noise survey in Central London and reported that traffic noise in urban roads was mainly governed by the total number of vehicles, traffic composition, road gradients, and road surfaces. To study the contribution from the first two categories, they performed over 100 curbside measurements of traffic noise in urban roads. All measurement points selected were on level roads, where traffic was flowing smoothly at speeds between 32 kmph (20 mph) and 48 kmph (30 mph). The points were situated away from traffic lights, pedestrian crossings, and crossroads. Narrow roads with flanking tall buildings were avoided. In 1973, Lewis[2] investigated the noise generated by single vehicles in freely flowing traffic. He indicated that vehicles traveling in such a condition formed two distinct groups, petrol or diesel, with respect to noise emissions and proposed a simple regression model for each group. In 1976, Delany *et al.*[3] introduced the procedure for predicting L_{10} noise level due to road traffic. The associated memorandum,[4] entitled ''the Calculation of Road Traffic Noise,'' was published by the U.K. Department of the Environment. Gilbert[5] investigated L_{10} noise levels in urban streets. He presented two regression models for predicting curbside traffic noise, but only for proportions of heavy vehicles ranging from 0% to 30%. In 1978, Barry and Reagan[6] provided design guidance for the prediction of highway noise and introduced the first version of the Federal Highway Administration's Traffic Noise Model using L_{10} noise level. At about the same time, the Ontario Ministry of Transportation and Communications completed development of a predictive equation based on the L_{eq} concept.[7] In 1983, Jones *et al.*[8] used computer methods to simulate road traffic noise in various restricted flow conditions. They showed that L_{10} noise level adjacent to the accelerating traffic streams is generally found to be greater than that adjacent to decelerating streams. In 1989, Chew[9] considered the façade effect from buildings flanking one side of expressways and proposed a modified traffic noise prediction model based on the regression form as suggested by CRTN. Chew and Lim[10] reported that the presence of buildings on both sides of a road could increase L_{10} noise level by more than 10 dBA when comparing to no buildings on both sides under the same traffic condition. At distances of more than 20 m from the buildings, this façade effect would be negligible. In India, Chakrabarty *et al.*[11] measured road traffic noise at 24 traffic junctions of Calcutta during the summer of 1993. They developed regression equations for predicting the ambient noise levels $L_{eq(h)}$ by using the equivalent number of light/heavy vehicles per hour and the distance from the traveled traffic centerline. In Nigeria, Saadu *et al.*[12] used direct regression technique to obtain a regression model for road traffic noise in urban areas.

Most of the prediction models including U.K.'s CRTN, U.S.'s TNM, and their modified versions were developed using the noise levels measured in widely opened areas and for vehicles traveling at a high speed (mostly 50 km/h or above). The measurement points were located at certain distances from highways while there was no physical barrier or only small physical objects such as low-rise apartments in the vicinity. The traffic and environmental conditions of urban areas in Hong Kong, and other new commercial centers

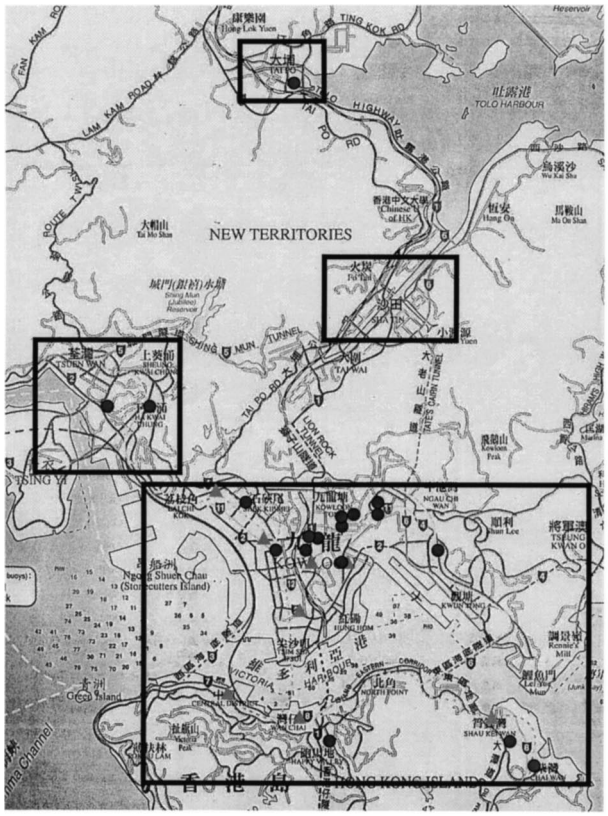


FIG. 1. Road map of Hong Kong.

in Asia such as Guangzhou and Shanghai, are very different. They consist of a large percentage of heavy vehicles—vans, buses, and lorries and tall buildings flanking both sides of urban roads. In fact, the Hong Kong Government[13] reported that there were 496 606 licensed vehicles and 1885 kilometers of roads, making Hong Kong one of the cities with the highest possible vehicle densities in the World in 1999. In this work, we analyze a set of roadside traffic noise data collected along 18 major roads in 1999 and derive a multiple-regression model for predicting L_{10} noise level due to road traffic. The applicability of this prediction model is validated by another set of noise data obtained in the winter months of 2000.

II. METHOD OF NOISE MEASUREMENT

Urban areas mean the places with high density and diverse development, including a mixture of such elements as major trades, commercial activities, and residential premises. The most well known urban areas in Hong Kong are Central, Admiralty, Wan Chai, Causeway Bay, Tsim Sha Tsui, and Mong Kok located at both sides of Victoria Habour. A road map of Hong Kong is shown in Fig. 1 in which urban areas are enclosed by the boxes with solid thick lines. Figure 2 shows the details of one representative site. It should be noted that continuous tall buildings (10 stories or more) are flanking on both sides of the roads.

The measurement sites are located along 18 major roads as shown in Fig. 1. These sites are representative as many people live or study under similar environments. Each of them was sufficiently far (at least 20 m and some of them up to 100 m) from traffic lights, bus stops, and/or crossroads. Most vehicles cruised at constant speeds at 80% of these road sections, except at very close to junctions or stops where they need to decelerate or accelerate. In each site, a B&K 2231 IEC-651 Type I sound level meter was used to measure the statistical noise levels. The sound level meter was calibrated by a B&K 4230 acoustical calibrator that produces a pure tone signal of 94 dB at 1 kHz before and after each measurement. It was mounted on a stand at a height of 1.2 m above ground and was located at 1 m from the façade of the selected building. In literature, Oakes and Tomlinson[14] compared noise measurements made at curbside and at 1 m from façades in urban streets where the pavement width varied from 3 to 5 m. Their results showed that in central urban areas with continuous building façades traffic noise levels at the façades do not differ significantly from levels at the curbside. Each measurement lasted for 30 min and the A-weighted statistical noise levels including L_{10}, L_{50}, L_{90}, $L_{\mathrm{eq,30min}}$, and $L_{\max}$ were measured. During the same period

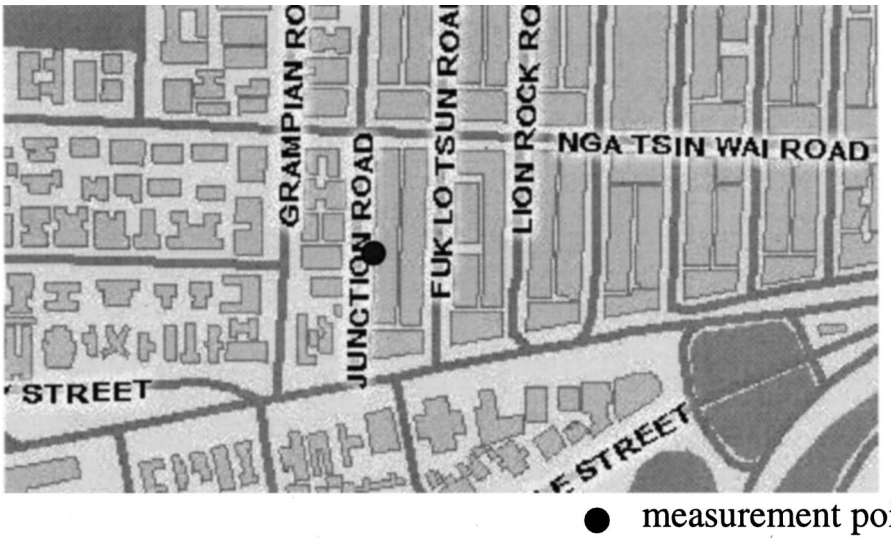


FIG. 2. Location of measurement point along a major road.

TABLE I. Site locations and data collected (in 1999) (Map No. NM=New Territories, KM=Kowloon, HM=Hong Kong Island).

Site no.	HK map no.	Road	L_{eq} (dBA)	L_{10} (dBA)	L_{50} (dBA)	L_{90} (dBA)	L_{max} (dBA)	Far lane (Veh/h)	Near lane (Veh/h)	Far lane (HV/h)	Near lane (HV/h)	Hourly total veh. Q	Hourly heavy veh. H	% of heavy veh. P	Ave. speed (km/h) V
1	KM10	Carpenter Road	69.4	73.4	64.9	58.4	90.6	330	210	144	96	540	240	44.4	25.6
2	NM06	Nam Wan Road	76.8	79.4	75.9	73.4	85.2	402	468	162	216	870	378	43.4	30.7
3	KM05	Junction Road	75	79.3	70.3	62.8	94.9	594	300	174	72	894	246	27.5	25.7
4	HM09	Chai Wan Road	75.9	79.4	69.9	60.4	93	372	576	66	132	948	198	20.9	35.2
5	NM36	Hing Fong Road	73.9	77.4	71	65.2	94.1	636	558	372	264	1194	636	53.3	29.5
6	HM10	Chai Wan Road	70.2	73.6	68.1	63.6	83.2	594	600	258	264	1194	522	43.7	42.8
7	KM04	Tong Mei Road	77	81.8	70.3	65.8	92.2	210	1038	126	534	1248	660	52.9	26.1
8	KM05	Junction Road	77.4	81.9	73.4	60.9	99.6	894	528	390	216	1422	606	42.6	29.1
9	KM09	Argyle Street	74.9	78.8	73.3	65.3	88.5	1416	912	276	120	2328	396	17.0	31.7
10	KM10	Choi Hung Road	82.1	85.6	80.5	75.6	94.2	1326	1032	342	354	2358	696	29.5	30.9
11	KM10	Choi Hung Road	86.4	89.1	85.1	80.1	100	1218	1146	570	678	2364	1248	52.8	28.9
12	KM03	Tai Po Road	76.1	79.3	74.8	67.8	86.6	1068	1374	486	588	2442	1074	44.0	36.7
13	NM34	Texaco Road	85.7	88.9	82.9	78.4	103.3	1092	1488	594	516	2580	1110	43.0	26.9
14	HM14	Wong Nai Chung Rd	74.1	76.9	71.4	64.4	100.8	1362	1362	342	342	2724	684	25.1	40.7
25	KM09	Ma Tau Wai Road	79.4	83.8	73.8	67.3	96.7	1248	1650	402	702	2898	1104	38.1	24.7
16	KM09	Argyle Street	82.8	86.1	80.6	75.6	96.3	1806	1584	528	342	3390	870	25.7	38.7
17	KM16	Kwun Tong Road	88.8	91.4	87.9	84.9	98.5	1890	2064	954	900	3954	1854	46.9	45.6
18	KM09	Waterloo Road	85.4	88.1	84.6	81.1	99.7	2616	2220	853	617	4836	1470	30.4	36.9

of time, the number of light vehicles and the number of heavy vehicles were recorded simultaneously at the near lane(s) as well as the far lane(s). The classifications of light vehicles and heavy vehicles in accordance with the Hong Kong Environmental Protection Department's guideline are given as follows:

- Light vehicles including private cars, motor-cycles, taxi, and light good vehicles.
- Heavy vehicles including private and public light buses, private and public buses, lorries, trucks, container vehicles and all other vehicles with an unladen weight exceeding 1525 kg.

Since most prediction models would be used to predict an hourly L_{10} noise level due to road traffic, the number of light vehicles and the number of heavy vehicles counted were extrapolated to hourly flow. The total number of vehicles (or the total traffic flow), Q, was then determined by summing the hourly numbers of light vehicles, L, and heavy vehicles, H (i.e., $Q=H+L$). The average speed of vehicles was calculated by recording the time of vehicles passing over a known distance of 25 m. All the measurements were taken place at the daytime period of 07:00–19:00. It should be noted that measurements would be inhibited when traffic congestion that could cause abnormally high engine noise by the start–stop operation of vehicles was observed.

III. RESULTS AND DATA ANALYSIS

Table I presents the measured noise levels and the calculated average speed, the total number of vehicles, and the number of heavy vehicles. It shows that the measured L_{10} noise level ranged from 73.4 dBA to 91.4 dBA. The total number of vehicles ranged from 540 to 4836 per hour. The number of heavy vehicles ranged from 198 to 1854 per hour. From the information obtained from the Hong Kong Department of Transport,[15] the average age of light vehicles was 4.5 years while the average age of heavy vehicles was around 7 years. The calculated vehicle speed in these urban roads was between 25.6 km/h and 45.6 km/h in which most vehicles cruised at constant speeds and over-takings rarely happened on the roads. Twelve out of 18 major roads had a percent of heavy vehicles exceeding 30%. It should be noted that the average vehicle density could be obtained by using the total number of vehicles and the calculated vehicle speed. For examples, there were 540 vehicles at an average speed of 25.6 km/h at Site No. 1 and the vehicle density of this road was 540/25.6=21 vehicles/km. At Site No. 18, the calculated vehicle density was 4836/36.9=131 vehicles/km.

Experience[2–10] showed that L_{10} noise level depends on the logarithmic functions of traffic data. In this research, we first assumed the logarithmic values of the hourly total traffic flow (Log Q), the number of heavy vehicles (Log H), the number of light vehicles (Log L), and the average speed of vehicles (Log V) to be the most significant factors. These parameters were then analyzed by the analysis of variance (ANOVA)[16,17] in which the significance of each parameter could be ranked quantitatively. The ANOVA results were substantiated by using Tukey's studentized method. It should be noted that three categories were subdivided for each group of factors according to their value in ascending order as shown in Table II in both ANOVA and Tukey's studentized method.

TABLE II. Categorization of parameters for ANOVA test.

Parameters	First group	Second group	Third group
Total traffic flow Q, vehicles/h	$Q<1000$	$1000\leqslant Q<2000$	$2000\leqslant Q$
Heavy vehicles H, vehicles/h	$H<400$	$400\leqslant H<800$	$800\leqslant H$
Light vehicles L, vehicles/h	$L<600$	$600\leqslant L<1200$	$1200\leqslant L$
Average speed V, km/h	$V<30$	$30\leqslant V<40$	$40\leqslant V$

TABLE III. *f*-statistics results for the significant factors.

Factors	Degree of freedom		Mean square		*f*-value	*p*-value
	Treatment	Error	Treatment	Error		
Log Q	2	15	95.25	19.73	4.83	0.024
Log H	2	15	133.35	14.65	9.10	0.003
Log L	2	15	60.60	24.34	2.49	0.116
Log V	2	15	3.24	32.00	0.10	0.904

The hypothesis of ANOVA was, ''the logarithmic function of the total traffic flow, the number of heavy vehicles, the number of light vehicles or the average speed of vehicles has an effect on L_{10} noise level:''

$$H_0: \ \mu_{\log Q_1} = \mu_{\log Q_2} = \mu_{\log Q_3}$$

(same for $\log H$, $\log L$, and $\log V$)

H_a: not all equal.

The ANOVA output from SAS software is summarized in Table III. From the *F*-distribution table for a numerator, denominator degree of freedom of 2, 15, respectively and an α-type I error of 0.05, $F_{(0.05,2,15)}$ is equal to 3.68. Statistically speaking, the hypothesis would be rejected if one (or more) *p*-value was less than 0.05 (or *f*-value to be greater than 3.68). Table III shows that some of the null hypotheses failed and some parameters should have an effect on the measured L_{10} noise levels. Further analyses using Tukey's studentized method ($p<0.05$) indicated that $\log Q$ and $\log H$ were the two most significant factors of urban traffic noise. They were followed by $\log L$ and $\log V$.

A multiple regression analysis of L_{10} noise levels using $\log Q$ and $\log H$ gives

$$L_{10}(Q,H) = 6.9 \log Q + 8.5 \log H + 35.3 \quad (R = 0.76). \tag{1}$$

The correlation (*R*-value) between the measured L_{10} noise levels and the regenerated L_{10} noise levels is 0.76. In order to check whether a better regression model was available, multiple-regression analyses were performed in taking $\log Q$, $\log H$, $\log L$, and $\log V$ into consideration. The resulting multiple regression formulas are

$$L_{10}(Q,H,L,V) = 10.6 \log Q + 7.1 \log H - 0.04 \log L - 17 \log V + 53.1 \quad (R = 0.79), \tag{2}$$

$$L_{10}(Q,H,V) = 10.6 \log Q + 7.1 \log H - 17 \log V + 53 \quad (R = 0.79). \tag{3}$$

The correlation between the measured L_{10} data and the regenerated L_{10} data improved from 0.76 to 0.79. By inspecting the coefficients in Eqs. (2) and (3) carefully, one can observe that the coefficient of $\log L$ in Eq. (2) is very small while the coefficients of other parameters in these two equations are almost identical. As a result, $\log L$ can be eliminated and Eq. (2) reduces to Eq. (3). The mean difference and the root mean square difference between the measured L_{10} data and the regenerated L_{10} data by using Eq. (3) is shown in Table IV. The CRTN predictions based on the traffic data collected are shown in the last column of Table IV for comparison. The predicted L_{10} noise level using the CRTN model is

$$L_{10}(\text{hourly}) = 10 \log Q + 3.3 \log(V + 40 + 500/V) + 10 \log(1 + 5P/V) - 26.6, \tag{4}$$

where *P* is the percentage of heavy vehicles. Distance attenuation and the façade effect were added into the calculated

TABLE IV. Diffeences between the measured and regenerated L_{10} values.

Site no.	Measured L_{10}	Regenerated L_{10} using Eq. (3)	Difference	Regenerated L_{10} using Eq. (4) w/att.	Difference
1	73.4	74.9	−1.5	80.0	−6.6
2	79.4	77.2	2.2	78.9	0.5
3	79.3	77.3	2.0	80.3	−1.0
4	79.4	74.6	4.8	81.9	−2.5
5	77.4	80.5	−3.0	83.8	−6.4
6	73.6	77.2	−3.6	82.9	−9.3
7	81.8	81.8	0.0	84.2	−2.4
8	81.9	81.3	0.6	83.7	−1.8
9	78.8	81.6	−2.8	82.5	−3.7
10	85.6	83.6	2.0	84.5	1.1
11	89.1	85.9	3.2	86.8	2.3
12	79.3	83.8	−4.5	86.0	−6.6
13	88.9	86.5	2.4	86.5	2.4
14	76.9	82.2	−5.3	84.5	−7.6
15	83.8	87.6	−3.8	86.8	−3.0
16	86.1	84.3	1.8	85.5	0.6
17	91.4	86.1	5.3	88.5	2.9
18	88.1	87.9	0.2	87.6	0.5
Mean			0.0		−2.3
Sample standard deviation			3.2		3.8

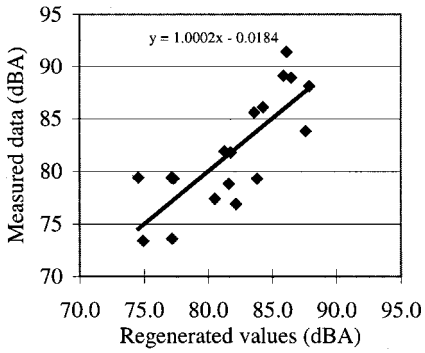


FIG. 3. Scattering diagram plotting the measured L_{10} noise levels against the regenerated L_{10} noise levels using Eq. (3) and the recorded traffic data.

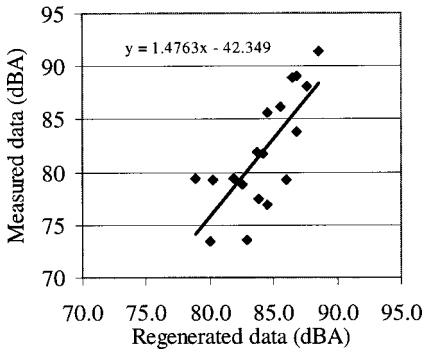


FIG. 4. Scattering diagram plotting the measured L_{10} noise levels against the regenerated L_{10} noise levels using CRTN and the recorded traffic data.

CRTN values according to the text advised.[4] Figures 3 and 4 show the plots of the regenerated L_{10} values versus the measured L_{10} data.

The mean differences in Table IV indicated that Eq. (3) is better than Eq. (4). This was confirmed by observing the best-fit lines shown in Figs. 3 and 4, respectively, because the slope of the line in Fig. 3 is much closer to $+1$ when comparing to that of Fig. 4. The calculated results showed that the CRTN model could not give accurate L_{10} predictions in urban areas. In most cases, the CRTN model overestimated L_{10} noise level by 2–6 dB.

Equation (3) shows that L_{10} noise level normally increases by 3 dBA when the total traffic flow increases by two times. The effect of increasing the number of heavy vehicles has a compounded effect on the overall L_{10} noise level because a heavy vehicle is much noisier than a light vehicle. The coefficient of the logarithmic function of the average vehicle speed was negative. It is because we have an exceptionally high proportion of heavy vehicles (around 40%) and a relatively low speed (25.6–45.6 km/h) in urban areas in Hong Kong. Dixon and Phillips[18] reported that the noise generated by a diesel engine would slightly decrease when the engine speed changed from 2000 to 3000 rpm (corresponding to 25 km/h to 40 km/h) and would increase at speeds higher than 50 km/h or 3000 rpm.

IV. MODEL VALIDATION

In the winter months of 2000, roadside traffic noise measurements were carried out in another 8 urban areas as shown in Fig. 1. The method of noise measurement and the equipment used were identical to the ones described in Sec. III. Moreover, a sound recording system including a SONY MZ-R75 mini-disk recorder, a B&K 4165 condenser microphone and a B&K 5935 dual microphone supply was used to record the traffic noise at each of the measurement sites. A SONY PC-1E digital video camera was used to record traffic conditions. The traffic data, the measured L_{10} noise levels, and the predicted L_{10} noise levels based on Eq. (3) using the recorded traffic information are shown in Table V. In this set of recorded data, the total number of vehicles ranged from 780 to 3144 vehicles per hour. The number of heavy vehicles ranged from 260 to 1124 vehicles per hour. The calculated vehicle speed was between 26.9 km/h and 36.9 km/h. Six out of eight major roads had a percent of heavy vehicles exceeding 30% (around 40% in average). One can observe that the maximum difference between the measured L_{10} noise levels and the predicted L_{10} noise levels using Eq. (3) is 3.6 dBA while the mean difference between these two sets of data is 0.4 dBA. It should be noted that the re-analysis of measurement data using the recorded sound and digital images confirmed the measured L_{10} noise levels, the traffic counts, and the calculated vehicle speed to be accurate.

V. CONCLUSION

A roadside traffic noise survey was conducted in heavily built-up urban areas in Hong Kong in 1999. Relatively high L_{10} noise levels ranging from 73.4 dBA to 91.4 dBA were

TABLE V. Traffic data, the measured L_{10} noise levels, and the predicted L_{10} noise levels using Eq. (3).

Site no.	Road	Hourly total veh. Q	Hourly heavy veh. H	% of heavy veh. P	Ave. speed (km/h) V	Measured L_{10} (dBA)	Predicted L_{10} using Eq. (3) (dBA)	Difference
N1	Hennessy Road	984	552	56.1	26.9	82.9	79.9	−3.0
N2	Des Voeux Road	780	492	63.1	26.3	81.9	78.6	−3.3
N3	Lai Chi Kok Road	1088	332	30.5	35.9	79.4	76.7	−2.7
N4	Cheung Sha Wan Rd	2692	1031	38.3	31.7	84.9	85.2	0.3
N5	Nathan Road	1604	524	49.2	30.8	80.9	81.0	0.1
N6	Waterloo Road	1412	260	18.4	30.4	78.4	78.3	−0.3
N7	Kwun Tong Road	3144	1124	35.8	34.9	81.9	85.5	3.6
N8	Ma Tau Chung Rd	1920	560	29.1	36.9	78.4	80.7	−2.3
Mean								−0.9
Sample standard deviation								2.4

recorded in those areas where the total traffic flow was between 540 and 4836 vehicles per hour. Twelve out of the 18 selected roads had a percent of heavy vehicles exceeding 30%. Statistical analysis using the ANOVA and Tukey test ($p<0.05$) reveals that the total traffic flow and the number of heavy vehicles are the most significant factors. A multiple regression model showing L_{10} noise level to be logarithmic functions of the total traffic flow, the number of heavy vehicles and the average speed of vehicles was obtained. The applicability of this model was validated by another set of roadside traffic noise data obtained in the winter of 2000. It is recommended that with the advance of digital equipment, portable mini-disk recorder and digital video camera shall be deployed in future noise measurements. This set of audiovisual recording equipment can allow engineers to have a ''feel'' again in the laboratory environment and provide a great opportunity for data re-analysis.

ACKNOWLEDGMENTS

The authors wish to thank P. S. Ng for his helpful discussions and useful comments on this paper. Part of the measured noise and traffic data was provided by the Government of Hong Kong Special Administration Region. The findings and opinions in the paper belong solely to the authors, and do not reflect any government's opinions.

[1] R. J. Stephenson and G. H. Vulkan, ''Traffic noise,'' J. Sound Vib. **7**, 242–262 (1968).
[2] P. T. Lewis, ''The noise generated by single vehicles in freely flowing traffic,'' J. Sound Vib. **30**, 191–206 (1973).
[3] M. E. Delany, D. G. Harland, R. A. Hood, and W. E. Scholis, ''The prediction of noise levels L_{10} due to road traffic,'' J. Sound Vib. **48**, 303–325 (1976).
[4] The UK Department of Environment, *Calculation of Road Traffic Noise* (Her Majesty's Stationary Office, London, 1975).
[5] D. Gilbert, ''Noise from road traffic (interrupted flow),'' J. Sound Vib. **51**, 171–181 (1976).
[6] T. M. Barry and J. A. Reagan, ''FHWA highway traffic noise prediction model,'' Report No. FHWA-RD-77-108, US Federal Highway Administration, Washington, DC (1978).
[7] J. J. Hajek, ''L_{eq} traffic noise prediction model,'' Transportation Research Record No. 648, Transportation Research Board, National Academy of Sciences, pp. 48–52 (1977).
[8] R. R. K. Jones, D. C. Hothersall, and R. J. Salter, ''Techniques for the investigation of road traffic noise in regions of restricted flow by the use of digital computer simulation methods,'' J. Sound Vib. **75**, 307–322 (1981).
[9] C. H. Chew, ''Prediction of traffic noise from expressways—Part I: buildings flanking one side of expressways,'' Appl. Acoust. **28**, 203–212 (1989).
[10] C. H. Chew and K. B. Lim, ''Façade effects on the traffic noise from the expressway,'' Appl. Acoust. **41**, 47–62 (1994).
[11] D. Chakrabarty, S. C. Santra, and A. Mukherjee, ''Status of road traffic noise in Calcutta metropolis, India,'' J. Acoust. Soc. Am. **101**, 943–949 (1997).
[12] A. A. Saadu, R. O. Onyeonwu, E. O. Ayorinde, and F. O. Ogisi, ''Road traffic noise survey and analysis of some major urban centers in Nigeria,'' Noise Control Eng. J. **46**, 146–158 (1998).
[13] Hong Kong Information Services Department (2000). Hong Kong Annual Report 1999, Chapter 14: Transport.
[14] B. Oakes and M. A. Tomlinson, ''A note on the measurement of traffic noise in congested urban situations,'' Appl. Acoust. **6**, 319–322 (1973).
[15] Hong Kong Department of Transport (2001), ''The statistics on number of registered/licensed vehicles by year of manufacture,'' private communication.
[16] J. L. Devore, *Non-linear and Multiple Regression. Probability and Statistics for Engineering and the Sciences*, 3rd ed. (Brooks/Cole Publishing Company, **CITY**, 1991), pp. 501–557.
[17] R. P. Cody and J. K. Smith, *Applied Statistics and the SAS Programming Language*, 4th ed. (Prentice Hall, New York, 1997).
[18] J. Dixon and A. V. Phillips, ''Power unit low-frequency airborne noise,'' in *IMechE Conference 1998*, edited by the European Conference on Vehicle Noise and Vibration (Professional Engineering Publishing Limited, London, 1998), pp. 71–85.

Distribution of risk factors for hearing loss: Implications for evaluating risk of occupational noise-induced hearing loss

Mary M. Prince
Industrywide Studies Branch, Division of Surveillance, Hazard Evaluations, and Field Studies, National Institute for Occupational Safety and Health, 4676 Columbia Parkway, Cincinnati, Ohio 45226

(Received 4 April 2001; revised 3 March 2002; accepted 24 May 2002)

This paper presents an analysis of hearing threshold levels among 2066 white male workers employed in various U.S. industries studied in the 1968–72 NIOSH Occupational Noise and Hearing Survey (ONHS). The distribution of hearing threshold levels (HTL) is examined in relation to various risk factors (age, prior occupational noise, medical conditions) for hearing loss among a population of noise exposed and control (low noise-exposed) industrial workers. Previous analyses of a subset of these data from the ONHS focused on 1172 highly ''screened'' workers. An additional 894 male workers (609 noise-exposed and 285 controls), who were excluded for various reasons (i.e., nonoccupational noise exposure, otologic or medical conditions affecting hearing, prior occupational noise exposure) have been added to examine hearing loss in an unscreened population. Data are analyzed by age, duration of exposure, and sound level (8-h TWA) by individual test frequency. Results indicate that hearing threshold levels are higher among unscreened noise-exposed and control workers relative to screened workers. Analysis of risk factors such as nonoccupational noise exposure, medical conditions, and type of industry among unscreened controls indicated that these factors were not significantly associated with increased mean HTLs or risk of material impairment over and above what is expected due to age. Age-specific mean hearing threshold levels (and percentiles of the distribution) among the unscreened ONHS control population may be used as a comparison population of low-noise exposed white male industrial workers for evaluating the effectiveness of hearing conservation programs for workers less than 55 years of age. To make valid inferences regarding occupational noise-induced hearing loss, it is important to use hearing data from reference (control) populations that are similar with respect to the degree of subject screening, type of work force (blue vs white collar), and the distribution of other risk factors for hearing loss. [DOI: 10.1121/1.1494993]

PACS numbers: 43.50.Qp, 43.64.Wn [MRS]

I. INTRODUCTION

Current U.S. noise regulations require that employers in the manufacturing and mining industries of the U.S. enroll employees exposed to daily average exposures of 85 dBA or more in a hearing conservation program and to occasionally monitor their noise levels (OSHA, 1983; MSHA, 1999). These noise regulations do not cover noise-exposed workers in small businesses (less than 100 workers) or in industries such as construction, agriculture, transportation, and public utilities, where excessive noise exposure and hearing loss is likely to be prevalent (NIOSH, 1988a, b, 1990).

The burden of hearing loss due to occupational noise, and perhaps other risk factors, is likely to be greatly underestimated, due to the lack of a systematic national health surveillance system to collect statistics on hearing loss and occupational noise exposure. Hence, relatively imprecise estimates exist of the number of noise-exposed workers and the prevalence of occupational noise-induced hearing loss (NIHL) in the U.S. workforce. It has been estimated that about 9 million U.S. workers are occupationally exposed to noise levels exceeding or equal to 85 dBA on a daily basis (Simpson and Bruce, 1981), with more than half of these workers engaged in manufacturing and utility. Approximately 1 million workers are estimated to suffer from work-related hearing loss, primarily from manufacturing industries (Weeks *et al.*, 1991). The high prevalence of occupational noise exposure and the fact that long-term NIHL is a preventable condition provides researchers and public health officials a unique opportunity to develop and test interventions that may prevent NIHL for millions of individuals.

Carefully designed epidemiological studies examining the relationship of contemporary workplace noise exposure and the distribution of noise-induced hearing loss among adult U.S. worker populations have not been conducted recently. Moreover, representative U.S. national data on the burden of hearing loss from occupational noise relative to other risk factors for hearing loss are either not available (Ries, 1994) or are limited to particular states or sectors of industry (Reilly *et al.*, 1998). Such large-scale data collection efforts tend to be expensive, and when conducted, often omit important occupational information (such as job title, noise exposure level, number of years employed in noisy jobs) that can be linked to audiometric testing data. Another limiting factor in evaluating occupational hearing loss in relation to noise and other risk factors is the greater availability of hearing protection devices (HPD) in most manufacturing settings in the U.S. (Davis and Sieber, 1998; Royster and Royster, 1984). Therefore, utilizing existing data is an economical way to gather information needed for planning future studies and evaluating the impact of changes in expo-

TABLE I. Distribution of study subjects by industry group.

Industry type	Total population ($N=2066$)		Exposed ($N=1401$)		Controls ($N=665$)	
	No.	%	No.	%	No.	%
Printing	906	43.9	603	43.0	303	45.6
Steel fabrication	355	17.2	160	11.4	195	29.3
Woodworking	270	13.1	211	15.1	59	8.9
Tunnel patrol	171	8.3	146	10.4	25	3.8
Paper bag mfg.	103	5.0	82	5.9	21	3.2
Aluminum fabrication and processing	123	6.0	105	7.5	18	2.7
Quarry	69	3.3	63	4.5	6	0.9
Hydroelectric	38	1.8	0	0.0	38	5.7
Trucking	31	1.5	31	2.2	0	0.0

sure, hearing protection use, and other hearing loss risk factors in the population.

This paper analyzes one of the key studies of U.S. workers from several industries conducted by the National Institute for Occupational Safety and Health (NIOSH; formerly Bureau of Occupational Safety and Health) during 1968–1972. This study remains historically and scientifically relevant because (1) a subset of the data (1172 ''screened'' white male population) were analyzed to form the basis of the current Occupational Safety and Health Administration (OSHA) noise standard (NIOSH, 1972; Lempert and Henderson, 1973; Prince *et al.*, 1997) and (2) it is one of the few existing data sets with raw data on hearing threshold levels, TWA noise levels, risk factor, demographic, and industry type.

Data from the larger NIOSH ONHS study, which includes workers excluded from the original analyses, have never been extensively evaluated or published elsewhere. Generalizing results from the prior risk analysis to an unscreened population may allow inferences to be drawn for working populations that are more representative of workers enrolled in industrial hearing conservation programs. The main objectives of the analysis are to examine risk profiles of the unscreened population to the highly subgroup screened group and to describe the characteristics of low noise-exposed populations for use as comparative data in analysis of contemporary hearing data from noise-exposed industrial workers. In this paper, the ''screened'' population will refer to the original 1172 workers analyzed previously (NIOSH, 1972; Prince *et al.*, 1997), while those excluded from analysis are referred to as the ''excluded'' population ($N=894$). The total population of workers examined in this analysis, formed by pooling the screened and excluded ONHS subpopulations, is referred to as the ''unscreened'' population.

II. METHODS

A comprehensive description of this study was previously published in an NIOSH technical report (Lempert and Henderson, 1973). Pertinent study methods provided from that report are summarized below.

A. Study population

The study population was drawn from volunteer companies responding to publicized announcements of the study from industrial hygiene conferences and Public Health Service distributions of the procedures and goals of the Occupational Noise and Hearing Survey (ONHS). All plants requesting noise and hearing evaluations were considered for the study if the following criteria were met: (1) factory or occupational noise conditions were relevant to development of a noise standard for general industry; (2) the work force represented a wide range of years of exposure to noise (based on walk-through noise surveys). Among the 13 plants subsequently selected for this study, 9 different industrial groups are represented in the survey. For plants having less than 500 employees, the entire work force was tested. In larger plants, a random sample of employees were selected. Employees from each plant who worked in offices or other quiet areas were also included in the survey as a comparison or control group. Among the 2066 workers selected for noise monitoring, 61% of the study subjects worked in printing and steel fabrication (Table I).

B. Data collection procedures

1. Noise characterization

Preliminary surveys of each plant provided information needed to develop a sampling strategy. Noise level surveys were conducted to assess daily noise exposures of workers. The following items were obtained for each plant: location and type of operation or work performed, general noise characteristics (e.g., impulsive, steady state, low frequency), temporal characteristics (e.g., continuous, fluctuating, intermittent), and overall noise levels using the ''A,'' ''C,'' and ''linear'' scales of the sound-level meter. Tape-recorded noise samples were played through octave band and third-octave band real-time analyzers for the band with center frequencies of 31.5, 63, 125, 250, 500, 1000, 2000, 4000, and 8000 Hz. Recorded tapes were played through an octave band filter set through a third-octave band, real-time analyzer.

Sequential sampling of field measurements of dBA levels at intervals of 15 seconds throughout the sample period

of 10 min were developed for areas where sound levels showed evidence of random variations with time. The Bruel and Kjaer (B&K) type 4420 distribution analyzer, in conjunction with the type 2305 level recorder, was used to obtain the probability distribution of sound levels over selected time intervals. Tape recordings were taken in the field to permit analysis in the laboratory. If impact sounds were present, measurements of peak-pressure levels were made using the B&K 2204S impact meter. If the impacts occurred so rapidly as to blend together, then the noise was regarded as being essentially continuous.

To determine daily noise exposure patterns for a worker or worker group, workmen and supervisors were interviewed. In many cases, time-study charts were prepared, segmenting the workday into a succession of exposures at specific noise levels and for specified durations. Discussions with both management and workers were necessary to determine changes in workers' noise exposures over the course of many years. Consideration was given to variations in occupational noise conditions due to machinery replacement or relocation and also to changes in work routine and location of workers. No mass use of hearing protection devices was observed in the companies surveyed (Prince *et al.*, 1997).

2. Audiometric and risk factor evaluation

All audiometric tests were done in a Rudmose Audiometric Travel Lab model RA-113. This audiometric van housed an acoustically isolated, sound-deadened chamber in which six persons could be tested simultaneously. Audiometric equipment consisted of a Rudmose RA-108 pure-tone six-person air-conduction audiometer that produced test tones at frequencies 0.5, 1, 2, 3, 4, and 6 kHz, presented first to the left ear and then to the right ear of each subject. Standard procedures for audiometer calibration and hearing tests were implemented as described by Lempert and Henderson (1973). Ambient noise levels in the testing booths were within the limits specified in the ANSI S3.1-1960 (R1971) standard (ANSI, 1960).

Each noise-exposed worker was tested before the beginning of this work shift to minimize temporary hearing threshold shifts from exposure to noise at the worksite immediately before the test. Test scheduling usually required subjects to arrive 30–75 min early; otoscopic examinations and administration of an questionnaire were also done at the time of testing. Non-noise-exposed workers were tested at any time during the work shift, since their pretest noise exposures were not considered significant enough to produce a temporary threshold shift. A questionnaire administered at the time of the hearing test collected information on other factors that might affect hearing loss risk, noise exposure, and interpretation of audiometric test results such as

(1) Previous job history of noise exposure—number of years worked in noisy environments.
(2) Military history, including amount of time spent using a weapon, in actual combat experience, and in routine daily exposure to nonweapon-type noise (e.g., aircraft engines, armored vehicles).
(3) Nonoccupational noise exposure, including civilian firearm use (number of rounds/year, with and without hearing protection), motorbike riding, mechanized farming, piloting an airplane, machine workshop activity, sport or drag car racing, and noisy hobbies (e.g., rock music).
(4) Medical history that included whether they ever had experienced severe head trauma, chronic ear infection, hereditary deafness, Meniere's disease, use of ototoxic drugs, ear surgery, concurrent severe head colds, or tinnitus at the time of audiometric testing.

3. Exclusion criteria

Given the study's primary aim to establish a valid statistical relationship between occupational noise exposure and hearing loss that would be applicable to general industry, NIOSH investigators excluded data from the analysis sample if (1) workers' noise exposure histories or validity of audiometric data were uncertain, and (2) observed hearing loss may have been caused by factors other than occupational noise.

Based on an evaluation of questionnaire data, audiometric test results, and otoscopic examinations, ''screened sample'' was constructed using the following exclusion criteria:

(1) Previous job history involving 2 or more years occupational noise exposure;
(2) Military history showing: (a) 100 days or more of weapons-type noise; (b) 1 or more years of actual combat experience; and/or (c) 2 or more years of routine daily exposure to nonweapon-type noise (e.g., aircraft engines, armored vehicles).
(3) Nonoccupational noise exposure, including civilian firearm use (1000 rounds/year for 1 or more years or 500 rounds/year for 5 or more years with no hearing protection use), and any noisy hobbies previously listed above if participation was at least 3 times per week for 1 or more years.
(4) History of any of the medical conditions or ototoxic drugs that could affect hearing.
(5) Otoscopic examination indicating congenital or acquired ear malformations, almost total occlusion of ear canal by cerum, perforated or severely scarred tympanic membrane or otitis media.
(6) Audiometric irregularities such as (a) suspected conductive loss in one or both ears; (b) large unilateral hearing losses (e.g., loss in one ear is 40 dB greater than the other ear at two or more test frequencies); and (c) suspected subject response to tinnitus rather than tone presentation.
(7) Significant pretest noise (significant noise exposure within 14 h prior to testing).

Approximately 1287 (35%) of the total 3699 workers were excluded because they met one or more of the above criteria. No more than three exclusions per worker were assigned. These exclusion procedures produced a sample of 2412 used to estimate the impact of industrial noise exposures on the occupational groups included in the individual

TABLE II. Exclusion criteria by exposure group.

Exclusion criteria categories	Total population ($N=894$)		Exposed ($N=609$)		Controls ($N=285$)	
	Number	Percent	Number	Percent	Number	Percent
Noise exposure history	**574**	**64.2**	**395**	**64.9**	**179**	**62.8**
Previous job noise	358	40.0	252	41.4	106	37.2
Pre-test noise[a]	21	2.4	19	3.1	2	0.7
Military Noise	134	15.0	85	14.0	49	17.2
- *weapons*	*112*	*12.5*	*72*	*11.8*	*40*	*14.0*
- *non-weapon*	*22*	*2.5*	*13*	*2.1*	*9*	*3.2*
Leisure activity noise	61	6.8	39	6.4	22	7.7
- *weapons*	*30*	*3.4*	*17*	*2.8*	*13*	*4.5*
- nonweapon	*31*	*3.5*	*22*	*3.6*	*9*	*3.2*
Medical history	**91**	**10.2**	**58**	**9.5**	**33**	**11.6**
Otologic abnormality	**67**	**7.5**	**39**	**6.4**	**28**	**9.8**
Multiple causes	**162**	**18.1**	**117**	**19.2**	**45**	**15.8**

[a]Statistically significant difference between exposed and control groups for pretest noise exposure ($X_1^2=4.95$, $p=0.03$).

plant surveys. Additional exclusion criteria were applied to the remaining workers ($N=2412$) to produce a ''composite ONHS population'' that could ''accurately determine the risk to hearing as a function of noise level'' (Lempert and Henderson, 1973). In particular, data were excluded for workers with insufficient noise exposure data or noise consisting of discrete impact sounds or noise with highly variable and unpredictable levels. All maintenance workers were excluded because of exposure uncertainty. Women were excluded due to inadequate sample size for drawing conclusions concerning the relationship between noise level and hearing loss (Lempert and Henderson, 1973). This remaining population was comprised of 1172 predominately white male workers, whose data have been previously analyzed (NIOSH, 1972; Prince *et al.*, 1997).

Based on a reconstruction of these exclusion criteria, the 2066 workers analyzed for this paper represent predominately white male workers with well-defined noise exposure data but whom have not been screened for conditions related to hearing loss commonly found in the general population of workers.

C. Analysis of the unscreened ONHS data

Demographic (age, industry type, duration employed or exposed), noise exposure (8-h TWA sound levels), and hearing health characteristics (mean and median HTLs, percent impairment) of the unscreened ONHS population are presented, as well as a breakdown of the population by exposure (control versus noise-exposed) and screening status (i.e., whether the subset of the unscreened population represents the 1172 screened or the 894 excluded workers). Analysis of the distribution of mean and median HTLs by age among the low noise-exposed ''control'' group for this population was conducted for purposes of comparison to the ANSI S3.44 (ANSI S3.44, 1996) and ISO 1999 (ISO, 1990) standards as well as other published reports of hearing levels among industrial workers (Berger *et al.*, 1977; Driscoll and Royster, 1984; Royster and Thomas, 1979; Royster *et al.*, 1980) and nonindustrial populations (Morrell *et al.*, 1996). Predicted mean HTL by age and frequency among the screened and unscreened ONHS control populations were examined using a least-squares regression approach (Draper and Smith, 1966), where linear and quadratic (second-order effects) of age were tested.

The effects of age and other risk factors (nonoccupational exposure, previous job noise exposure, medical conditions, military service) on risk of material hearing impairment [binaural pure tone averages (PTA) of >25 dB for frequency averages of 1–4 kHz, 0.5–2 kHz, and 3, 4, and 6 kHz; denoted as PTA-1234, PTA-512, PTA-346] among controls are examined using logistic regression methods (Breslow and Day, 1980). Statistical analyses were conducted using SAS (SAS Institute, Inc., 1989) and graphical displays of the data were produced using S-PLUS (MathSoft, 1997).

III. RESULTS

A. General population characteristics

Table II shows the number of workers who failed the criteria for each category (expressed as percentage of total) or who failed at least one of the criteria. No statistically significant differences between exposed and controls were observed for most criteria except pretest noise exposure. The proportion of noise-exposed workers with pretest noise exposure was higher than the proportion in controls [3.1% vs 0.7%, ($X_1^2=4.95$, $p=0.03$)]. However, examination of the mean and distribution of HTLs by frequency and ear do not show large differences (<5 dB HL) between the exposed and controls workers initially excluded due to previous noise exposure.

General population characteristics are presented for the screened ($N=1172$), excluded ($N=894$), and unscreened ONHS population ($N=2088$) in Table III. The mean age of the ONHS population is about 40 years. The mean age in screened controls was significantly lower than the mean age in the excluded control group. Among the screened population, the control group had a higher proportion of workers under age 29 than the exposed group. Given the selection criteria described for the study, the population represents a

TABLE III. Population characteristics by exposure status.

Variable	Total unscreened exposed population ($N=1401$)		Total unscreened control population ($N=665$)		Screened				Excluded			
					Controls ($N=380$)		Exposed ($N=792$)		Controls ($N=285$)		Exposed ($N=609$)	
Age	#	%	#	%	#	%	#	%	#	%	#	%
<29	254	18.1	165	24.8	114	30.0	150	18.9	51	17.9	104	17.1
29–35	256	18.3	128	19.3	86	22.6	170	21.5	42	14.7	86	14.1
35–43	315	22.5	131	9.7	68	17.9	171	21.6	63	22.1	144	23.7
44–50	262	18.7	111	16.7	49	12.9	124	15.7	62	21.8	138	22.7
50+	314	22.4	130	19.6	63	16.6	177	22.3	67	23.5	137	22.5
Mean±s.d.	40.5±11.6		38.8±11.9		37.0±11.9		39.9±11.8		41.2±11.6		41.3±11.2	
- Range	13–71		17–69		17–65		18–71		18–67		13–67	
- 10th, 90th percentiles	(25, 56)		(24, 56)		(23, 55.5)		(25, 57)		(25, 57)		(24, 55)	
Duration exposed[a]	#	%	#	%	#	%	#	%	#	%	#	%
<1	56	4.0	41	6.2	25	6.6	17	2.1	16	5.6	39	6.4
1–2	86	6.1	61	9.2	34	8.9	46	5.8	27	9.5	40	6.6
2–4	210	15.0	168	25.3	99	26.1	109	13.8	69	24.2	101	16.6
5–10	390	27.8	140	21.0	76	20.0	215	27.1	64	22.5	175	28.7
11–20	373	26.6	153	23.0	91	23.9	209	26.4	62	21.7	164	26.9
21+	286	20.4	102	15.3	55	14.5	196	24.8	47	16.5	90	14.8
Mean±s.d.	12.3±9.7		9.9±9.3		9.9±9.3		13.6±10.4		9.9±9.4		10.6±8.6	
- Range	0.25–46		0–41		0–40		0–46		0–41		0–42	
- 10th, 90th percentiles	(1, 27)		(1, 25)		(1, 25)		(2, 29)		(1, 25)		(1, 24)	
Daily noise level (dBA)	#	%					#	%			#	%
80–87	683	48.8	<80 dBA		<80 dBA		336	42.4			347	57.0
88–92	350	25.0					222	28.0			128	21.0
⩾92	368	26.3					234	29.6	<80 dBA		124	22.0
Mean±s.d.	89.0±4.2						89.5±4.2				88.5±4.1	
- Range	80–102						80–102				80–102	
- 10th, 90th percentiles	(86, 94)						(86, 94)				(86, 94)	

[a]For controls, ''duration exposed'' is duration of employment.

relatively stable workforce. The mean duration of noise exposure is about 12 years, with the screened population having slightly higher duration of exposure. For controls, the mean duration employed was about 10 years. With regard to noise exposure levels, a larger proportion of workers in the excluded group were exposed to 8-h time weighted average (TWA) sound levels of 80–87 dB as compared to the screened group ($X_2^2=29.2,\ p=0.001$).

As shown in Fig. 1, binaural hearing levels among low noise-exposed controls was higher at all frequencies in the unscreened control group than screened controls. However, differences in control HTL for these populations are relatively small among workers younger than 36 years. These differences were most pronounced between older workers at the highest audiometric frequencies. Figure 2 shows the median hearing threshold levels (HTL) and 10th and 90th percentiles by age and dBA level for the unscreened and unscreened ONHS populations. Similar patterns with noise exposure and age are observed in both groups, namely an increase in median HTL with increasing age and noise exposure. Comparable graphs for the screened population can be found in Fig. 1 of Prince *et al.* (1997).

The distribution of hearing threshold levels by exposure status and industry type for the unscreened ONHS population was examined and generally showed typical patterns of loss by frequency, age, and exposure status. Mean ages between controls and exposed groups were not significantly different across all industries. As expected, mean HTLs and rates of impairment were higher among the exposed versus the controls. Mean ages of workers were highest for printing (40.8 years in controls and 43.8 years in exposed) and quarry (44.8 years in controls and 42.3 years in exposed), and lowest among wood-working and paper bag manufacturing (mean ages 32 and 36 years, respectively). As to patterns of hearing loss across industry, all but the printing industry jobs showed significant noise notches among exposed workers (relative to controls) at 3 and 4 kHz. The distribution of overall HTLs by industry types suggest that some of the variability is due to differences in age, exposure years, and noise levels across these groups. Adjustment for age, exposure years, and noise levels is likely to produce mean HTLs that are fairly consistent across industry categories. The effects of industry type and other factors on age-related hearing loss are explored further in the logistic regression analysis using the low noise-exposed (''control'') population.

B. Effects of age: Analysis of control populations

Predicted hearing threshold levels by age were generated for binaural averages for each test frequency and selected frequency averages (Fig. 3). Quadratic effects of age in the unscreened ONHS population were significant for the low to mid-frequencies (0.5, 1, 2, and 3 kHz), but not for the higher

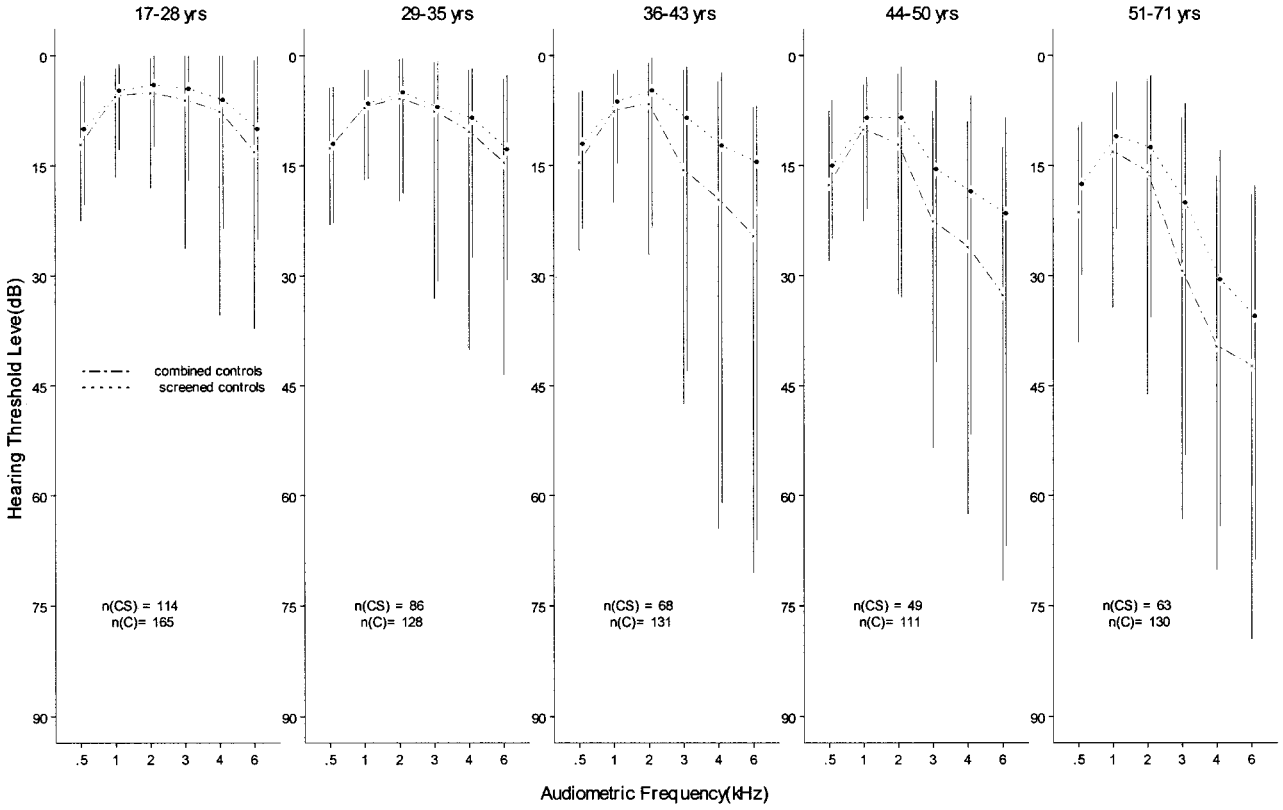


FIG. 1. Comparison of median, 10th, and 90th percentile binaural HTL among the screened and unscreened (combined) controls.

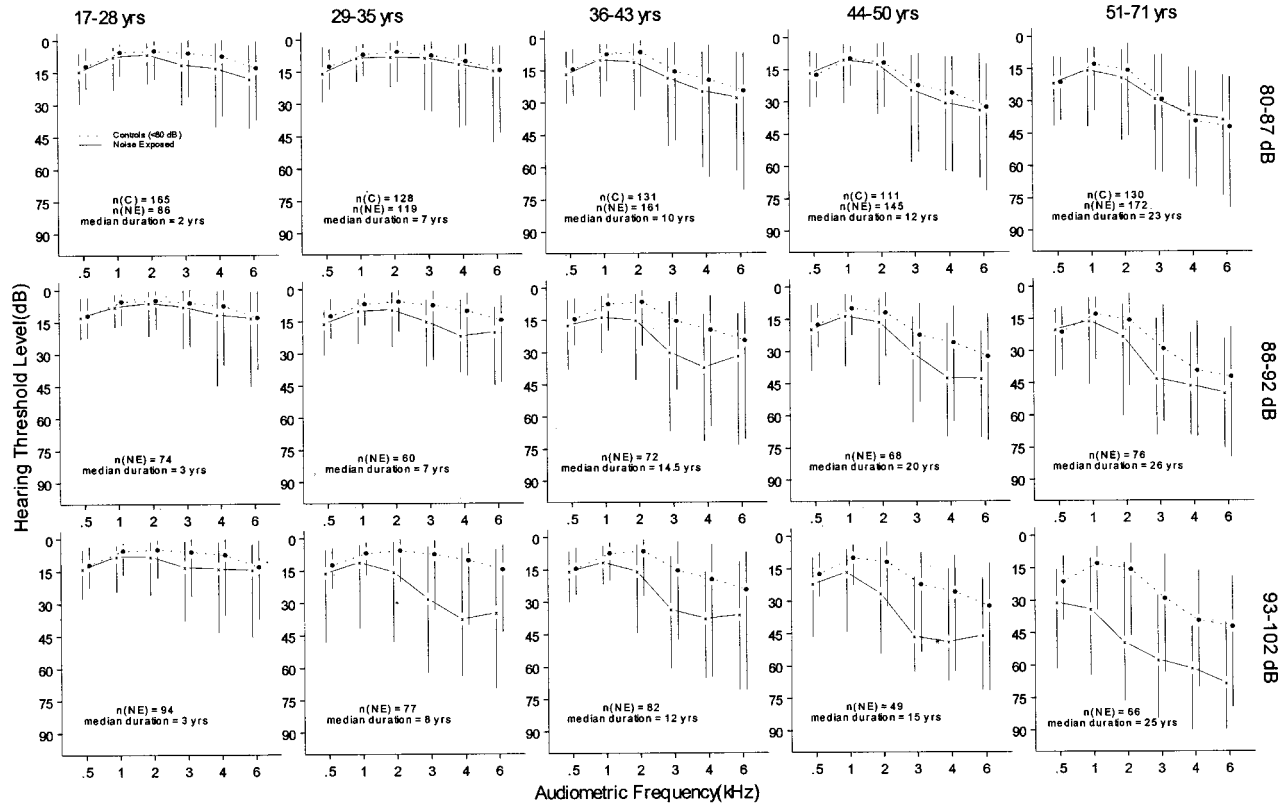


FIG. 2. Comparison of median, 10th, and 90th percentiles binaural HTL by age and dB sound level for the unscreened ONHS population by exposure status.

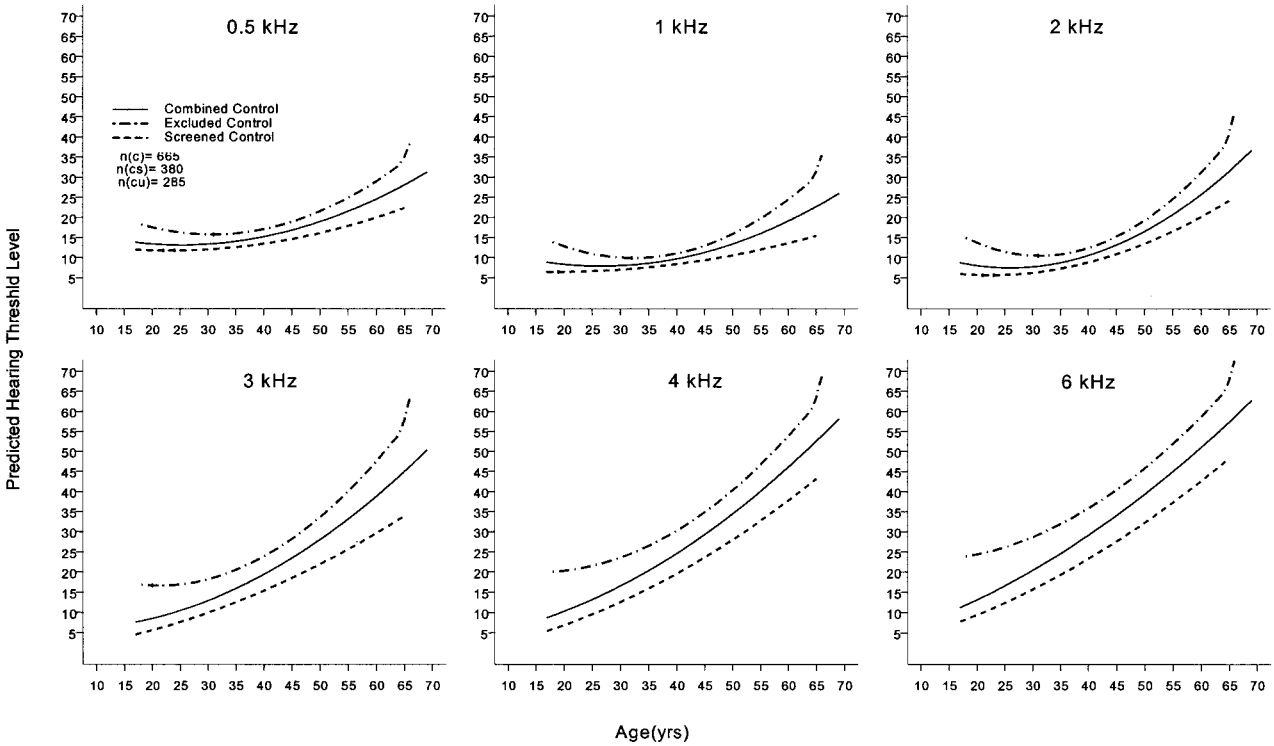


FIG. 3. Predicted hearing threshold levels by age and test frequency.

frequencies (4 and 6 kHz). The excluded control group always had higher HTLs than the screened control group for each frequency.

The means for the age categories (17–29, 30–39, 40–49, 50–59, and 60–69) were used to represent age for comparison to ANSI S3.44, 1996 and ISO 1999 (1990). Except for 500-Hz HTLs, values for this population are similar to those in annex C for white males. HTLs at 500 Hz. are higher than those observed in other populations (Driscoll and Royster, 1984; Robinson, 1970; Robinson and Sutton, 1975; Passchier-Vermeer, 1968; Morrell *et al.*, 1996), most likely due to high noise levels in the audiometric test booths. The unscreened control group had higher median HTLs and 10th and 90th percentiles than the screened control group. A comparison of median HTL among controls in the unscreened ONHS population to those based on annex B (ANSI S3.44, 1996) indicate that except for age 60, median HTLs (binaural average for each test frequency) are similar (Fig. 4).

For the screened control group, comparisons to annex A and to the Baltimore Longitudinal Study on Aging [BLSA] (Morrell *et al.*, 1996) showed higher median HTLs for 500 Hz at all ages (30, 40, 50, and 60 years) examined but relatively similar median HTL for test frequencies 1, 2, and 3 kHz. The differences between the BLSA and NIOSH screened controls decrease with increasing age for frequencies 1, 2, and 3 kHz, but generally increase with increasing age for 0.5 and 4 kHz. The observed differences of 3–9 dB (higher HTLs in the NIOSH screened controls vs BLSA) across age groups for 1, 2, and 3 kHz may be due to differences in ANSI audiometric test standards, audiometric test equipment (manual versus automatic), and procedures. Larger differences (11–20 dB across age categories) in median HTL at 4 kHz was observed between the BLSA screened population (Morrell *et al.*, 1996) and NIOSH screened controls.

C. Effect of other risk factors: Analysis of control populations

In this analysis, we examined age-adjusted odds ratios for material hearing impairment among low noise-exposed (control) workers in the ONHS unscreened population relative to various risk factors collected as part of the study. The results of the logistic regression analysis indicate that after adjustment for age, factors such as previous job exposure, nonoccupational sources of noise (leisure and military), and medical history or otologic abnormalities were not significant contributors of increased hearing impairment in the low noise-exposed unscreened ONHS population (Table IV). Although elevated rates of impairment were not statistically significant for any of the factors examined, slight elevations were noted for military weapon exposure and nonleisure weapon exposure for the frequencies most sensitive to noise-induced hearing loss (3, 4, and 6 kHz). For impairment associated with the speech frequencies (1–4 kHz and 0.5–2 kHz), elevated risk was observed for otologic abnormalities and military and leisure weapon noise exposure (0.5–2 kHz only). After adjustment for age and other factors, the type of industry was not associated with increased impairment rates among low-noise-exposed workers.

IV. DISCUSSION

The analysis focused on examining the distribution of hearing levels by various risk factors and evaluated age ef-

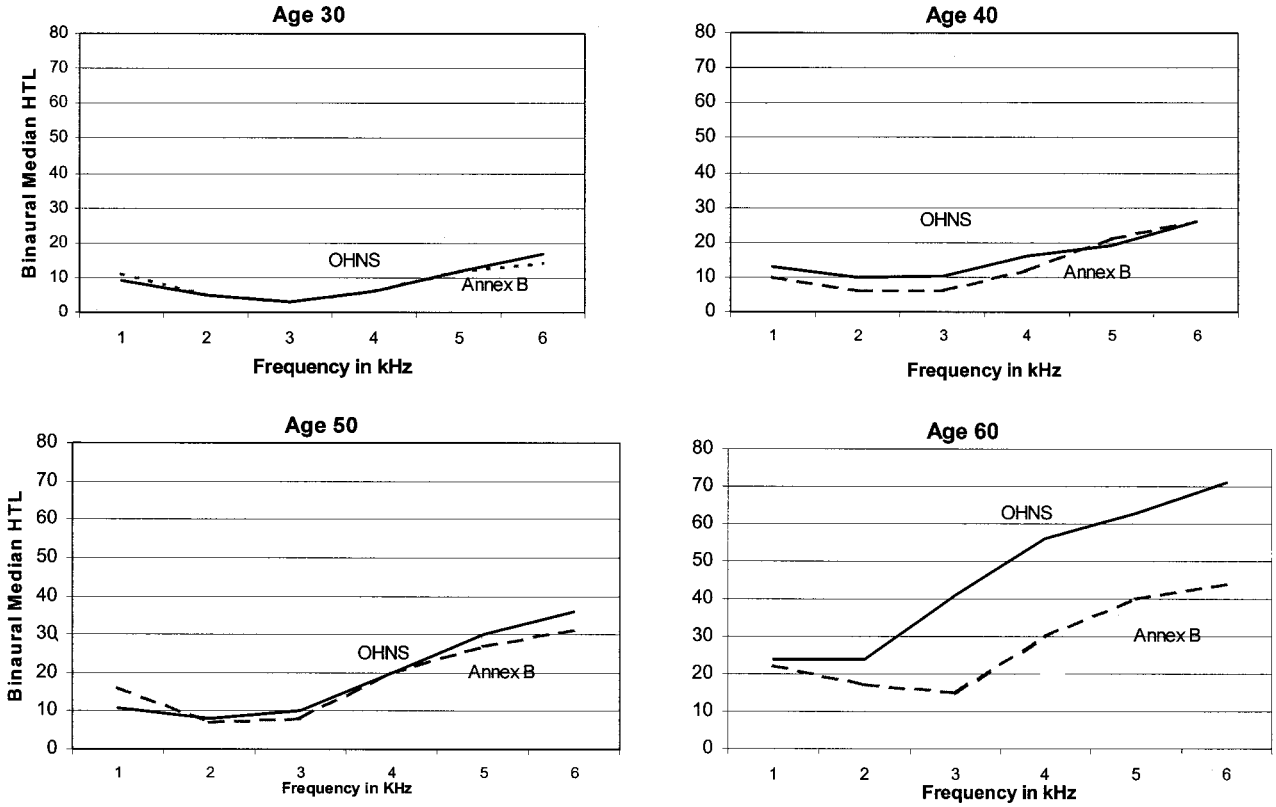


FIG. 4. Binaural average median HTL by test frequency: ONHS versus ANSI S3.44, Annex B.

fects among unscreened low-noise-exposed industrial workers for future use in evaluation of hearing levels among noise-exposed workers enrolled in hearing conservation programs. Results indicate that screened populations have lower (better) average hearing threshold levels than unscreened populations. The usual variability in hearing threshold levels was observed by age, duration exposed, noise level, and test frequency for both the screened and unscreened populations. As had been observed in other studies, variability was greatest for older age groups, higher test frequencies, and increasing noise level. Examination of age-adjusted odds ratios for various definitions of material impairment indicated that factors such as nonoccupational noise, previous noisy job exposure, and medical conditions were not associated with significant increases in material impairment. Moreover, the analyses of mean hearing threshold levels among the low-noise-exposed unscreened workers indicated that most of the increase is due to age rather than the identified risk factors for hearing loss. The type of industry from which low-noise-exposed workers are drawn was not significantly associated with increased mean hearing threshold levels or rates of impairment after adjusting for age.

TABLE IV. Age-adjusted odds ratios for hearing impairment associated with various factors among unscreened controls.

	Definition of material impairment (>25 dB)					
	PTA-1234		PTA-346		PTA-0.5,2,1	
Risk factor	Odds ratio	95% CI[a]	Odds ratio	95% CI[a]	Odds ratio	95% CI[a]
Previous job noise exposure	0.7	0.3, 1.6	0.7	0.3, 1.5	0.8	0.2, 2.6
Military noise	1.4	0.6, 3.1	1.1	0.5, 2.4	1.5	0.5, 4.0
- weapon	1.6	0.7, 3.7	1.2	0.5, 2.9	1.8	0.7, 5.0
- non-weapon	0.7	0.1, 4.1	0.7	0.1, 3.1	—	—
Leisure noise	0.9	0.3, 2.7	1.7	0.5, 2.4	1.0	0.2, 4.2
- weapon	0.6	0.1, 3.0	0.9	0.2, 4.5	1.4	0.2, 8.4
- non-weapon	1.2	0.3, 5.0	2.5	0.7, 9.6	0.6	0.1, 5.7
Medical history	1.1	0.4, 2.9	0.7	0.3, 1.9	1.1	0.3, 4.5
Otologic abnormalities	1.4	0.5, 3.9	0.6	0.2, 1.9	3.0	0.8, 10.9

[a]95% confidence limits: Lower, Upper. *P*-values associated with each factor were not statistically significant at the $p=0.05$ level. Interpretation of odd ratio is as follows: for the example of military noise, the odds or risk of material impairment (for PTA-1234 definition) among workers with military noise exposure is 1.4-fold (or 40%) higher than for workers with no military noise exposure after adjustment for all other factors in the model.

A. Implications for planning public health interventions and epidemiologic studies

Better data on the distribution of NIHL and its risk factors in the population are necessary to plan effective research protocols for testing and evaluating public health interventions. A review of the scientific literature indicates that there is generally weak and inclusive evidence that occupational hearing conservation programs prevent noise-induced hearing loss (Dobie, 1995). A key component often missing for an adequate evaluation of hearing conservation program (HCP) effectiveness is an appropriate reference population. Such populations are useful in separating the effects due to aging and other nonoccupational causes of hearing loss from occupational factors. However, there is a lack of contemporary longitudinal data for working populations not exposed to occupational noise. Currently, comparisons can be made against available populations such as the Baltimore Longitudinal Study on Aging [BLSA] (Brant and Fozard, 1990; Pearson *et al.*, 1995; Morrell and Brant, 1991; Morrell *et al.*, 1996) or statistical models (ISO 1999, 1990; ANSI S3.44, 1996).

The BLSA, a key longitudinal study on aging, is a highly selective white collar population (e.g., predominately white males, high socioeconomic status, primarily older study subjects). A comparison of the median and percentiles of HTL to the BLSA data (Morrell *et al.*, 1996) indicated that the overall spread of the distribution was similar but that there were significant differences in median HTL (11–20 dB across age categories) at 0.5 and 4 kHz. The reason for the disparity between these populations may be due to differences in ANSI audiometric standards, (ANSI, 1989 for BLSA vs ANSI, 1969 for ONHS), audiometer type (manual versus automated), audiometric test procedures (high background booth levels), or increased variability due to differences in unknown risk factors between industrial (blue collar) and nonindustrial (white collar) populations. Therefore, comparisons to actively working, noise-exposed blue collar workers may not be valid. It is likely that comparisons to a highly screened, white collar population would produce higher predicted hearing threshold levels attributable to occupational noise exposure.

The ideal comparison, a low-noise-exposed population from the same facility, is usually not available and, if it can be identified from the HCP database, the workers with less than 80 dBA exposure have intermittent exposure to hazardous noise levels. A comparison of hearing threshold data among workers with high noise exposure (85 dBA or greater) vs this internal ''low-noise'' (under 80 dBA) group would underestimate the effects of noise damage. Therefore, one goal of this analysis was to evaluate whether the unscreened ONHS control data may be used as a comparison database for purposes of evaluating hearing health outcomes and impact of noise-reduction strategies among contemporary workers enrolled in industrial hearing conservation programs. The utility of the ONHS data in this regard is important because widespread use of HPDs is not a confounding factor in this population.

Based on the current analysis, there are some caveats regarding use of these data for this purpose. First, the representativeness of the data relative to contemporary working populations is limited to white males working in certain industries. The U.S. workforce today is much more diverse with regard to gender, nationality, and race than the workforce of the 1970's. It is therefore likely that the distribution of risk factors for hearing loss (medical, social, health care, recreational noise sources, etc.) could be quite different, resulting in more variability in the underlying distribution of hearing threshold levels in the general population. For example, studies examining the impact of various risk factors (e.g., age, blood pressure, and alcohol and cigarette consumption) among noise-exposed workers (McBride, 1993; Tomei *et al.*, 1996) and non-noise-exposed workers (Brant *et al.*, 1996; Gates *et al.*, 1993; Noriyuki *et al.*, 2000) found that after adjusting for age, only a few medical factors such as abnormal ear conditions and hypertension were associated with the development of hearing loss. Gender and racial differences in the prevalence of such risk factors may also affect the burden of hearing loss in the population due to factors other than occupational noise exposure.

Second, the control population from ONHS may not be representative of white male workers in the general U.S. population (which would draw from a wider variety of industries). In the population analyzed for this paper, 61% of the sample was derived primarily from two industries (printing, 43% and steel fabrication, 17%). The sample from the printing industry had the smallest difference between controls and exposed for mean HTLs by frequency, while the largest differences by exposure status occurred in the steel fabrication group. Despite the limited range of industries represented, there appears to be a sufficiently wide range of variability for key variables such as age, noise exposure, length of employment, and level of noise exposures to provide valid inferences of the effect of noise on hearing loss without the confounding effect of HPD use. This is evidenced by the remarkable consistency of the median HTLs for most age groups among the ONHS control population compared to annex B of the ANSI and ISO standards (ANSI S3.44, 1996; ISO 1999, 1990). Annex B represents unscreened low-noise-exposed workers derived from a more representative U.S. population-based sample (i.e., a nonindustrial low-noise-exposed comparison population) over the period of 1960–62. As shown in Fig. 4, the main divergence in median HTLs between the ONHS industrial control population and annex B occurs at age 60 (55–65), where median HTLs for the ONHS unscreened control group are consistently higher (20–35 dB) for binaural averages over 3, 4, and 6 kHz than annex B. This difference for the older age group may be attributable to (a) increased statistical variability in the distribution of HTLs due to small sample size among older workers ($n=130$, aged 50+ years) in the ONHS control population relative to annex B, and (b) differential selection bias of older workers due to the cross-sectional nature of the ONHS survey, resulting in older workers having worse hearing than the general population.

Data comparisons of the distribution of HTLs for older workers (aged 55 years and older) from the ONHS population should be avoided. It is suggested that annex B or C of ANSI S3.44 (ANSI S3.44, 1996) be used as a comparison

group for unscreened, noise-exposed occupational cohorts. For screened occupational groups, the screened ONHS population can be used or annex A from the ISO 1999 (ISO 1999, 1990) and ANSI S3.44 (ANSI S3.44, 1996) standards. Both standards draw their data from hearing and noise levels from populations during the 1960's and 1970's. The BLSA data (Morrell *et al.*, 1996) may be used as an additional data source, particularly if the industrial population includes white collar workers, or to estimate lower limits of expected hearing loss (best-case scenario) among medically screened workers exposed to noise on an intermittent basis. The impact of the observed differences in the distribution of HTL on estimating risk of hearing loss in unscreened populations will be evaluated in an upcoming paper examining the same set of ONHS data.

The data analyzed for this paper demonstrates the need for careful planning of future studies whose aim is to estimate the burden of noise-induced hearing loss among occupational cohorts or the general population. These studies should avoid ''convenience samples'' of volunteer industries and implement study designs so that the distribution of age and other important demographic factors is balanced between exposed and control populations to avoid spurious inferences regarding associations between hearing health outcomes and various risk factors.

The collection of more precise hearing and noise data using state-of-the-art instrumentation in epidemiological studies is a crucial foundation for planning public health and intervention campaigns aimed at reducing noise-induced hearing loss in the population, such as that highlighted in *Healthy People 2010*, which identifies and tracks important public health objectives for the nation (U.S. DHHS, 2000).

ACKNOWLEDGMENTS

This paper is dedicated to the memory of Dr. Daniel Johnson, a dedicated colleague, mentor, and friend, who encouraged development of this work and provided input on earlier versions of this manuscript. The author also wishes to thank the journal peer reviewers for their thoughtful and insightful comments and Ms. Bing Xue, Xiangdong Zhou, and Ruishan Wu for their technical support in producing graphical and tabular displays of the data.

ANSI (**1960**). ANSI S3.1-1960, ''American National Standard Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms'' (American National Standards Institute, New York).

ANSI (**1969**). ANSI S3.6-1969, ''American National Standard Specification for Audiometers'' (American National Standards Institute, New York).

ANSI (**1989**). ANSI S3.6-1989, ''American National Standard Specification for Audiometers'' (American National Standards Institute, New York).

ANSI (**1996**). ANSI S3.44-1996, ''American National Standard Determination of Occupational Noise Exposure and Estimation of Noise-induced Hearing Impairment'' (American National Standards Institute, New York).

Berger, E. H., Royster, L. H., and Thomas, W. G. (**1977**). ''Hearing levels of nonindustrial noise exposed subjects.'' J. Occup. Med. **19**, 664–670.

Brant, L. J., and Fozard, J. L. (**1990**). ''Age changes in pure-tone hearing thresholds in a longitudinal study of normal human aging,'' J. Acoust. Soc. Am. **88**, 813–820.

Brant, L. J., Gordon-Salant, S., Pearson, J. D., Klein, L. L., Morrell, C. H., Metter, E. J., and Fozard, J. L. (**1996**). ''Risk factors related to age-associated hearing loss in the speech frequencies.'' J. Am. Acad. Audiol. **7**, 152–160.

Breslow, N. E., and Day, N. E. (**1980**). ''Classical Methods of Analysis of Grouped Data,'' in *Statistical Methods in Cancer Research: Vol. I—The Analysis of Case-control Studies* (International Agency for Research on Cancer, Lyon, France), IARC Publication No. 32, Chap. 4, pp. 140–148.

Davis, R. R., and Sieber, W. K. (**1998**). ''Trends in hearing protector usage in American manufacturing from 1972–1989,'' Am. Ind. Hyg. Assoc. J. **59**, 715–722.

Dobie, R. A. (**1995**). ''Prevention of noise-induced hearing loss,'' Arch. Otolaryngol. Head Neck Surg. **121**, 385–391.

Draper, N. R., and Smith, H. (**1966**). *Applied Regression Analysis* (Wiley, New York).

Driscoll, D. P., and Royster, L. H. (**1984**). ''Comparisons between the median hearing threshold levels for an unscreened black nonindustrial noise exposed population (NINEP) and four presbycusis data bases.'' Am. Ind. Hyg. Assoc. J. **45**, 577–593.

Gates, G. A., Cobb, J. L., D'Agostino, R. B., and Wolf, P. A. (**1993**). ''The relation of hearing in the elderly to the presence of cardiovascular disease and cardiovascular risk factors,'' Arch. Otolaryngol. Head Neck Surg. **119**(2), 156–161.

ISO 1999 (**1990**). ''Acoustics—Determination of occupational noise exposure and estimation of noise-induced hearing impairment,'' International Organization for Standardization.

Lempert, B. L., and Henderson T. L. (**1973**). ''Occupational Noise and Hearing 1968 to 1972: A NIOSH Study,'' U.S. Department of Health, Education, and Welfare, Public Health Service, Center for Disease Control, National Institute for Occupational Safety and Health, Division of Laboratories and Criteria Development, Cincinnati, OH.

MathSoft (**1997**). S-PLUS 4 *Guide to Statistics*, Data Analysis Products Division, MathSoft, Seattle, WA.

McBride, D. (**1993**). ''Hearing conservation in the mining industry. Evaluation of a risk factor questionnaire.'' Occup. Med. **43**, 85–92.

Morrell, C. H., and Brant, L. J. (**1991**). ''Modeling hearing threshold levels in the elderly,'' Stat. Med. **10**, 1453–1464.

Morrell, C. H., Gordon-Salant, S., Pearson, J. D., Brant, L. J., and Fozard, J. L. (**1996**). ''Age- and gender-specific reference ranges for hearing level and longitudinal changes in hearing level,'' J. Acoust. Soc. Am. **100**, 1949–1967 (Part 1).

MSHA (**1999**). ''Health Standards for Occupational Noise Exposure; Final Rule,'' Mine Safety and Health Administration, 30 CFR Part 62, 64 Fed. Reg., 49548-49634, 49636-49637.

NIOSH (**1972**). ''NIOSH criteria for a recommended standard: Occupational exposure to noise,'' Cincinnati, OH: U.S. Department of Health, Education, and Welfare, Public Health Service, Centers for Disease Control, National Institute for Occupational Safety and Health, DHSS(NIOSH) Publication No. HIM 73-11001.

NIOSH (**1988a**). *National Occupational Exposure Survey (NOES), Field Guidelines. Vol. I*, Cincinnati, OH: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 88-106.

NIOSH (**1988b**). *National Occupational Exposure Survey (NOES), Analysis of Management Interview Responses. Vol. III*, Cincinnati, OH: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 89-103.

NIOSH (**1990**). *National Occupational Exposure Survey (NOES), Sampling Methodology. Vol. II*, Cincinnati, OH: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 89-102.

Noriyuki, N., Mitsuharu, O., Nakamura, K., Suzaki, K., and Tatar, K. (**2000**). ''Cigarette smoking and risk for hearing impairment: A longitudinal study in Japanese male office workers.'' J. Occup. Environ. Med. **42**(11), 1045–1049.

OSHA (**1983**). ''Occupational Noise Exposure: Hearing Conservation Amendment; Final Rule,'' Occupational Safety and Health Administration, 290 CFR 1910.95; 48 Fed. Reg., 9738-9785.

Passchier-Vermeer, W. (**1968**). ''Hearing loss due to exposure to steady-state broadband noise,'' Report No. 35 and Supplement to Report No. 35, Institute for Public Health Engineering, The Netherlands.

Pearson, J. D., Morrell, C. H., Gordon-Salant, S., Brant, L. J., Mettor, E. F., Klien, L. L., and Fozard, J. L. (**1995**). ''Gender differences in a longitudinal study of age-associated hearing loss,'' J. Acoust. Soc. Am. **97**, 1196–1205.

Prince, M. M., Stayner, L. T., Smith, R. J., and Gilbert, S. J. (**1997**). ''A re-examination of risk estimates from the NIOSH Occupational Noise and Hearing Survey (ONHS),'' J. Acoust. Soc. Am. **101**, 950–963.
Reilly, M. J., Rosenman, K. D., and Kalinowski, D. J. (**1998**). ''Occupational noise-induced hearing loss surveillance in Michigan,'' J. Occup. Environ. Med. **40**, 667–674.
Ries, P. W. (**1994**). ''Prevalence and characteristics of persons with hearing trouble: United States 1990–1991,'' Vital Health Stat. **10**, 1–75.
Robinson, D. W. (**1970**). ''Relations between hearing loss and noise exposure,'' in *Hearing and Noise in Industry*, edited by W. Burns and D. W. Robinson (Her Majesty's Stationery Office, London), pp. 100–151.
Robinson, D. W., and Sutton, G. O. (**1975**). ''A Comparative Analysis of Data on the Relation of Pure-tone Audiometric Thresholds to Age.'' NPL Acoustics Report AC84, England, April 1978.
Royster, L. H., Royster, J. D., and Thomas, W. G. (**1980**). ''Representative hearing levels by race and sex in North Carolina industry,'' J. Acoust. Soc. Am. **68**, 551–566.
Royster, L. H., and Thomas, W. G. (**1979**). ''Age effect hearing levels for a white nonindustrial noise exposed population (ninep) and their use in evaluating industrial hearing conservation programs,'' Am. Ind. Hyg. Assoc. J. **40**, 504–511.
Royster, L. H., and Royster, J. D. (**1984**). ''Hearing protection utilization: Survey results across the USA,'' J. Acoust. Soc. Am. Suppl. 1 **76**, S43.
SAS Institute, Inc. (**1989**). SAS Institute, Inc., SAS/STAT® *User's Guide, Version 6, Fourth Edition, Volume 1*, Cary, NC: SAS Institute, Inc., 1989.
Simpson, M., and Bruce, R. (**1981**). ''Noise in America: The Extent of the Noise Problem,'' U.S. Environmental Protection Agency, EPA Report No. 550/9-81-101.
Tomei, F., Papaleo, B., Baccolo, T. P., Tomao, E., Alfi, P., and Fantini, S. (**1996**). ''Chronic noise exposure and the cardiovascular system in aircraft pilots.'' Med. Lav **87**(5), 394–410.
U.S. DHHS (**2000**) U.S. Department of Health and Human Services. Healthy People 2010: *Understanding and Improving Health*. 2nd ed. Washington, DC: U.S. Government Printing Office, November.
Weeks, J. L., Levy, B. S., and Wagner, G. R. (Eds.) (**1991**). *Preventing Occupational Disease and Injury* (American Public Health Association, Washington, D.C.).

Rating, ranking, and understanding acoustical quality in university classrooms

Murray Hodgson[a)]
School of Occupational and Environmental Hygiene and Department of Mechanical Engineering, University of British Columbia, 2206 East Mall, Vancouver, British Columbia V6T 1Z3, Canada

(Received 11 July 2001; revised 14 February 2002; accepted 5 May 2002)

Nonoptimal classroom acoustical conditions directly affect speech perception and, thus, learning by students. Moreover, they may lead to voice problems for the instructor, who is forced to raise his/her voice when lecturing to compensate for poor acoustical conditions. The project applied previously developed simplified methods to predict speech intelligibility in occupied classrooms from measurements in unoccupied and occupied university classrooms. The methods were used to predict the speech intelligibility at various positions in 279 University of British Columbia (UBC) classrooms, when 70% occupied, and for four instructor voice levels. Classrooms were classified and rank ordered by acoustical quality, as determined by the room-average speech intelligibility. This information was used by UBC to prioritize classrooms for renovation. Here, the statistical results are reported to illustrate the range of acoustical qualities found at a typical university. Moreover, the variations of quality with relevant classroom acoustical parameters were studied to better understand the results. In particular, the factors leading to the best and worst conditions were studied. It was found that 81% of the 279 classrooms have ''good,'' ''very good,'' or ''excellent'' acoustical quality with a ''typical'' (average-male) instructor. However, 50 (18%) of the classrooms had ''fair'' or ''poor'' quality, and two had ''bad'' quality, due to high ventilation-noise levels. Most rooms were ''very good'' or ''excellent'' at the front, and ''good'' or ''very good'' at the back. Speech quality varied strongly with the instructor voice level. In the worst case considered, with a quiet female instructor, most of the classrooms were ''bad'' or ''poor.'' Quality also varies with occupancy, with decreased occupancy resulting in decreased quality. The research showed that a new classroom acoustical design and renovation should focus on limiting background noise. They should promote high instructor speech levels at the back of the classrooms. This involves, in part, limiting the amount of sound absorption that is introduced into classrooms to control reverberation. Speech quality is not very sensitive to changes in reverberation, so controlling it for its own sake should not be a design priority. [DOI: 10.1121/1.490363]

PACS numbers: 43.55.Gx, 43.55.Hy, 43.55.Ka [MK]

I. INTRODUCTION

Nonoptimal classroom acoustical design directly affects speech perception by students and instructors and, thus, reduces student learning efficiency. This is particularly true for students who have a hearing loss or are working in a second language. Furthermore, it may cause voice problems for the instructor. In order to promote improved classroom acoustical quality on the University of British Columbia (UBC) campus, and to advise the university on which classrooms should be prioritized for renovation, and why, a study was undertaken to rate and rank all 279 unrestricted-access university classrooms according to their acoustical quality for unaided speech. Speech-reinforcement systems were not considered in this study.

Acoustical quality for speech was quantified by Speech Intelligibility (SI), the percentage of speech material spoken by a standard speaker that would be expected to be correctly recognized by average, normal-hearing listeners working in their first language. A number of physical correlates of SI exist; Speech Transmission Index (STI) was used here, since it is often measured. Ignoring factors related to instructor accent or enunciation, STI and SI at a listener position in a classroom depend on two main factors—the speech-signal to background-noise level difference in decibels, and the room reverberation. The speech-signal level difference depends on the instructor voice level and on the classroom's acoustical characteristics—in particular, how the speech level decreases with distance from the instructor to the listener. The background-noise level comprises noise from the ventilation system and in-class equipment (such as projectors), in-class student-activity noise and noise originating outside the classroom. In this study, noise from in-class equipment and from outside the classroom was assumed negligible. Reverberation depends mainly on classroom size and on the amount of sound absorption—including that contributed by the classroom's occupants—that it contains. With respect to speech, reverberation is best quantified by the early-to-late energy ratio.[1] However, it is more usual to use reverberation time or early decay time, which, in this study, are assumed identical. While opinions on the subject vary, it is assumed here that, for excellent speech conditions in fairly quiet classrooms, reverberation times in the furnished, occupied room should be in the range of about 0.2–1.0 s, increas-

a)Electronic mail: hodgson@mech.ubc.ca

ing with classroom volume.[2] Moreover, the speech-to-noise level difference should exceed a value of about 15 dB, above which no improvement in quality occurs; given typical instructor speech levels, it is often recommended that total classroom A-weighted background-noise levels should not exceed 35 dB.[3]

II. METHODOLOGY

A. STI/SI prediction

Speech intelligibility can be measured; however, tests are time consuming and involve human subjects. Alternatively, in this work, it was calculated from the STI using a regression equation fitted to pairs of corresponding STI and ''short-sentence'' SI values taken from Barnett and Knight:[4]

$$\mathrm{SI} = -270.9\mathrm{STI}^4 + 817.4\mathrm{STI}^3 - 923.3\mathrm{STI}^2 + 476.8\mathrm{STI} - 0.009. \quad (1)$$

Following the approach of Bradley,[1] STI was calculated from the A-weighted speech signal to background-noise level difference (SNA) and the 1 kHz early decay time (EDT1). This allowed existing empirical models, which predict these two quantities, to be employed. STI was determined from SNA and EDT1 using the procedure described by Steeneken and Houtgast.[5] First, they are individually converted to values of a modulation transfer function. The resulting values are then multiplied, converted to apparent signal-to-noise level differences, clipped and converted to STI values. This procedure thus constitutes a simplified version of the Steeneken and Houtgast[5] octave-band STI method, based on A-weighted levels and 1 kHz EDTs.

At any position *r*, SNA can be determined from measured or predicted values of the A-weighted speech level $[\mathrm{SLA}(r)]$ and the background-noise level (BGNA):

$$\mathrm{SNA}(r) = \mathrm{SLA}(r) - \mathrm{BGNA}. \quad (2)$$

The following procedure was used here to obtain predicted speech levels that both vary in a realistic way with a source/receiver distance in individual classrooms,[6] which are derived from vocal output powers that vary with the prevailing acoustical conditions,[7] and that predict voice levels representative of UBC instructors.[7] Two empirical models[7] were combined as follows.

(i) First, A-weighted instructor power levels, $\mathrm{LWA}_{\mathrm{emp}}$ in dB, were predicted using Eq. (4) in Ref. 7:

$$\mathrm{LWA}_{\mathrm{emp}} = 54.8 + 0.5\mathrm{SANA} + 0.016V - 9.6\log(A_0), \quad (3)$$

in which SANA is the total A-weighted student-activity-noise level in dB, V is the room volume in m^3, and A_0 is the total occupied-room absorption in m^2.

(ii) Second, SLA_u intercepts, I_u in dB, and slopes, s_u in dB/dd (dd=distance doubling)—subscript ''*u*'' refers to the unoccupied classroom—were determined for all the unoccupied classrooms. Data for 37 of the classrooms formed the database from which intercept and slope regression models were developed.[8] For these classrooms the measured values were used. For all other classrooms, the quantities were predicted using the regression models—Eqs. (4b) and (5b) in Ref. 8, as follows:

$$I_u = 65.79 - 0.0105\mathrm{LW} + 1.5198\mathrm{fwdist} - 1.4061\mathrm{absorb} - 4.3186\mathrm{upseat}, \quad (4)$$

$$s_u = -1.208 - 0.0877L + 1.1401\mathrm{basic}, \quad (5)$$

in which L and W are the classroom length and width, respectively, in m, and fwdist is the distance of the speech source from the nearest room surface (usually the front wall), in m. The parameter *absorb* indicates the amount of ceiling and/or wall absorption, and is set, for example, to 1.0 with a full-coverage ceiling absorption. Parameter *upseat* is zero if the seats are nonabsorptive and 1.0 if they are padded and, therefore, sound absorptive. Parameter *basic* is 1.0 if the room contains no sound-absorbing features, and zero otherwise. These formulas were developed, assuming vocal output levels corresponding to an average adult speaking at midway between a normal and a raised voice.[8,9] The total A-weighted output power level corresponding to these levels can be easily estimated ($\mathrm{LWA}_{\mathrm{av/nr}} = 72.8$ dB). Of course, if the output power level changes, the intercept (SLA_u at 1 m), but not the slope, changes by the same amount.

(iii) Thus, for a given classroom, predicted intercepts were adjusted by an amount equal to the difference between the power levels predicted by Eq. (3) and that corresponding to the level used to predict the intercept by Eq. (4):

$$I_u' = I_u - (\mathrm{LWA}_{\mathrm{av/nr}} - \mathrm{LWA}_{\mathrm{emp}}). \quad (6)$$

(iv) Finally, speech levels, $\mathrm{SLA}_u(r)$ in dB, at any source/receiver distance, r in m, were calculated from the resulting adjusted intercept I_u and the slope s_u predicted by Eq. (5), as follows:

$$\mathrm{SLA}_u(r) = I_u + s_u \ln(r)\ln(2). \quad (7)$$

Unoccupied SLA_u's were then corrected to the occupied condition (SLA_o—subscript ''*o*'' refers to the occupied condition) on the assumption of 70% classroom occupancy, typical of university classrooms, using diffuse-field theory:

$$\mathrm{SLA}_o(r) = \mathrm{SLA}_u(r) + 10\log\frac{\left(\frac{1}{4\pi r^2} + \frac{4}{A_u}\right)}{\left(\frac{1}{4\pi r^2} + \frac{4}{A_o}\right)}, \quad (8)$$

in which $A_o = A_u + A1_p \cdot O \cdot C$ is the occupied-classroom absorption, in m^2, with $A1_p = 0.81\ \mathrm{m}^2$ the average 1 kHz absorption per person,[5] O the classroom occupancy, and C the classroom seating capacity. A_u is found from the dimensions and absorption-coefficient contributions of the absorbent classrooms' surface features, as described in Ref. 8. Diffuse-field theory was used here despite the fact that it is known to predict steady-state levels inaccurately,[6,8] for want of a better method, and on the assumption that it predicts changes in levels due to small changes in absorption with reasonable accuracy.

BGNA_o was calculated both without and with the contribution of student-activity noise. With it included, BGNA_o was calculated as the energetic sum of the A-weighted ventilation noise (VNA_o) and the A-weighted student-activity noise (SANA):

$$\mathrm{BGNA}_o = 10\log[10^{\mathrm{VNA}_o/10} + 10^{\mathrm{SANA}/10}]. \quad (9)$$

VNA was measured in all classrooms when unoccupied, and was corrected to the occupied condition using diffuse-field reverberant-field theory [as in Eq. (8), without the $1/r^2$ direct-field terms—this was used under the same assumption as discussed above]. SANA was predicted using empirical formulas developed in previous work:[7]

$$\mathrm{SANA} = 83.0 + \log(O \cdot C) - 34.4\log(A_0) + 0.081A_0. \quad (10)$$

As for $\mathrm{EDT1}_u$, measured data was available for 76 of the classrooms when unoccupied. Values for the remaining classrooms were predicted using an empirical method developed using these data.[8] The resulting values were corrected to the occupied condition ($\mathrm{EDT1}_o$) for the assumed occupancy [using Eq. (3) in Ref. 8, which assumes 70% occupancy, generalized to any occupancy].

As detailed below, speech intelligibility was predicted at typical listener positions throughout each occupied classroom, assuming four different instructor voice levels.

B. Classroom data collection

In order to obtain complete data for use in predicting classroom SLA, EDT1, and, thus, STI and SI, the 203 university classrooms that had not been measured prior to this study were visited, and the following data recorded.

(i) classroom dimensions, from which volumes and surface areas were calculated;
(ii) acoustical features, including absorbent-surface features, from which the average 1 kHz surface-absorption coefficient $\alpha 1$ was determined;
(iii) A-weighted ventilation-noise levels were measured at two to four typical positions in the student seating area, with the ventilation system—if present—in "normal" operation. The levels were averaged and applied at all receiver positions in the predictions. These levels must be considered as indicative, not precise; while levels from constant-air–volume systems remain nominally constant, they can vary with the condition of the system's air filters, or if building occupants modify air-outlet diffusers or block openings. Newer classrooms often have variable-air–volume systems generating noise that can vary with time. It is not known exactly which classrooms involved such systems, nor the exact state of the systems at the time of noise measurements.

Table I shows the ranges, averages, and standard deviations of the values of a number of physical and acoustical descriptors of the 279 classrooms, confirming that they represented a wide range of dimensions and acoustical characteristics.

TABLE I. Ranges, averages, and standard deviations of physical and acoustical descriptors of the 279 university classrooms.

Descriptor	Minimum	Maximum	Average	Std. dev.
Capacity	10	498	64	64
Length (m)	4.2	25.2	9.1	3.4
Width (m)	2.9	23.4	9.0	3.1
Height (m)	2.5	9.0	3.2	0.7
Volume (m^3)	55	4166	313	373
V/S (m)	0.6	2.8	0.9	0.2
Floor area (m^2)	20	534	89	66
$\alpha 1$	0.06	0.37	0.19	0.07

C. Speech-quality prediction

1. Receiver positions and classroom averaging

Early decay times in the classrooms were assumed to be independent of position. However, speech levels vary with receiver position relative to the instructor. Thus, SLA and, therefore, SNA, STI, and SI vary with position. In particular, in university lecture rooms in which the instructor is at the front, these quantities generally decrease from the front to the back of the room. In this project, in order to study the variability of speech quality, the classroom seating was divided into nine subsections, designated A–F, as illustrated in Fig. 1. The students were assumed to be uniformly distributed over the seating area. In classrooms with rectangular floor plans, the nine subsections were assumed to have the same areas and numbers of seats. On the other hand, classrooms with diamond-shaped seating areas, for example, have more students sitting at medium distances from the instructor than at the front or back, relative to the rectangular configuration. In this case, it was assumed that the numbers of students in areas B and F were 50% smaller than in the rectangular case. Speech intelligibilities were predicted at the centers of each area; the source, and the area centers were assigned generalized dimension-related coordinates, as shown in Fig. 1. The range (average) of the source/receiver distances were, for example, 1.2 to 7.0 m (2.5 m) for area A and 3.8 to 19.5 m (7.3 m) for area F; the values for the average distance to the centers of all areas were 2.7 to 13.6 m

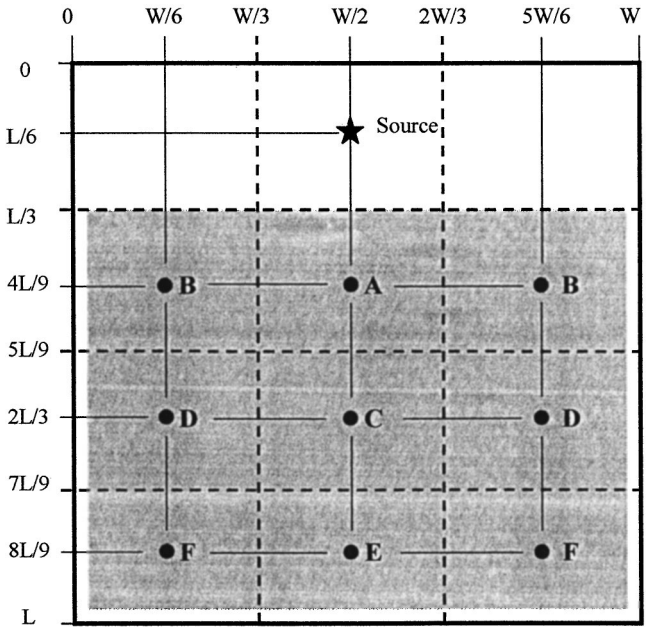


FIG. 1. Illustration of the nine classroom seating areas and their centers, with generalized coordinates.

TABLE II. Subjective quality ratings and associated STI and SI ranges.

STI range	SI range	Quality rating
⩾0.80	⩾98%	"excellent"
⩾0.65	⩾96%	"very good"
⩾0.50	⩾93%	"good"
⩾0.40	⩾88%	"fair"
⩾0.30	⩾80%	"poor"
<0.30	<80%	"bad"

(5.1 m). Classroom-average SI_{av}'s were calculated by averaging the resulting area SI's as appropriate to the particular classroom seating configuration.

2. Voice levels and occupancy

SI_{av} was first predicted for four instructor voice levels and 70% occupancy, typical of UBC classrooms. The four voice levels corresponded to an average male instructor ("typical" case), an average female instructor (average male levels minus 2.6 dB[7]), a quiet female instructor (average female levels minus two standard deviations, assumed equal to 8 dB[7]), and a loud male instructor (average-male levels plus two standard deviations—"best" case).

Predictions were also done for a "worst-case" scenario with the quiet female instructor and 10% occupancy. Detailed results are presented for the "typical" case; only summary results are presented for the other cases.

3. Qualitative quality descriptors

To obtain a better understanding of the significance of the STI/SI values predicted in university classrooms, subjective quality ratings were assigned to ranges of STI/SI values, as shown in Table II.[3] These ratings have not been validated and must be considered as tentative.

Note that Bradley, Reich, and Norcross[10] found that the just-noticeable difference in the perception of speech in noise, with intelligibility quantified by STI, was 0.03. According to Eq. (1), this corresponds to a just-noticeable difference in the SI perception of between about 3% at SI values corresponding to "bad" quality, decreasing to about 0.4% at SI values corresponding to high quality. The authors also suggest that an STI change of 0.1 likely represents a "practically significant change;" this corresponds to changes in SI varying between about 15% and 1% over the range from "bad" to "excellent" quality.

III. RESULTS

A. Early decay time

Figure 2 shows the predicted distribution of the EDT1 values in both the unoccupied and 70%-occupied cases. Unoccupied $EDT1_u$ varied from 0.27 to 2.84 s. The 70%-occupied values varied from 0.24 to 1.05 s, with the majority of classrooms having values between 0.4 and 0.6 s. Seven occupied classrooms exceeded 0.8 s; seven had values less than 0.3 s. These highest and lowest values are explained by the input data for these classrooms, shown in Table III. Either the corresponding capacities, volumes, and, more importantly for reverberation, volume-to-surface-area ratios, were generally much higher and lower than average, respectively, and/or their average surface-absorption coefficients were, respectively, generally lower and higher than average due to low and high amounts of sound-absorbing surface features.

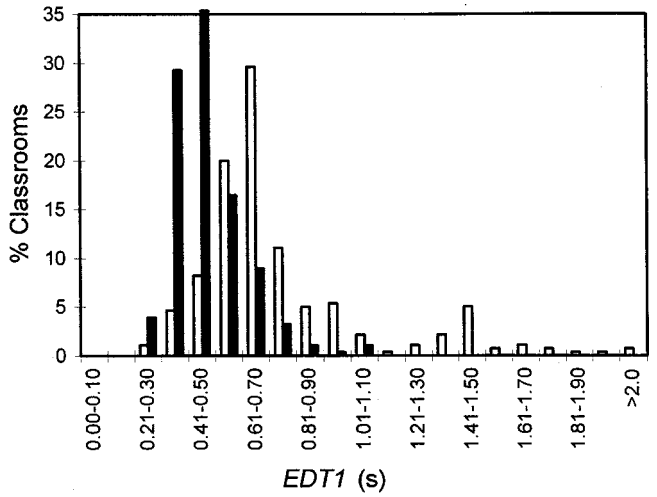


FIG. 2. Distribution of predicted 1 kHz early-decay times in 279 unoccupied (white columns) and 70%-occupied (black columns) university classrooms.

B. Background-noise level

Figure 3 shows the predicted distribution of the 70%-occupied A-weighted ventilation-noise levels, as well as that of the total background-noise levels ($BGNA_o$), with student-activity noise included. VNA_o varied from 27.5 to 51.0 dB, averaging 39 dB. It was typically 2 dB lower than in the unoccupied classrooms. Student-activity noise levels varied from 35.0 to 47.8 dB. $BGNA_o$ varied from 38.0 to 51.3 dB, and was between 0.2 and 16.0 dB higher than ventilation-noise levels alone. Ventilation levels in well over half of the classrooms exceeded the often-recommended limit of 35 dB; with student-activity noise included, no classroom met the criterion.

TABLE III. Input data for the classrooms with the seven highest (H) and seven lowest (L) predicted $EDT1_o$'s.

Name	*C*	*V*	*V*/*S*	*Carpet*	*absorb*	*upseat*	$\alpha 1$	$EDT1_o$
H1	498	4166.6	2.80	1	0.1	1	0.24	1.05
H2	51	427.9	1.09	0	0	0	0.09	1.01
H3	100	1237.3	1.76	1	0.45	0	0.18	1.01
H4	40	414.9	1.13	0	0.35	0	0.13	0.91
H5	24	203.9	0.88	0	0	0	0.09	0.90
H6	100	629.4	1.23	0	0.1	0	0.10	0.90
H7	94	479.0	1.13	0	0	0	0.08	0.86
Avg., highest 7	130	971.5	1.34	0.14	0.36	0.14	0.16	0.95
L1	16	67.3	0.62	1	1	0	0.23	0.29
L2	13	98.5	0.70	1	0.7	1	0.31	0.28
L3	21	92.8	0.69	1	0.7	0.8	0.29	0.28
L4	13	77.2	0.66	1	1	0.5	0.29	0.28
L5	14	75.6	0.66	1	1	0.5	0.29	0.27
L6	10	63.7	0.63	1	1	0.7	0.31	0.25
L7	24	66.8	0.64	1	1	0.5	0.29	0.24
Avg., lowest 7	16	77.4	0.66	1.0	0.91	0.57	0.28	0.27
Avg., all 279	64	314.7	0.93	0.30	0.78	0.08	0.19	0.48

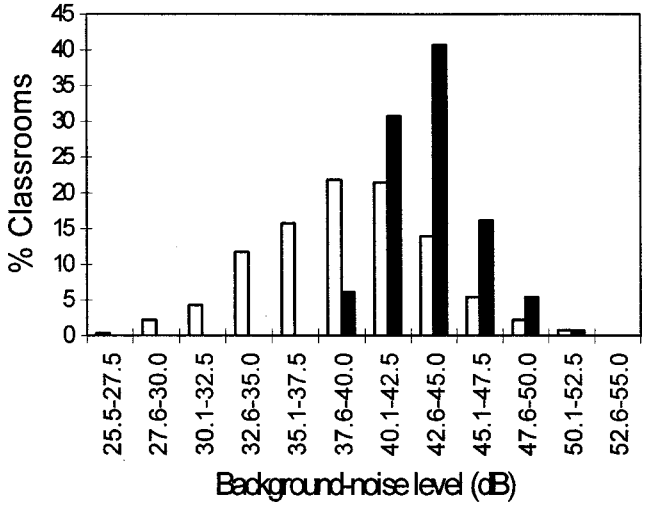


FIG. 3. Distribution of predicted A-weighted background-noise levels in 279, 70%-occupied university classrooms: (white columns) ventilation noise; (black columns) total (ventilation and student-activity) noise.

C. Typical (average-male) instructor

1. Speech level

Figure 4 shows the predicted distribution of A-weighted speech levels at the fronts (areas A) and backs (areas F) of the 70%-occupied university classrooms with a ''typical'' (average-male) instructor. At the front, speech levels varied from 41.3 to 55.9 dB, with most values between 46 and 51 dB. At the back, they varied from 37.9 to 54.5 dB, with most values between 42 and 47 dB. These speech levels are consistent with those measured in university classrooms during lectures.[7]

2. Speech-signal to background-noise level difference

Figure 5 shows the predicted distribution of A-weighted speech-signal to background-noise level differences at the fronts (area A) and backs (area F) of 70%-occupied university classrooms with an average-male instructor, and with student-activity noise not included. At the front, the differences varied from −6.2 to 25.7 dB, with most values between 6 and 12 dB, below the value required for excellent speech intelligibility. At the back, level differences varied from −9.6 to 22.9 dB, with most values lying between 3 and 9 dB, well below the optimal value. Of course, level differences are lower when student-activity noise is included. Values are consistent with those measured in university classrooms during lectures.[7]

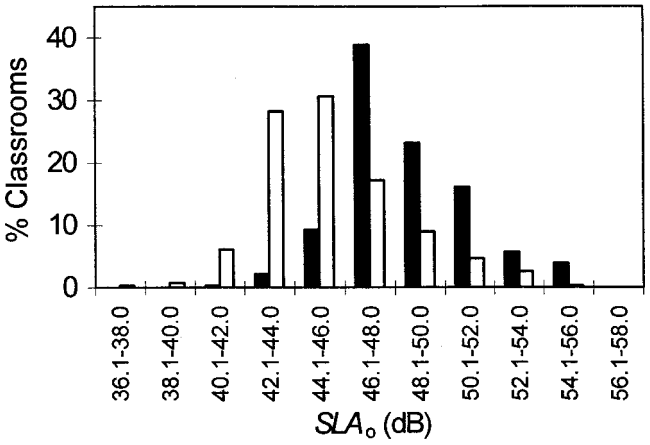


FIG. 4. Distribution of predicted A-weighted speech levels in areas A (black columns) and F (white columns) at the fronts and backs of 279, 70%-occupied university classrooms with a ''typical'' instructor.

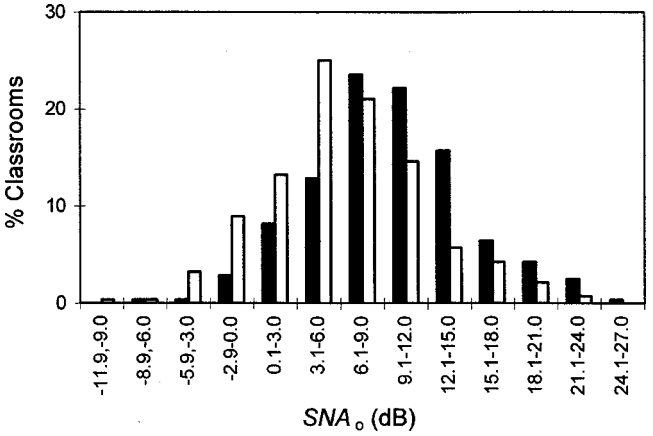


FIG. 5. Predicted A-weighted speech-signal to background-noise level differences in areas A (black columns) and F (white columns) at the fronts and backs of 279, 70%-occupied university classrooms with a ''typical'' instructor.

3. Speech intelligibility

Table IV summarizes the predicted speech-intelligibility results. It shows that, with a typical instructor, the acoustical quality of most university classrooms is fairly high—with ''excellent,'' ''very good,'' or ''good'' quality—in all seats. However, a significant proportion of the classrooms have low quality −31 (12%) classrooms are ''fair,'' ''poor,'' or ''bad'' at the front, 52 (19%) on average, and 78 (28%) at the back.

Table V lists the university classrooms with the five ''best'' and five ''worst'' room-average SI_{av}'s, for the case of a ''typical'' instructor. Also shown are the values of a number of relevant quantities, their average values for the five best and five worst classrooms, and their average values for all 279 classrooms. The average ''best'' classroom has lower values of all dimension-related factors (e.g., capacity, volume, and floor area) than on average for all classrooms. The average ''worst'' classroom has values that are greater than, or approximately equal to, the overall averages. Consequently, source/receiver distances are less than and greater than average in the best and worst classrooms, respectively. Both the best and worst classrooms had typical sound-absorbing surface treatments and, thus, typical average surface-absorption coefficients ($\alpha 1$). Unoccupied early decay times ($EDT1_u$) were lower than average in the best class-

TABLE IV. Number (percentage) of classrooms with a ''typical'' speaker in each quality category, as predicted.

Case	''bad''	''poor''	''fair''	''good''	''very good''	''excellent''
Front (area A)	2(1)	2(1)	27(10)	126(45)	120(43)	2(1)
Room-average	2(1)	13(5)	37(13)	160(57)	65(23)	2(1)
Back (area F)	5(2)	19(7)	54(19)	148(53)	51(18)	2(1)

TABLE V. Data for the five "best" (B) and five "worst" (W), and the average, classrooms with a "typical" instructor (predicted room-average SI_o).

Name	C	V	S_{floor}	$\alpha 1$	EDT1	I'_u	s_u	VNA	$SLA_{o,A}$	$SNA_{o,A}$	$SLA_{o,F}$	$SNA_{o,F}$	$SI_{o,avg}$
B1	16	67.3	26.4	0.23	0.29	54.1	−1.84	31.0	51.2	20.2	48.4	17.4	98.3
B2	12	60.4	23.5	0.23	0.30	54.5	−1.65	31.5	52.9	21.4	50.2	18.7	98.3
B3	16	70.3	23.4	0.19	0.34	55.6	−1.92	33.4	52.1	18.7	49.3	15.9	97.7
B4	35	135.9	54.4	0.23	0.32	50.0	−1.73	31.1	47.7	16.6	44.4	13.3	97.6
B5	63	206.4	68.8	0.19	0.38	51.2	−2.05	27.8	46.3	18.5	43.0	15.2	97.5
Avg, best 5	28	108.1	39.3	0.21	0.33	53.1	−1.84	31.0	50.0	19.1	47.1	16.1	97.9
W1	40	269.8	96.3	0.19	0.49	51.4	−2.25	47.3	46.4	−0.9	43.0	−4.3	83.6
W2	56	409.0	112.4	0.16	0.66	53.6	−2.26	48.9	48.3	−0.7	44.8	−4.2	82.0
W3	64	420.1	114.8	0.16	0.64	53.7	−2.27	49.4	48.1	−1.3	44.6	−4.8	80.3
W4	62	246.4	88.2	0.24	0.36	49.4	−1.96	51.0	45.7	−5.2	42.3	−8.7	66.7
W5	85	393.7	109.0	0.34	0.47	49.5	−2.17	51.0	44.9	−6.2	41.4	−9.6	58.3
Avg, worst	61	347.8	104.1	0.22	0.52	51.5	−2.18	49.5	46.7	−2.9	43.2	−6.3	74.2
Avg, all 279	64	314.7	88.7	0.19	0.47	52.6	−1.93	39.2	48.6	9.4	45.4	6.2	93.3

rooms, but average in the worst classrooms. Occupied values were 34% lower and 9% higher than average in the best and worst classrooms, respectively. Speech levels at 1 m were 0.5 dB higher and 1.1 dB lower, respectively, in the best and worst classrooms. Speech levels decreased with distance from the speaker at 0.1 dB/dd more slowly and 0.3 dB/dd faster than average, respectively, in the best and worst rooms. Thus, speech levels at both the fronts $[SLA_o(A)]$ and backs $[SLA_o(F)]$ were about 1.5 dB higher and 2 dB lower than average in the best and worst classrooms, respectively. Background-noise levels (VNA_o) in the best and worst classrooms were 8 dB lower and 10 dB higher than average, respectively. Thus, signal-to-noise level differences were about 10 dB higher and 12 dB lower at both the fronts and backs of the best and worst classrooms, respectively. The resulting room-average speech intelligibilities were 4.6% higher and 18.9% lower, respectively.

This simple analysis suggests that high speech intelligibility is, in general, obtained by designing a classroom to have small dimensions, low reverberation, and low background noise; furthermore, the design should promote high speech levels. However, some individual classrooms in each category have characteristics that are different from the average. Two of the best classrooms have lower intercepts, one a lower slope, than average. Thus, they have lower than average speech levels, but level differences that are still high. Three of the worst classrooms have lower than average $EDT1_o$. However, they also have lower than average level differences due to high noise levels, and are of low quality.

4. Factors affecting quality

The relationship between classroom design parameters and speech quality can be analyzed in more detail by plotting individual design-parameter values against room-average speech intelligibility for the 279 classrooms. Only the results with student-activity noise not included are discussed here. Of course, such a univariate analysis ignores possible interactions between parameters. Figure 6 shows the relationship between predicted SI_{av} and room volume. Clearly there is no correlation—classrooms of high quality can be small or large. Moreover, the largest classrooms are generally of high quality; smaller classrooms are among those of the lowest and highest qualities.

Figure 7 shows the relationship between predicted SI_{av} and occupied classroom early-decay time. Again, there is no correlation.

Figure 8 shows the correlation between SI_{av} and room-average speech-to-noise level difference. Clearly, there is a strong positive correlation—quality increases with level difference. Also shown in this figure are the speech-quality ranges. The level-difference ranges corresponding to the various qualities are approximately as follows: 18.4 to 19.6 (average=19.0) dB ("excellent"); 8.3 to 23.8 (average =13.4) dB ("very good"); 2.5 to 18.5 (average=7.2) dB ("good"); −1 to 6.9 (average=0.9) dB ("fair"); −3.1 to 0.7 (average=−2.0) dB ("poor"); and −8.5 to −3.6 (average =−6.6) dB ("bad"). The lowest value in each range represents the approximate noise level required to ensure each level of quality. Similarly, the average values of each range are the approximate levels typically associated with each quality. These results support the assertion that a speech-to-noise level difference of about 15 dB is required for high speech quality.

Let us now look at the correlation between predicted SI_{av} and the determinants of speech-to-noise level difference. Figure 9 shows the relationship between SI_{av} and occupied-classroom room-average speech levels. There is no correlation. Figure 10 shows the same result for occupied ventilation-noise level. Clearly, there is a strong negative

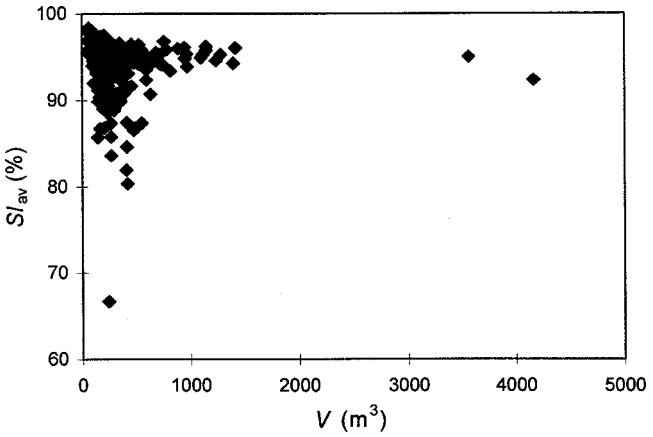


FIG. 6. Predicted variation of room-average speech intelligibility (for a "typical" instructor) with room volume, for 279 university classrooms.

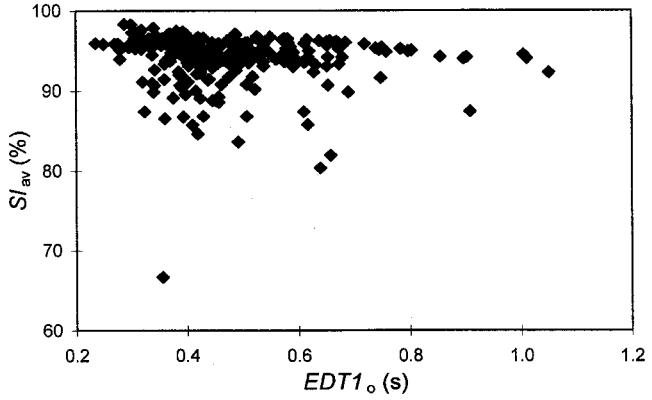


FIG. 7. Predicted variation of room-average speech intelligibility (for a ''typical'' instructor) with 1 kHz occupied-classroom early decay time, for 279 university classrooms.

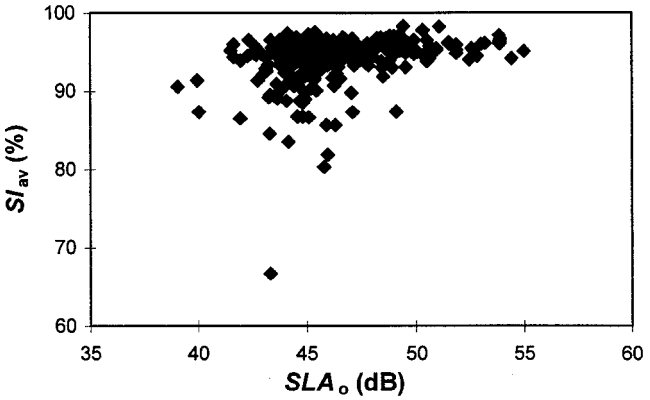


FIG. 9. Predicted variation of room-average speech intelligibility (for a ''typical'' instructor) with A-weighted occupied-classroom speech level, for 279 university classrooms.

correlation; quality generally decreases with increased noise level. An analysis similar to that done for the speech-to-noise level difference shows that the ranges of noise levels found in classrooms of the various qualities are as follows: 31.0 to 31.5 dB (''excellent''); 27.5 to 40.2 dB (''very good''); 31.3 to 46.3 dB (''good''); 39.3 to 47.3 dB (''fair''); 42.2 to 48.9 dB (''poor''); 49.4 to 51.0 dB (''bad''). The noise level must not exceed 27.5 and 31.5 dB, respectively, to ensure ''very good'' and ''good'' qualities. Noise levels of 34.0, 39.5, and 43.4 dB, respectively, typically correspond to ''very good,'' ''good,'' and ''fair'' qualities. The first of these three results is consistent with the maximum background-noise level of 35 dB, which is often recommended for high speech quality.

With an aim to considering possible interactions between the parameters, and on the expectation that room size is a dominant parameter, the classroom data were stratified into three size/volume ranges: small ($V<200\,\mathrm{m}^3$); medium ($200\leqslant V\leqslant 400\,\mathrm{m}^3$); and large ($V>400\,\mathrm{m}^3$). SI_{av} was correlated with the same parameters as above. Interestingly, the results were virtually identical in all cases, with one possible exception. Particularly in the medium and large classrooms, speech levels tended to increase with volume, as might be expected. In any case, in general, classroom size is apparently not a dominant factor.

The results reported here suggest strongly that classroom size and reverberation are not significant factors in determining speech quality, but that noise level and signal-to-noise level difference, are important factors.

D. Other voice levels and occupancy

Table VI shows the number (percentage) of the 279 university classrooms with the various predicted room-average qualities in the cases of 70% occupancy and the four instructor voice levels studied. Speech intelligibilities were calculated by simply changing speech levels in all classrooms by $+8$, -2.6, and -10.6 dB, respectively, relative to levels for the average male instructor, as discussed above. Of course, this resulted in identical decreases in the signal-to-noise level differences, and corresponding changes in speech intelligibility. Clearly, the instructor output level has a strong effect on speech quality. In the best case, with a loud, male instructor, all but 2 classrooms have at least ''good'' quality, and 13 are ''excellent.'' On the other hand, in the worst case, with a quiet female instructor, the vast majority of the classrooms have no better than ''fair'' quality; 43% are ''bad,'' with room-average speech intelligibilities as low as zero.

Also shown are the results for the extreme case of the

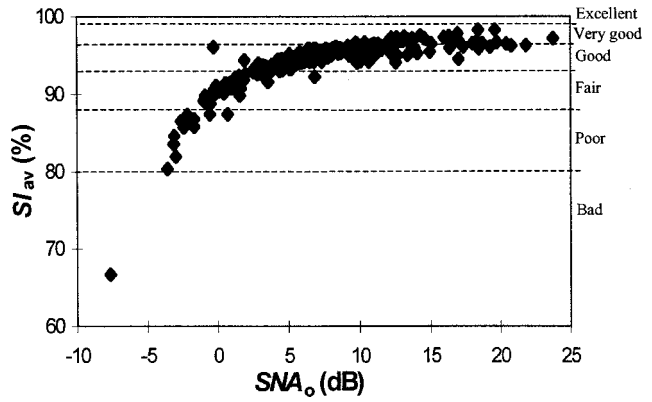


FIG. 8. Predicted variation of room-average speech intelligibility (for a ''typical'' instructor) with occupied-classroom speech-to-noise level difference, for 279 university classrooms.

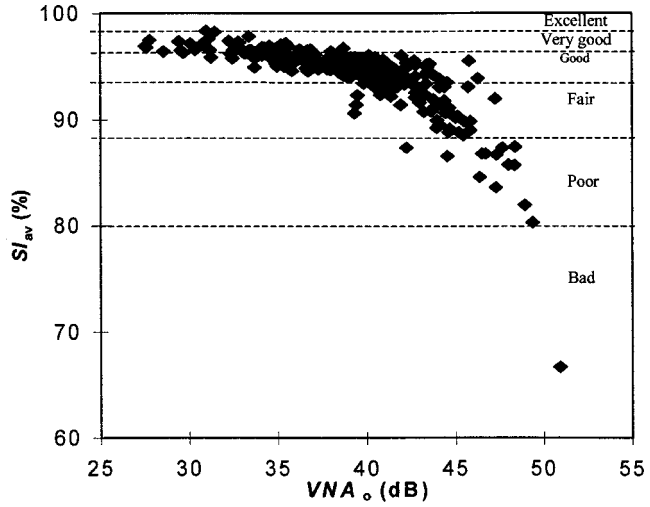


FIG. 10. Predicted variation of room-average speech intelligibility (for a ''typical'' instructor) with A-weighted occupied-classroom ventilation noise level, for 279 university classrooms.

TABLE VI. Number (percentage) of classrooms in each room-average quality category, for four instructors and two occupancies, as predicted.

Case	''bad''	''poor''	''fair''	''good''	''very good''	''excellent''
Loud male, 70% occupancy	0(0)	0(0)	2(1)	64(23)	200(72)	13(5)
Average male, 70% occupancy	2(1)	13(5)	37(13)	160(57)	65(23)	2(1)
Average female, 70% occupancy	12(4)	26(9)	66(24)	139(50)	34(12)	2(1)
Quiet female, 70% occupancy	120(43)	71(25)	59(21)	26(9)	3(1)	0(0)
Quiet female, 10% occupancy	149(53)	72(26)	42(15)	14(5)	2(1)	0(0)

quiet, female instructor and 10% occupancy. Decreased occupancy resulted in increased EDT1 and decreased SNA and, thus, decreased SI_{av}.

The instructor voice level and occupancy had little effect on the order of the classrooms with respect to speech quality, only on the absolute quality.

IV. CONCLUSION

The objective of rating and ranking university classrooms for acoustical quality has been achieved using room-average speech intelligibility (SI_{av}) as the quality measure. This involved the application of empirical models developed from classroom measurement data for predicting SI, and a simplified STI-calculation procedure. Clearly these models must be validated experimentally in future work. SIs were predicted at nine positions in each university classroom when 10% or 70% occupied, for four instructor voice-output levels. With a ''typical'' instructor, most university classrooms are of fairly high quality at all positions; however, a number are of lower quality. Quality varies strongly with vocal output, with increased output increasing quality. It also varies with occupancy, with decreased occupancy resulting in decreased quality. Classroom quality is not correlated with classroom size, speech level, or early decay time. However, it is strongly correlated with the background-noise level and signal-to-noise level difference.

The results suggest that high speech quality can be achieved in classrooms of any size if they are well designed. Moreover, at least in the case of speakers with typical or strong voices, speech-reinforcement systems are not necessary.

Speech-reinforcement systems can, however, be used to increase instructor speech levels, particularly at the back of the classroom. This may be particularly warranted in the case of quiet instructors.

ACKNOWLEDGMENTS

The author would like to acknowledge Mr. John Lane, UBC Physical Access Coordinator, who commissioned and partially funded this work. He would also like to acknowledge the considerable contributions to this work of Tsun Kei (Charles) Ng and Dr. Susan Kennedy of UBC. Thanks also to Michael Noble, BKL Consultants Ltd., and to Barry McKinnon, McSquared System Design Group, Inc. for their comments on the work at various stages.

[1] J. S. Bradley, ''Speech intelligibility studies in classrooms,'' J. Acoust. Soc. Am. **80**, 846–854 (1986).
[2] M. R. Hodgson and E.-M. Nosal, ''Effect of noise and occupancy on optimum reverberation times for speech intelligibility in classrooms,'' J. Acoust. Soc. Am. **111**, 931–939 (2002).
[3] B. Seep, R. Glosemeyer, E. Hulce, M. Linn, P. Aytar, and R. Coffeen, ''Classroom acoustics—a resource for creating learning environments with desirable listening conditions,'' report published by the Acoustical Society of America, 2000.
[4] P. W. Barnett and R. D. Knight, ''The common intelligibility scale,'' Proc. Inst. Acoust. **17**, 199–204 (1995).
[5] H. J. M. Steeneken and T. Houtgast, ''A physical method for measuring speech-transmission quality,'' J. Acoust. Soc. Am. **67**, 318–326 (1980).
[6] M. R. Hodgson, ''Experimental investigation of the acoustical characteristics of university classrooms,'' J. Acoust. Soc. Am. **106**, 1810–1819 (1999).
[7] M. Hodgson, R. Rempel, and S. Kennedy, ''Measurement and prediction of typical speech and background-noise levels in university classrooms during lectures,'' J. Acoust. Soc. Am. **105**, 226–223 (1999).
[8] M. R. Hodgson, ''Empirical prediction of speech levels and reverberation times in classrooms,'' J. Build. Acoust. **8**, 1–14 (2001).
[9] American National Standard ''Methods for Calculation of the Speech Intelligibility Index,'' ANSI S3.5-1997.
[10] J. S. Bradley, R. Reich, and S. G. Norcross, ''A just noticeable difference in C_{50} for speech,'' Appl. Acoust. **58**, 99–108 (1999).

A model cochlear partition involving longitudinal elasticity

Taha S. A. Jaffer,[a] Hans Kunov, and Willy Wong
Institute of Biomaterial and Biomedical Engineering, 4 Taddle Creek Road, Toronto, Ontario M5S 3G9, Canada and Edward S. Rogers Department of Electrical and Computer Engineering, 10 King's College Road, Toronto, Ontario M5S 3G9, Canada

(Received 13 June 2001; revised 3 December 2001; accepted 28 May 2002)

This paper addresses the issue of longitudinal stiffness within the cochlea. A one-dimensional model of the cochlear partition is presented in which the resonant sections are coupled by longitudinal elastic elements. These elements functionally represent the aggregate mechanical effect of the connective tissue that spans the length of the organ of Corti. With the plate-like morphology of the cochlear partition in mind, the contribution of longitudinal elasticity to partition dynamics is appreciable, though weak and nonlinear. If the elasticity is considered Hookian then the nonlinearity takes a cubic form. Numerical solutions are presented that demonstrate the compressive nature of the partial differential nonlinear equations and their ability to produce realistic cubic distortion product otoacoustic emissions. Within the framework of this model, some speculations can be made regarding the dynamical function of the phalangeal processes, the sharpness of active cochlear mechanics, and the propogation of pathology along the partition. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1496078]

PACS numbers: 43.64.Bt, 43.64.Kc [LHC]

I. INTRODUCTION

The model cochlea is represented by a long tapered fluid-filled canal, split into two chambers by a viscoelastic membrane known as the cochlear partition (CP). The CP is a complex sandwich of tissues that is responsible for the mechano-neural transduction of sound. In most theoretical models, each resonant section of the CP is considered unconnected to its nearest neighbors. From a histological perspective, this is not true as the CP is a contiguous epithelial tissue that contains a variety of cellular interconnections. The motivation to describe various active, otoacoustic, perceptual, and pathological auditory phenomena has spurred the modification of the cochlear traveling wave model to include active dynamics (Neely and Kim, 1983), nonlinear dynamics (Hall, 1974), additional resonant modes (Allen, 1980), and structural coupling (Geisler and Sang, 1995). In this paper, we develop another modification: the inclusion of elastic tissue that longitudinally couples one resonant region on the CP to another.

Longitudinal stiffness is often found among the discarded features of most cochlear models. The perception is that within the CP the transverse stiffness of the basilar membrane dominates other membrane dynamics (Mammano and Nobili, 1993). This assumption is based on a study performed by Völdrich (1978) to verify the stiffness experiments carried out by von Békésy (1960). von Békésy deflected portions of the basilar membrane in cadaver cohleae using a blunt probe and found that the indentations were elliptical near the stapes and circular near the helicotrema. These indentations would have been narrow and radially oriented if the longitudinal stiffness was negligible in comparison to the transverse stiffness of the basilar membrane. However, in the cadaver cochlea the longitudinal stiffness was found to be appreciable. These studies also demonstrated that the basilar membrane is not under static tension like the surface of a balloon. Though von Békésy's work was seminal to our understanding cochlear transduction, the technology did not exist to extend his findings to include living cochleae.

Two decades later, Völdrich (1978) performed a follow-up study using live guinea pig cochleae. He found that the local deflections were radially oriented, narrow, and confined to the vicinity of the probe. In contrast to von Békésy, he found that the deflections did not increase along the partition; the indentations remained narrow and local. These findings suggested that the primary source of stiffness in living cochleae is transverse, presumably with its origins in the collagenous fibers of the basilar membrane. This study provides the experimental justification for neglecting longitudinal stiffness in models of functional living cochlea. Not only is such an assumption theoretically appealing, it also carries practical value in that the underlying mechanics of the CP are greatly simplified.

Another two decades later Naidu and Mountain (2001) also studied longitudinal stiffness in mongolian gerbil cochleae. Their observations challenge the view established earlier by Völdrich that the longitudinal stiffness within the CP is negligible. Their recent experiments quantified local mechanical deflections of the CP using videomicroscopy. The results obtained are consistent with the earlier findings of von Békésy, not those of Völdrich. For instance, the basilar membrane exhibits longitudinal coupling that clearly increases in strength from base to apex. In addition, they also observed that the inclusion of the organ of Corti increases the overall longitudinal stiffness of the CP. These observations challenge the prevalent view that the longitudinal stiffness within the living CP is negligible, though still maintaining its relative weakness compared to the transverse stiffness of the basilar membrane.

In this paper we assume that the longitudinal stiffness is

[a]Electronic mail: jaffer@ecf.utoronto.ca

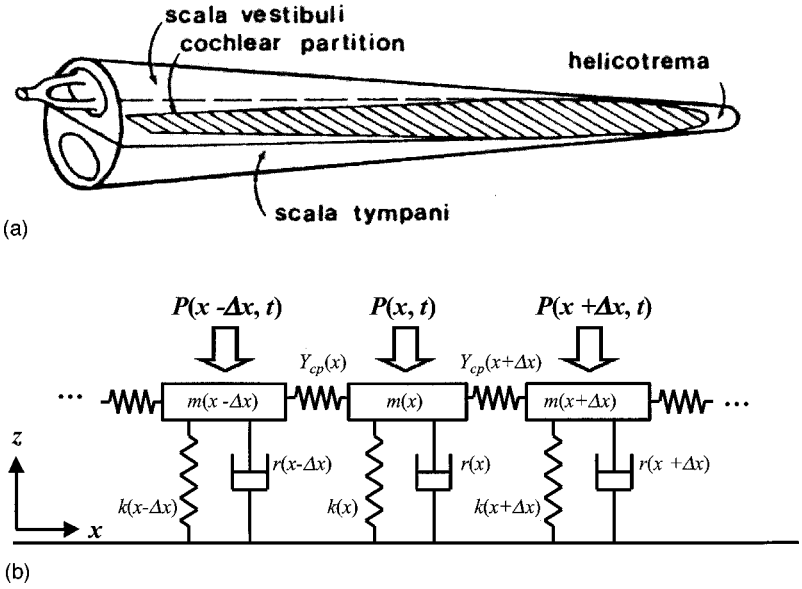


FIG. 1. The traveling wave model of the cochlea including longitudinal elasticity. (a) An idealization of an uncoiled cochlea separated into two fluid-filled chambers by the plate-like cochlea partition. (b) The cochlear partition is considered to be a series of viscoelastic plates where the stiffness longitudinally is considered small in comparison to the transverse stiffness, $k(x)$. The classic model of the cochlear partition is similar to the one shown with the exception that the springs interconnecting the plates are taken away. The diagram shows the proposed model presented in this paper.

considered weak and can be described elastically. The addition of longitudinal mechanics give rise to nonlinearities in systems with transverse dynamics (Marion and Thornton, 1995). To a first-order approximation, these nonlinearities take a cubic form. Thus, one consequence of adding longitudinal stiffness into the cochlear model may be the production of measurable distortion product otoacoustic emissions (DPOAEs). Otoacoustic emissions (OAEs) are known to have both active and passive components (Yates, 1995). For instance, DPOAEs have been found to persist under hypoxic, pharmacological, and necrotic trauma (Lonsbury-Martin *et al.*, 1987; Weir, 1988; Kemp and Brown, 1984; and Schmeidt and Adams, 1981). The common feature in each of these traumatized cases was the uncompromised contiguity of the sensory epithelium. We suggest that the stiffness mechanics related to this contiguity can help explain the nonphysiologically vulnerable DPOAEs.

Furthermore, weak coupling nonlinearities are capable of producing other observable low level effects such as entrainment, instability, and frequency dispersion (Marion and Thornton, 1995). These phenomena have been noted in other OAE recordings, but are thought to originate in the activity of the outer hair cells (Wilson and Sutton, 1981; and van Dijk *et al.*, 1994). There have already been suggestions that weak linear coupling among the resonators of the cochlea may be the source of spontaneous OAEs (Fritz *et al.*, 1996). Longitudinal coupling in the cochlea is not a novel concept and has been used to explain a variety of auditory phenomena under a variety of physiological postulates (von Békésy, 1960; Zwislocki and Kletsky, 1980; Jau and Geisler, 1983; Geisler and Sang, 1995; Wickesberg and Geisler, 1986; and Naidu and Mountain, 2001, among others). Thus, consideration of the nonlinear coupling induced through longitudinal stiffness within the cochlear model may provide a description of other unexplained low level auditory phenomena.

The purpose of this paper is to present a one-dimensional model of the CP in which elastic tissue connects the resonant viscoelastic plates of the cochlea. The connective tissue will be modeled by springs, oriented in the longitudinal direction, perpendicular to the transverse motion of the CP, and attached to both (left and right) neighbors. We infer, based on the finding that the organ of Corti introduces further longitudinal stiffness, that the primary source of longitudinal tissue connectivity is the reticular lamina. Framed in this model, an order of magnitude analysis suggests that the reticular lamina can have an appreciable effect on partition dynamics. Furthermore, under the assumption that the cochlear connective tissue behaves in a lossless Hookian fashion, a preliminary numerical solution will demonstrate the weakly compressive nature of this nonlinearity and the cubic distortion product otoacoustic emissions that can result.

II. THE CLASSICAL TRAVELING WAVE MODEL OF THE COCHLEA

A. Hydrodynamics of the traveling wave

The cochlea is a coiled structure carved out from the temporal bone of the skull. For modeling purposes, the cochlea is considered uncoiled with rigid walls tapering from base to apex. The cochlear canal is split into three fluid-filled chambers, the scala vestibuli (upper), the scala tympani (lower), and the scala media (middle). The cochlear partition (CP) is an aggregate description of the structures found in the scala media, which separates the tympanic and vestibular canals. Thus the cochlear model consists of upper and lower fluid-filled chambers separated by the CP [as seen in Fig. 1(a)]. The CP at rest, sits upon the longitudinal x axis, with transverse motion parallel to the z axis. Consequently, the CP transverse displacement (essentially that of the basilar membrane) is herein denoted as $z(x,t)$, where t represents time.

The fluids within the cochlea are assumed to be compositionally similar, acting in a linear, lossless, and incompressible fashion (Viergever, 1980). In the one-dimensional model, the cochlear fluid flows parallel to the CP, where the motion of the CP in the longitudinal direction (along the x axis) is assumed to be negligible. Following the treatment

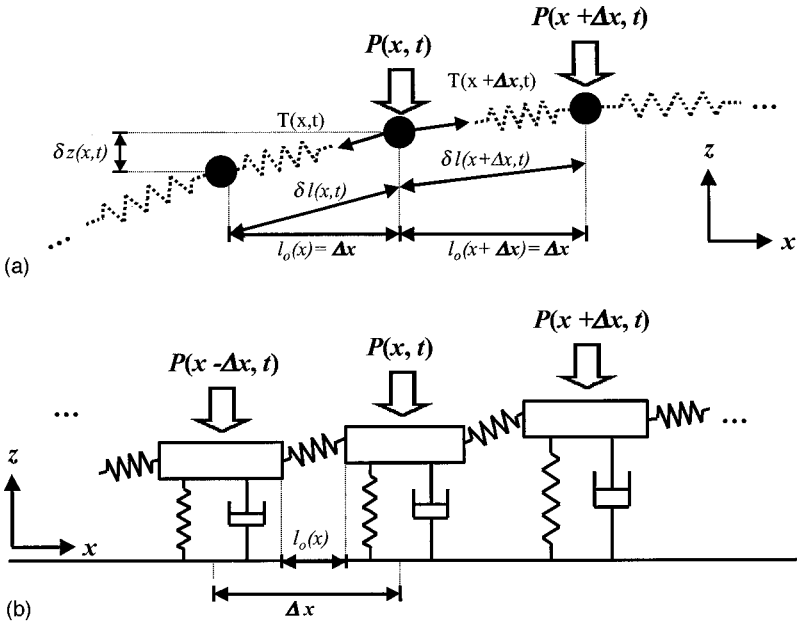


FIG. 2. Schematic of the longitudinal elastic coupling model in motion. (a) The tension between plates is caused by the addition of spring-like components between the plates. The transverse component of the elastic tension is considered in the classic model of viscoelastic motion at any point of the cochlear partition. (b) To simplify the derivation, the plates of the cochlear partition are treated as point-like masses such that the direction of the tensile forces produced by the longitudinal springs follow the slope of the propagating wave. Note the assumption of purely vertical motion insists that the horizontal spacing between the elements remain constant in time.

given by Dallos (1973), the cochlear hydrodynamics are described by

$$\frac{\partial^2 P}{\partial x^2}(x,t) = 2\frac{\rho b(x)}{S(x)}\frac{\partial^2 z}{\partial t^2}(x,t), \tag{1}$$

where $P(x,t)$ is the fluid pressure difference between the scala tympani and scala vestibuli, ρ is the fluid density, $b(x)$ is the width of the CP, and $S(x)$ is the cross-sectional area of the chambers.

B. Cochlear partition dynamics

There are many sophisticated models of the CP found in the literature today. As shown in Fig. 1(b), one of the simplest models involves the description of the CP as a series of viscoelastic plates, each of which behaves as a driven damped oscillator. Each point on the CP is driven by the pressure difference $P(x,t)$. The equation that describes the motion of the CP is given by

$$L_{\mathrm{CP}}[z(x,t)] = P(x,t), \tag{2}$$

where the linear operator $L_{\mathrm{CP}}[z(x,t)]$ is defined as

$$L_{\mathrm{CP}}[z(x,t)] = m(x)\frac{\partial^2 z}{\partial t^2}(x,t) + r(x)\frac{\partial z}{\partial t}(x,t) + k(x)z(x,t), \tag{3}$$

such that $m(x)$ represents the mass, $r(x)$ represents the viscous damping, and $k(x)$ represents the stiffness of the cochlear partition, all measured per unit area.

There is one term missing from Eq. (3) and that is the linear shear viscosity between the resonators. Fundamentally, there is no reason to ignore shear viscosity as it may dominate any other shearing forces within the cochlear partition, such as those produced through longitudinal elasticity (Mammano and Nobili, 1993). However, in this paper we restrict ourselves to the simple model described in Eq. (3). We argue that even for shear viscosity dynamics in the range of the inertial and viscoelastic mechanics, the nonlinearity we derive should still be capable of producing cubic distortions. For example, assume that our nonlinearity is weak compared to the shear viscosity. In a linear world, when one dynamic term is significantly greater than another, the smaller of the two can safely be ignored. In a nonlinear world this is not necessarily the case. Weak nonlinearities can be manifested in the modulation and intermodulation harmonics observed, at frequencies other than the ones at which the system is responding linearly.

As a consequence of its linearity, the cochlear shear viscosity is incapable of frustrating the production of distortion harmonics from weaker nonlinearities. The shear viscosity can only dominate the nonlinear dynamics occurring at the primary stimulus frequencies, not at the distortion frequencies. Thus, we feel that there is no need to add another linear term in our model in this initial exploratory treatment. Inclusion only complicates the numerical model. This simple linear basis outlined in Eq. (3) will make the following treatment and computation transparent at the cost of ignoring the importance of shear viscosity in linear mechanics. Future investigations should be able to extend our derivation to more complicated cochlear models with ease, especially those containing shear viscosity. The exploration of the interplay between the different shearing mechanisms may provide for a fruitful avenue of future research.

III. PARTITION DYNAMICS INCLUDING LONGITUDINAL ELASTICITY

A. Additional simplifications and assumptions

The model being considered is the one shown in Fig. 2(b) and can be modified to include frictional effects. To begin with, assume that the tissue connecting the plates of the CP can be modeled with springs. Springs have the useful property that the tensile forces they impart are proportional to their extension and parallel to their orientation. The derivation can be simplified further with the assumption that the plates forming the CP can be initially treated as point masses,

as shown in Fig. 2(a). The benefit of this approach is clarity; the free body analysis is simpler since the plate-like structure and rotational motion are not considered. Later, this constraint shall be relaxed to include the proper geometry of the arrayed plates.

B. Model of elastic mechanics

The elastic mechanics can be described by adding tensile forces at the ends of the cochlear plate, herein denoted as $T(x,t)$, and directed along the slope of the partition. Following the treatment of transverse waves on a string (Wallace, 1984), a longitudinal force balance combined with the assumption of purely vertical motion yields

$$\frac{\partial}{\partial x}\left[\frac{T(x,t)}{\sqrt{1+\left(\frac{\partial z}{\partial x}(x,t)\right)^2}}\right]\simeq 0. \tag{4}$$

In a similar manner, a transverse force balance yields

$$L_{\mathrm{CP}}[z(x,t)]=\frac{1}{b(x)}\frac{\partial}{\partial x}\left[\frac{T(x,t)\frac{\partial z}{\partial x}(x,t)}{\sqrt{1+\left(\frac{\partial z}{\partial x}(x,t)\right)^2}}\right]+P(x,t). \tag{5}$$

Combining Eqs. (4) and (5), the simple CP dynamics described in Eq. (3) can be augmented to include a nonlinear term $D[z(x,t)]$ such that

$$L_{\mathrm{CP}}[z(x,t)]=D[z(x,t)]+P(x,t), \tag{6}$$

where the nonlinear distortion is given by

$$D[z(x,t)]=\frac{1}{b(x)}\frac{T(x,t)}{\sqrt{1+\left[\frac{\partial z}{\partial x}(x,t)\right]^2}}\frac{\partial^2 z}{\partial x^2}(x,t). \tag{7}$$

The tension, $T(x,t)$, can be modeled in any appropriate manner. In the simplest case where the tension is considered constant, the partial differential equations combining Eqs. (1) and (6) involve a nonlinearity related to the gradient of the displacement, $\partial z/\partial x$. Thus, elastically coupling the sections of the CP results in a nonlinearity, even if the coupling mechanism is itself linear. The nonlinearity is a consequence of the orientation of the components, not their material or mechanical properties.

C. The lossless Hookian model of tissue connectivity

We consider the interconnective tissue that spans the length of the CP to behave in a Hookian fashion. This is valid under the assumption that the change in length of the interconnecting springs is small compared to their rest lengths (as will be demonstrated later). Hookian springs have the attractive feature that they only generate tensile forces when extended. They do not exhibit static tension, which agrees with observation (von Békésy, 1948). Furthermore, for small extensions Hookian springs produce weak tensile forces. This is also consistent with observations of fresh basilar membrane preparations (Völdrich, 1978; Naidu and Mountain, 2001). With these two thoughts in mind, it seems reasonable, as a first approximation, to consider lossless Hookian elasticity to describe the longitudinal connective tissue.

Let $A_{\mathrm{CP}}(x)$, $Y_{\mathrm{CP}}(x)$, and $\tau_{\mathrm{CP}}(x)$, respectively, be the cross-sectional area, Young's modulus, and the effective thickness of the interconnecting CP tissue. If $\delta l_{\mathrm{PM}}(x,t)$ represents the change in the length of the tissue between the point masses, and the rest length of the tissue is taken to be equal to Δx, then Hooke's law states that the tensile forces obey

$$T(x,t)=\frac{A_{\mathrm{CP}}(x)Y_{\mathrm{CP}}(x)}{\Delta x}\delta l_{\mathrm{PM}}(x,t). \tag{8}$$

Using the Pythagorean theorem to determine the change in length with respect to the rest length, the tensile force produced in a Hookian tissue obey, in the continuous case,

$$\lim_{\Delta x\to 0} T(x,t)=\tau_{\mathrm{CP}}(x)b(x)Y_{\mathrm{CP}}(x)\left[\sqrt{1+\left(\frac{\partial z}{\partial x}\right)^2}-1\right]. \tag{9}$$

Thus, in the case of Hookian tissue elasticity without considering plate geometry, the distortion (7) reduces to

$$D[z(x,t)]=\tau_{\mathrm{CP}}(x)Y_{\mathrm{CP}}(x)\left[1-\frac{1}{\sqrt{1+\left(\frac{\partial z}{\partial x}\right)^2}}\right]\frac{\partial^2 z}{\partial x^2}. \tag{10}$$

D. Adjusting the dynamics to account for plate lengths

We now relax the assumption that the CP behaves as an array of point masses connected by springs [Fig. 2(a) versus Fig. 2(b)]. If the plate-like morphology is taken into account, then the rest lengths of the springs will decrease. Consequently, the strain the springs experience will increase, such that the gradient and Laplacian of the displacement, $z(x,t)$, will be heightened by a factor G_Δ defined as

$$G_\Delta=\frac{\Delta x}{l_0(x)}, \tag{11}$$

where $l_0(x)$ is the actual effective rest length of the tissue connecting the resonant plates defined for the region between the ends of neighboring plates. Substituting Eq. (11) into Eq. (10) results in the Hookian distortion,

$$D[z(x,t)]=G_\Delta\tau_{\mathrm{CP}}(x)Y_{\mathrm{CP}}(x)\left[1-\frac{1}{\sqrt{1+\left(G_\Delta\frac{\partial z}{\partial x}\right)^2}}\right]\frac{\partial^2 z}{\partial x^2}. \tag{12}$$

E. Adjusting dynamics for motion of the reticular lamina

One last consideration needs to be made. It has been shown that the reticular lamina and basilar membrane attain

different amplitudes when vibrating to a tonal stimulus. Mammano and Ashmore (1993) have shown that the reticular lamina displacement is larger than that of the basilar membrane by a factor of approximately 5. Their data suggest an even larger difference for some frequencies of stimulation. The motion of the reticular lamina is important in the framework of our model since we postulate that it is a source of significant longitudinal stiffness. If the forces generated at the reticular lamina by the longitudinal stiffness are assumed to be effectively conveyed to the basilar membrane through the outer hair cell and Deiter cell complex, then we would expect the effect of the distortion at the level of the basilar membrane to magnify. The gradient and Laplacian at the level of the reticular lamina will be heightened by a factor G_r defined as

$$G_r = \frac{z_{\mathrm{RL}}(x,t)}{z(x,t)}, \tag{13}$$

where the basilar membrane displacement is essentially the displacement of the CP, $z(x,t)$.

This gain has an equivalent effect to G_Δ. Rather than the reduction of the rest lengths, the height differential is increased. The effect is to increase the effect of the gradient and Laplacian by another factor G_r. If we define the combined gain G as

$$G = G_\Delta G_r, \tag{14}$$

then distortion nonlinearity takes the form

$$D[z(x,t)] = G\tau_{\mathrm{CP}}(x) Y_{\mathrm{CP}}(x) \left[1 - \frac{1}{\sqrt{1 + \left(G \frac{\partial z}{\partial x} \right)^2}} \right] \frac{\partial^2 z}{\partial x^2}. \tag{15}$$

F. The weak cubic nonlinearity

To simplify expression (12) a Taylor approximation can be employed to yield

$$\left[1 - \frac{1}{\sqrt{1 + G^2 \left(\frac{\partial z}{\partial x} \right)^2}} \right] = \frac{1}{2} \left(G \frac{\partial z}{\partial x} \right)^2 - \frac{3}{8} \left(G \frac{\partial z}{\partial x} \right)^4 + \cdots. \tag{16}$$

If we assume $|G(\partial z/\partial x)| \ll 1$ (which will be demonstrated later) then the Hookian distortion (12) takes the nonlinear cubic form

$$D[z(x,t)] \simeq \epsilon(x) \left[\frac{\partial z}{\partial x}(x,t) \right]^2 \frac{\partial^2 z}{\partial x^2}(x,t), \tag{17}$$

where

$$\epsilon(x) = G^3 \tau_{\mathrm{CP}}(x) Y_{\mathrm{CP}}(x). \tag{18}$$

For ease of reference, define $\epsilon(x)$ as the coefficient of tissue connectivity. Therefore, at least in form, the addition of longitudinal elasticity appears to produce a cubic nonlinearity with proportionality factor $\epsilon(x)$. In principal, our hypothesis concerning the production of distortion product otoacoustic emissions appears strengthened. However, though the dynamics reduce to a cubic nonlinearity, they may be negligible in comparison to the inertial, viscous, and stiffness properties of the CP as introduced in Eq. (3). If negligible, then the above-mentioned derivation constitutes only a mild curiosity.

IV. RECONSIDERING THE NEGLIGIBILITY OF LONGITUDINAL ELASTICITY

A. Longitudinal coupling in the cochlear partition

The longitudinal elasticity is considered weak in comparison to the inertial and viscoelastic properties of the CP (Mammano and Nobili, 1993). There are three primary sources of longitudinal elasticity: the basilar membrane, the tectorial membrane, and the reticular lamina. Both the basilar membrane and tectorial membrane are composed of an amorphous acellular ground substance, which is considerably less stiff than the fibers of the basilar membrane (Steele *et al.*, 2000).

B. The reticular lamina

In contrast, the reticular lamina is formed by the ''tight'' hexagonal tiling of the tops of the outer hair cells and the phalangeal processes of the Dieter's cells (Mammano and Nobili, 1993). Functionally, the tight cellular junctions serve to separate the endolymphatic and perilymphatic fluids (Santi, 1988). Due to this peculiar morphology, the reticular lamina is considered to be radially stiff, but somewhat compliant to longitudinal stretching. As has been shown in Eq. (17), if the tissue is weakly Hookian and is governed by transverse dynamics, then the longitudinal elasticity is proportional to a spatial nonlinearity which involves the product of the traveling wave gradient and Laplacian. The gradient is generally smaller than 10^{-2} and the dynamics appear to be negligible (Mammano and Nobili, 1993). However, such a conclusion would be premature without any estimate of the proportionality constant (18), $\epsilon(x)$, or the Laplacian of the displacement, $(\partial z/\partial x)(x,t)$.

It has also been argued that the structure of the RL is more suited to bending and buckling under strain. This argument depends strongly on the subcellular structures providing the coupling between the phalangeal processes and the outer hair cells. Actin-based or other stiff collagenous materials, if found, would suggest that the reticular lamina would not be prone to buckling. The stiffer the interconnection the less likely the buckling is to occur. However, these physiological studies have not been carried out and the question as to the structure and mechanics between the plates of the RL is still uncertain. Under these conditions, and until conclusive measurements are carried out, we feel justified in exploring the conjecture that the interconnections behave in some elastic manner.

C. Composition and morphology of reticular lamina

Both the composition and morphology of the reticular lamina suggest that the effects of longitudinal elasticity may be greater than previously thought. Compositionally, the reticular lamina is formed by the junction of cells. Cellular junctions are known to be considerably stiffer than acellular

material. Such biological tissue, like healthy and diseased aortic tissue (Vorp *et al.*, 1998), have Young's moduli in the range 10^5-10^7 kPa (Fletcher, 1992); several orders of magnitude greater than the 200 and 30 kPa found in the connective ground material of the basilar and tectorial membrane (Steele *et al.*, 2000). For example, Reissner's membrane is a thin sheet of cells formed by tight cellular junctions with a Young's modulus on the order of 60 MPa (Steele *et al.*, 2000).

Furthermore, the elastic modulus of microtubules, found throughout the pillar cells, phalangeal processes, and Deiter's cells have elastic moduli in the range of 1–3 GPa, suggesting that the organ of Corti as a whole may offer appreciable longitudinal stiffness. If the tight connections between the plates of the reticular lamina are like that of Reissner's membrane, then the longitudinal elastic modulus is several orders of magnitude greater than that of the basilar membrane or tectorial membrane. This estimate is still several orders of magnitude smaller than the 3 GPa elastic modulus of the collagenous fibers of the basilar membrane and so the model we pursue maintains the dominance of the transverse stiffness over the longitudinal stiffness (Steele *et al.*, 2000).

Additionally, the plate-like morphology of the reticular lamina suggests that the strain experienced by the connective tissue is larger than the continuous membrane models. Consequently, the estimates of the Laplacian and gradient of the displacement will increase by a factor of G_Δ^3, effectively inflating the estimate of the coefficient of tissue connectivity. Even under the thin tissue assumption, the inclusion of these observations challenge the underlying assumption that the organ of Corti functions as a free viscoelastic strip because the proportionality constant is effectively larger than previously thought.

We continue our line of thought by reconsidering the negligibility of longitudinal elasticity in CP dynamics. To do so in a quantitative manner, an order of magnitude analysis is constructed on the model parameters given in Eqs. (17) and (18). First, the negligibility of longitudinal tissue elasticity of the basilar membrane is verified. The same analysis is then performed for the reticular lamina to determine the significance of its longitudinal elastic contribution.

V. ORDER OF MAGNITUDE ANALYSIS

A. Basilar membrane longitudinal elasticity

The basilar membrane is generally thought to be the source of transverse stiffness in the cochlear partition. The longitudinal elasticity of the basilar membrane is dominated by a ground acellular gel-like substance that surrounds the stiff radially oriented fibers (Santi, 1988). From images of the guinea pig cochlea, the thickness of the basilar membrane is taken to be on the order of 10 μm (Echteler, 1994). In the longitudinal direction, Young's modulus is taken to be approximately 200 kPa (Steele *et al.*, 2000). The wavelength of the cochlear traveling wave is taken to vary between 100 μm at the base (Ren, 2001) to 1 mm near the apex (Viergever, 1980), though it may be smaller in the region of resonance if active mechanics are involved. The addition of G_Δ and G_r are not necessary here; they apply at the level of the reticular lamina with its plate-like composition and heightened displacement. The maximum displacement of the basilar membrane ranges between 10 and 100 nm (Johnstone *et al.*, 1986) and can be even higher for large pressure stimuli in excess of 100 dB (SPL). For this analysis, assume that the maximum displacement is about 10 nm. With these estimates in hand, the gradient can be shown to be on the order of

$$\left|\frac{\partial z}{\partial x}\right| \sim \frac{z_{\max}}{\lambda(x)} < 10^{-4}, \tag{19}$$

whereas the Laplacian is at most

$$\left|\frac{\partial^2 z}{\partial x^2}\right| \sim \frac{z_{\max}}{\lambda^2(x)} < 10^{-6}\frac{1}{\mu\mathrm{m}}. \tag{20}$$

Therefore, the distortion introduced by including longitudinal elasticity in the basilar membrane will be in the range

$$\|D_{\mathrm{bm}}[z(x,t)]\| \sim \frac{\tau_{\mathrm{bm}} Y_{\mathrm{bm}} z_{\max}^3}{\lambda^4(x)} \in [10^{-9}\ \mathrm{Pa}, 10^{-5}\ \mathrm{Pa}]. \tag{21}$$

This result can be compared to the inertial and viscoelastic dynamics using similar analysis. Using reasonable values for mass, resistance, and stiffness given for human cadaver cochleae (see Appendix C), the inertial, viscous, and stiffness terms yield

$$\|m(x)\ddot{z}(x,t)\| \sim m(x) z_{\max}\omega^2 \in [10^{-2}\ \mathrm{Pa}, 10^{3}\ \mathrm{Pa}], \tag{22}$$

$$\|r(x)\dot{z}(x,t)\| \sim r(x) z_{\max}\omega \in [10^{-3}\ \mathrm{Pa}, 10^{2}\ \mathrm{Pa}], \tag{23}$$

$$\|k(x)z(x,t)\| \sim k(x) z_{\max} \in [10^{-3}\ \mathrm{Pa}, 10^{3}\ \mathrm{Pa}]. \tag{24}$$

Living cochleae could have viscoelastic dynamics much smaller than suggested above. It is clear that the viscoelastic dynamics are five to seven orders of magnitude larger than the longitudinal stiffness dynamics of the basilar membrane. These results are summarized in Fig. 3. Hence, the longitudinal elasticity introduced by consideration of the ground substance between the radial fibers of the basilar membrane is negligible compared to the transverse stiffness of the basilar membrane. This suggests that the primary source of stiffness in the basilar membrane is transverse, which agrees well with the literature, and lends credibility to the approach. Therefore, the additional longitudinal elastic mechanics would have a weakly perturbative effect on the linear dynamics.

B. The reticular lamina longitudinal elasticity

The same analysis can also be carried out for the reticular lamina. Though a thin layer, we suggest that the reticular lamina is longitudinally stiffer than the basilar membrane. The reticular lamina is formed by tight but thin junctions between the tops of the motile outer hair cells and supportive Dieter cell phalangeal processes. For this reason, we assume that Young's modulus for the reticular lamina is similar to that of Reissner's membrane, with a value of 60 MPa, approximately. The effective thickness of the reticular lamina, τ_{RL}, should be reflective of the actual degree of contact between the junctions of the cuticular plates. Based on various

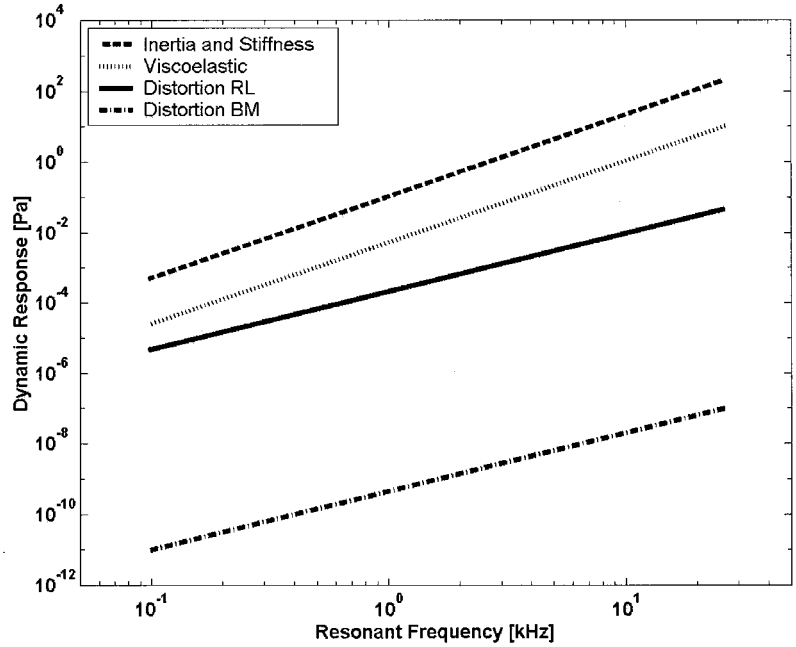


FIG. 3. Order of magnitude comparison of dynamical terms in one-dimensional cochlear model. Order of magnitude estimates of the dynamic terms are evaluated at the resonant frequencies of the longitudinal positions. The transverse stiffness and viscoelastic dynamics compare well with the dynamics of the distortion caused by the reticular lamina. The distortion caused by longitudinal stiffness in the basilar membrane is several orders of magnitude smaller than the smallest dynamic, and is considered negligible in comparison. Parameters used are listed in the appendix and detailed in the paper.

sources (Santi, 1988; Lim, 1986), τ_{RL} should be on the order of 1 μm. In addition, we estimate G_{Δ} to be approximately 10 and G_r to be approximately 5.4. With these estimates, an order of magnitude estimate of the elastic distortion in the reticular lamina yields

$$\|D_{\mathrm{RL}}[z(x,t)]\| \sim G_{\Delta}^3 G_r^3 \frac{\tau_{\mathrm{RL}} Y_{\mathrm{RL}} z_{\max}^3}{\lambda^4} \in [10^{-5}\ \mathrm{Pa}, 10^{-1}\ \mathrm{Pa}]. \quad (25)$$

From an order of magnitude analysis, the effect of the distortion is two or three orders of magnitude smaller than the viscoelastic dynamics. However, DPOAEs are recorded in the range of -10 to 10 dB(SPL) for sound stimuli with levels of 50 to 70 dB(SPL). This implies that there is a three to four order of magnitude difference between the stimulus and recorded DPOAE pressures. The cubic distortion produced by the reticular lamina is well within this range, whereas that of the basilar membrane is several orders of magnitude smaller than this.

The Hookian approximation for Eq. (25) is also validated by this analysis. Consider

$$\left| G \frac{\partial z}{\partial x} \right| \sim 10^{-3} \ll 1.$$

This relation serves as the basis for taking the Taylor approximation that led to the cubic nonlinear form of the distortion. The Hookian description of the tissue is determined by the strain experienced by the connective tissue. For wavelike motion, the strain is proportional to the gradient. Since the gradient is much smaller than one, the Hookian approximation for the tissue dynamics is justified.

From this order of magnitude analysis, the distortion at the level of the reticular lamina appears to be appreciable. Though weaker than the linear inertial and stiffness dynamics, it is still within an order of magnitude of the linear viscoelastic properties of the CP (see Fig. 3). Note that these estimates were based on order of magnitude estimates of the physiological parameters $Z_{\max}$, Y_{RL}, τ_{RL}, G_{Δ}, or G_r—and could be larger. For instance, if the maximum displacement of the cochlear partition is taken to be 50 nm instead of 10 nm, the estimates for both the reticular lamina and basilar membrane distortion are amplified by two orders of magnitude and become comparable to the inertial dynamics. The distortions are pronounced for the low and mid-frequencies in the range 200 Hz–2 kHz. The dynamics are only three magnitudes smaller for frequencies in the 10–k 20 kHz range. Therefore, though the elastic modulus of the reticular lamina is two to three orders of magnitude smaller than that of the collagenous fibers of the basilar membrane, the longitudinal stiffness dynamics are not negligible.

VI. NUMERICAL SOLUTION AND COMPUTATIONAL METHOD

To study the effect of the nonlinear distortion (17) on cochlear dynamics, a computer model of the human auditory system was constructed. The model was based on the cochlear traveling wave presented by Dallos (1973) under the physical assumptions and anatomical considerations detailed by Viergever (1977, 1978, 1980). These considerations are for the cadaver human cochlea such that no active mechanics are included. This model is encapsulated in the linear CP mechanics (3) and linear cochlear hydrodynamics (1). In our numerical model, the linear CP mechanics is replaced by the nonlinear expression (6) relating the linear model and the nonlinear distortion produced when longitudinal elasticity is considered. Diependaal's (1988) adaptation of the Matthews (1980) linear second-order middle ear model was used as the interface between incoming sound and the nonlinear cochlea so that sound pressures could be estimated at the ear drum. All parameters were chosen to reflect human cochlear and middle ear characteristics (as noted in Appendix D).

The computational framework used to simulate the cochlear nonlinear partial differential equations is that pre-

sented by Diependaal *et al.* (1987). The method was designed for one-dimensional models of the CP including active and nonlinear mechanical properties. The two coupled nonlinear partial differential equations can be decomposed into a large system of ordinary nonlinear differential equations using Galerkin's principle. The nonlinearity was approximated using standard discretizations of the gradient and Laplacian of the CP displacement wave form (Strauss, 1992). The resulting nonlinear system of differential equations was integrated in time using a variable-step fourth-order Runge–Kutta scheme with an error tolerance set to a displacement of 1 fm. As a trade-off between speed and accuracy, the model human cochlea was discretized into 350 elements. Stability was a concern since the nonlinearity involved is spatial, whereas Diependaal's technique appears to be suited for exploring temporal nonlinearities. Though perhaps not the ideal method for accurately computing stable numerical solutions of nonlinear partial differential equations, this technique did produce stable solutions for the weak nonlinearities described in this paper.

VII. RESULTS AND DISCUSSION

A. Weakly compressive behavior

Using the above-presented numerical solution technique, a cursory exploration into the nature of the nonlinearity given in Eq. (17) was undertaken. For the connective tissue of the reticular lamina we assumed a Young's modulus ranging between 60 MPa and 1 GPa. The uncertainty in the estimates of G are the primary source of variation and the tissue connectivity ϵ could be larger. A 1 kHz tone was presented to the model at a variety of sound levels. The partition displacement was observed and its maximum value was recorded over 220 ms. For all values of ϵ, the nonlinearity demonstrated a weak compression of the displacement amplitude for sound levels ranging from 50 to 70 dB(SPL), but had negligible effect for lower sound levels. This is shown in Fig. 4(a), where the maximum partition displacement is plotted against sound level. The elastic modulus of the tissue was taken to be 100 MPa to demonstrate the trend. At 70 dB(SPL), there was a 10% reduction in partition amplitude, which if compared to the linear model would result in a 1 dB(SPL) drop in hearing level. This is not a very large reduction and demonstrates that the nonlinearity has a small impact upon normal cochlear function. For pathologically large values of tissue connectivity, the drop is even larger, though difficult to quantify because of numerical instability.

Under normal circumstances the nonlinearity is weakly compressive, however under pathological cases of hardened tissue, hearing ability should be compromised. Higher tissue connectivities were not explored because of instability in the numerical solution, however the downward exponential trend is visible in Fig. 4(b), where we have extended the range of the tissue modulus to that of the basilar membrane. We are not suggesting that the reticular lamina has longitudinal stiffness comparable to the basilar membrane, only that compressive effects of longitudinal elasticity is only appreciable for abnormally large (possibly pathological) values of membrane stiffness. This agrees with intuition, since an extremely stiff longitudinal connection should produce no motion in the cochlear partition regardless of the outer hair cell active mechanism. Thus the model suggests that pathological hardening of the tissue between resonant elements in the cochlear partition should result in hearing loss even if the underlying outer hair cells and basilar membrane are functioning normally. Therefore, the addition of a pathological level of longitudinal stiffness should lead to hearing loss.

B. Distortion product otoacoustic emissions

Our hypothesis centered on the idea that the addition of longitudinal elastic mechanics could produce cubic distortion product otoacoustic emissions. To test this assertion, two 70 dB(SPL) tones at 1000 and 1210 kHz were presented to the model. The elastic modulus was set at 100 MPa. The output of the model was sampled at 10 kHz for 1 s. The spectrum was estimated using the Welch periodogram. The noise in the model is the result of model error, that is the error in the Runge–Kutta procedure used to evolve the solution forward in time.

The results of this simulation, presented in Fig. 5, support the hypothesis presented in this paper. For the frequencies of the two primaries, clear cubic distortion product otoacoustic emissions (DPOAEs) were observed at frequencies 790, 1420, 3210, and 3420 Hz representing the $2f_1-f_2$, $2f_2-f_1$, $2f_2+f_1$, and $2f_1+f_2$, respectively. The largest was the $2f_1-f_2$ DPOAE, sitting five standard deviations above the noise floor at -2 dB(SPL), whereas the others were below the four standard deviation level. Curiously, the $2f_1-f_2$ DPOAE is known to be the largest harmonic produced in such distortion tests taken in humans and other mammals (Probst *et al.*, 1991). This might be a coincidence, but appears to hold over a broad range of tissue connectivity values. The longitudinal tissue elasticity may provide an explanation for this previously inexplicable phenomenon in human DPOAE production. Furthermore, DPOAEs were obtained for f_1 primaries up to 2 kHz. Thereafter the numerical solution became unstable.

In addition to the cubic distortions observed a number of other harmonics were also visible three standard deviations above the noise floor. For example, significant spectral energy was observed for 103, 394, and 1570 Hz. The nature of these distortions has yet to be categorized, be they harmonic, intermodulation, or the result of some chaotic process. If the longitudinal elasticity in living cochlea is large enough to affect cochlear dynamics in the manner suggested in this paper, then the observation of these other harmonics can help distinguish the consequences of this nonlinearity compared to any other DPOAE producing nonlinearity conjectured to exist in the cochlea.

C. Speculations on active processes in the cochlea

We would like to point out a difference between a true mathematical cubic that behaves passively and the cubic term that we developed in this paper. When simple resonators are coupled to one another, the additional mechanical term introduced to the motion of one resonator is related to the motion of the other resonator. However, the motion is

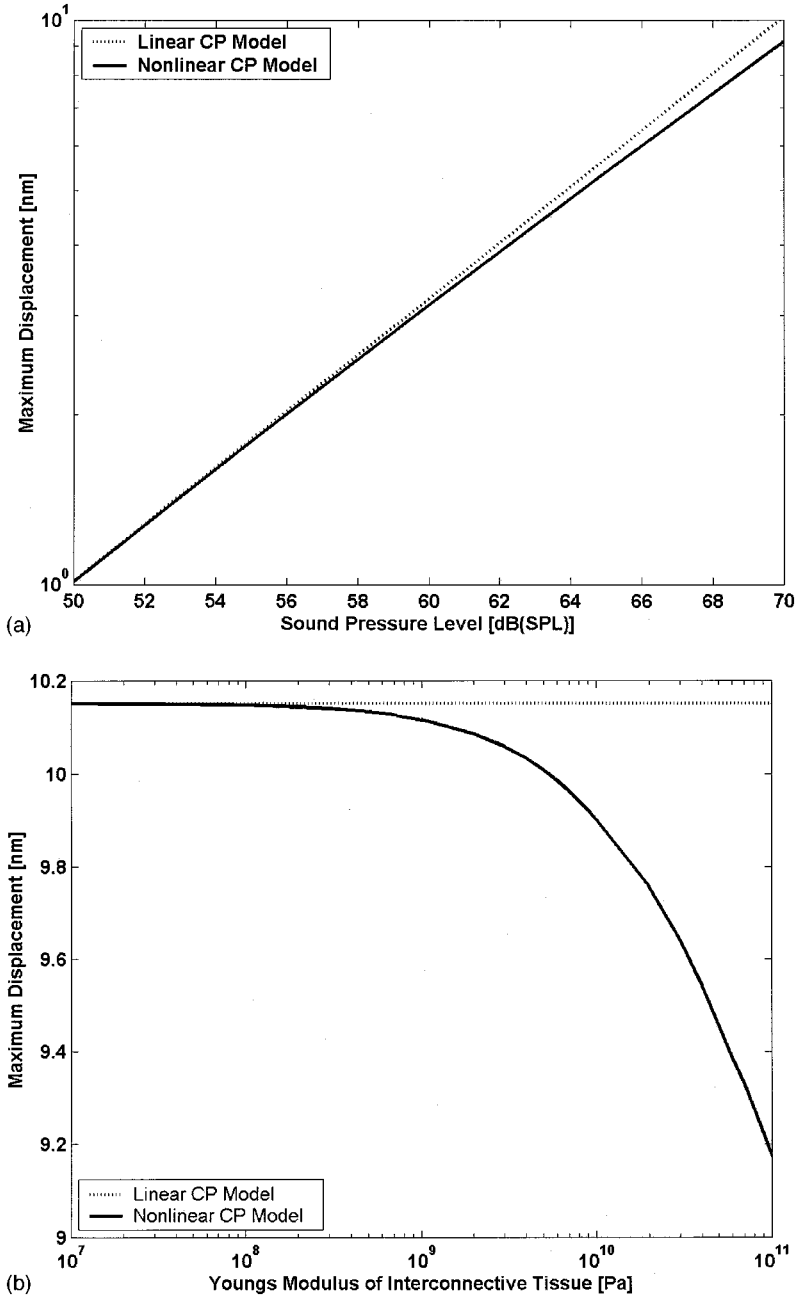


FIG. 4. (a) The compressive nature of the tissue connectivity nonlinearity is demonstrated for a tissue with Young's modulus of 100 MPa. Though not evident due to the logarithmic scale, at 70 dB SPL, the partition loses approximately 10% of its maximum displacement as compared to the linear system. (b) The maximum displacement of the cochlear partition in the nonlinear model is plotted against increasing tissue elastic modulus. The horizontal line is the maximum displacement seen in the linear model. The larger the stiffness of the tissue, the greater the reduction in partition displacement. These results were obtained for a 1 kHz tone at 70 dB SPL.

time-dependent, and so it could be argued that the coupling has introduced an active nonlinearity; that is, energy from a source other than the self-energy of a resonator influences its motion. It is not clear that the mechanical nonlinearity we propose should behave like a typical passive nonlinearity because the motion of any given resonator is coupled to the motion of its nearest neighbors. Energy is allowed to pass from one to another. For any given cochlear resonator, energy is brought in externally through the coupling. This clouds the notion of an active nonlinearity, since active models generally assume no longitudinal coupling, and use local time-varying nonlinearities to describe external injections of energy.

D. Restrictions on outer hair cell activity

The dynamic associated with the elastic coupling presented in Eq. (25) is related to the Laplacian of the traveling wave of CP displacement. Active cochlear mechanics are supposed to ''sharpen'' the cochlear tuning curve at a given location for a specific tone. The distortion could increase dramatically if the quality factor of the resonant members is sufficiently high to produce cusp-like behavior in the spatial wave form of the traveling wave. Sharp displacements lead to second derivatives whose numerical values are very large. This sharpness should mainly occur at the characteristic frequency where the high frequency regions of the CP would be most susceptible since the Laplacian of the traveling wave is largest there.

In such regions, the dynamics associated with the elastic longitudinal coupling would get very large in proportion to the Laplacian, effectively magnifying the effect of the longitudinal stiffness. If the purpose of the outer hair cells is to beneficially sharpen the response of the CP for a given tone, then this activity would increase the stress experienced by

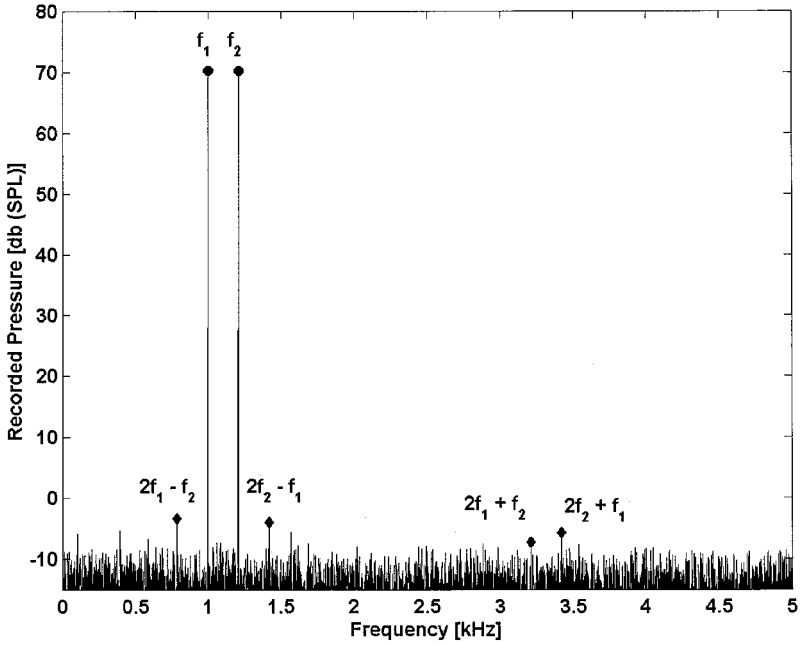


FIG. 5. The two primaries were given at 70 dB(SPL) with frequencies 1 and 1.21 kHz. The cubic distortion product $2f_1-f_2$ (790 Hz) is visible five standard deviations above the noise floor. The level of the cubic distortion is approximately -2 dB (SPL), which is within range for a normal hearing individual. The other cubic distortions (1420, 3210 Hz) are three standard deviations above the noise floor, as are a number of other unrelated harmonics (103, 394, and 1570 Hz). The elastic modulus was set to 100 MPa.

the partition structure. Sharp behavior would lead to excessive longitudinal stress in the CP and present a danger to the integrity of the CP, in that the basilar membrane or reticular lamina could tear. Such tears would be detrimental to the functioning of the outer hair cells because such damage allows the ionically rich fluids of the scalae to mix. The consideration of longitudinal stiffness may impose restrictions upon the postulated forms of active mechanics of the outer hair cells.

E. Propagation of pathology

The longitudinal elastic coupling should be significant in any region of the cochlear partition that vibrates ''sharply.'' Under normal circumstances, this occurs primarily at the characteristic frequency, not at the base or apex where the traveling wave tends to flatten out. Consider an abnormal situation where pathological regions along the CP exist. These regions are immobile as opposed to the surrounding healthy tissue. Such conditions would lead to large displacement differentials, increasing the longitudinal stress. The consequence of this is that the damage could propagate along the CP over time because the increased strain increases the likelihood of a tear occurring in the neighboring regions. Hence, this model suggests that once cochlear conductive pathology begins, it continues. The model also suggests that DPOAE may also arise from such pathology and may be identifiable in a clinical setting.

F. The protective function of the phalangeal processes

Within the framework of the model put forward in this paper, suggestions can be made as to the function of the phalangeal processes and their curious cross-linked structure. The cross-linking between the tops of the outer hair cells and the neighboring Deiter cells may be an evolutionary adaptation to protect the contiguous sensory epithelium from tearing. Such damage could be caused by excessive stress in the connecting tissue due to the high resonance induced by the active processes of the outer hair cells. The stiff phalangeal processes act to couple the motion of one section to that of another, but the coupling is restrictive. In other words, the motion of one region of the CP must be similar to its neighbors.

The benefit of this adaptation is that large variations in CP displacement cannot occur along any two neighboring points. From the point of view of our model, this would ensure that singularities in the second derivative would have difficulty occurring and alleviate some of the risk involved with sharp tuning. Furthermore, if there was some localized trauma, the cross-linked structure would ensure that large displacement differentials could not occur and further the trauma. Therefore, we suggest that in order to involve active mechanics in the sensory mechanical process the cross-linked structure of the phalangeal processes has been evolved to maintain epithelial contiguity under sharp resonant conditions.

G. Identification of cochlear conductive presbycusis

Our primary motivation for pursuing a model containing longitudinal stiffness stems from our need to understand how the pathological hardening of cochlear tissue affects cochlear function. Of immediate concern to us is conductive presbycusis—hearing loss associated with a thickening of the basilar membrane (Nadol, Jr., 1979; Bohne *et al.*, 1990). Recent studies have implicated gene activity to hearing loss (Willems, 2000). For example the COCH gene has been implicated in a form of hearing loss in which mucopolysaccharides are deposited within the cochlea and vestibule. Some day a genetic abnormality may be found which leads to protein deposits or some other stiffening mechanism affecting the cellular structure of the organ of Corti. Under these circumstances, the clinical detection of conductive presbycusis and its discrimination from other forms of hearing pathology

would be valuable. We have demonstrated that longitudinal elastic dynamics within the CP involve a cubic nonlinearity and so we suggest that tissue stiffening may be identifiable through the use of DPOAEs.

There are studies that have measured the $2f_1-f_2$ emission in mouse models of presbycusis (Parhem *et al.*, 1999 for example). In these models, DPOAE levels decline with age. Histological examination suggests that there could be any number of sources for the hearing loss, though in the cases studied, sensory degeneration appears to be the primary culprit. From our model, we know that DPOAEs can be elicited in cochlea with longitudinally stiffened tissue. Our preliminary results suggest the following conjecture. If there exists an individual with a poor audiogram (or steady-state brain stem response) who is still producing DPOAEs, then perhaps that person suffers from cochlear conductive hearing loss. If we can identify these cases before the condition becomes mixed with other sensory, neural, or metabolic degradation (Schuknecht and Gacek, 1993), then perhaps medical intervention can remedy the pathology before hearing loss gets worse. The confirmation of such a conjecture would increase the clinical relevance of DPOAE-based hearing tests.

VIII. CONCLUSIONS AND FUTURE WORK

Considering the complexity of the structural composition of the tissue comprising the organ of Corti, it seems reasonable to postulate a passive longitudinal elastic coupling mechanism at the level of the cochlear partition. This paper has detailed a model of the cochlear partition in which a weak elastic longitudinal coupling has been included between the resonant viscoelastic plates of the cochlea. The connective tissue between plates was modeled by springs oriented in the longitudinal direction, perpendicular to the transverse motion of the cochlear partition. Under the assumptions of Hookian behavior, it has been shown that the nature of such a dynamic is not negligible, though weak, and exhibits a cubic nonlinearity.

Modern cochlear models include nonlinear dynamics, resonant modes, and structural coupling to enhance the traveling wave model. Such modifications have suggested explanations of various otoacoustic and perceptual phenomena. The addition of longitudinal elastic dynamics produce cubic distortion product otoacoustic emissions where the $2f_1-f_2$ intermodulation component is largest. The weak nonlinear dynamics presented in this paper may produce other observable low level otoacoustic emissions, such as the transiently evoked and spontaneous variety. Other chaotic nonlinear phenomena may also be exhibited such as entrainment, beating, and frequency dispersion. Future work will explore such responses of this model.

ACKNOWLEDGMENT

This work was supported by the Natural Sciences and Engineering Research Council of Canada, Grant No. 4316-98.

APPENDIX A: MODEL BOUNDARY CONDITIONS

The boundary conditions for this nonlinear model extend the set of boundary conditions for the traveling wave model. At the base of the cochlear model, the pressure gradient driving the fluid is attributed solely to the motion of the stapes since the cochlear fluids behave incompressibly. At the apex, there is no pressure difference between the two chambers since the fluids mingle through the helicotremal opening. The traveling wave boundary conditions are given as

$$\frac{\partial P}{\partial x}(0,t)=2\rho\frac{S_{\mathrm{OW}}}{S_0}\frac{\partial u_{\mathrm{st}}}{\partial t}(t), \tag{A1}$$

$$\frac{\partial P}{\partial x}(L_c,t)=0, \tag{A2}$$

where S_{OW} is the area of the oval window, S_0 is the area of vestibular chamber at the base of the cochlea, L_c is the length of the CP, and $u_{\mathrm{st}}(t)$ is the velocity of the stapes.

The addition of longitudinal elasticity necessitates the addition of two more boundary conditions. First consider the apex of the cochlea where the CP is untethered, and so the end is free. This final segment of the CP experiences no strain since there both ends are not being pulled and thus the segment has no change in length. Since the strain is proportional to the gradient of CP displacement we suggest that this gradient at the cochlear apex is equal to zero. Thus an additional helicotremal boundary condition must be observed, that is,

$$\frac{\partial z}{\partial x}(L_c,t)=0. \tag{A3}$$

The boundary condition at the cochlear base is motivated by insights gained through consideration of the model developed in this paper. We assume that the tissue forming the organ of Corti seemlessly incorporates itself into the bony shelf that protrudes into the cochlear duct. The CP does not behave as a pinioned elastic strip where only the continuity of the partition displacement is a concern. The gradient of the displacement must also be zero otherwise a discontinuity in the gradient would develope at the cochlear base. If such a discontinuity occurs then a singularity in the Laplacian develops. Based on the model presented in this paper, the longitudinal tensile forces generated at such a boundary would be so high that they would tear the cochlear partition from the base. Any dynamics involving longitudinal coupling would subsequently explode and damage the CP. To ensure that this does not happen, then minimally the gradient at the base has to be constrained to zero so that the Laplacian is not singular. Thus, an additional boundary condition at the cochlear base must be observed, namely,

$$\frac{\partial z}{\partial x}(0,t)=0. \tag{A4}$$

APPENDIX B: MODEL OF THE MIDDLE EAR

We can couple a simplified model of the middle ear to the nonlinear cochlear model developed. This is a beneficial addition because we can stimulate the model with ear canal

pressures and so simulate experiments. Otherwise, the stimulus to the cochlear model would be in terms of the stapes displacement which is more difficult to interpret and apply to experimental observations. Following Diependaal (1988), we can approximate the middle ear as

$$T_m S_m P_e(t) + S_{\mathrm{OW}} P(0,t) = [T_m^2 m_m + m_s]\ddot{\xi}_{\mathrm{st}}(t) + T_m^2 r_m \ddot{\xi}_{\mathrm{st}}(t) + k_m \xi_{\mathrm{st}}(t),$$

where $\xi_s(t)$ represents the displacement of the stapes. All other symbols are defined in Appendix B. Under this model of the middle ear the basal boundary condition of the cochlear model becomes

$$\frac{\partial P}{\partial x}(0,t) = 2\rho \frac{S_{\mathrm{OW}}}{S_0} \frac{\partial^2 \xi_{\mathrm{st}}}{\partial t^2}(t), \tag{B1}$$

where the time rate of change of the stapedial velocity $u_{\mathrm{st}}(t)$ is replaced by the acceleration of stapes. The observed literature values of T_m, S_m, S_{OW}, m_m, m_s. The resistance, r_m, and stiffness of the middle ear, k_m, are chosen such that resonant frequency of the middle ear is approximately 1.3 kHz and damping ratio is approximately 0.7.

APPENDIX C: LIST OF SYMBOLS

Symbol	Description	Units
$A_{\mathrm{CP}}(x)$	Effective longitudinal cross-sectional area of the CP	$\mu\mathrm{m}^2$
$A_{\mathrm{RL}}(x)$	Effective longitudinal cross-sectional area of the reticular lamina	$\mu\mathrm{m}^2$
$b(x)$	Width of CP	$\mu\mathrm{m}$
$D[z(x,t)]$	Nonlinear distortion operator for the CP	Pa
$D_{\mathrm{BM}}[z(x,t)]$	Nonlinear distortion operator defined for the basilar membrane	Pa
$D_{\mathrm{RL}}[z(x,t)]$	Nonlinear distortion operator defined for the tectorial membrane	Pa
$\delta l(x,t)$	Change in distance between plates of CP	$\mu\mathrm{m}$
$\delta l_{\mathrm{PM}}(x,t)$	Change in distance between point masses	
Δx	Distance between centers of resonant plates of CP	$\mu\mathrm{m}$
$\epsilon(x)$	Coefficient of tissue connectivity	Pa $\mu\mathrm{m}$
G	Effective gain in CP displacement gradient	
G_{RL}	Gain in CP displacement gradient due to motion of reticular lamina	
G_{Δ}	Gain in CP displacement gradient due to spacing between CP plates	
$k(x)$	Effective stiffness of CP per unit transverse area of CP	$\mathrm{ng}/(\mathrm{ms}^2\,\mu\mathrm{m}^2)$
k_m	Estimated stiffness of the middle ear	$\mathrm{ng}/\mathrm{ms}^2$
$l_0(x)$	Distance between edges of CP plates when CP at rest	$\mu\mathrm{m}$
$\lambda(x)$	Wavelength of traveling cochlear wave	$\mu\mathrm{m}$
$m(x)$	Effective mass of CP per unit transverse area of CP	$\mathrm{ng}/\mu\mathrm{m}^2$
m_m	Mass of the malleus and incus combined	ng
m_s	Mass of the stapes	ng
$r(x)$	Effective viscoelastic damping of CP per unit transverse area of CP	$\mathrm{ng}/(\mathrm{ms}\,\mu\mathrm{m}^2)$
r_m	Middle ear resistance	ng/ms
ρ	Cochlear fluid density	$\mathrm{ng}/\mu\mathrm{m}^3$
$P(x,t)$	Fluid pressure difference within cochlea across the CP	Pa
$P_e(t)$	Air pressure at the tympanic membrane	Pa
$S(x)$	Longitudinal cross-sectional area of the scalae	$\mu\mathrm{m}^2$
S_m	Cross-sectional area of the ear tympanic membrane	$\mu\mathrm{m}^2$
t	Time	ms
$T(x,t)$	Tension developed within longitudinal tissue	pN
T_m	Ossicular chain lever ratio	
$\tau_{\mathrm{CP}}(x)$	Longitudinal cross-sectional area of the CP	$\mu\mathrm{m}^2$
x	Position along CP as measured from cochlear base	$\mu\mathrm{m}$
Y_{BM}	Measured longitudinal Young's modulus of the basilar membrane	Pa
$Y_{\mathrm{CP}}(x)$	Estimated longitudinal Young's modulus of the CP	Pa
Y_{RL}	Estimated longitudinal Young's modulus of the reticular lamina	Pa
Y_{TM}	Measured longitudinal Young's modulus of the tectorial membrane	Pa
$z(x,t)$	Effective CP displacement	$\mu\mathrm{m}$
$Z_{\max}$	Maximum CP displacement at 70 dB (SPL)	$\mu\mathrm{m}$

APPENDIX D: PARAMETERS OF THE MODEL OF THE EAR

Symbol	Value	Source
$m(x)$	$7.7\times10^{-1}e^{5.0\times10^{-5}x}$ ng/μm^2	Viergever, 1980
$r(x)$	$6.3\times10^{1}e^{-2.1\times10^{-4}x}$ ng/(μm^2 ms)	Viergever, 1980
$k(x)$	$2.1\times10^{4}e^{-3.7\times10^{-4}x}$ ng/(μm^2 ms^2)	Viergever, 1980
$b(x)$	$8\times10^{1}e^{-5.0\times10^{-5}x}$ μm	Viergever, 1980
$S(x)$	$2.5\times10^{6}e^{-5.0\times10^{-5}x}$ μm^2	Dallos, 1973
$\tau_{\mathrm{RL}}(x)$	1 μm	Santi, 1988
$\tau_{\mathrm{BM}}(x)$	10 μm	Echteler, 1994
$Y_{rm}(x)$	60 MPa	Estimated
$Y_{\mathrm{BM}}(x)$	200 kPa	Steele *et al.*, 2000
$\lambda(x)$	$1\times10^{2}e^{-6.6\times10^{-5}x}$ μm	Estimated from Viergever, 1980 and Ren, 2001
G_Δ	10	Estimated from Santi, 1988
G_r	5.4	Mammano and Ashmore, 1993
T_m	1.3	Dallos, 1973
S_m	5.5×10^{7} μm^2	von Békésy, 1960
S_{OW}	3.2×10^{6} μm^2	Viergever, 1980
m_m	4×10^{7} ng	Geisler, 1998
m_s	2×10^{6} ng	von Békésy, 1960
r_m	4.7×10^{8} ng/ms	Estimated
k_m	4.6×10^{9} ng/ms	Estimated

Allen, J. B. (**1980**). ''A cochlear micromechanical model of transduction,'' in *Psychophysical, Physiological, and Behavioural Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen (Delft University Press, Delft, 1980), pp. 85–95.

Allen, J. B., and Neely, S. T. (**1992**). ''Micromechanical models of the cochlea,'' Phys. Today **45**, 40–47.

Bohne, B. A., Gruner, M. M., and Harding, G. W. (**1990**). ''Morphological correlates of aging in the chinchilla cochlea,'' Hear. Res. **48**, 79–91.

Dallos, P. (**1973**). *The Auditory Periphery: Biophysics and Physiology* (Academic, New York).

Dallos, P. (**1992**). ''The Active Cochlea,'' J. Neurosci. **12**, 4575–4585.

de Boer, E. (**1980**). ''Auditory physics: Physical principles in hearing theory. I,'' Phys. Rep. **62**, 88–174.

Diependaal, R. J. (**1988**). *Nonlinear and Active Cochlear Models*, Doctoral dissertation, Delft University of Technology (Delft University Press, Delft).

Diependaal, R. J., Duifhuis, H., Hoogstraten, H. W., and Viergever, M. A. (**1987**). ''Numerical methods for solving one-dimensional cochlear models in the time domain,'' J. Acoust. Soc. Am. **82**, 1655–1666.

Echteler, S. M., Fay, R. R., and Popper, A. N. (**1994**). ''Structure of the mammalian Cochlea,'' in *Comparative Hearing: Mammals*, edited by R. R. Fay and A. N. Popper (Springer, New York).

Feather, N. (**1961**). *Vibrations and Waves* (Edinburgh University Press, London).

Fletcher, N. H. (**1992**). *Acoustic Systems in Biology* (Oxford University Press, New York), 313 pp.

Fritze, W., Steurer, M., and Fritze, P. (**1996**). ''Some ideas on the origin of otoacoustic emissions,'' Biochem. Biophys. Res. Commun. **227**, 494–498.

Fung, Y. C. (**1981**). *Biomechanics: Mechanical Properties of Living Tissues* (Springer, New York).

Geisler, C. D. (**1998**). *From Sound to Synapse* (Oxford University Press, New York).

Geisler, C. D., and Sang, C. (**1995**). ''A cochlear model using feed-forward outer-hair-cell forces,'' Hear. Res. **86**, 132–146.

Hall, J. L. (**1974**). ''Two-tone distortion products in a nonlinear model of the basilar membrane,'' J. Acoust. Soc. Am. **56**, 1818–1828.

Jau, Y. C., and Geisler, C. D. (**1983**). ''Results from a cochlear model utilizing longitudinal coupling,'' in *Mechanisms of Hearing*, edited by W. R. Webster and L. M. Aiken (Monash University Press, Melbourne), pp. 169–176.

Johnstone, B. M., Patuzzi, R., and Yates, G. K. (**1986**). ''Basilar membrane measurements and the travelling wave,'' Hear. Res. **22**, 147–153.

Kemp, D. T. (**1978**). ''Stimulated acoustic emissions from within the human auditory system,'' J. Acoust. Soc. Am. **64**, 1386–1391.

Kemp, D. T. (**1986**). ''Otoacoustic emissions, travelling waves, and cochlear mechanisms,'' Hear. Res. **22**, 95–104.

Kemp, D. T., and Brown, A. M. (**1984**). ''Ear canal acoustic and round window electrical correlates of $2f_1-f_2$ distortion generated in the cochlea,'' Hear. Res. **13**, 39–46.

Kim, D. O. (**1986**). ''An overview of nonlinear and active cochlear models,'' in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer, New York), pp. 239–249.

Kletsky, E. J., and Zwislocki, J. J. (**1979**). ''Cochlear-microphonic versus hair-cell tuning in the cochlea,'' J. Acoust. Soc. Am. Suppl. 1 **65**, S84.

Lim, D. J. (**1986**). ''Functional structure of the organ of Corti: A review,'' Hear. Res. **22**, 117–146.

Long, G. R., Tubis, A., and Jones, K. L. (**1991**). ''Modeling synchronization and suppression of spontaneous otoacoustic emissions using Van der Pol oscillators: Effects of aspirin administration,'' J. Acoust. Soc. Am. **89**, 1201–1212.

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,'' Hear. Res. **28**, 173–189.

Mammano, F., and Ashmore, J. F. (**1993**). ''Reverse transduction measured in the isolated cochlea by laser Michelson interferometry,'' Nature (London) **365**, 838–841.

Mammano, F., and Nobili, R. (**1993**). ''Biophysics of the Cochlea: Linear approximation.'' J. Acoust. Soc. Am. **93**, 3320–3331.

Marion, J. B., and Thornton, S. T. (**1995**). *Classical Dynamics of Particles and Systems* (Harcourt Brace and Company, San Diego).

Matthews, J. W. (**1980**). ''Mechanical modelling of nonlinear phenomena observed in the peripheral auditory system,'' Doctoral dissertation, Washington University, St. Louis, MO.

Nadol, Jr., J. B. (**1979**). ''Electron microscopic findings in presbycusic degeneration of the basal turn of the human cochlea,'' Otolaryngol.-Head Neck Surg. **87**, 818–836.

Naidu, R. C., and Mountain, D. C. (**1998**). ''Measurements of the stiffness map challenge a basic tenet of cochlear theories,'' Hear. Res. **124**, 124–131.

Naidu, R. C., and Mountain, D. C. (**2001**). ''Longitudinal coupling in the basilar membrane,'' submitted to J. Assoc. Res. Otolaryngol.

Nash, W. A. (**1972**). *Theory and Problems of Strength of Materials*, 2nd ed. (McGraw-Hill, New York).

Neely, S. T., and Kim, D. O. (**1983**). ''An active cochlear model showing sharp tuning and high sensitivity,'' Hear. Res. **9**, 123–130.
Parham, K., Sun, X.-M., and Kim, D. O. (**1999**). ''Distortion product otoacoustic emissions in the CBA/J mouse model of presbycusis,'' Hear. Res. **134**, 29–38.
Patuzzi, R. (**1996**). ''Cochlear micromechanics and macromechanics,'' in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), Chap. 4.
Pinsky, P. M., and Datye, D. V. (**1991**). ''A microstructurally-based finite element model of the incised human cornea,'' J. Biomech. **24**, 907–909; **24**, 911–922.
Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). ''A review of otoacoustic emissions,'' J. Acoust. Soc. Am. **89**, 2027–2067.
Ren, T. (**2001**). ''Direct measurement of the traveling wave using a scanning laser interferometer in sensitive gerbil cochlea,'' Assoc. Res. Otolaryngol. **24**, 555.
Ren, T., and Nuttall, A. (**2000**). ''Measurement of the basilar membrane vibration at the basal turn in the sensitive gerbil cochlea,'' Assoc. Res. Otolaryngol. **23**, 717.
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.
Santi, P. A. (**1988**). ''Cochlear microanatomy and ultrastructure,'' in *Physiology of the Ear*, edited by A. F. Jahn and J. Santos-Sacchi (Raven Press, New York), pp. 173–199.
Schmiedt, R. A., and Adams, J. C. (**1981**). ''Stimulated otoacoustic emissions in the ear of the gerbil,'' Hear. Res. **5**, 295–305.
Schuknecht, H. F., and Gacek, M. R. (**1993**). ''Cochlear pathology in presbycusis,'' Ann. Otol. Rhinol. Laryngol. **102**, 1–16.
Slepecky, N. (**1996**). ''Structure of the mammalian cochlea,'' in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), pp. 44–129.
Steele, C. R., Baker, G., Tolomeo, J., and Zetes, D. (**2000**). ''Cochlear Mechanics,'' in *The Biomedical Engineering Handbook*, 2nd ed., edited by J. D. Bronzino (CRC Press, Boca Raton, FL).
Strauss, W. A. (**1992**). *Partial Differential Equations* (Wiley, New York).
Ulfendahl, M. (**1997**). ''Mechanical responses of the mammalian cochlea,'' Prog. Neurobiol. (Oxford) **53**, 331–380.
van Dijk, P., Wit, H. P., Talmadge, C. L., and Long, G. R. (**1994**). ''Correlation between amplitude and frequency fluctuations of spontaneous otoacoustic emissions,'' J. Acoust. Soc. Am. **96**, 153–169.
Viergever, M. A. (**1977**). ''A two-dimensional model for the cochlea. II. The heuristic approach and numerical results,'' J. Eng. Math **11**, 11–28.
Viergever, M. A. (**1978**). ''On the physical background of the point-impedance characterization of the basilar membrane in cochlear mechanics,'' Acustica **39**, 292–297.
Viergever, M. A. (**1980**). *Mechanics of the Inner Ear—A Mathematical Approach*, Doctoral dissertation, Delft University of Technology (Delft University Press, Delft).
Völdrich, L. (**1978**). ''Mechanical properties of the basilar membrane,'' Acta Oto-Laryngol. **86**, 331–335.
Völdrich, L., and Ulehlova, L. (**1987**). ''Cochlear micromechanics,'' Acta Oto-Laryngol. **103**, 661–664.
von Békésy, G. (**1948**). ''On the elasticity of the cochlear partition,'' J. Acoust. Soc. Am. **20**, 227.
von Békésy, G. (**1960**). *Experiments in Hearing* (McGraw-Hill, New York).
Vorp, D. A., Raghavan, M. L., and Webster, M. W. (**1998**). ''Mechanical wall stress in abdominal aortic aneurysm: Influence of diameter and asymmetry,'' J. Vasc. Surg. **27**, 632–639.
Wallace, P. R. (**1984**). *Mathematical Analysis of Physical Problems* (Dover, Toronto).
Weir, C. C., Pasanen, E. G., and McFadden, D. (**1988**). ''Partial dissociation of spontaneous otoacoustic emissions and distortion products during aspirin use in humans,'' J. Acoust. Soc. Am. **84**, 230–237.
Wickesberg, R. E., and Geisler, C. D. (**1986**). ''Longitudinal stiffness coupling in a 1-dimensional model of the peripheral ear,'' in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer, Berlin), pp. 113–120.
Willems, P. J. (**2000**). ''Mechanism of disease: Genetic causes of hearing loss,'' New England J. Med. **342**, 1107–1110.
Wilson, J. P., and Sutton, G. J. (**1981**). ''Acoustic correlates of tonal tinnitus,'' in *Tinnitus*, edited by D. Evered and L. Lawrenson (Pitman, London).
Yates, G. K. (**1995**). ''Cochlear structure and function,'' in *Hearing*, 2nd ed., edited by B. C. J. Moore (Academic, San Diego).
Zwislocki, J. J., and Kletsky, E. J. (**1980**). ''Micromechanics in the theory of cochlear mechanics,'' Hear. Res. **2**, 505–512.

Inner hair cell loss and steady-state potentials from the inferior colliculus and auditory cortex of the chinchilla[a)]

Sally Arnold[b)]
Department of Speech-Language Pathology, Buffalo State College, 1300 Elmwood Avenue, Buffalo, New York 14222-1095

Robert Burkard
Center for Hearing and Deafness, State University of New York at Buffalo, Buffalo, New York 14214-3007

(Received 30 October 2001; accepted for publication 24 May 2002)

Steady-state evoked potentials were measured from unanesthetized chinchillas both before and after carboplatin-induced selective inner hair cell loss. Recordings were made from both the inferior colliculus (IC) and the auditory cortex (AC). The steady-state potential was measured in the form of the envelope following response (EFR), obtained by presenting a two-tone stimulus ($f1 = 2000$ Hz; $f2 = 2020$, 2040, 2080, 2160, or 2320 Hz), and measuring the magnitude of the Fourier coefficient at the $f2-f1$ difference frequency. From the IC, precarboplatin, EFR amplitude vs difference tone frequency showed a bandpass pattern, with maximum amplitude at either 160 or 80 Hz, depending upon stimulus level. Postcarboplatin, the preferred difference frequency was 80 Hz for all stimulus levels. From the AC, EFR amplitude versus difference tone frequency also showed a bandpass pattern, with the maximum amplitude at 80 Hz both pre- and postcarboplatin. EFR amplitude from the IC was decreased for some conditions postcarboplatin, while the amplitude from the AC showed no significant change.
[DOI: 10.1121/1.1494991]

PACS numbers: 43.64.Ri, 43.64.Qh [LHC]

I. INTRODUCTION

While the role of outer hair cells (OHCs) in auditory functioning has been extensively studied, the role of inner hair cells (IHCs) is less known, because selective IHC lesions have been hard to achieve. With the recent discovery that the chemotherapeutic agent carboplatin selectively destroys IHCs in the chinchilla (Hofstetter *et al.*, 1997; Takeno *et al.*, 1994a, 1994b), it has become possible to study the effects of IHC loss in this species.

Several studies in carboplatin-treated chinchillas have shown that the auditory-nerve compound action potential (CAP) decreases proportionally to the loss of IHCs (Qiu *et al.*, 2000; Takeno *et al.*, 1994a; Wang *et al.*, 1997). An interesting question is, how does this loss of afferent input affect processing at higher levels of the auditory system? In the inferior colliculus (IC), the onset potential elicited by transient tone-burst or click stimuli is generally reduced postcarboplatin (Burkard *et al.*, 1997; Qiu *et al.*, 2000), although results are somewhat variable between animals (McFadden *et al.*, 1998; Qiu *et al.*, 2000). Surprisingly, the cortical response to tone-burst stimuli shows no change with IHC loss, and even an increase in amplitude in some cases (Qiu *et al.*, 2000).

While the response to transient stimuli has been studied in carboplatin-treated animals, the response to continuous or *steady-state* stimuli has not. Since both types of evoked responses are used clinically, it would potentially be useful to measure the steady-state response in carboplatin-treated chinchillas to expand our knowledge of the effects of IHC loss. In this study, steady-state responses were measured from the IC and the auditory cortex (AC) of the chinchilla, pre-and postcarboplatin. Steady-state responses were measured in the form of the *envelope following response* (EFR), which was elicited by the monaural presentation of a two-tone stimulus ($f1$ and $f2$) (Dolphin *et al.*, 1994). Although the $f2-f1$ difference tone (DT) of the two signals is not present in the acoustic signal, it is observed in the evoked response, and presumably represents a neural response to the periodicity of the stimulus envelope. In this study, the EFR was measured across a range of DT frequencies, because the function has not been fully defined in the chinchilla.

A steady-state evoked potential sometimes used clinically for threshold estimation is the amplitude modulation following response (AMFR) (Aoyagi *et al.*, 1994b; Galambos *et al.*, 1981; Griffiths and Chambers, 1991; Lins *et al.*, 1995; Picton *et al.*, 1998), which is comparable to the EFR in that it measures the phase-locked response to the stimulus envelope. An understanding of how the loss of IHCs and afferent input to the central auditory nervous system affects processing of steady-state signals could have clinical implications for the diagnosis of IHC lesions.

II. METHODS

A. Subjects

Six young-adult chinchillas were used as subjects. Following induction of anesthesia using ketamine/xylazine,

[a)]Portions of this work were presented in ''Effect of IHC loss on the steady-state and onset evoked potentials recorded from the inferior colliculus and auditory cortex of the chinchilla,'' at Association for Research in Otolaryngology, St. Petersburg, FL, February, 2001.

[b)]Electronic mail: arnoldsa@bscmail.buffalostate.edu

tungsten electrodes were chronically implanted in the right inferior colliculus and right auditory cortex, with a ground electrode placed anteriorly. Following surgery, the animals were allowed to recover for at least 1 week prior to data collection.

All data collection was done both prior to and 4–5 weeks following carboplatin injection (75 mg/kg i.p.). For the evoked response recording, unanesthetized chinchillas were placed in an animal restraint device (Snyder and Salvi, 1994) inside a sound-attenuating chamber. Evoked responses were recorded during a single session, lasting approximately 1 hour.

B. Stimuli and electrophysiological recording

Electrophysiological recording was accomplished using a Tucker Davis Technologies system. Stimuli were delivered via two Etymotic ER2 earphones, held in place with a baby foam insert. The left ear, contralateral to the recording electrodes, was stimulated in all cases. Recordings were made separately from the IC and the AC electrodes to eliminate any possibility of electrode cross talk. The order of recording site was randomized across animals.

Prior to data collection, all stimuli were calibrated by inserting the earphones into a 0.2-cc volume, along with an ER-7C probe microphone. The voltage out of the microphone was then measured and converted to pSPL.

Steady-state responses were measured in the form of the envelope following response (EFR). To obtain the EFR, two tones (234.8-ms duration, 2-ms rise/fall time, fixed phase) were presented to two Etymotic ER2 earphones, mixed acoustically with a foam insert in the left ear canal. The EFR was measured using $f1 = 2000$ Hz and $f2 = 2020$, 2040, 2080, 2160, and 2320 Hz. The order of stimulus pairs was randomized across animals. Stimuli were presented at a rate of just less than 3 Hz (ISI = 334.8 ms), from 80 to −10 dB pSPL, in 10-dB steps. Responses were recorded over a 249.8-ms time window (Tucker Davis Technologies DB4 bioamplifier, 10 000 gain, 10–8000-Hz filter), then averaged in the time domain for 50 sweeps.

To measure the EFR, the Fourier component at the $f2-f1$ difference tone (DT) frequency was obtained by cropping (25–229.8 ms) and windowing (10-ms cosine squared at onset and offset) in the time domain, and computing an FFT using 2048 frequency domain points. The amplitude of the Fourier component at the DT frequency was obtained by manual cursoring of the response spectrum by the first author. Data were obtained at all stimulus levels in each animal, even if the amplitude at the DT frequency was below the noise floor.

C. Cochleograms

Following the second data collection session, the animals were sacrificed, and the cochleas were dissected out and prepared using standard surface preparation techniques as described in Phillips *et al.* (2001), and stained with succinate dehydrogenase, which stains mitochondria. Cochleograms were prepared by counting the number of missing IHCs and OHCs, which were then expressed as percent loss as a function of distance from the apex and as a function of frequency (Greenwood, 1990).

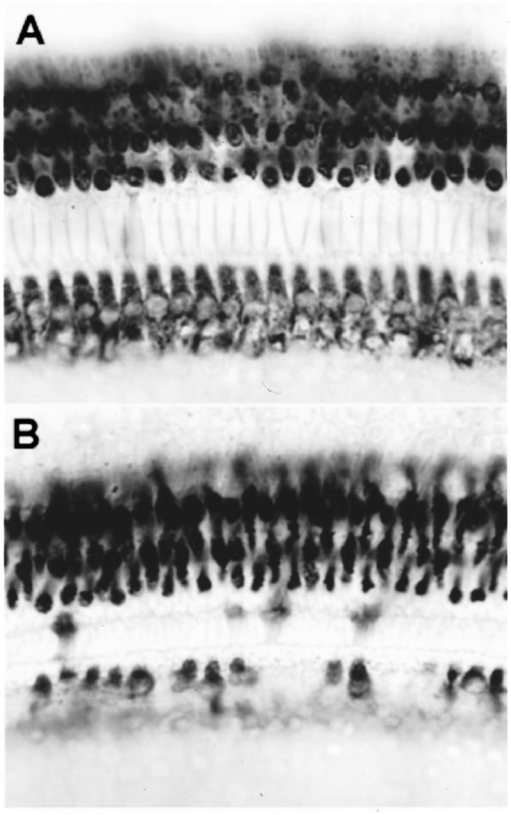


FIG. 1. Photomicrographs of the Organ of Corti, taken from the middle turn of (a) control animal #6660 (not used in the electrophysiologic investigations of the present study) and (b) from carboplatin-treated chinchilla #6272 (original magnification: 400x).

III. RESULTS

A. Hair cell loss

A photomicrograph from a chinchilla not treated with carboplatin is shown in Fig. 1(A). There is one continuous row of stained IHCs along the bottom of the photomicrograph (the entire cell body is seen in this view), and three rows of darkly stained OHCs are seen along the top of the photomicrograph. Postcarboplatin, all animals showed substantial IHC loss across the cochlea, with negligible loss of OHCs. This pattern of damage can be seen in Fig. 1(B), which is a photomicrograph of a section of the organ of Corti from chinchilla #6272. This photomicrograph shows that the three rows of OHCs are well preserved (i.e., the three rows of darkly stained OHCs can be seen to extend the length of the section). A patchy IHC loss can be seen, i.e., a few stained IHCs are observed in patches, with only background staining seen between the stained IHCs.

Figure 2 shows the left cochleograms (the stimulated ear, contralateral to the recording sites) for the six chinchillas. Note that the degree of IHC loss was variable across animals. Chinchilla #6267 sustained the least amount of IHC loss, approximately 15%–55%, which was fairly evenly distributed across the cochlea. Chinchillas #6266, #6271, and

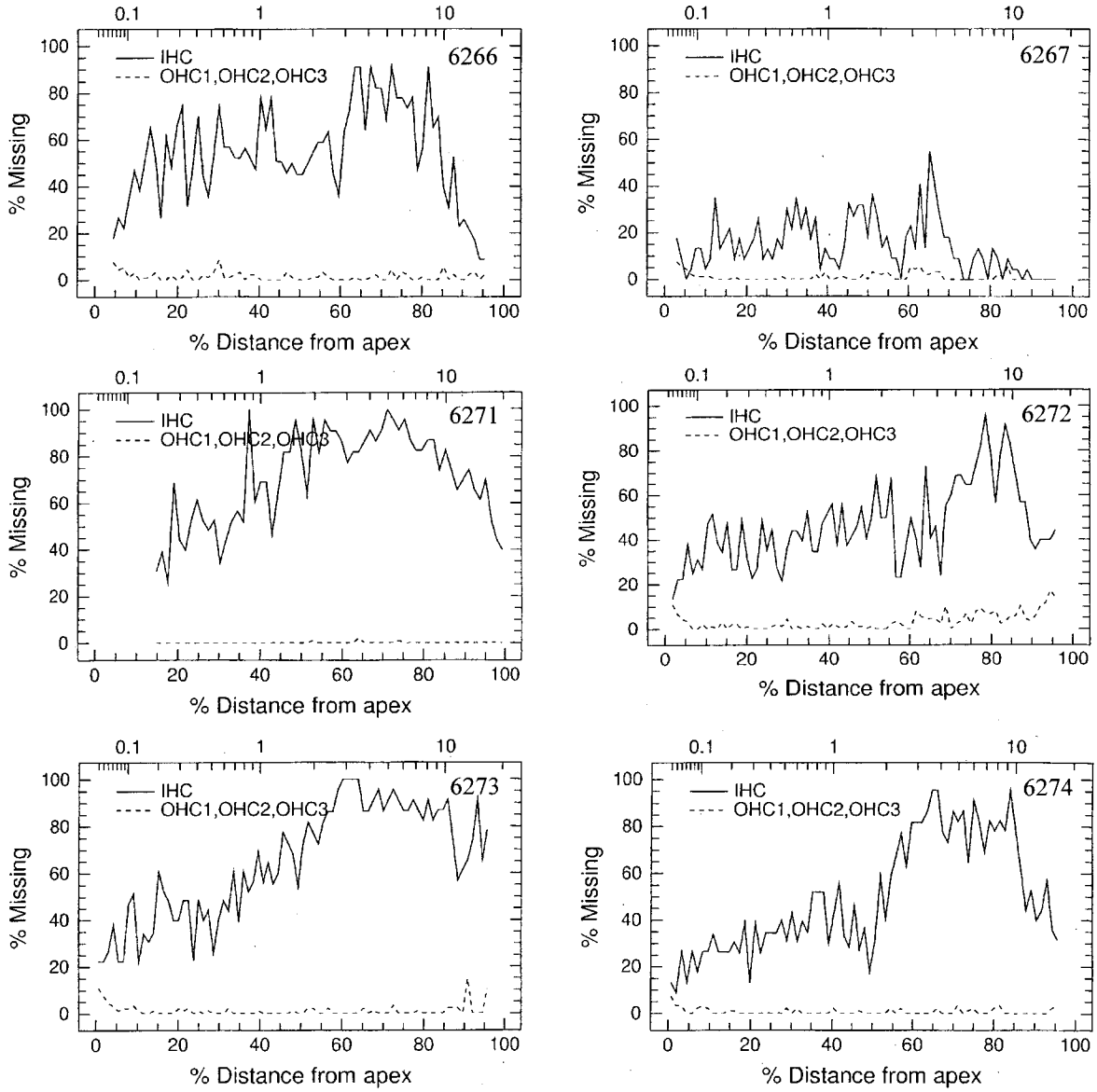


FIG. 2. Left cochleograms from each of the six chinchillas. Percent missing is shown separately for the IHCs (solid lines) and OHCs (dashed lines).

#6273 all showed the greatest amount of IHC loss, which was moderate in the apical region, increasing to approximately 90% or greater in the basal region. The average cochleogram for the six chinchillas (see Fig. 3) shows an inner hair cell loss of approximately 25%–40% in the apical region, increasing to 60%–70% in the basal region, with minimal or no loss of outer hair cells. The average IHC loss was roughly 50% in the frequency region of the stimulus pairs used for EFR testing (~2 kHz).

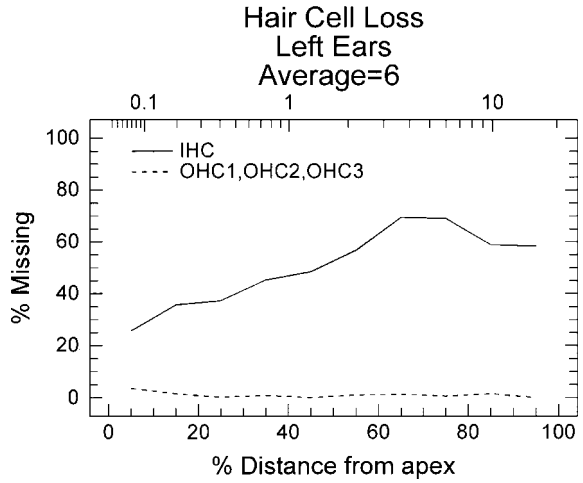


FIG. 3. Average left cochleogram for the six animals.

B. Envelope following responses

1. Raw data

Figure 4 shows an example of raw EFR data, measured from the IC for a two-tone stimulus ($f1 = 2000$ Hz, $f2 = 2160$ Hz) at 60 dB pSPL. Both time- and frequency-

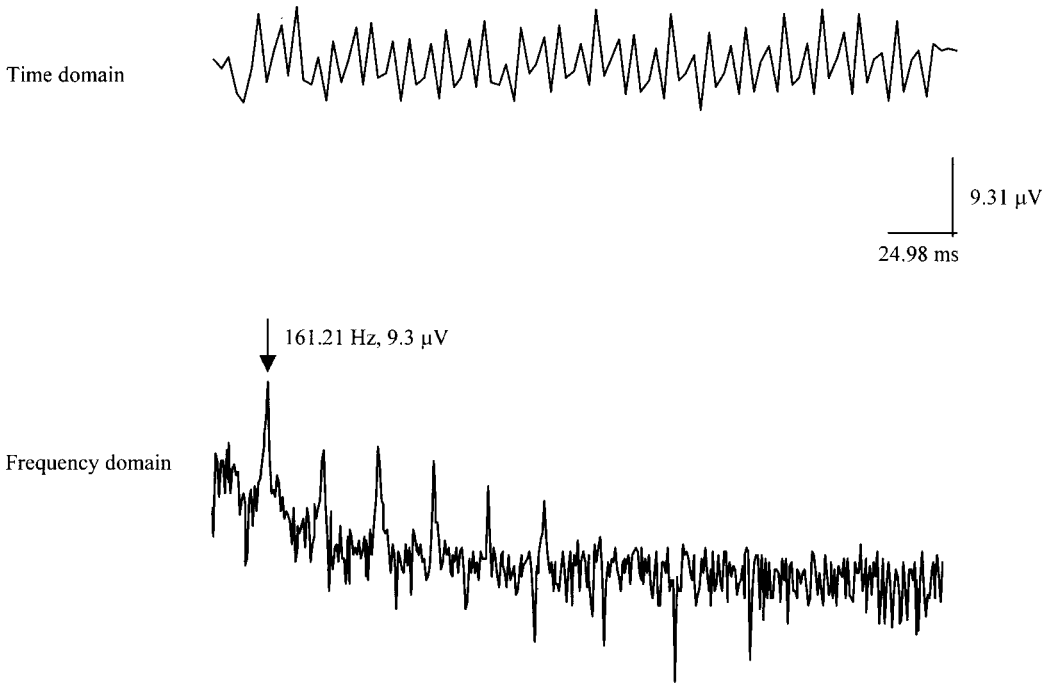


FIG. 4. Example of EFR from chinchilla #6274 to two-tone stimulus (2000 Hz, 2160 Hz) presented at 60 dB pSPL. The upper plot shows the time domain response, the lower plot shows the frequency domain response.

domain representations are shown. The EFR is represented by the peak in the spectrum at ~160 Hz (arrow). In this particular example, peaks are also visible at harmonics of the EFR.

2. EFR amplitude from the IC

The amplitude of the EFR measured from the IC, averaged across the six animals, is shown for each DT frequency in Fig. 5. Precarboplatin data are shown in Fig. 5(a), while the postcarboplatin data are shown in Fig. 5(b).

Precarboplatin, the mean amplitude vs DT curve shows a bandpass function, with the maximum amplitude at either 160 Hz (for 80, 70, and 30 dB pSPL), or at 80 Hz (for 60, 50, and 40 dB pSPL). Amplitude falls off rather sharply below 80 Hz and above 160 Hz. Following carboplatin, maximum EFR amplitude was seen at DT=80 Hz for all stimulus levels.

In order to more clearly visualize the effects of carboplatin upon EFR amplitude, the mean pre- and postcarboplatin I/O functions are shown for each DT separately in Figs. 6(a)–(e). Before carboplatin, the mean amplitude of the EFR from the IC generally increases monotonically with increasing stimulus level for all DTs, except for 80 Hz, which plateaus at high stimulus levels. For the 80-dB pSPL stimulus level, the largest EFR amplitude occurs at 160 Hz (14.6 μV), followed by 80 Hz (12.6 μV), 40 Hz (10.4 μV), 20 Hz (4.9 μV), and 320 Hz (3.5 μV). Mean EFR amplitude from the IC decreases postcarboplatin for moderate to high stimulus levels, except at DT=80 Hz, 80 dB pSPL, due to the rollover in the prestimulus I/O function.

To test for the significance of the difference between pre- and postcarboplatin amplitudes, a two-way repeated measures ANOVA was computed for each DT frequency. Amplitude values over the range of 40–80 dB were included in the analyses. The factors analyzed were effect of carboplatin and effect of stimulus level. As expected, the main effect of stimulus level was significant for all DT frequencies ($p<0.05$), and is not of interest here. At both DT=160 Hz and DT=320 Hz, the main effect of carboplatin upon amplitude was significant [160 Hz: $F(1,5)=23.296$, $p=0.005$], [320 Hz: $F(1,5)=21.105$, $p=0.006$]. At DT=80 Hz, the main effect of carboplatin was not significant [$F(1,5)=1.752$, $p=0.243$], but there was a significant carboplatin-by-level interaction [$F(4,20)=9.030$, $p<0.001$]. Pairwise multiple comparisons using the Tukey test indicated that the effect of carboplatin was significant for the 60-dB pSPL stimulus level only ($p<0.05$). The main effect of carboplatin was not significant for DT=20 Hz [$F(1,5)=5.240$, $p=0.071$] or DT=40 Hz [$F(1,5)=3.245$, $p=0.132$].

3. EFR amplitude from the AC

The mean EFR amplitude vs DT frequency functions from the cortex are shown in Fig. 7 for the precarboplatin 7(a) and postcarboplatin 7(b) conditions. From the AC, EFR amplitude versus difference tone frequency also shows a bandpass pattern, with the maximum amplitude at 80 Hz both pre- and postcarboplatin. Note that response amplitude decreases sharply above 80 Hz. At 320 Hz, amplitudes are very small and are essentially in the noise floor.

Figures 8(a)–(e) show the mean EFR input–output functions from the AC pre- and postcarboplatin, for each DT frequency separately. The mean amplitude of the EFR from the AC is generally smaller than from the IC. For the 80-dB pSPL stimulus level, precarboplatin, the largest amplitude occurs at 80 Hz (8.8 μV), followed by 40 Hz (7.3 μV), 20 Hz (5.9 μV), 160 Hz (2.6 μV), and 320 Hz (0.2 μV).

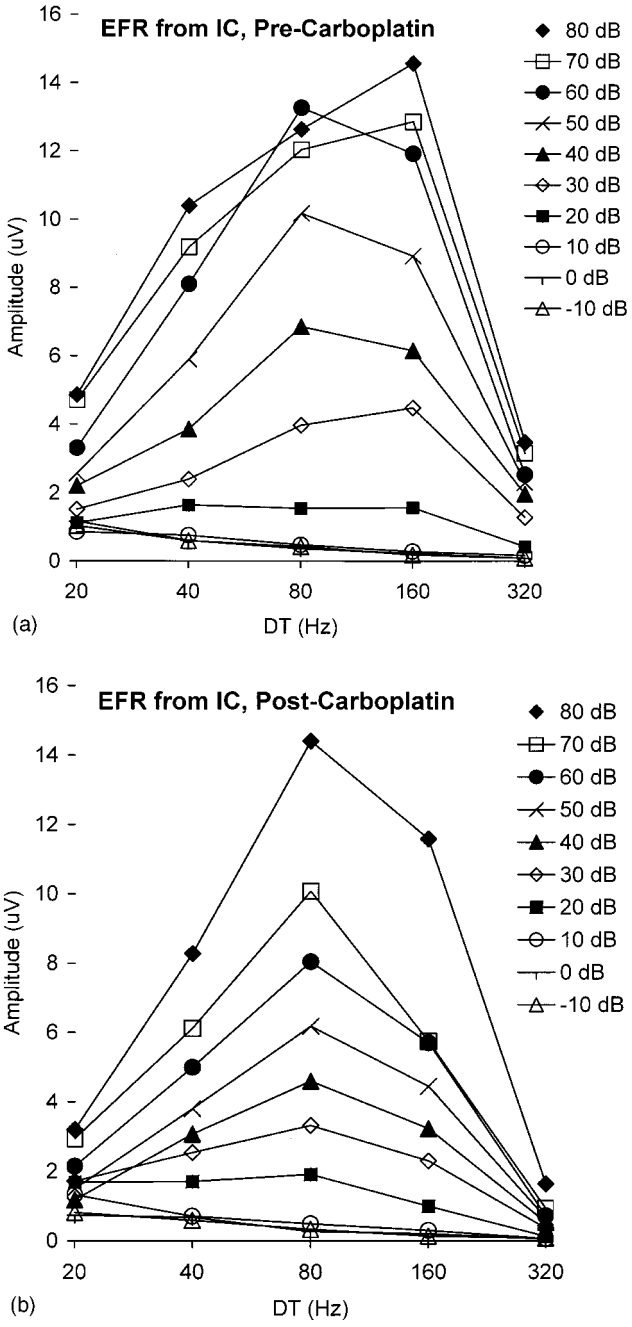


FIG. 5. (a-b). Mean EFR amplitude from the IC, plotted as a function of difference frequency (DT). Pre-carboplatin data is shown in Fig. 5a; post-carboplatin data in Fig. 5b.

Due to small response amplitudes, it is difficult to visualize any trends in EFR amplitude postcarboplatin at 160 and 320 Hz. At 20 Hz, the pre- and postamplitude curves appear to be overlapping. For 40 and 80 Hz, response amplitudes were actually slightly larger postcarboplatin, except at the 80-dB pSPL signal level. As for the IC, a two-way repeated measures ANOVA was used to statistically analyze the effect of carboplatin upon EFR amplitude from the cortex. For all DT frequencies, the main effect of carboplatin was not statistically significant [20 Hz: $F(1,5)=0.0133$, $p=0.913$; 40 Hz: $F(1,5)=0.0113$, $p=0.920$; 80 Hz: $F(1,5)=0.646$, $p=0.458$; 160 Hz: $F(1,5)=0.854$, $p=0.398$; 320 Hz: $F(1,5)=1.382$, $p=0.293$].

IV. DISCUSSION

A. Preferred DT frequency for the EFR from the IC

In the present study, for the IC, the amplitude of the EFR was maximum at DT=160 or 80 Hz, depending upon stimulus level, and decreased at lower and higher DT frequencies. In a previous study (Arnold and Burkard, 1998), we measured the EFR from the IC, also in unanesthetized chinchillas, using an $f1$ frequency of 2000 Hz, and DT frequencies of 100, 200, 300, 400, 500, 600, 800, and 1000 Hz. It was found that the response was largest at 100 Hz, and decreased systematically at higher DT frequencies. The present results define the function with more resolution, showing the bandpass nature of the response. The variability of peak amplitude location suggests that the actual maximum may be somewhere between 80 and 160 Hz. Alternatively, it could be due to between- or within-subject variability or to a stimulus level effect. Further data collection, using more precise frequency resolution, and including measures of test–retest reliability, might clarify this issue.

While we measured the EFR using two-tone stimulus pairs, phase locking to the stimulus envelope can also be measured using sinusoidally amplitude-modulated (SAM) stimuli. The waveform envelopes of two-tone stimuli and SAM stimuli are similar, showing periodic amplitude fluctuations at the difference frequency (for two-tone stimuli) or the modulation frequency (for SAM stimuli). Dolphin and colleagues (Dolphin *et al.*, 1994) showed that the envelope following response to two-tone stimuli and SAM stimuli show virtually identical characteristics when the frequency difference between the two tones is equal to the modulation frequency of the SAM stimuli. They speculated that both measurements are manifestations of the same phase-locking phenomenon.

Responses of single units in the IC to SAM stimuli have been measured in a variety of animal species. The preferred modulation frequency for phase locking to the stimulus envelope has been found to be between 20–40 Hz in the rat (Rees and Moller, 1983), between 30–100 Hz in the cat (Langner and Schreiner, 1988), between 50–100 Hz in the guinea pig (Rees and Palmer, 1989), and 87 Hz in the rabbit (Batra *et al.*, 1989). Our present data from the chinchilla IC, showing a preferred rate that is somewhat higher, are perhaps related to species differences. Alternatively, it may be that the near-field EFR is a reflection of maximum synchronization of synaptic potentials, rather than based upon the discharge rate of action potentials, which is measured in the single-unit studies.

B. Preferred DT frequency for the EFR from the AC

EFR amplitude across DT frequency for the AC also showed a bandpass function, with maximum amplitude occurring at a rate of 80 Hz. This preferred rate for phase locking to the stimulus envelope in the cortex is higher than cortical single-unit studies using amplitude-modulated stimuli would suggest. For example, in the guinea pig, cortical neurons typically did not follow AM frequencies above 20 Hz (Creutzfeldt *et al.*, 1980), while in cat auditory cortex,

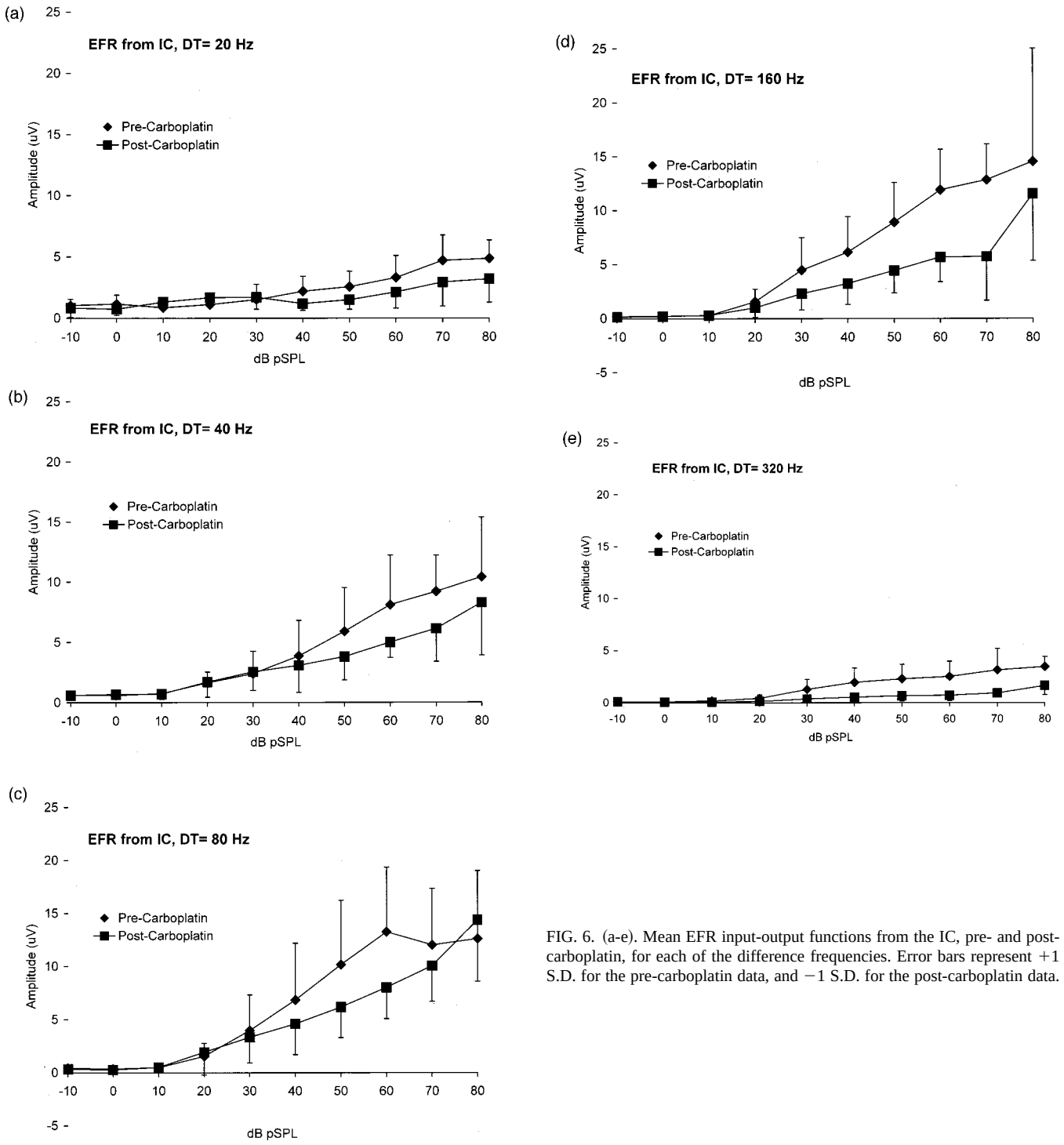
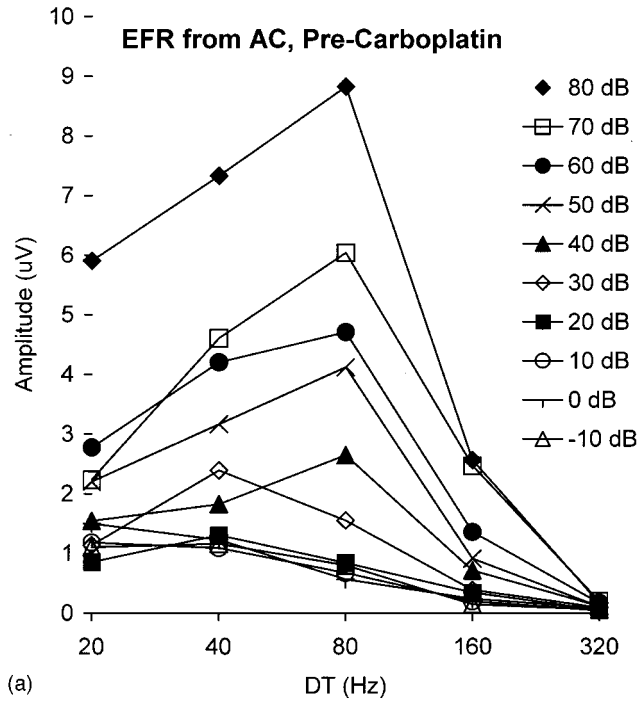


FIG. 6. (a-e). Mean EFR input-output functions from the IC, pre- and post-carboplatin, for each of the difference frequencies. Error bars represent +1 S.D. for the pre-carboplatin data, and −1 S.D. for the post-carboplatin data.

the best modulation frequency ranged from 5–31 Hz, depending upon cortical location (Schreiner and Urbas, 1988). In another study from cat auditory cortex (Eggermont, 1994), the best modulation frequency was 4 Hz for SAM tones, and 8–10 Hz for AM noise bursts and click trains.

One explanation for these differences in cortical responsiveness to the stimulus envelope may be the lack of anesthesia in the present study. Previous research has shown that cortical neurons can respond in a time-locked fashion at high rates in unanesthetized animals. In a recent study, Burkard and colleagues (Burkard *et al.*, 1999) recorded near field from the cortex in unanesthetized chinchillas. Time-locked activity in the form of cortical-evoked potentials for conventional averaging and maximum length sequences (MLS) were measured. The authors noted that they were able to obtain time-locked activity at rates as high as 500 Hz. In an earlier study (Ribaupierre *et al.*, 1972), single-unit responses from the cortex in unanesthetized cats were measured to click trains of various rates. The highest rate to which most units would time lock was between 50–100 Hz, but some units responded to rates as high as 200 Hz. In a similar study, only using *anesthetized* cats (Schreiner and Raggio, 1996), it



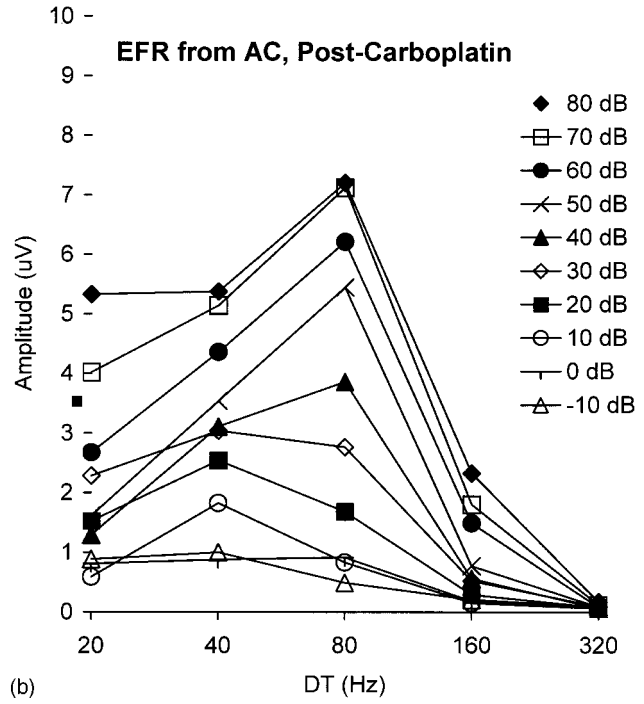


FIG. 7. (a-b). Mean EFR amplitude from the AC, plotted as a function of difference frequency. Pre-carboplatin data is shown in Fig. 7a; post-carboplatin data in Fig. 7b.

was found that cortical units would respond in a time-locked fashion only at rates up to about 10 Hz on the average. These studies support the idea that anesthesia strongly influences the time-locked activity of cortical neurons, and that the cortex can respond in a time-locked fashion at substantially higher rates without anesthesia.

In the previously mentioned research measuring cortical responsiveness to SAM stimuli, two of the studies showing low best-modulation frequencies used anesthetized animals (Eggermont, 1994; Schreiner and Urbas, 1988), lending support to this idea. However, in the third study (Creutzfeldt *et al.*, 1980), in which guinea pig cortical neurons did not respond above a modulation rate of 20 Hz, the animals were tested in an *unanesthetized* state, which argues against this idea.

As mentioned above, an alternative interpretation of these differences is that our near-field responses reflect maximum synchronization of synaptic events, in contrast to the single-unit studies, which reflect maximum synchronization of action potentials.

C. EFR amplitude and IHC loss

In the IC, maximum EFR amplitude shifted downward from a DT frequency of 160 to 80 Hz for high stimulus levels following IHC loss. If this is a real shift and not merely a reflection of test variability, then the loss of IHCs and/or afferent input to the central auditory nervous system appears to modestly degrade the ability to follow periodicity in the stimulus waveform. Alternatively, this change might suggest that in the IC, postcarboplatin responses are similar to precarboplatin responses at lower stimulus levels, where modulation rate transfer functions peak at 80 Hz.

To our knowledge, no data exist concerning the effect of IHC loss specifically upon phase locking of single-unit responses within the auditory system. However, Wang and colleagues, recording from single auditory-nerve fibers in carboplatin-treated chinchillas with IHC loss, did show a decrease in the driven discharge rate of fibers. For units with characteristic frequencies below 3000 Hz, the average maximum driven discharge rate for the carboplatin-treated group was reduced by 20% compared to an untreated control group (Wang *et al.*, 1997). Whether this decrease in firing rate for individual neurons would affect phase locking of an ensemble of neurons is unclear.

The preferred EFR rate for the cortex was 80 Hz both pre- and postcarboplatin, and therefore appears to be largely unaffected by IHC loss. Perhaps this is because the cortical EFR is already maximal at 80 Hz precarboplatin, and the modest reduction in maximum IC-EFR from 160 to 80 Hz is thus not reflected in the data from the cortex.

In general, EFR amplitude from the IC decreased postcarboplatin, and this decrease was statistically significant for DT = 160 and 320 Hz, as well as for 80 Hz at the 60-dB pSPL stimulus level. On the other hand, EFR amplitude from the cortex showed no significant change postcarboplatin, and at 40 and 80 Hz actually showed a slight, although not statistically significant, increase. This result is surprising, considering that there was substantial loss of IHCs in the frequency region of the stimulus and that IC amplitude *did* decrease.

In a previous study from our laboratory, Qiu *et al.* (2000) measured onset-type evoked potentials from the round window, IC, and auditory cortex in a group of chinchillas with carboplatin-induced IHC loss. The authors found that the auditory-nerve compound action potential (CAP) from the round window was reduced to a degree that was roughly proportional to the degree of IHC loss. The response

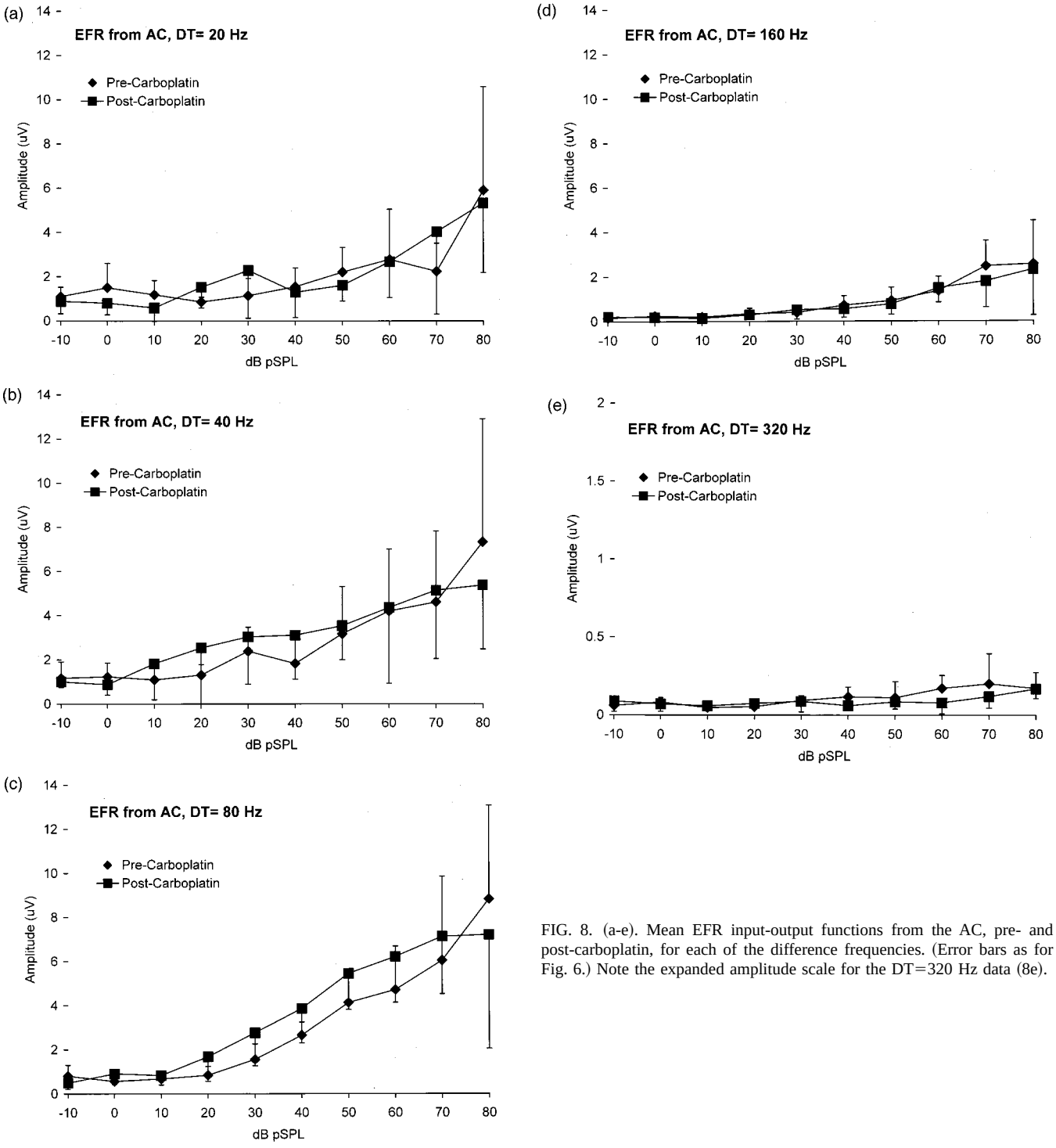


FIG. 8. (a-e). Mean EFR input-output functions from the AC, pre- and post-carboplatin, for each of the difference frequencies. (Error bars as for Fig. 6.) Note the expanded amplitude scale for the DT=320 Hz data (8e).

from the IC was generally reduced also, but to a lesser degree than the CAP. Interestingly, the response from the cortex in most cases was either unchanged or actually enhanced in the presence of IHC loss. These findings suggested a mechanism in the central auditory nervous system to compensate for the loss of peripheral input. In a follow-up study, Salvi *et al.* (2000) showed that enhancement of cortical responses in the presence of IHC loss may be due to a loss of GABA-A mediated inhibition. We did not measure the EFR from the round window because we were concerned about contamination with the cochlear microphonic (CM). As Qiu *et al.* (2000) acknowledged, with major IHC loss the cochlear potentials from the OHCs remain large, while the output of the auditory nerve is diminished in amplitude. Since there are clear CM responses to two-tone stimuli at the difference tone frequency, we would not have been able to obtain the EFR from the auditory nerve without CM contamination. Because of this limitation, we were not able to compare EFR responses at three levels of the auditory system (periphery, IC, cortex). However, the present pattern of results, showing that responses from the cortex were less affected than responses from the IC, suggests that the enhancement phenomenon may apply to the processing of steady-state signals as well as transient signals.

D. Clinical implications: Origins of the steady-state response

A steady-state response used clinically for threshold estimation is the amplitude-modulation following response (AMFR) (Aoyagi *et al.*, 1994b; Galambos *et al.*, 1981; Griffiths and Chambers, 1991; Lins *et al.*, 1995; Picton *et al.*, 1998). Using scalp electrodes and sinusoidally amplitude-modulated signals, the threshold of the response to the modulation frequency is measured for various carrier frequencies. The threshold thus obtained reflects sensitivity in the region of the carrier tone frequency, and holds an advantage over traditional auditory brainstem response (ABR) threshold testing in that the stimuli are longer in duration, and hence more frequency specific. The AMFR was originally described using a modulation frequency of 40 Hz (Galambos *et al.*, 1981). However, a modulation rate better suited for clinical use is 80 Hz, because the response is not affected by sleep (Aoyagi *et al.*, 1994a; Cohen *et al.*, 1991), and can be recorded more reliably in infants (Levi *et al.*, 1993). Since it has been shown that the responses to two-tone and SAM signals are similar (Dolphin *et al.*, 1994), our findings with the EFR have relevance to the clinical use of the AMFR.

One theory about the origin of the scalp-recorded AMFR suggests multiple sources in the auditory system, depending upon modulation frequency, i.e., the response to low-modulation frequencies is generated cortically, while the response to higher frequencies has a brainstem origin (Cohen *et al.*, 1991; Kuwada *et al.*, 1986; Levi *et al.*, 1993). On the other hand, a recent lesion study (Kiren *et al.*, 1994) found no change in the AMFR for modulation frequencies from 20–200 Hz with cortical ablation, appearing to rule out the cortex as a site of origin. Also in the Kiren *et al.* study, the AMFR was attenuated with IC lesions, suggesting that the IC is a major contributor to the AMFR. Our EFR data, showing that the response from the IC was maximal at 80–160 Hz, while the cortical response was maximal at 80 Hz and dropped off sharply at higher frequencies, tend to support the idea of multiple sources, with a trend towards a decrease in best modulation frequency as one ascends the auditory system. However, at 80 Hz, the modulation frequency commonly used clinically, our data support a dual generator, IC and cortex, since we measured robust responses from both these locations at a rate of 80 Hz. However, it remains possible that our cortical data reflect thalamic contributions that could also have been present in the Kiren *et al.* data.

It is important to note that the present study was conducted in awake, unanesthetized animals, and the effect of sleep and/or anesthesia on the EFR amplitude/DT curves is unknown at the present time. Since in a clinical situation the AMFR is often measured in sleeping subjects, future research should focus on repeating the present experimental paradigm in anesthetized animals.

ACKNOWLEDGMENTS

Work supported by NIH NIDCD DC03600. Yuqing Guo is thanked for surgically implanting the electrodes, and for helping with one of the figures. Haiyan Jiang is thanked for generating the cochleograms, and Dalian Ding is thanked for providing the photomicrographs.

Aoyagi, M., Furuse, H., Yokota, M., Kiren, T., Suzuki, Y., and Koike, Y. (**1994a**). ''Detectability of amplitude-modulation following response at different carrier frequencies,'' Acta Oto-Laryngol., Suppl. **511**, 23–27.

Aoyagi, M., Kiren, T., Furuse, H., Fuse, T., Suzuki, Y., Yokota, M., and Koike, Y. (**1994b**). ''Pure-tone threshold prediction by 80 Hz amplitude-modulation following response,'' Acta Oto-Laryngol., Suppl. **511**, 7–14.

Arnold, S., and Burkard, R. (**1998**). ''The auditory evoked potential difference tone and cubic difference tone measured from the inferior colliculus of the chinchilla,'' J. Acoust. Soc. Am. **104**, 1565–1573.

Batra, R., Kuwada, S., and Stanford, T. R. (**1989**). ''Temporal coding of envelopes and their interaural delays in the inferior colliculus of the unanesthetized rabbit,'' J. Neurophysiol. **61**, 257–268.

Burkard, R., Trautwein, P., and Salvi, R. (**1997**). ''The effects of click level, click rate, and level of background masking noise on the inferior colliculus potential (ICP) in the normal and carboplatin-treated chinchilla,'' J. Acoust. Soc. Am. **102**, 3620–3627.

Burkard, R. F., Secor, C. A., and Salvi, R. J. (**1999**). ''Near-field responses from the round window, inferior colliculus, and auditory cortex of the unanesthetized chinchilla: Manipulations of noise-burst level and rate,'' J. Acoust. Soc. Am. **106**, 304–312.

Cohen, L. T., Rickards, F. W., and Clark, G. M. (**1991**). ''A comparison of steady-state evoked potentials to modulated tones in awake and sleeping humans,'' J. Acoust. Soc. Am. **90**, 2467–2479.

Creutzfeldt, O., Hellweg, F. C., and Schreiner, C. (**1980**). ''Thalamocortical transformation of responses to complex auditory stimuli,'' Exp. Brain Res. **39**, 87–104.

Dolphin, W. F., Chertoff, M. E., and Burkard, R. (**1994**). ''Comparison of the envelope following response in the Mongolian gerbil using two-tone and sinusoidally amplitude-modulated tones,'' J. Acoust. Soc. Am. **96**, 2225–2234.

Eggermont, J. J. (**1994**). ''Temporal modulation transfer functions for AM and FM stimuli in cat auditory cortex. Effects of carrier type, modulating waveform, and intensity,'' Hear. Res. **74**, 51–66.

Galambos, R., Makeig, S., and Talmachoff, P. J. (**1981**). ''A 40 Hz auditory potential recorded from the human scalp,'' Proc. Natl. Acad. Sci. U.S.A. **78**, 2643–2647.

Greenwood, D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2604.

Griffiths, S. K., and Chambers, R. D. (**1991**). ''The amplitude modulation-following response as an audiometric tool,'' Ear Hear. **12**, 235–241.

Hofstetter, P., Ding, D., Powers, N., and Salvi, R. (**1997**). ''Quantitative relationship of carboplatin dose to magnitude of inner and outer hair cell loss and the reduction in distortion product otoacoustic emission amplitude in chinchilla,'' Hear. Res. **112**, 199–215.

Kiren, T., Aoyagi, M., Furuse, H., and Koike, Y. (**1994**). ''An experimental study on the generator of amplitude-modulation following response,'' Acta Oto-Laryngol., Suppl. **511**, 28–33.

Kuwada, S., Batra, R., and Maher, V. L. (**1986**). ''Scalp potentials of normal and hearing-impaired subjects in response to sinusoidally amplitude-modulated tones,'' Hear. Res. **21**, 179–192.

Langner, G., and Schreiner, C. E. (**1988**). ''Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms,'' J. Neurophysiol. **60**, 1799–1822.

Levi, E. C., Folsom, R. C., and Dobie, R. A. (**1993**). ''Amplitude-modulated following response (AMFR): Effects of modulation rate, carrier frequency, age, and state,'' Hear. Res. **68**, 42–52.

Lins, O. G., Picton, P. E., Picton, T. W., Champagne, S. C., and Durieux-Smith, A. (**1995**). ''Auditory steady-state responses to tones amplitude-modulated at 80-110 Hz,'' J. Acoust. Soc. Am. **97**, 3051–3063.

McFadden, S. L., Kasper, C., Ostrowski, J., Ding, D., and Salvi, R. J. (**1998**). ''Effects of inner hair cell loss on inferior colliculus evoked potential thresholds, amplitudes, and forward masking functions in chinchillas,'' Hear. Res. **120**, 121–132.

Phillips, D. P., Hall, S. E., Guo, Y., and Burkard, R. F. (**2001**). ''Sensitivity of unanesthetized chinchilla auditory system to noise burst onset, and the effects of carboplatin,'' Hear. Res. **155**, 133–142.

Picton, T. W., Durieux-Smith, A., Champagne, S. C., Whittingham, J., Moran, L. M., Giguere, C., and Beauregard, Y. (**1998**). ''Objective evaluation

of aided thresholds using auditory steady-state responses,'' J. Am. Acad. Audiol **9**, 315–331.
Qiu, C., Salvi, R., Ding, D., and Burkard, R. (**2000**). ''Inner hair cell loss leads to enhanced response amplitudes in auditory cortex of unanesthetized chinchillas: Evidence for increased system gain,'' Hear. Res. **139**, 153–171.
Rees, A., and Moller, A. R. (**1983**). ''Responses of neurons in the inferior colliculus of the rat to AM and FM tones,'' Hear. Res. **10**, 301–330.
Rees, A., and Palmer, A. R. (**1989**). ''Neuronal responses to amplitude-modulated and pure-tone stimuli in the guinea pig inferior colliculus, and their modification by broadband noise,'' J. Acoust. Soc. Am. **85**, 1978–1994.
Ribaupierre, F., Goldstein, M., and Yeni-Komshian, G. (**1972**). ''Cortical coding of repetitive acoustic pulses,'' Brain Res. **48**, 205–225.
Salvi, R., Wei, S., Wang, J., Guo, Y., Ding, D., and Burkard, R. (**2000**). ''Diminished GABA-A mediated inhibition in auditory cortex following carboplatin-induced inner hair cell loss,'' presented at the 2000 Midwinter Meeting of the Association for Research in Otolaryngology, St. Petersburg, Florida.
Schreiner, C., and Raggio, M. (**1996**). ''Neuronal responses in cat primary auditory cortex to electrical cochlear stimulation. II. Repetition rate coding,'' J. Neurophysiol. **75**, 1283–1300.
Schreiner, C. E., and Urbas, J. V. (**1988**). ''Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields,'' Hear. Res. **32**, 49–63.
Snyder, D. L., and Salvi, R. J. (**1994**). ''A novel chinchilla restraint device,'' Lab Anim. **23**, 42–44.
Takeno, S., Harrison, R. V., Ibrahim, D., Wake, M., and Mount, R. J. (**1994a**). ''Cochlear function after selective inner hair cell degeneration induced by carboplatin,'' Hear. Res. **75**, 93–102.
Takeno, S., Harrison, R. V., Mount, R. J., Wake, M., and Harada, Y. (**1994b**). ''Induction of selective inner hair cell damage by carboplatin,'' Scanning Microsc. **8**, 97–106.
Wang, J., Powers, N. L., Hofstetter, P., Trautwein, P., Ding, D., and Salvi, R. (**1997**). ''Effects of selective inner hair cell loss on auditory-nerve fiber threshold, tuning and spontaneous and driven discharge rate,'' Hear. Res. **107**, 67–82.

Acoustic intensity, impedance and reflection coefficient in the human ear canal

B. L. Farmer-Fedor and R. D. Rabbitt[a)]
Department of Bioengineering, 2480 Merrill Engineering Building, 50 South Central Campus Drive, University of Utah, Salt Lake City, Utah 84112

(Received 10 January 2002; revised 1 April 2002; accepted 20 May 2002)

The sound power per unit cross-sectional area was determined in human ear canals using a new method based on measuring the pressure distribution (P) along the length of variable cross-section acoustic waveguides. The technique provides the pressure/power reflection coefficients ($R/\mathcal{R}$) as well as the acoustic intensity of the *nonplanar* incident wave (I_+, *the acoustic input to the ear*) and the *nonplanar* outgoing wave (I_-, *the acoustic output of the ear*). Results were compared to the classical acoustic impedance (Z) and associated *plane-wave* power reflection coefficient ($\mathcal{R}_Z$). Performance of the method was investigated theoretically using horn equation simulations and evaluated experimentally using pressure data recorded in nonuniform waveguides. The method was applied in normal-hearing young adults to determine ear-canal position- and frequency-dependence of $I_\pm$, $\mathcal{R}$, and $\mathcal{R}_Z$ using random phase broadband stimuli (1–15 kHz; ~75 dB SPL). Reflection coefficient ($\mathcal{R}$) measurements at two different locations within individual human ear canals exhibited a position dependence averaging $\Delta\mathcal{R}\sim 0.1$ (over 6 mm distance)—a difference consistent with predictions of inviscid acoustics in nonuniform waveguides. Since this position dependence was relatively small, an ''optimized'' position-independent reflection coefficient was defined to facilitate practical application and intersubject comparisons. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1494445]

PACS numbers: 43.64.Yp, 43.64.Bt, 43.20.Mv [LHC]

I. INTRODUCTION

The sound transmission behavior of the ear canal is often discussed in terms of slender waveguide theory, which breaks the three-dimensional (3D) acoustical problem down into propagation, reflection and decay of a series of cross-sectional acoustic modes (Morse and Ingard, 1968). Slender waveguide theory reproduces much of the known behavior of the ear canal including the existence of 3D evanescent modes near the tympanic membrane (TM) and near the concha superimposed on a one-dimensional (1D) ''plane-wave'' that propagates along the entire length of the canal (Shaw and Teranishi, 1967; Lawton, 1979; Stinson *et al.*, 1982; Stinson, 1985; Rabbitt and Holmes, 1988). Evanescent cross-sectional modes can be ignored in the central *core region* of the canal where the pressure distribution has been described as the superposition of an incident wave traveling toward the tympanic membrane and an outgoing wave traveling toward the concha (Shaw and Teranishi, 1967; Lawton, 1979; Stinson *et al.*, 1982; Stinson, 1985; Rabbitt and Freidrich, 1991). The outgoing wave represents the sum of a passive reflection plus an active otoacoustic emission originating within the cochlea. The sound pressure measured at any point within the ear canal contains contributions from all three of these pressure components—the incident sound, passive reflection and cochlear emission. An important problem in auditory research involves identification and separation of these three components on the basis of acoustical measurements in the ear canal. Proper identification is a necessary prerequisite for quantification of (i) the acoustic input delivered to the ear, (ii) the fraction of the input passively transmitted, and (iii) active emissions by the cochlea.

Both time domain and frequency domain approaches have shown some promise in distinguishing acoustic inputs, passive reflections and active emissions based on acoustical measurements in the ear canal. Time domain methods take advantage of the differences in latency and emission nonlinearity to distinguish the pressure components (Kemp, 1978; Keefe, 1997; Berlin, 1998). Frequency domain methods generally distinguish the incident from outgoing sound waves using impedance measurements (Rosowski *et al.*, 1984; Keefe *et al.*, 1992, 1993; Voss and Allen, 1994; Lynch *et al.*, 1994; Huang *et al.*, 1997, 2000; Margolis, 2001) or standing wave patterns (Stinson, 1985; Shaw and Teranishi, 1967), and distinguish nonlinear cochlear emissions using distortion products or level-dependence (Lonsbury-Martin *et al.*, 1990; Burns *et al.*, 1998). The focus of the work reported here was to develop an alternative method to measure the incident wave acoustic intensity (I_+, *the acoustic input to the ear*) and the outgoing wave acoustic intensity (I_-, *the acoustic output of the ear, i.e., the reflection and emission*) on the basis of multiple pressure measurements within the human ear canal. The method is distinct from previous approaches in accommodating the nonuniform ear canal morphology, avoiding source calibration and avoiding sealing the ear canal. The method is also suitable for analysis of the nonlinear input–output behavior of the ear. We begin by developing the theoretical basis of the method followed by a description of the experimental approach, application in nonuniform

a)Author to whom correspondence should be addressed. Electronic mail: r.rabbitt@utah.edu

 0001-4966/2002/112(2)/600/21/$19.00

horns mimicking cross-sectional properties of the ear canal and, finally, application in normal-hearing young adults.

II. THEORETICAL BACKGROUND

A. Horn equations

In the present work we are interested in determining components of sound energy transmission in the ear canal on the basis of acoustic pressure measurements. The problem is simplified by restricting attention to the core region of the ear canal where higher-order evanescent modes are relatively small and the pressure distribution is well suited for analysis using the 1D curvilinear horn (CH) theory (Agulló *et al.*, 1999; Keefe and Barjau, 1999). Assuming homogeneous thermodynamic properties of air in the canal, the CH theory takes the form

$$\frac{1}{c^2}\frac{\partial^2 P}{\partial T^2}=\frac{1}{A_C}\frac{\partial}{\partial S}\left(A_I\frac{\partial P}{\partial S}\right), \tag{1}$$

where $P(S,T)$ is the pressure distribution along the length of the ear canal, T is time and c is the speed of sound. S is the lengthwise curvilinear coordinate defined in Appendix B. The generalized area functions $A_C(S)$ and $A_I(S)$ are related to the compliance and inertance of a differential section of the horn and can be determined directly from the horn geometry and the 3D theory of inviscid acoustics (Agulló *et al.*, 1999; Keefe and Barjau, 1999). When the area functions are selected properly, the solution of Eq. (1) is equivalent to the solution of the 3D equations of acoustics. It is notable that the classical Webster horn equation (Webster, 1919) applied previously to the ear canal (Khanna and Stinson, 1985; Stinson and Lawton, 1989; Rabbitt, 1988) is equivalent to the CH theory if and only if the area function appearing in the Webster equation is related to $A_C(S)$ and $A_I(S)$ in a frequency-dependent manner as defined in Appendix B [Eq. (B4); also see Agulló *et al.*, 1999; Keefe and Barjau, 1999].

Pressure distributions defined by the CH or Webster horn equations are sometimes called ''plane waves'' to distinguish them from trapped, nonpropagating, evanescent modes present near the concha and TM (Rabbitt and Freidrich, 1991; Rabbitt and Holmes, 1988). As pointed out by Stinson and Khanna (1989) and Agulló *et al.* (1999), it is important to note that the CH and Webster horn equations are not actually plane-wave models, but rather are one-independent-variable (1D) models. The pressure distributions obeying the CH equation also satisfy the 3D theory of inviscid acoustics and therefore expand and contract with the cross-sectional area of the canal. Fluid particles near the wall move tangent to the wall while fluid particles near the center move tangent to the lengthwise curvilinear coordinate S. The corresponding velocity potential is 3D due to the fact that it is a function of cross-sectional position and lengthwise coordinate, S. The influence of the 3D pressure distribution on the area functions (A_I and A_C) is described for axisymmetric horns by Agulló *et al.* (1999). Solutions of the CH equation reduce to ''plane waves'' only if the cross-sectional area is constant (in this case, $A_I=A_C=A$). Because of this, equations and results obtained for plane waves do not necessarily hold for wave propagation in the ear canal. It is shown below, for example, that the reflection coefficient for nonplanar waves in the ear canal is not identical to the approximate value obtained on the basis of plane-wave acoustic impedance.

A great deal of insight into ear canal acoustics can be gained through study of the CH and Webster models. This is facilitated by a simple transformation of the dependent variable and conversion to the Fourier domain using

$$P(S,T)=\tilde{P}(S)\sqrt{\frac{A_I(S_o)}{A_I(S)}}e^{i\omega T}, \tag{2}$$

where S_o is a reference position near the TM. Inserting this into the CH equation [Eq. (1)] provides the following second-order differential equation governing the pressure distribution $\tilde{P}$ in the Fourier domain,

$$\frac{d^2\tilde{P}}{dS^2}+\Lambda^2\tilde{P}=0, \tag{3}$$

where the function Λ^2 is given by

$$\Lambda^2(S)=\left(\frac{A_C}{A_I}\kappa^2+\left(\frac{1}{2A_I}\frac{dA_I}{dS}\right)^2-\frac{1}{2A_I}\frac{d^2A_I}{dS^2}\right). \tag{4}$$

The acoustic wavenumber is $\kappa^2=(\omega/c)^2$. Note that Λ^2 reduces to the constant κ^2 for uniform horns or plane waves. For some special horn geometries Λ^2 is a simple analytic function and exact solutions of Eq. (3) can be found (Keefe and Barjau, 1999; Pierce, 1994; Benade and Jansson, 1974; Salmon, 1946a, b). In the ear canal $\Lambda^2(S)$ is a subject-specific function of position and may be positive, negative or zero depending upon frequency and spatial location. If $\Lambda^2 \gg 0$, solutions of Eq. (3) appear in the form of two traveling waves, one traveling toward the TM and the other traveling toward the concha. In contrast, if $\Lambda^2 \ll 0$, solutions of Eq. (3) correspond to evanescent pressure distributions that do not propagate as traveling waves. The case when $\Lambda^2=0$ falls between these two extremes and mathematically corresponds to a *turning point* (Holmes, 1995). In simplified terms, turning points represent positions where the pressure distribution would be predicted to undergo a smooth transition from an ideal traveling wave to an ideal evanescent mode or visa-versa. This may occur at numerous locations within the ear canal, particularly as the frequency is reduced.

The possibility of turning points leads to the question: Does the ear canal support 1D traveling waves, evanescent waves, or both? Some insight can be gained by setting $\Lambda^2=0$ and solving Eq. (4) for the critical frequencies at which turning points occur,

$$\omega_{cr}=\mathrm{Re}\left(c\sqrt{\frac{1}{2A_C}\frac{d^2A_I}{dS^2}-\frac{1}{4A_IA_C}\left(\frac{dA_I}{dS}\right)^2}\right), \tag{5}$$

with units rad/s. Figure 1(a) shows ω_{cr} as a function of position, with 0 located at the TM, estimated on the basis of human ear-canal geometries [using areas previously reported by Stinson and Lawton (1989), reproduced in Fig. 1(b)]. To compute these critical frequencies we curve fit the area data for each canal with a seventh-order polynomial to find a smooth version of $A(S)$ prior to computing spatial derivatives. We have also assumed $A_I \sim A_C \sim A$, corresponding to a

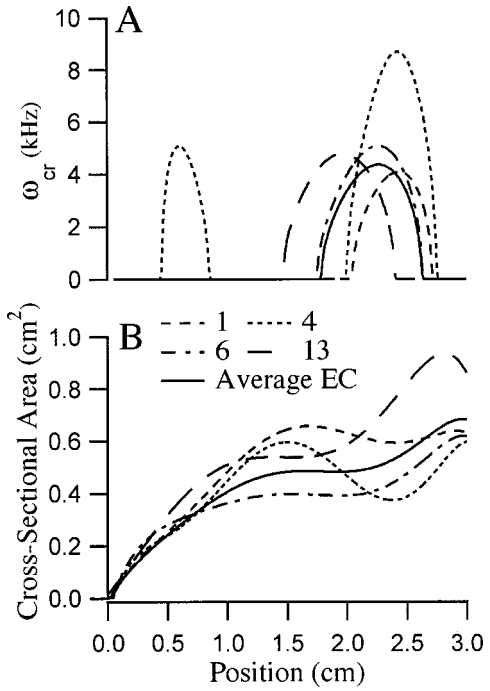


FIG. 1. Critical frequency versus position. (a) Critical frequencies represent *turning points*, which separate propagating wave regions (above each curve) from evanescent zones (area below each curve) in the ear canal. (b) Corresponding ear canal cross-sectional areas for which the critical frequencies in (a) were calculated (based on Stinson and Lawton, 1989). In this case, we assumed $A_I \sim A_C \sim A$ to calculate ω_{cr} using Eq. (5).

simple Webster approximation; therefore the critical frequencies shown are only approximate. As a crude guideline one might view the area below each curve in Fig. 1(a) as indicating evanescent regions and the area above each curve as indicating propagating wave regions. Turning points, however, do not result in an immediate transition from a traveling wave to an evanescent mode, but rather involve a slow transition. For a first-order turning point the transition is governed by Airy's equation (see Appendix A and Holmes, 1995). If the evanescent region spans only a short length of the canal, then a propagating wave entering the region would not be significantly attenuated but, rather, would pass through the region with slightly decreasing amplitude and nearly constant phase. In other words, a short evanescent region would stretch the wavelength and introduce a small local reflection (i.e., slight attenuation). Critical frequency estimates in Fig. 1 indicate that evanescent regions are relatively short in length in human ear canals and, therefore, one might expect their existence to primarily influence the effective wavelength and, to a lesser extent, the local reflection of pressure waves along the length of the canal. Nevertheless, the expected existence of nonpropagating evanescent regions leads us to question under what conditions the standard plane-wave formulas used for acoustic impedance, acoustic intensity and the reflection coefficient apply in the ear canal.

B. Pressure reflection coefficient

The fact that waves satisfying the CH equation are nonplanar, in general, indicates that the reflection coefficient in the ear canal may be different than would be predicted on the basis of ideal plane waves. To address this point, we developed a technique to use ear-canal pressure data to search for nonplanar traveling waves of the form

$$P(S,T) = P_{\text{in}} + P_{\text{out}} = \Phi\{P_+ e^{i\Psi_+} + P_- e^{-i\Psi_-}\} e^{i\omega T}, \quad (6)$$

where the pressure, P, is represented as a sum of an incident wave P_{in} and an outgoing wave P_{out}. In this expression, $\Phi = \Phi(S)$ is the amplitude function and $\Psi = \Psi_{\pm}(S)$ are the phase functions for the incident and outgoing waves, respectively (note that the phase functions do not influence the magnitude of the pressure if they are real-valued). Equation (6) is a solution of the CH equation and the 3D theory of acoustics provided the amplitude and phase functions are selected properly—a task achieved here by nonlinear optimization applied to human ear canal pressure data.

The form of Eq. (6) was motivated by a high-frequency asymptotic solution of the CH horn equation, Eqs. (1) and (3), obtained using the method of Wentzel, Kramers and Brillouin (WKB, see Appendix A). A detailed description of the WKB method as applied to Eq. (3) is provided by Holmes (1995) (also see Keefe and Barjau, 1999). The complex-valued constants P_- and P_+ in Eq. (6) give the amplitude of the outgoing and incident waves evaluated at a reference position $S = S_o$ [$\Phi(S_o) = 1$]. The WKB analysis gives explicit forms for $\Psi_{\pm}$ and Φ, valid at high frequencies, in terms of the ear-canal cross-sectional area functions, frequency and acoustical properties of air (see Appendix A). Unfortunately, the first-order WKB approximation provides consistently good results in human ear canals only at high frequencies (above 4–12 kHz depending upon individual canal geometry). This is due primarily to a singularity in the WKB amplitude function that occurs near turning points [i.e., near $\Lambda^2(S) = 0$, see Fig. 1]. Since multiple turning points are expected to occur in real ear canals at low frequencies, and since ear canal geometries are not known *a priori*, a matched asymptotic analysis is impractical. Nevertheless, the general form of Eq. (6) remains a valid approximation over the entire frequency range provided we select the appropriate amplitude and phase functions. These functions are frequency dependent and approach the WKB forms only in the high frequency limit. In the present work we estimate Φ and $\Psi_{\pm}$ on the basis of ear-canal pressure measurements. The method does not assume any particular form of Λ, nor does it require estimation of A_C or A_I.

The advantage of writing the pressure in the form of Eq. (6) is that it clearly separates the acoustic input to the ear, $P_{\text{in}} = \Phi P_+ e^{i(\Psi_+ + \omega T)}$ from the acoustic output of the ear, $P_{\text{out}} = \Phi P_- e^{i(-\Psi_- + \omega T)}$. The ratio of these two components is the *pressure reflection coefficient R*,

$$R = \frac{P_-}{P_+} e^{-i(\Psi_+ + \Psi_-)}. \quad (7)$$

This is valid regardless of the specific functional forms of Φ and $\Psi_{\pm}$ and holds for both plane waves and nonplanar waves. The exponential terms are defined to evaluate to one at the reference point S_o, so the pressure reflection coefficient at the reference point near the TM is simply $R(S_o) = P_-/P_+$. The power reflection coefficient is $\mathcal{R} = RR^*$, where the superscript * indicates the complex conjugate.

Note that $\mathcal{R}$ *is a constant along the length of the ear canal if and only if the phase functions* $\Psi_{\pm}$ *are real-valued*. Based on the WKB analysis in Appendix A we expect $\Psi_{\pm}$ to be real-valued when $\omega > \omega_{cr}$ and therefore we expect $\mathcal{R}$ to be constant along the length of the ear canal at high frequencies.

It is interesting to note that the reflection coefficient obtained using the plane-wave acoustic impedance does not reproduce the more general pressure reflection coefficient given by Eq. (7). It is shown below that the standard plane-wave technique fails due to the assumed relationship between $\mathcal{R}$ and plane-wave acoustic impedance.

C. Acoustic impedance

A relatively common technique used to estimate $\mathcal{R}$ (or R) in the ear canal is through measurement of acoustic impedance, Z. The 1D specific, i.e., area-normalized, acoustic impedance is defined in the frequency domain as

$$Z = \frac{\tilde{P}}{\tilde{V}}, \tag{8}$$

where $\tilde{P}$ is pressure and $\tilde{V}$ is the acoustic particle velocity (Pierce, 1994). The circumflex ($\sim$) denotes the frequency domain. It is easy to show from the theory of inviscid linear acoustics that the acoustic impedance looking toward the tympanic membrane can be computed from the ear-canal pressure distribution using

$$Z = \rho c \zeta = i\omega\rho \left(\frac{d \ln(\tilde{P})}{dS} \right)^{-1}, \tag{9}$$

where ρ is density, c is the speed of sound in air and ζ is the nondimensional ratio of specific impedance (Z) to the characteristic impedance of air (ρc). The sign of Eq. (9) has been selected to provide the input impedance of the ear, i.e., the impedance viewed in the $-S$ direction. We note that the sign of Eq. (9) differs from Pierce (1994), but provides the same result due to the opposite direction of S. This expression assumes that the particle velocity is tangent to the lengthwise curvilinear S coordinate and that the pressure is recorded along the curved S coordinate. In previous work by numerous investigators (Keefe *et al.*, 1993; Voss and Allen, 1994; Lynch *et al.*, 1994; Huang *et al.*, 2000) the impedance has routinely been converted to a pressure reflection coefficient R_Z using the equation

$$R_Z = \frac{\zeta - 1}{\zeta + 1}, \tag{10}$$

where the associated power reflection coefficient is $\mathcal{R}_Z = R_Z R_Z^*$. Equation (10) is clearly correct for propagating plane waves in uniform ducts (Pierce, 1994). It is notable, however, that Eq. (10) does not give the same result as Eq. (7) simply because it was derived for uniform plane waves and therefore is only an approximation for variable cross-section waveguides such as the ear canal even if the impedance Z is measured exactly.

A consistent formula for a variable-cross section horn giving R in terms of the pressure distribution along the curved centerline can be derived directly from Eq. (6). This is done by taking the first derivative of Eq. (6) and combining the result with the original Eq. (6) to solve for the pressure reflection coefficient, R:

$$R = \frac{P_-}{P_+} e^{-i2\Psi} = \frac{d \ln(\Phi)/dS + (i\hat{\kappa} - d \ln(\tilde{P})/dS)}{-d \ln(\Phi)/dS + (i\hat{\kappa} + d \ln(\tilde{P})/dS)}, \tag{11}$$

where we have introduced the effective wavenumber as the derivative of the phase function, $\hat{\kappa} = d\Psi/dS$. For simplicity we have taken the incident and outgoing phase functions to be identical in form, i.e., $\Psi = \Psi_+(S;\omega) = \Psi_-(S;\omega)$ (see Sec. V). The notation $\Psi(S;\omega)$ indicates that S is the independent variable along the curved centerline and the semicolon denotes that the frequency, ω, appears as an implicit parameter. *Equation (11) reduces to* R_Z [*Eq. (10)*] *if and only if* $d\Phi/dS = 0$ and $\hat{\kappa} = \kappa$. These conditions hold for straight uniform tubes, or ideal plane waves, but do not hold for the Webster equation or the CH equation and do not hold in the ear canal. Therefore, $R_Z \neq R$ in the ear canal. In fact, it is not possible to compute R from Z without additional information.

The impedance, Z, is relatively easy to measure at locations within the ear canal, making it relatively easy to determine R_Z. In order to determine R using Eq. (11) it is also necessary to know quantities giving $\hat{\kappa}$ and $d \ln(\Phi)/dS$. The effective wavenumber $\hat{\kappa}$ is expected to be close to κ, at least above ω_{cr}; hence we take κ as a first estimate in a nonlinear optimization method. To provide an initial estimate for $d \ln(\Phi)/dS$ we note that $(d/dS)(P_-/P_+) = 0$, which must be true since P_+ and P_- are not functions of position. To take advantage of this observation, we multiply Eq. (11) by $e^{i2\Psi}$ and take the derivative with respect to S. Noting that $(d/dS)(P_-/P_+) = 0$, $2ie^{i2\Psi} \neq 0$, and that the function $d \ln(\Phi)/dS$ is real-valued, we obtain

$$\frac{d \ln(\Phi)}{dS} \approx \frac{\mathrm{Im}\left(\left(\frac{d \ln(\tilde{P})}{dS} \right)^2 + \frac{d^2 \ln(\tilde{P})}{dS^2} + \hat{\kappa}^2 - \frac{d \ln(\tilde{P})}{dS} \frac{1}{\hat{\kappa}} \frac{d\hat{\kappa}}{dS} \right)}{\mathrm{Im}\left(2 \frac{d \ln(\tilde{P})}{dS} - \frac{1}{\hat{\kappa}} \frac{d\hat{\kappa}}{dS} \right)}. \tag{12}$$

It can be verified by inserting the traveling wave expressions into Eq. (12) that $d \ln(\Phi)/dS$ is equal to zero for plane waves but clearly is nonzero in variable cross-section waveguides such as the ear canal. Upon substitution of Eq. (12) into Eq. (11), the problem is reduced to a nonlinear parameter estimation exercise to extract $\hat{\kappa}$ and R from ear canal pressure data. Parameter estimation methods used in the present study are provided in Sec. III.

D. Effective Webster area estimate

The phase of the pressure distribution in the ear canal can be reproduced to reasonable accuracy by assuming the effective wavenumber $\hat{\kappa}$ is a real-valued constant, albeit dis-

tinct from the acoustic wavenumber κ (Stinson and Lawton, 1989; Stinson and Khanna, 1989). With this approximation, Eq. (12) reduces to

$$\frac{d\ln(\Phi)}{dS} \approx \frac{d\ln(\bar{\Phi})}{dS} = \mathrm{Re}\left(\frac{d\ln(\tilde{P})}{dS}\right) + \frac{1}{2}\frac{d}{dS}\left\{\ln\left(\mathrm{Im}\left|\frac{d\ln(\tilde{P})}{dS}\right|\right)\right\}. \tag{13}$$

Although the actual amplitude function, $\Phi(S;\omega)$, varies with frequency, this approximate amplitude function $\bar{\Phi}(S)$ is independent of frequency and unique for each cross-sectional area function. This allows $\bar{\Phi}(S)$ to be estimated from noisy pressure data using

$$\frac{d\ln(\bar{\Phi})}{dS} = \sum_{k=1}^{K}\left[\mathrm{Re}\left(\frac{d\ln(\tilde{P}_k)}{dS}\right) + \frac{1}{2}\frac{d}{dS}\left\{\ln\left(\mathrm{Im}\left|\frac{d\ln(\tilde{P}_k)}{dS}\right|\right)\right\}\right] W_k\left(\sum_{k=1}^{K} W_k\right)^{-1}, \tag{14}$$

where the subscript "k" denotes the frequency, and W_k is a weighting parameter taken here to be proportional to the spatially averaged pressure magnitude. For the Webster horn equation, the effective area function $A_W(S)$ is related to $\bar{\Phi}(S)$ according to

$$A_W(S) \approx \frac{A_o}{\bar{\Phi}(S)^2}, \tag{15}$$

where the boundary conditions are $A_o = A_W(S_o)$ and $\bar{\Phi}(S_o) = 1$. It is interesting to note that Eq. (15) is an exact expression for the Webster horn equation and reproduces the area estimate of Hudde (1983) under the additional assumption $\hat{\kappa} = \kappa$. It extends to the more general CH equation only if the Webster area is related to the CH area functions as defined in Appendix B. We therefore refer to this estimated area as the "*effective Webster area function* A_W" as opposed to the actual cross-sectional area of the horn.

E. Acoustic intensity

The sound power transmitted by the incident and outgoing waves in Eq. (6) can be determined in the form of *instantaneous acoustic intensity*—a vector quantity equal to the sound power propagation per unit area. It is computed as the pressure times the particle velocity. The 3D theory of inviscid linear acoustics gives the instantaneous acoustic intensity $\vec{I}$ as

$$\vec{I} = P\vec{V} = \frac{-P}{\rho}\int_T \vec{\nabla} P(\xi)\, d\xi, \tag{16}$$

where P is the pressure and $\vec{V}$ is the acoustic particle velocity. The magnitude of the instantaneous acoustic intensity along the lengthwise coordinate of the ear canal can be written in the frequency domain as

$$I = \left|\frac{i\tilde{P}}{\rho\omega}\frac{d\tilde{P}}{dS}\right|. \tag{17}$$

This quantity can be computed for the total pressure ($\tilde{P} = \tilde{P}_{\mathrm{in}} + \tilde{P}_{\mathrm{out}}$), for the incident wave ($\tilde{P}_{\mathrm{in}} = P_+\Phi(S)e^{i\hat{\kappa}S}$), or for the outgoing wave ($\tilde{P}_{\mathrm{out}} = P_-\Phi(S)e^{-i\hat{\kappa}S}$) to obtain I_{tot}, I_+ or I_-, respectively. Using pressures in the form of Eq. (6), the magnitudes of the acoustic intensities of the incident (+) and outgoing (−) waves are

$$I_\pm = \frac{P_\pm P_\pm^* \Phi^2}{\rho\omega}\sqrt{\left(\frac{d\ln(\Phi)}{dS}\right)^2 + \hat{\kappa}^2}, \tag{18}$$

where we have assumed real-valued phase functions (and $\hat{\kappa} = d\Psi/dS$, see Sec. V). This expression reduces to the standard plane wave form, and is constant with spatial position, only if the amplitude function is one ($\Phi = 1$) and the effective wavenumber is equal to the acoustic wavenumber ($\hat{\kappa} = \kappa$)—conditions which hold only for ideal plane waves. In the ear canal the intensities vary along the length of the canal due to the variable cross-sectional area functions.

The ratio of the outgoing intensity to the incident intensity defines the *power reflection coefficient* $\mathcal{R}$ and is found from the acoustic intensities defined by Eq. (18) to be

$$\mathcal{R} = \frac{I_-}{I_+} = \left(\frac{P_-}{P_+}\right)^2 = RR^*. \tag{19}$$

Once again we find the magnitude of the reflection coefficient to be independent of position providing that the phase functions $\Psi_\pm$ appearing in Eq. (6) are real-valued. This acoustic intensity derived result is identical to that based on the pressure reflection coefficient in Eq. (7). The result holds even when the amplitude function varies with position, as would be the case in human ear canals due to their variable cross-sectional areas.

Equation (18) provides the acoustic intensities of the incident and outgoing waves taken independently. To determine the total instantaneous acoustic intensity, I_{tot}, Eq. (17) is applied directly to the total pressure field. It is significant to note that the total pressure $\tilde{P}(S)$ forms standing waves in the ear canal due to the interference of incident and outgoing waves. Because of this, I_{tot} also exhibits standing waves, but at half the wavelength. This standing wave effect is not present in the incident and outgoing intensities given by Eq. (18).

III. METHODS

A. Experimental methods

Four males and two females, ages 22–34, participated in the present study under informed consent and methods approved by the University of Utah Institutional Review Board (IRB, 5735-96). Hearing thresholds were determined using air conduction pure-tone audiometry. All subjects were within the normal range of hearing (<10 dB HL from 250 to 8000 Hz). Pressure measurements in the ear canal used to determine acoustic impedance and reflection coefficients were then taken in these subjects while they were alert and in the supine position.

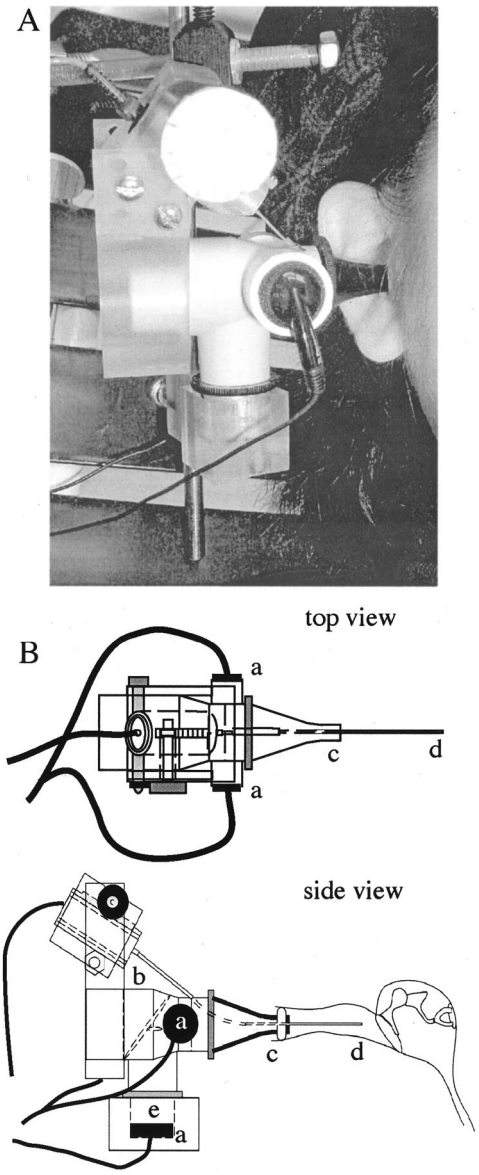


FIG. 2. Experimental apparatus. An otoscope was modified to present sound stimuli through two stereo earphones and one mono earphone (a) and to measure pressure using a probe microphone (b) with a silicone probe tip attached (d). The speculum of the otoscope (c) was placed in the concha and the probe tip inserted into the ear canal under visual guidance using the otoscope light. After probe tube insertion, the light was removed and the third earphone inserted into the light hole (e) during pressure measurements, which were taken at 2-mm intervals along the length of the ear canal. (A) Photo of the modified otoscope prior to insertion of the third earphone. (B) Sketches of the top and side views of the measurement apparatus.

The measurement apparatus, shown in Fig. 2(A) and sketched in Fig. 2(B), consisted of an otoscope (NOTOCO Earscope Model 1202) modified to support two stereo earphones and one mono earphone (Fig. 2, object a; stereo: Panasonic RP-HV147, mono: Comtek SM-N) for sound delivery and a probe microphone (Fig. 2, object b; Etymotic Research ER-7C) for sound pressure measurement. The stereo earphones were situated in short tubes attached to the side of the otoscope. The source geometry was designed to minimize cross-sectional evanescent modes in the sound delivered to the ear canal entrance. All three sources were excited in temporal synchrony (in phase). The position of each of the earphones was adjusted manually within its support tube in order to minimize destructive acoustic interaction and thereby maximize the centerline pressure and plane wave delivered to the ear canal. The role of the third earphone was primarily to minimize the muffler effect of the otoscope cavity near 2 kHz. Measurement of cross-sectional pressure distributions at several locations within 8.5-mm-diam straight tubes showed only small perturbations in pressure ($<3\%$), indicating that plane waves were indeed being delivered by the device. A rack-and-pinion micromanipulator was fixed to the otoscope to allow the microphone probe tip (Fig. 2, object d) to be indexed along the length of the ear canal. The microphone probe tip was positioned by fitting the speculum of the otoscope (Fig. 2, object c) loosely in the concha and securing the apparatus in place with a head strap. A light was then placed in the appropriate opening of the otoscope (Fig. 2, location e) to provide illumination for viewing insertion of the probe tip into the ear canal. During measurements, the light was removed and the otoscope light hole was filled by the third earphone speaker.

Prior to pressure measurements in each ear, the silicone probe tube was replaced with a clean one. Although each microphone tube assembly had a slightly different transfer function, calibration was not necessary to determine R, which is a dimensionless ratio of pressures. Probe tube dependent microphone calibration would be needed for studies addressing absolute pressure and/or absolute acoustic intensity. Following probe tube replacement, the measurement apparatus was secured in place and the silicone probe tip was placed to within ~ 2 mm of the TM under visual guidance through the otoscope. The initial probe tip position was used as the reference, therefore $S_o \approx 2$ mm. By convention, the ear canal curved coordinate "S" and microphone probe tip position "X" both increased as the distance from the TM increased. Since the probe tip was not forced to move exactly along the curvilinear centerline of the ear canal, the microphone position only provided an estimate of the canal coordinate [i.e., $S = S(X) \sim X$]. Data recorded at indexed positions along the length of the ear canal provided the pressure distribution as a function of microphone probe tip position (X).

The input stimulus was random-phase noise, which was digitally constructed to have a relatively flat power spectrum from 1 to 15 kHz (WaveMetrics, IgorPro), and downloaded (National Instruments, GPIB) to an arbitrary waveform generator (Tektronix, AFG5102). The sound stimulus was presented 12 times per second at approximately 75 dB SPL. Broadband stimuli were delivered to the stereo earphones. The same waveform was low-pass filtered with a passive cut-off frequency of 3 kHz and delivered to the third earphone in the otoscope light hole. A stimulus-phase-locked TTL pulse was output by the waveform generator and used to trigger digital sampling of the microphone voltage. Time-domain pressure waveforms were amplified to span the range of the analog to digital converter, filtered, and sampled at 50 kHz (Tektronix, DSA602). The probe tip was indexed an average of nine times in 2-mm intervals along the core region of the ear canal (~ 18 mm total). The 256 stimulus repetitions were averaged in the time domain (DSA602) and

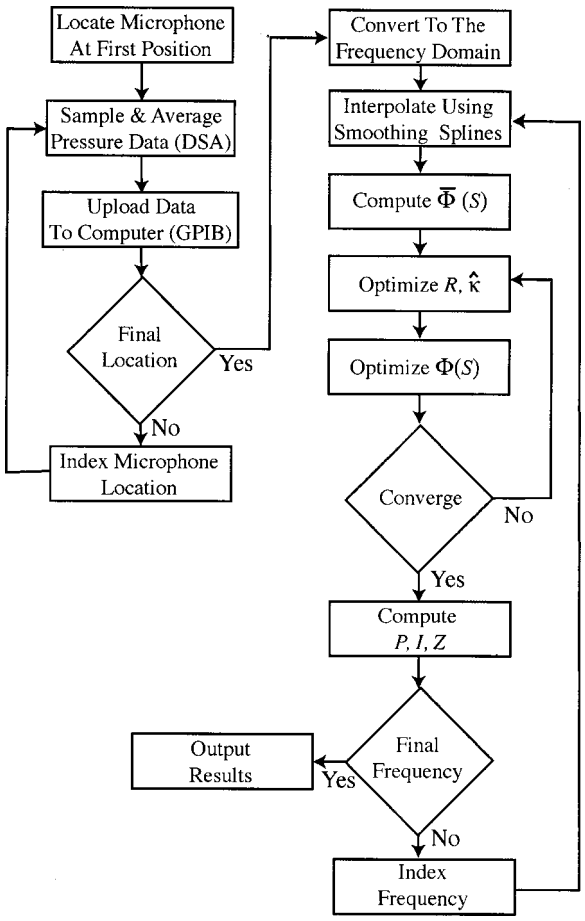


FIG. 3. Methods flow chart. The measurement steps taken for this study are outlined down the left side of the figure and the analysis steps are outlined down the right side. These steps were followed to estimate $\Phi(S)$, $R(S_o)$ and $\hat{\kappa}$ and to calculate the inverse pressure, acoustic intensities and ear canal termination impedance. The same analysis procedure was followed using theoretical data, data collected in physical models and data collected in human ear canals so that results would be comparable.

results were uploaded to a PC (Apple, Macintosh; National Instruments, GPIB) for processing.

The experimental procedure and data analysis are outlined in the flow chart of Fig. 3.

B. Data analysis

Data analysis was performed off-line using custom software written in the IgorPro environment (WaveMetrics, Lake Oswego, OR). Periodic, 4096 point, time-domain pressure data were transformed to the frequency domain (FFT) and analyzed one frequency at a time as outlined on the right side of Fig. 3. The real and imaginary components of the pressure, measured at ten positions along the core region of the canal, were interpolated to 5000 spatial locations using smoothing splines (Reinsch, 1967) to create a smooth function appropriate for numerical differentiation. The spline function, $\hat{P}(S)$, was determined by minimization of the function (shown for the real part)

$$\int_{S_0}^{S_N}\left(\frac{d^2\,\mathrm{Re}(\hat{P})}{dS^2}\right)^2 dS, \tag{20}$$

with noise accounted for by subjecting the minimization to the constraint

$$\sum_{n=1}^{N}\left(\frac{\mathrm{Re}(\hat{P}(S_n)-\tilde{P}_n)}{\sigma}\right)^2 \leqslant 1, \tag{21}$$

where $\hat{P}(S_n)$ is the value of the smooth spline at point S_n and $\tilde{P}_n$ is the pressure data at the same point. An identical procedure was used for the imaginary part. The parameter σ appearing in this expression is an estimate of the standard deviation of the dependent variable arising due to uncertainty in the position of the microphone tip and in the pressure itself. The standard deviation of repeated pressure measurements, once converted to the frequency domain, was relatively flat across the frequency spectrum and averaged $\sigma_1 = 0.4$ mV (~ 1.3 mPa). A more significant source of error was due to uncertainty in S_n due to estimating the centerline coordinate on the basis of the probe microphone position X_n. Some of this error was accounted for by optimizing the wavenumber $\hat{\kappa}$ following Stinson and Khanna (1989). Optimizing $\hat{\kappa}$ accounts for a linear dependence of the centerline coordinate S_n on the microphone position X_n (i.e., $S = bX$, *where b is a constant*) and curvature of the isopressure surfaces, but does not account for nonlinearity of the functional relationship between S and X. Assuming a standard deviation of 0.5 mm due to unaccounted for nonlinearities leads to a standard deviation in pressure less than $\sigma_2 = 0.04\kappa|\tilde{P}|$. Present results combined these two sources of error using $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$. This standard deviation was used in Eq. (21).

Numerical differentiation was applied to the smoothed splines to determine $f_s(S) = d\ln(\hat{P}_s)/dS$. The amplitude function $\Phi(S;\omega)$ was initialized using $\bar{\Phi}(S)$, which was computed by substitution of f_s into Eq. (14). The effective Webster cross-sectional area function $A_W(S)$ was estimated using Eq. (15), but it is important to note that this area function was not used in the analysis of the acoustic intensity or the reflection coefficient. Splines were decimated by sampling at 100 equally spaced points ($S = S_n$) to generate a discrete set of smoothed data (f_n, $n = 1,...,100$) for display and parameter estimation. Increasing the number of sampling points to >100 took longer to analyze and did not change the results. The experimental data were cast in the form $f(S) = d\ln(\tilde{P})/dS$, and fit with the theoretical curve based on Eq. (11):

$$\frac{d\ln(\tilde{P})}{dS} = i\hat{\kappa}\left(\frac{1-(P_-/P_+)e^{-i2\Psi}}{1+(P_-/P_+)e^{-i2\Psi}}\right) + \frac{d\ln(\Phi)}{dS}. \tag{22}$$

Equation (22) was subsequently used for parameter estimation. The Levenberg–Marquardt method was applied to minimize the difference between the theory $f_s(S)$ and the data f_n through optimization of the complex-valued pressure reflection coefficient $R(S_o) = P_-/P_+$ and the real-valued wavenumber $\hat{\kappa}$ (Press *et al.*, 1986). Estimated parameters were subsequently substituted into Eq. (6) to generate the estimated pressure waveforms. The amplitude function was then scaled to match the amplitude of the pressure data with linear interpolation applied between data points. The updated amplitude function was substituted into Eq. (11) to obtain a

new expression for $f(S)$. Levenberg–Marquardt optimization was then reapplied to find a new estimate for the reflection coefficient $R(S_o)$ and wavenumber $\hat{\kappa}$. As illustrated in Fig. 3, this process was repeated to converge upon final estimates of $\Phi(S)$, $R(S_o)$ and $\hat{\kappa}$. The convergence criteria required a change in $R(S_o)$ less than 10^{-3}. A 95% confidence interval for the magnitude of R was estimated using the covariance matrix of the curve-fit to estimate the variance in R and a two-tailed Student t-test at the last iteration of the procedure. Since the optimization method is nonlinear, the confidence interval is only approximate and should be viewed as a relative measure for comparison of results computed at different frequencies and between subjects.

For each ear, the spline-interpolated data were used to compute an impedance-based reflection coefficient, R_{Zs}, from Eqs. (9) and (10) (spline-interpolated data indicated using "s" subscript). The required numerical derivatives were computed using the data and the phase of R_{Zs} was adjusted to the ear canal reference point (S_o) using the acoustic wavenumber κ for each Fourier component following previous methods (Lynch *et al.*, 1994; Voss and Allen, 1994; Huang *et al.*, 2000). In addition, a smoothed version of the impedance-based reflection coefficient, R_Z, was computed on the basis of the optimized traveling wave pressure distribution. For this, the procedure described in the previous paragraphs was applied to generate an optimized pressure distribution in the form of Eq. (6), and Eqs. (9) and (10) were then applied to the optimized waveform to compute R_Z.

C. Physical models

Two nonuniform horns were fabricated from 8-mm nominal i.d. glass tubing to compare results of the present method to the ASTM standing wave method (ASTM C384-95). The standard ASTM curve fitting procedure was modified to take advantage of Levenberg–Marquardt optimization with equal weighting for the real and imaginary parts of the pressure waveform. A flare at the exit of both horns was designed to provide reflection coefficients <0.9 at frequencies above 2 kHz. Although the tubing i.d. was chosen based on the average human ear canal diameter, it was not possible to match the termination impedance of the horns to that of a human ear canal using simple flare terminations. Horn "A" consisted of a 12-cm length of straight tubing at the entrance followed by a section of curved tubing with the same i.d. and a flared termination. Horn "B" consisted of two 12-cm lengths of straight tubing separated by a bulged section and a flared termination. Pressure distributions were measured along the length of the horns using measurement techniques described above with minor modifications. Due to the long length of the glass horns, stainless steel probe microphone tubes, ranging from 12 to 24 cm in length, were used in place of silicone probe tubes. As was the case in human ear canals, the position of the probe tip within the cross section was not controlled and therefore the probe tip only approximately followed the centerlines of the horns. Reflection coefficients were estimated for each adjacent set of ten recording sites, using the methods described above, to obtain the reflection coefficient as a function of position along the horn. A constant "optimized" reflection coefficient was also calculated for each section of each horn using the present methods. The modified ASTM standing wave method was applied to a subset of the same pressure data in the straight uniform sections for comparison (ASTM C384-95). Data were collected in an open laboratory and external transmission from the source around the outside of the horn to its exit was not controlled. Therefore, horn results are strictly for comparing methods and cannot be used to quantify the absolute termination impedance of the horns.

IV. RESULTS

A. Results using "Webster horn" simulated data

Prior to presentation of the human data, it is informative to investigate predictions of a simple horn equation regarding ear-canal position dependence of the reflection coefficient, impedance, and acoustic intensity in model ear canals. For this purpose, simulated data were generated by numerical solution of the Webster version of Eq. (1) using the Runge–Kutta method. Human cross-sectional area functions were taken from Stinson and Lawton (1989). To generate simulated data the Webster area was set equal to the ear canal cross-sectional area measured perpendicular to the centerline ($A_I=A_C=A_W=A$). The boundary condition at a reference point near the TM was approximated using the average termination impedance measured in human ear canals during our experiments. Numerical solutions were subsequently sampled at ten equally spaced points along the model ear canal and noise was added to generate "simulated data" for analysis. Data analysis methods described above were applied to the simulated data to compute the effective Webster cross-sectional area function, reflection coefficient, impedance and components of the acoustic intensity.

An initial approximation of the amplitude function $\Phi(S;\omega)$ was obtained using Eq. (14). The approximate amplitude function, $\overline{\Phi}(S)$, also provided a direct means to estimate the effective Webster cross-sectional area function A_W to within a scaling factor, A_o. The scaling factor was obtained here by matching the spatial average of the estimated area function to the known value. Figure 4(a) compares the estimated Webster area function [dashed curve, Eq. (15)] to the exact area function (solid curve) of model ear canal 13 used to simulate the pressure data. The estimated and actual curves are in close agreement. Similar results were obtained for simulated data generated using all of the human ear canal geometries shown in Fig. 1.

It is important to note that Eqs. (14) and (15) used to obtain the effective Webster area function [Fig. 4(a)] provided excellent agreement with the original area only because the Webster horn equation was used to generate the simulated data. The same procedure applied to data collected in horns, mimicking the nonuniform geometry of human ear canals, did not provide effective cross-sectional area functions equal to the physical cross-sectional area functions [see horn results below, Fig. 4(b)]. As in Fig. 4(a), the physical cross-sectional area (solid curve) was compared to the effective Webster area (dashed curve). These results indicate that the Webster horn equation will predict the ear canal pressure distribution only if the effective area function A_W is indi-

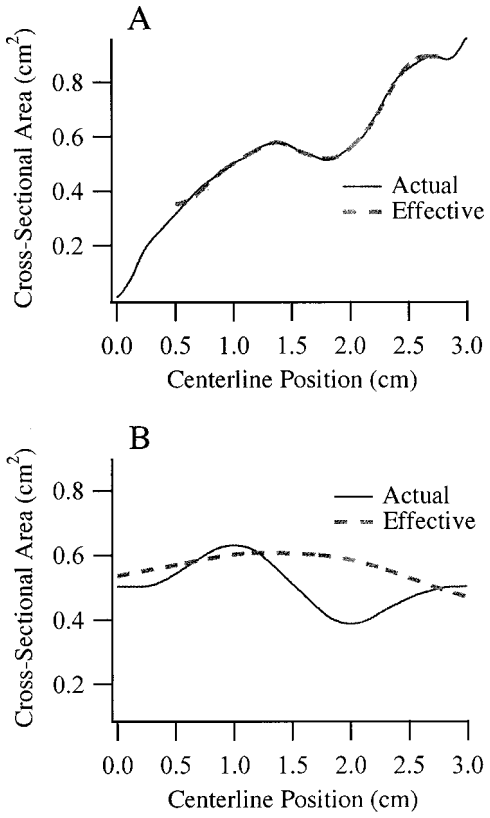


FIG. 4. Effective Webster cross-sectional area. (a) A cross-sectional area function (solid curve) from Stinson and Lawton (1989) was used to generate simulated ear canal pressure data (10 data points). Equations (14) and (15) were then used to obtain the effective Webster area function (dashed curve) from the simulated data. Differences between the two curves are the result of sampling the pressure at ten discrete points to create the simulated data. Increasing the number of sampling points eliminated these differences. (b) The cross-sectional area (solid curve) measured in a bulged tube was compared to the effective Webster area function (dashed curve). In real horns mimicking ear canal geometries, the effective Webster cross-sectional area function was not equal to the physical cross-sectional area function. See Appendix B for an explanation of the relationship between the effective Webster area function A_W and the geometrical cross-sectional area.

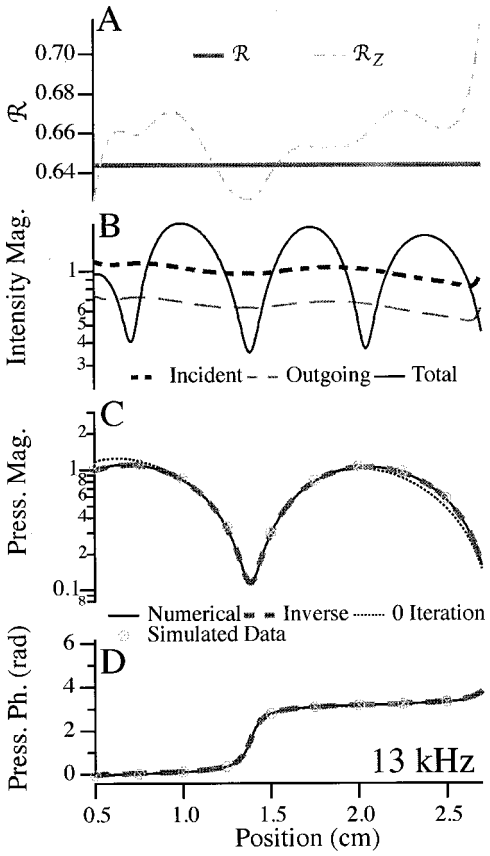


FIG. 5. Spatial distributions for the theoretical results. (a) Reflection coefficients were calculated using the present method ($\mathcal{R}$, solid line) and the impedance method ($\mathcal{R}_Z$, dot-dashed curve). (b) Total (solid curve), incident (short dashed curve), and outgoing (long dashed curve) acoustic intensity were calculated. Total acoustic intensity exhibited pronounced standing waves while the incident and outgoing acoustic intensities did not. (c) and (d) The magnitude and phase of the pressure were calculated numerically using the Runge–Kutta method (solid curve) and using the initial amplitude approximation, $\bar{\Phi}(S)$, (dotted curve) and final results (dashed curve) of the parameter optimization method. Simulated pressure data (open symbols) were obtained by sampling the numerical solution at ten equally spaced discrete points. These data were used in the parameter optimization steps described in Sec. III. The results of this analysis are shown for 13 kHz. The simulated data was created using human ear canal geometry #13 shown in Fig. 1.

rectly related to the physical cross-sectional area and selected to account for canal acoustics [see Appendix B, Eq. (B4)].

Figures 5–7 are examples from a model ear canal at three frequencies showing the pressure [(c) amplitude, (d) phase], acoustic intensity [(b)] and the reflection coefficient [(a)] as functions of distance from the TM. Pressure distributions were computed numerically for pure tones and estimated [(c, d), dashed] using the present inverse methods on the basis of ten simulated data points [(c, d), open circles]. Notice the excellent agreement between the inverse solution and the numerical solution in Fig. 5. Also shown is the zero iteration estimation (dotted) obtained using the approximation $\Phi=\bar{\Phi}$. Correspondence between the curves shows that the inverse solution satisfies the original differential equation and meets the boundary conditions to a high degree of accuracy. Therefore, Eq. (6) is an excellent approximation to the true solution of the horn equation. The fact that a frequency-specific effective area function A_W can be found to render the Webster equation equivalent to the CH equation (Appendix B) means that these results extend directly to the CH equation, albeit applied to a different geometry. In computing these results, the phase function was assumed to be real-valued. This resulted in an optimized power reflection coefficient $\mathcal{R}$ [Fig. 5(a), solid], which was a constant along the length of the model ear canal. Computing a reflection coefficient on the basis of the local impedance provided a position dependent result $\mathcal{R}_Z$ [(a), dot-dashed] with an average value differing from the optimized value of $\mathcal{R}$.

The optimized power reflection coefficient was also computed as the ratio of the intensity of the outgoing wave [(b), thin-long dashed] to the incident wave [(b), thick-short dashed]—a ratio that was constant along the length of the model ear canals due to the fact that the phase functions were restricted to be real-valued. It is notable that the intensity of the incident wave varied somewhat along the length due to changes in cross-sectional area, but was not plagued by standing waves present in the pressure distribution [(c), solid curve] and in the instantaneous total acoustic intensity [(b),

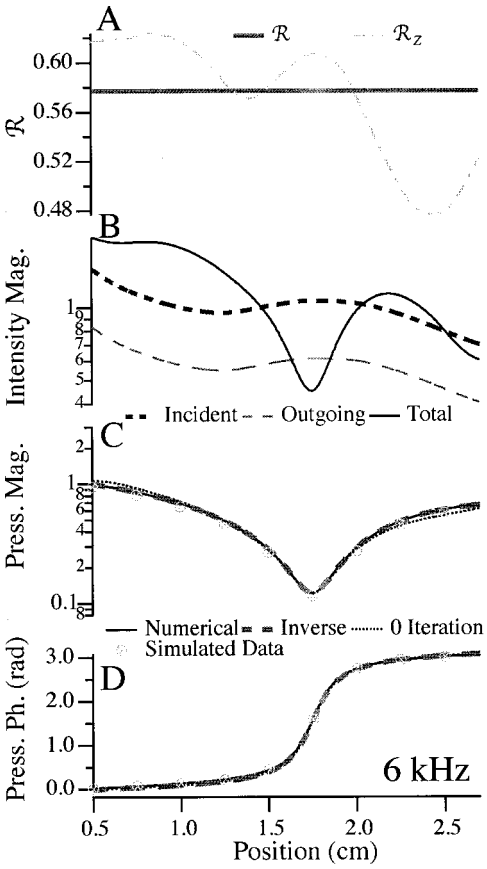


FIG. 6. Same as Fig. 5, except for 6 kHz.

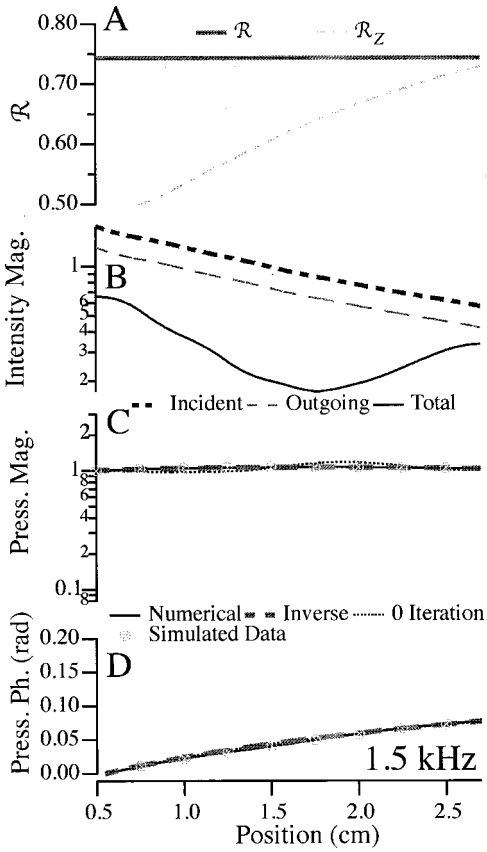


FIG. 7. Same as Fig. 5, except for 1.5 kHz.

solid curve]. Incident acoustic intensity is a true measure of the input to the ear (i.e., the stimulus) and provides the sound power of the incident wave per unit cross-sectional area.

Figures 6 and 7 show the same type of results as Fig. 5 but at lower frequencies. Aside from changes in wavelength, qualitatively similar results were obtained at all frequencies. It is clear by comparison of the numerical simulations to the inverse solutions that Eq. (6) provides an excellent approximation of the exact solution of the horn equation over the entire frequency range tested (1–15 kHz). This held true for all five model ear canal geometries tested. Although the simulated data shown in Figs. 5–7 were generated using the Webster horn equation, it is important to note that the same results could have been obtained using the 3D theory of acoustics or the CH equation using the appropriate area functions (see Appendix B). Therefore, we expect data in physical models and ear canals to resemble the simulated data shown in Figs. 5–7.

B. Results in physical models

To investigate the influence of ear canal curvature and changes in cross-sectional area, two physical models (horns) were fashioned from 8-mm nominal i.d. glass tubing to include a flared exit and a 12-cm length of uniform straight section at the entrance. The straight section allowed us to compare the present method to an ASTM standing wave method for determining reflection coefficients in tubes (see Sec. III). Horn "A" included a curved section and horn "B" included a bulged section. The curvature of horn A and the cross-sectional area of the bulged region of horn B are shown in Fig. 8, both as functions of position along the length of the tube. Also shown for comparison are one specific example and the average human ear-canal curvatures and cross-sectional areas based on Stinson and Lawton (1989). The change in cross-sectional area of horn B was representative of that measured by Stinson and Lawton (1989), while the curvature of horn A was within the range of human ear canals but below the average.

Figure 9 compares the reflection coefficient obtained by the present method ($\mathcal{R}$, solid) to the ASTM standing wave method ($\mathcal{R}_{sw}$, dashed; see Sec. III) in the straight section of a curved tube (right of the vertical dashed line) at two representative frequencies [(b), 9400 Hz; (c), 3430 Hz]. Results of the present method in the straight portion of the tube fell right on top of results using the standing wave method (see legend for specific $\mathcal{R}$ values), making it difficult to discern the difference between the dashed and solid lines in the figure. The present method was also applied to data collected in the curved section of the tube (left of the vertical dashed line) for comparison to the straight section. Horizontal lines (solid, left) show the optimized reflection coefficients (0.065 at 9400 Hz, 0.11 at 3430 Hz) obtained by the present method in the curved section. Identical results were obtained whether the probe was forced to follow the curved centerline of the tube or not, thus indicating that the method effectively accounted for sampling along a straight microphone trajectory rather than along the curved centerline of the horn. To estimate spatial dependence, regional reflection coefficients computed on the basis of ten adjacent data points surround-

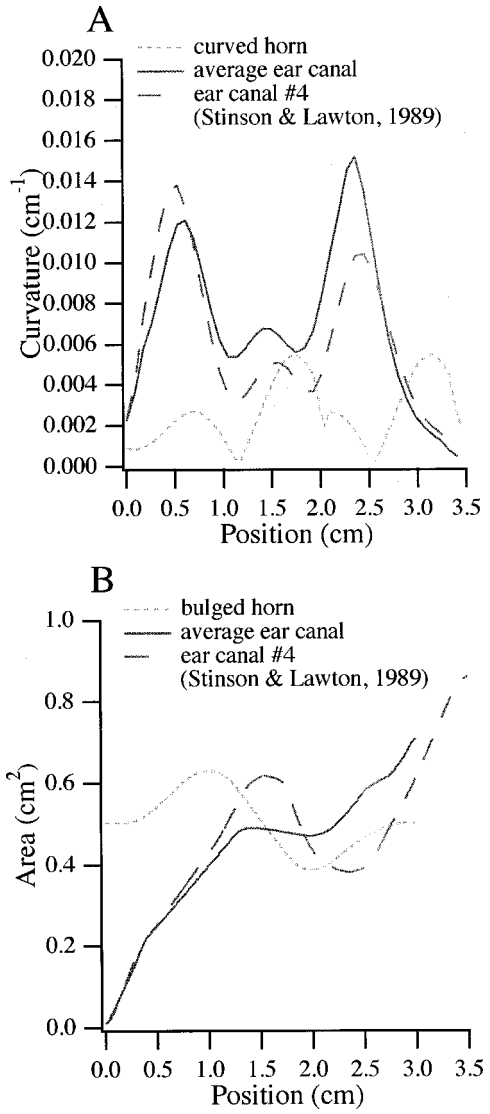


FIG. 8. Curvature and cross-sectional area of the physical models. (a) The curvature, i.e., 1/(radius of curvature), is plotted as a function of tube length for a curved horn (dotted curve) in which the intensities and reflection coefficient were calculated using the present method. The curvature of the model horn was about half that of an average ear canal (solid curve) and ear canal 4 (dashed curve) from Stinson and Lawton (1989). (b) The cross-sectional area of a bulged horn (dotted curve), in which the intensities and reflection coefficient were also determined by the present method, was comparable to ear canal 4 (dashed curve) from Stinson and Lawton (1989). The cross-sectional area of the average ear canal (solid curve), which exhibited smaller variations in cross-sectional area, is also shown for comparison.

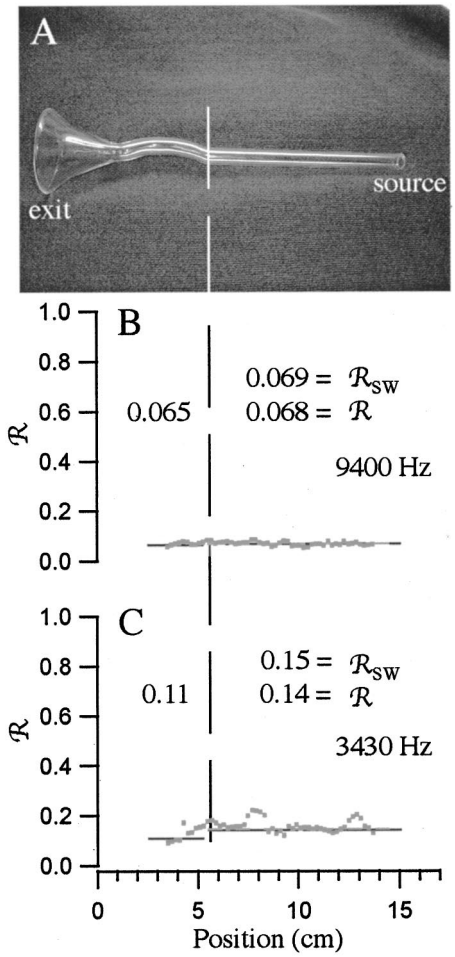


FIG. 9. Reflection coefficient in a curved horn. (a) Photo of a curved glass horn in which the reflection coefficient was calculated using the ASTM standing wave method and the present method. (b) and (c) In the straight portion of the tube (to the right of the dashed vertical line) reflection coefficients calculated from all the data using the ASTM standing wave method ($\mathcal{R}_{SW}$, dashed line) were indistinguishable from the "optimized" reflection coefficients calculated from all the data using the present method ($\mathcal{R}$, solid lines). In addition, reflection coefficients were calculated from every ten adjacent data points using the present method (dots), i.e., each dot represents $\mathcal{R}$ based on the surrounding 2-cm segment of the tube. Differences in $\mathcal{R}$ between the curved (to the left of the dashed vertical line) and straight portions of the tube were within measurement variability, indicating that the curvature did not cause local reflection or position sensitivity in $\mathcal{R}$ [(b) 9400 Hz $\Delta\mathcal{R}\sim 0$; (c) 3430 Hz $\Delta\mathcal{R}\sim 0.03$]. The reflection coefficients for 9400 and 3430 Hz are shown as representatives of the high and low frequency regions, respectively.

ing each spatial location are also shown [dots, (b) and (c)]. Each dot shows $\mathcal{R}$ calculated from the surrounding ten data points (18 mm) using the present method. It should be noted that the increased spread (vertical spread of dots) observed at low frequencies [Fig. 9(b) versus 9(c)] was associated with increased uncertainty in $\mathcal{R}$, as quantified in the parameter optimization procedure, and does not indicate actual position dependence (also see human results). Consistent with this, the standing wave reflection coefficient $\mathcal{R}_{sw}$ (Fig. 9, dashed) always fell within one standard deviation of $\mathcal{R}$ (dots) computed along the length of the tube. Curvature did not introduce any statistically significant spatial dependence in the reflection coefficient (95% confidence interval). Based on these results and the preceding theoretical analysis, one would not expect human ear canal curvature to cause any significant spatial variation in the power reflection coefficient $\mathcal{R}$.

In contrast to the curved horn, the bulged horn showed some evidence of partial traveling wave reflection that was manifested as position dependence in the reflection coefficient at low frequencies. This is illustrated in Fig. 10 at two example frequencies [(c) 9400 Hz; (d) 3430 Hz]. Horizontal lines compare results of the standing wave method ($\mathcal{R}_{sw}$, dashed) to the present method ($\mathcal{R}$, solid) on both the left and right sides of the bulge. As with the curved horn, the present method produced results nearly equivalent to the standing wave method in the straight regions of the tube (the curves overlay). Data at 3430 Hz illustrate a drop in the reflection

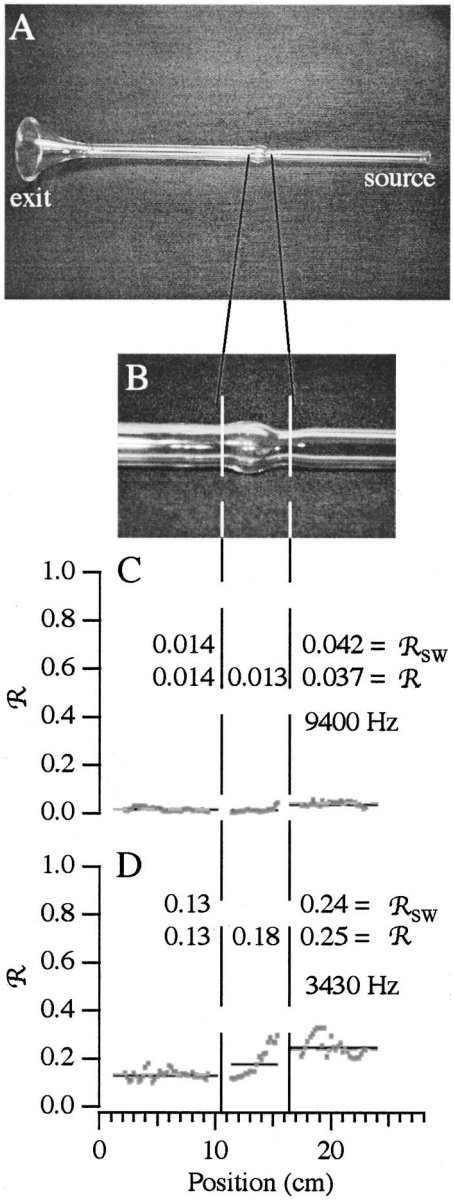


FIG. 10. Reflection coefficient in a bulged horn. (a) Photo of a bulged glass horn in which the reflection coefficient was calculated using the ASTM standing wave method and the present method. (b) Expanded photo of the bulged section of the horn shown between the two vertical dashed lines. (c) and (d) In the straight sections at both ends of the tube reflection coefficients calculated from all the data using the ASTM standing wave method ($\mathcal{R}_{SW}$, dashed line) were indistinguishable from the ''optimized'' reflection coefficients calculated from all the data using the present method ($\mathcal{R}$, solid lines). $\mathcal{R}$ was also calculated from every ten adjacent data points along the length of the horn using the present method (dots) such that each dot represents $\mathcal{R}$ based on the surrounding 2-cm segment of the horn. Influence of the nonuniform cross-sectional area is illustrated by the change in $\mathcal{R}$ occurring over the length of the bulge. Results indicate that the change in cross-sectional area introduced only slight position sensitivity in $\mathcal{R}$ at high frequencies [(c) example 9400 Hz $\Delta\mathcal{R}\sim 0.02$] but a more significant position sensitivity at low frequencies [(d), example 3430 Hz $\Delta\mathcal{R}\sim 0.12$].

coefficient on the exit side of the tube relative to the incident side. To further investigate spatial dependence, regional reflection coefficients were computed on the basis of ten adjacent data points surrounding each spatial location [dots, (c) and (d)]. Regional estimates illustrate a smooth transition through the region of changing cross-sectional area. Application of the present method to all data within the bulged region provided the ''optimized'' reflection coefficient in the region ($\mathcal{R}$, horizontal solid lines)—a value that fell very near to the average regional estimate (average dots versus line). These data indicate that the phase functions, Ψ, appearing in Eq. (7) had nonzero imaginary components in the region of the bulge associated with a local evanescent wave and partial reflection. This caused the magnitude of the reflection coefficient to vary with position through the bulged region. The change in reflection coefficient with position was $\Delta\mathcal{R}/\Delta X \sim 0.004$/mm at 9400 Hz and $\Delta\mathcal{R}/\Delta X \sim 0.02$/mm at 3430 Hz. This cross-sectional-area-induced position dependence, significant only at low frequencies, is expected to extend to human ear canals as described below.

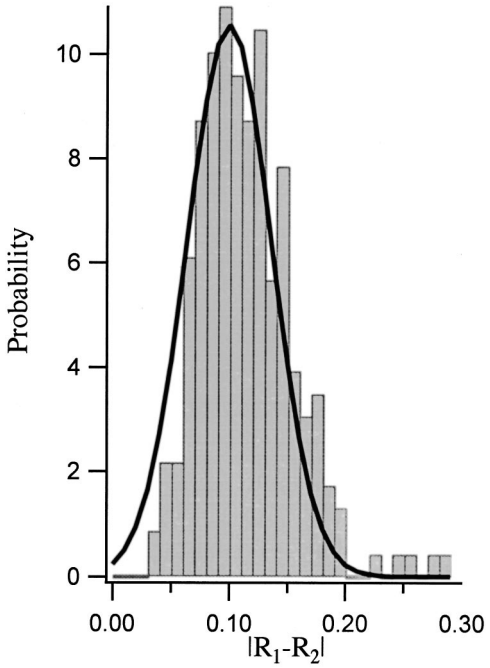


FIG. 11. Position dependence of $\mathcal{R}$ in human ear canals. Using the present method, the reflection coefficient in each ear canal tested was calculated using seven data points nearest the tympanic membrane ($\mathcal{R}_1$) and seven data points nearest the concha ($\mathcal{R}_2$). The centers of the two regions were 6 mm apart. A histogram of the differences calculated between the two regions ($\Delta\mathcal{R}=|\mathcal{R}_1-\mathcal{R}_2|$) for all six subjects and all frequencies tested is shown. The mean difference was $\sim$0.1 (1–15 kHz). This difference was most evident at low frequencies as was the case in the bulged horn results of Fig. 10.

C. Human results: Ear canal position dependence

Results in nonuniform bulged tubes showed a small, but significant, position dependence of the reflection coefficient that was most evident at low frequencies. Analysis of the CH equation indicated that this was due to local evanescent waves and reflections caused by the nonuniform geometry. To quantify the extent of position dependence of $\mathcal{R}$ in human ear canals, we applied the present method to data surrounding two different points—the first point was located at $X_1 \sim 6$ mm from the TM and the second located at $X_2 \sim 12$ mm from the TM ($\Delta X = 6$ mm). The difference in the power reflection coefficient ($\Delta\mathcal{R}=|\mathcal{R}_1-\mathcal{R}_2|$) was determined at each frequency and is reported for the population of human subjects as a probability density function in Fig. 11. The median difference in the power reflection coefficient (0.1)

divided by the distance between the measurement points was $\Delta\mathcal{R}/\Delta X \sim 0.017$/mm when summed over the entire frequency range (summary 7 subjects, 1–15 kHz). The median difference increased to $\Delta\mathcal{R}/\Delta X \sim 0.022$/mm when the analysis was restricted to stimulus frequencies below 3 kHz (summary 7 subjects, 1–3 kHz). This increase at low frequencies is consistent with predictions of the theory (above) and can be explained on the basis of local reflections associated with short evanescent regions.

In most human applications the question of interest is not ear-canal position dependence of the reflection coefficient but, rather, what is the dependence of $\mathcal{R}$ on the sound pressure level and the temporal/frequency properties of the stimulus waveform. For these types of applications it is advantageous to use a single, position-independent reflection coefficient that is optimized on the basis of the available ear canal pressure data. This corresponds to restricting the phase functions in the present method to be real-valued. Results reported above for simulated data demonstrate that this approach provides an excellent approximation for horn equations. We therefore applied the same approach using real-valued phase functions to determine optimum, position-independent reflection coefficients in human ear canals.

Figure 12 shows a human ear-canal example of the spatial pressure distribution [(c) amplitude, (d) phase], acoustic intensity distribution [(b)] and the power reflection coefficient [(a)] for the 13-kHz component of the stimulus. The spatial distributions recorded in human ear canals (Figs. 12–14) showed striking similarities to numerical solutions of the horn equation (Figs. 5–7; same presentation format). The raw pressure data are shown [(c, d), open circles] along with the zero iteration estimation obtained using $\Phi \approx \bar{\Phi}(S)$ [(c, d), dotted] and the optimized inverse solution [(c, d), dashed]. Consistent with the simulations, the method provided good agreement between the inverse solution and the human data even when the phase functions were assumed to be real valued. Using a real-valued phase function resulted in an optimized power reflection coefficient $\mathcal{R}$ that was independent of ear canal position [Fig. 12(a), solid line]. Also shown are plane-wave impedance-based reflection coefficients $\mathcal{R}_{Zs}$ [(a), dotted] and $\mathcal{R}_Z$ [(a), dot-dashed], both computed using Eq. (10). $\mathcal{R}_{Zs}$ was calculated directly from the spline-interpolated data while $\mathcal{R}_Z$ was calculated from the inverse pressure waveform [(c, d), dashed]. Because of this, $\mathcal{R}_Z$ was expected to be relatively close but not identical to $\mathcal{R}$, with differences due to nonuniform cross-sectional area effects in the amplitude function appearing in Eq. (6). Differences between $\mathcal{R}_Z$ and $\mathcal{R}_{Zs}$ include the additional influence of experimental uncertainties and using a real-valued phase function in the inverse solution. Figures 13 and 14 show the same type of results but for lower Fourier frequency components. It is notable that the plane-wave-based reflection coefficients ($\mathcal{R}_Z$, $\mathcal{R}_{Zs}$) varied significantly along the length of the ear canal, while a constant, position-independent, nonplanar reflection coefficient ($\mathcal{R}$) was found to be sufficient to describe the pressure data [also see Fig. 15(a) confidence intervals].

The traveling wave description can also be used to determine the components of acoustic intensity [total, incident and outgoing, Figs. 12(b), 13(b), and 14(b)]. Note that the

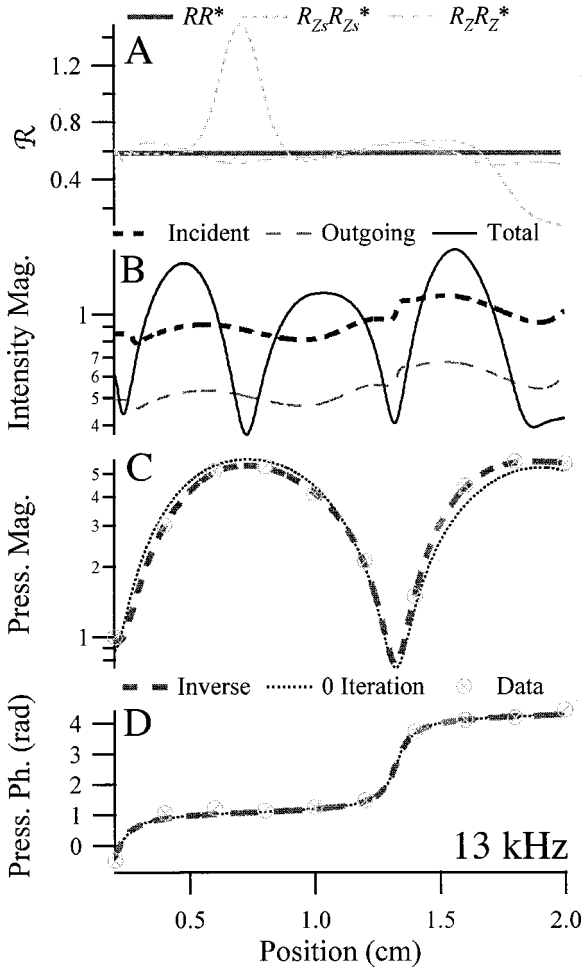


FIG. 12. Spatial distributions in a human ear canal. (a) The optimized reflection coefficient ($\mathcal{R}$, solid line) calculated using the present method was compared to reflection coefficients calculated using the impedance method with the raw spline-interpolated pressure data ($\mathcal{R}_{Zs}$, dotted curve) and with the parameter-optimized pressure data ($\mathcal{R}_Z$, dot-dashed curve). (b) The total (solid curve), incident (short dashed curve) and outgoing (long dashed curve) acoustic intensities are plotted as functions of position. The incident and outgoing intensities were relatively constant along the length of the canal and their ratio was constant. As in the simulated data, the total intensity exhibited standing waves. (c) and (d) The pressure magnitude and phase resulting from the parameter optimization method (dashed curve) and the measured pressure data (round symbols). The results of the zero-iteration in the parameter optimization procedure ($\Phi = \bar{\Phi}$) are also shown (dotted curve). Results shown for 13 kHz (Fig. 12) in human ear canals were qualitatively similar to those predicted by the horn equation [Eq. (3), see Figs. 5–7].

total intensity [(b), solid] exhibited pronounced standing waves at half the wavelength of the pressure standing waves [(c, d)]—as predicted by the one-dimensional wave theory and illustrated in the simulations. In contrast, the incident and outgoing intensities [(b), dashed] did not exhibit standing waves. Changes in the incident and outgoing intensities with ear-canal position were consistent with cross-sectional-area-dependent changes in the amplitude function. For the low frequency Fourier components, $\mathcal{R}_{Zs}$ and $\mathcal{R}_Z$ were closer to one another due to increasing signal to noise ratio, but still differed from the optimized value of $\mathcal{R}$.

D. Human results: Frequency dependence

Figure 15 shows the optimized reflection coefficient (solid) as a function of frequency for a single subject in the

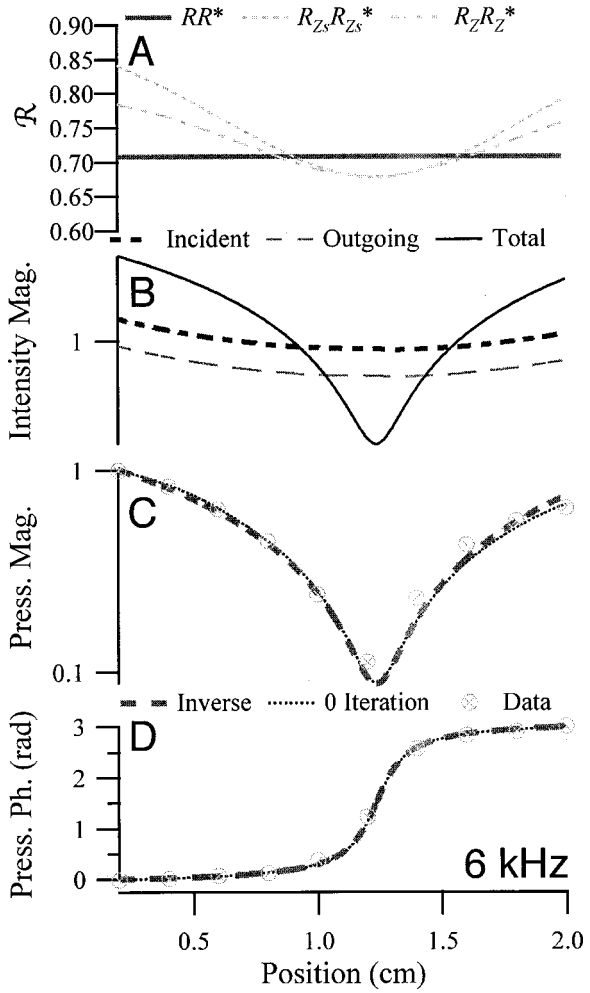


FIG. 13. Same as Fig. 12, except for 6 kHz.

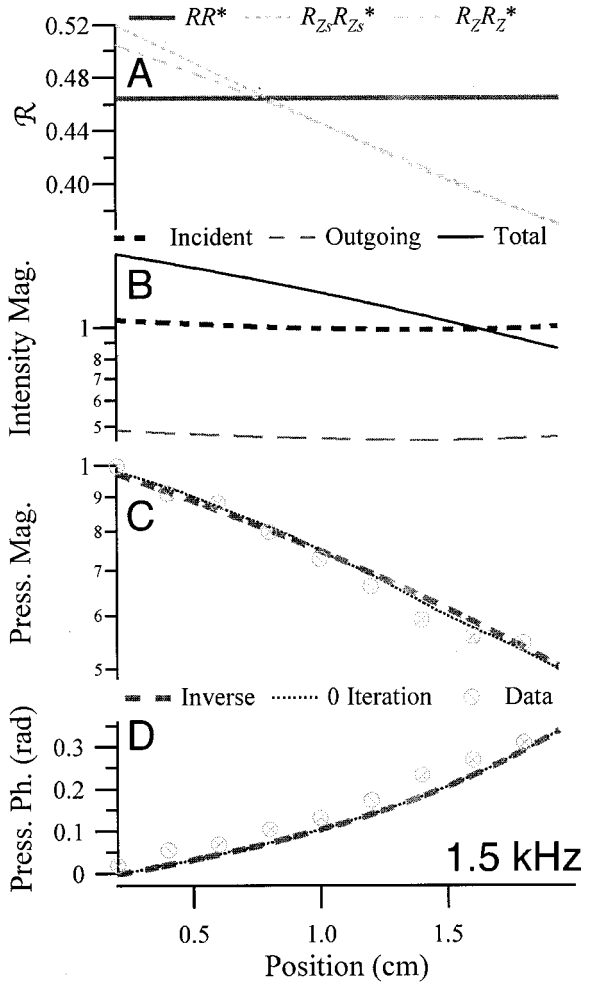


FIG. 14. Same as Fig. 12, except for 1.5 kHz.

form of magnitude [(b), power reflection coefficient $\mathcal{R}=RR^*$] and phase [(c), pressure reflection coefficient phase of R]. The top panel (a) provides a frequency-specific estimate of the 95% confidence interval of $\mathcal{R}$ (see Sec. III). Reflection coefficients computed on the basis of plane-wave impedance [Eq. (10)] were found to vary along the length of the ear canal; the total range is indicated in the figure as a shaded region for $\mathcal{R}_Z$ [(b, c), dark shaded region] and for $\mathcal{R}_{Zs}$ [(b, c), light shaded region]. The power reflection coefficient $\mathcal{R}$ generally fell within the range of $\mathcal{R}_Z$, which fell within the range of $\mathcal{R}_{Zs}$ [Fig. 15(b)]. As in Figs. 12(a), 13(a), and 14(a), $\mathcal{R}_{Zs}$, was calculated directly from the spline-interpolated data while $\mathcal{R}_Z$ was calculated from the inverse pressure waveform. Reflection coefficients estimated using plane-wave methods ($\mathcal{R}_{Zs}$ in particular) were found to exceed one for frequencies above 6 kHz at some locations in this subject's ear canal. However, the fact that $\mathcal{R}_{Zs}>1$ cannot be used to argue that the outgoing power exceeded the incident power due to the dubious plane-wave assumption and experimental uncertainties. In contrast to plane-wave estimates, the optimized power reflection coefficient, $\mathcal{R}$, based on nonplanar traveling waves did not exceed one under the conditions tested.

The phase of the pressure reflection coefficient R was relatively flat in the low frequencies, rolled off in the higher frequencies and was bounded by the phase of R_Z and R_{Zs} [Fig. 15(c)]. A transition in the phase where it changed from one relatively constant level to another relatively constant level occurred at a frequency near the minimum in the power reflection coefficient in all subjects (around 2.4 kHz in Fig. 15). This minimum is generally believed to be the fundamental resonance of the ear where effective inertial mass nearly balances the effective stiffness (Zwislocki, 1962; Rabbitt, 1988). Energy reflection at the resonance frequency would therefore be dominated by resistive dissipation corresponding to a minimum in the pressure reflection coefficient and a phase equal to zero. The method for computing phase contains an arbitrary constant that was adjusted in the present results to force the curves to pass through zero at the resonance frequency (between 2 and 3 kHz for the ear canals tested).

The optimized power reflection coefficient for each of the six subjects tested is shown in Fig. 16(a). In addition, the average power reflection coefficient is plotted as a function of frequency. In computing the averages, data points with a 95% confidence interval in $\mathcal{R}$ exceeding 0.1 were discarded as outliers ($n<0.1\%$ of total data points). Intersubject averaging brings out the common features at the expense of fine frequency structure present in individual ears. Some of the common features include an average ear canal resonance at

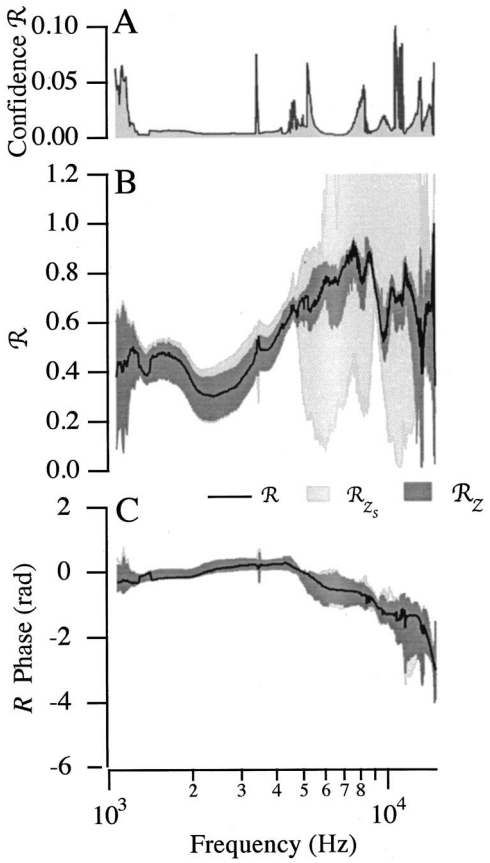


FIG. 15. $\mathcal{R}$ vs frequency: Single subject. The optimized reflection coefficient is shown as a function of frequency for a single subject in the form of magnitude $\mathcal{R}$ (B, solid) and phase of R (C, solid) for a representative human subject. (a) The 95% confidence interval for the power reflection coefficient $\mathcal{R}$. In other words, with 95% confidence $\mathcal{R}$ is the value plotted in (b) $\pm$ the "confidence $\mathcal{R}$" value. (b) Results are compared to the plane-wave impedance approximation [Eq. (10)], which resulted in position-dependent results. Two plane-wave reflection coefficients are shown: $\mathcal{R}_{Z_S}$ (light gray region) computed from the raw spline-interpolated pressure data, and $\mathcal{R}_Z$ (dark gray region) computed from the parameter-optimized inverse traveling wave pressure distribution. (c) The phase of the optimized pressure reflection coefficient R (solid) is shown relative to the range of phases obtained using the plane-wave impedance approximation, R_Z (dark gray region) and R_{Z_S} (light gray region). All phase results were shifted by a constant to set the phase at the resonance frequency equal to zero and reported relative to the fixed reference position $S_o = 2$ mm.

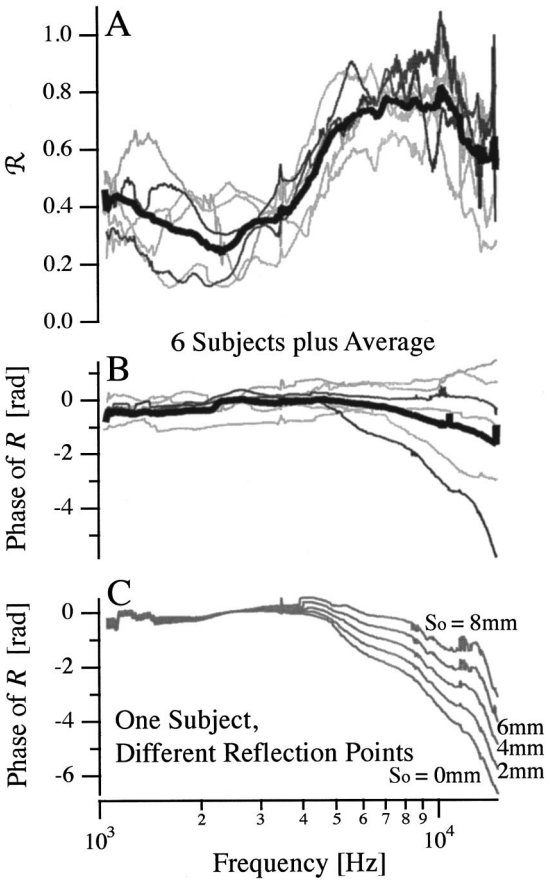


FIG. 16. $\mathcal{R}$ vs frequency: Population average. Notation follows Figs. 15(b) and (c). (a) $\mathcal{R}$ for each subject (thin curves) and the average $\mathcal{R}$ (thick curve) are plotted as functions of frequency. (b) The phase of the pressure reflection coefficient, R, for each subject (thin curves) and the average phase of R (thick curve) reported relative to the reference position $S_o = 2$ mm are plotted as functions of frequency. (c) This family of curves shows the influence on the phase of R of changing the reference location S_o (present methods) for a single subject. Note that the high-frequency phase flattens out as the reflection location moves medial to the first measurement location (S_o).

~ 2.5 kHz as indicated by the dip in the magnitude of $\mathcal{R}$ to ~ 0.28. At 1 kHz the average magnitude of $\mathcal{R}$ was ~ 0.4 and averaged ~ 0.7 at the upper end of the frequency range tested (15 kHz). Intersubject variability in R was relatively even across all frequencies (± 0.2), with slightly less variation in the mid frequencies (3–7 kHz).

The average pressure reflection coefficient phase is shown in Fig. 16(b). In each ear canal, the phase of R was relatively flat with a slight transition near the resonance frequency and a roll-off in the high frequencies.

The amount of high-frequency phase roll-off was strongly influenced by the location of the estimated reflection point (S_o), as shown in Fig. 16(c). Each phase curve in Fig. 16(c) was calculated with the same data using a different reflection point location. The top curve is for a reflection point 8 mm medial to the initial probe tip location ($S_o = 8$ mm). This would occur if the probe tip were 8 mm lateral to the TM at the initial measurement location or if the first measurement was near the TM and the actual reflection point was somewhere medial to the TM. The bottom curve is for $S_o = 0$ mm, i.e., the reflection point *at* the initial probe tip location. The initial probe tip location was typically within 2 mm of the tympanic membrane, corresponding to $S_o = 2$ mm. For this scenario, a reflection point 8 mm from the initial probe tip location would be well within the middle ear cavity. The phases used for analysis and plotted in Fig. 16(b) were calculated assuming $S_o = 2$ mm (visual estimate of first point distance from TM).

V. DISCUSSION

A. Generalized traveling-wave theory

The probable existence of turning points and evanescent regions in human ear canals led us to question the extent to which the planar traveling-wave theory can be applied. To address this question from a theoretical standpoint we compared a generalized traveling-wave approximation, which accounts for nonplanar waves, to numerical solutions of the horn equation in real ear canal geometries. Results

showed that the generalized traveling-wave equation [Eq. (6)] provides an excellent approximation for the horn equation in human ear canal geometries over the entire frequency range investigated (1–15 kHz). This held true even when the phase functions were restricted to be real valued. The fact that real-valued phase functions were sufficient indicated that evanescent zones in the ear canal may be sufficiently short to allow a majority of the energy to pass through with relatively small local reflections. In other words, the evanescent zones may primarily ''stretch'' the traveling waves by slowing or quickening changes in the real part of the phase function. The present analysis approximates stretching of waves in evanescent zones by allowing the effective wavenumber $\hat{\kappa}$ to differ from the acoustic wavenumber κ. This also accounts, in part, for the fact that the microphone trajectory in the present experiments did not follow the curved lengthwise coordinate of the canal.

It is notable that both the Webster and CH equations lead to Eq. (3). The only difference between the CH and Webster model, in this respect, involves the relationship between the area functions and the estimated amplitude and phase functions. This relationship is not needed to estimate the components of the acoustic intensity or the reflection coefficient and was not used in the present data analysis method.

B. Position dependence of the reflection coefficient

Huang *et al.* (2000) previously reported position dependence of the reflection coefficient in ear canals of domestic cats—a dependence that was particularly pronounced in some animals at discrete frequencies below 1 kHz. Consistent with our results, viscous dissipation was ruled out as a possible explanation. They suggested that a compliant ear canal wall or geometrical effects might be responsible. Present data and theoretical analysis indicate that the geometrical hypothesis is relevant to humans and may be particularly acute below ~ 3 kHz. Present data and analysis indicate that position dependence of the reflection coefficient, at least in humans, can be explained in terms of local evanescent regions that induce partial reflection of traveling waves. The existence of evanescent regions also explains why the acoustic wavenumber deviates from the optimum effective wavenumber ($\hat{\kappa}$) in ear canals (Stinson and Lawton, 1989; Stinson and Khanna, 1989). Evanescent regions, and therefore local reflections, are present when the frequency is near or below ω_{cr}, which occurs by and large at low frequencies in human ear canal geometries (Fig. 1). This explains why position dependence of the reflection coefficient in human ear canals became more pronounced as the frequency was reduced. Viscous dissipation would cause the opposite frequency dependence and therefore is inconsistent with the present data. Supporting this, we were unable to detect any significant dissipation in uniform tubes or in curved horns with cross-sectional areas and lengths mimicking typical human ear canals.

It is advantageous in practical applications to define a characteristic reflection coefficient that represents the behavior along the entire length of the ear canal. For this we applied a numerical procedure to search for a single *optimum* reflection coefficient at each frequency. Ear canal pressure distributions estimated by this method provided excellent fits to the data—both simulated (Figs. 5–7) and in human ear canals (Figs. 12–14). Frequency-specific 95% confidence intervals indicated $\mathcal{R}$ values were typically optimized to within 0.05 [Fig. 15(a)]. Since the method provides an estimate of the variance of $\mathcal{R}$, it allows for some quality control in the measurement of $\mathcal{R}$.

C. Reflection coefficient in human ear canals

The magnitude and phase of the average reflection coefficient for the current population are plotted in Fig. 16. The minimum in the magnitude of $\mathcal{R}$ and transition in the phase corresponds to a resonance that increases sound power transmission efficiency near 2.5 kHz (Fig. 16). The phase of $R(S_o)$ (R evaluated at $S=S_o$) was less than zero at low frequencies (below the resonance). A phase in the direction of $-\pi$ indicated that low-frequency ear-canal termination impedance was dominated by compliance, consistent with the literature (Lynch *et al.*, 1994; Keefe *et al.*, 1993; Voss and Allen, 1994). At high frequencies one might expect the ear-canal termination impedance to become mass dominated causing the phase of $\mathcal{R}$ to increase. Present average data, however, showed a relatively flat phase with a slight roll-off at high frequencies. These data are in reasonable agreement with Lynch *et al.* (1994) and Stinson and Khanna (1989), who suggested a resistive termination impedance at high frequencies.

One factor complicating the interpretation of reflection phase data involves the reference reflection point—a point typically taken to be at a fixed location in the ear canal at or near the tympanic membrane. The computed phase of R varies depending upon selection of the reference reflection point as shown in Fig. 16(c). As the reference point is moved deeper into the ear canal, the phase roll-off decreases, causing the phase reflection curve to flatten at high frequencies. This ambiguous reflection point is implicit in the phase of R at high frequencies and may be the source of disagreement regarding resistive versus mass termination behavior of the input impedance in human ear canals. Given the complex vibratory pattern of the tympanic membrane, acoustics of the middle ear and coupling to the cochlea, it is entirely possible that the ''reflection point'' may move with frequency. Existence of a fixed reference point is one of the shortcomings of lumped-parameter network models of the ear that may render such models inadequate for capturing the frequency dependence of ear-canal reflectance data.

D. Comparison to results using alternative methods

Figure 17 compares the average power reflection coefficient for the present population of human subjects to results of five previous studies (Stinson, 1990; Rabbitt and Dragicevic, 1991; Keefe *et al.*, 1993; Voss and Allen, 1994; Burns *et al.*, 1998). Reflection coefficients reported by all of these authors fall near to the present results. It is interesting to note that, like the present study, data reported by Lawton and Stinson (1986, analyzed in Stinson, 1990) and Rabbitt and Dragicevic (1991) were collected in open ear canals. In both cases, the standing wave analysis method used to ex-

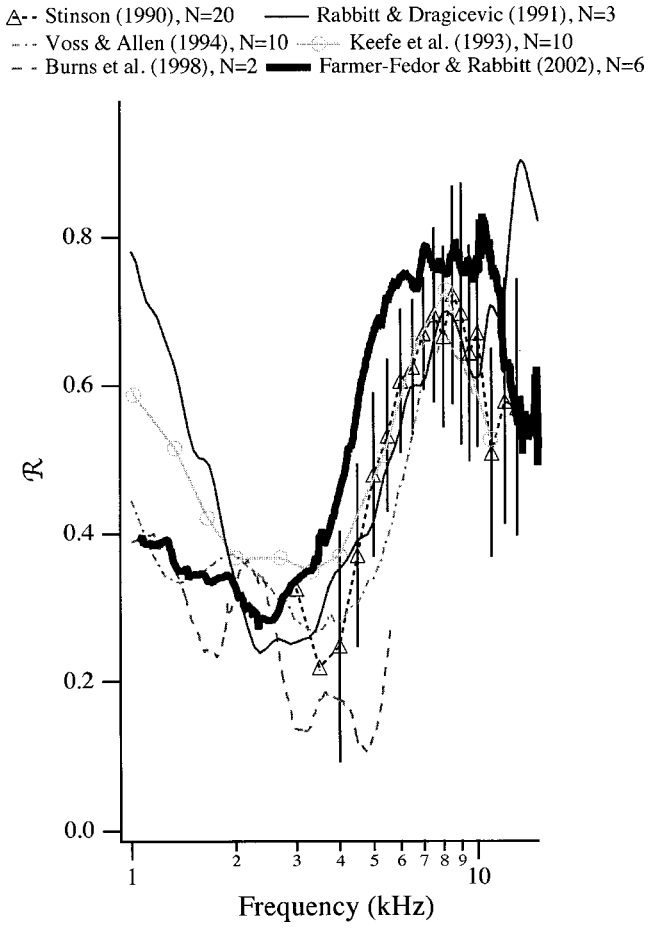


FIG. 17. $\mathcal{R}$ vs frequency as reported in the literature. The power reflection coefficients, $\mathcal{R}$, as a function of frequency reported by several investigators are presented for comparison.

tract $\mathcal{R}$ from their human data was actually a simplified version of the present approach [set Φ and $\hat{\kappa}$ in Eq. (11) to be constants and apply a simplified curve-fitting routine]. Therefore, it is not surprising that the range of present results overlap with the range of normal subjects reported by Stinson (1990) and that Rabbitt and Dragicevic (1991) provide a similar average $\mathcal{R}$ as well.

Comparisons to acoustic impedance-derived reflection coefficients (Keefe *et al.*, 1993; Voss and Allen, 1994; Burns *et al.*, 1998) are somewhat more complex for two fundamental reasons. First, typical acoustic impedance measurement systems record the pressure ($\tilde{P}$) and the acoustic volume velocity ($\tilde{Q}$). To determine the specific impedance it is necessary to convert the volume velocity to a particle velocity $\tilde{V}$ using an appropriate reference area (A_R; $\tilde{V}=\tilde{Q}/A_R$). In some ear canal measurements, A_R was the area at the measurement point and in others it was based on a population average or a subject-specific estimate (Voss and Allen, 1994; Lynch *et al.*, 1994; Huang *et al.*, 2000; Keefe *et al.*, 1993). Determination of the appropriate reference area to account for the complex area near the probe insertion and render the plane wave formula correct is in itself a significant challenge (Agulló *et al.*, 1988). This difficulty is completely avoided in the present method in that a reference area is not needed. Second, the acoustic impedance methods require that the ear canal be sealed during the measurement process. A sealed ear canal has a vastly different exit impedance than an open ear canal. Changes in the ear canal exit impedance can influence both the load on any active emissions and the pressure field (e.g., location of evanescent regions) within the canal. The potential influence of ear canal exit impedance on the reflection coefficient is discussed below. Both of these characteristics lead to differences between $\mathcal{R}_Z$ and $\mathcal{R}$ in an individual ear canal. Despite individual canal differences, the population averages of $\mathcal{R}_Z$ reported by Keefe *et al.* (1993), Voss and Allen (1994) and Burns *et al.* (1998) were similar to the population average of $\mathcal{R}$ obtained in the present analysis, particularly above 3 kHz.

Although there were differences in measurement conditions and exit impedance, current results were similar to previous reports in the overall variation of $\mathcal{R}$ with frequency. Most of the curves plotted in Fig. 17 show a decrease in $\mathcal{R}$ from 1 kHz to a minimum at the resonance frequency between 2 and 4 kHz. Owing to individual variations in ear responses, it is not surprising that each population average exhibits a resonance at a different frequency. From the resonant frequency, most of the curves increased to a peak near 7–8 kHz, then $\mathcal{R}$ decreased again at higher frequencies. Differences between reports are reasonable given established intersubject variations and differences in the measurement methods employed. One must also bear in mind that the mammalian ear is nonlinear; thus, its response and $\mathcal{R}$ depend upon the input. Because of this, reflection coefficient results are not expected to extend unmodified to other temporal waveforms, sound pressure levels, or populations.

E. Exit impedance of the ear canal

In development of the measurement apparatus, its effect on the exit reflection coefficient ($\mathcal{R}_{\text{exit}}$) of the ear canal was evaluated (i.e., the impedance looking out of the ear canal toward the pinna). $\mathcal{R}_{\text{exit}}$ was estimated from pressure data collected in an 8.5-mm-diam straight tube terminated by a KEMAR pinna fitted with the modified otoscope. The model ear exit reflection coefficient with the measurement apparatus in place was relatively low and averaged $\mathcal{R}_{\text{exit}} \sim 0.22$ ($\sigma = 0.18$, 1–15 kHz). This value was comparable to the exit reflection coefficient estimated from pressure data collected in the same tube terminated by the KEMAR pinna only. Some studies have shown that placing equipment in or on the ear will change the exit impedance of the ear (Zwicker, 1990; Vorländer, 2000). Our results indicated that placement of the modified otoscope in the concha maintained a relatively low ear canal exit impedance.

Impedance measuring systems that seal the ear canal impose an artificial boundary condition at the location of the probe. The impedance change caused by the probe would have quite different effects depending upon location of the probe relative to evanescent regions. Simply moving the probe would alter ear canal pressure patterns and local reflections caused by evanescent regions. This would cause the ''apparent'' reflection coefficient to change. Sealed probes also alter the load on any active emissions. The present experimental approach avoids sealing the ear canal and is designed to accommodate evanescent regions and emissions.

This methodological difference offers an alternative interpretation of the reflection coefficient position dependence and may also account for some of the quantitative differences between the present data and previous reports, particularly at low frequencies.

Additionally, the locations of peaks in I_{tot} standing waves and pressure standing waves depend upon ear-canal boundary conditions. In the present experiments the magnitude and location of the open exit boundary condition remained fixed even when $\mathcal{R}$ was measured in different ear-canal locations. This contrasts sealed systems, which introduce an impedance discontinuity at the site of the sealed probe and, thereby, alter the exit boundary condition and the standing wave patterns as the probe is moved. Therefore, the relatively small ear-canal position dependence reported here was not subject to potential artifact introduced by moving a sealed probe.

F. Acoustic intensity

A primary goal of the present work was to develop a method capable of measuring the incident acoustic intensity and the outgoing acoustic intensity in the nonuniform human ear canal. These components of intensity are distinct from the total intensity—a value dependent on both the incident and outgoing waves. Results in Figs. 5(b), 6(b), 7(b), 12(b), 13(b), and 14(b) show I_{tot} standing waves in the ear canal with a spatial wavelength one-half that of the corresponding pressure standing wave. It is easily validated that I_{tot} standing waves exist in uniform tubes and are caused by interference of waves traveling in two directions. For example, if the incident and outgoing waves were to have exactly the same magnitude, then I_{tot}, the product of pressure and particle velocity, would have a null at the location of the pressure null and a second null at the location of the pressure peak where the particle velocity is zero. Therefore, instantaneous I_{tot} measured at any individual point in the ear canal is not a good indicator of input to the ear. This can be partially rectified for periodic signals by using a time average intensity as noted below.

In contrast to I_{tot}, the incident acoustic intensity, I_+, does not form standing waves in the ear canal. It measures the sound power per unit ear canal cross-sectional area traveling toward the tympanic membrane (TM). The incident intensity tends to be greater near the TM than in the rest of the canal [Figs. 12(b), 13(b), and 14(b)]. Therefore, determination of the input to the ear using I_+ is somewhat dependent on the measurement location, but not nearly to the extent of ear canal pressure P or total intensity I_{tot}. In addition, measurement of I_+ is not sensitive to the acoustic load or to the ear canal termination impedance. Unlike P or I_{tot}, the incident intensity I_+ is a valid measure of the stimulus input to the ear over the entire auditory frequency range. Therefore, measurement of I_+ is important for experimental determination of acoustic stimulus in both clinical and research applications. For example, I_+ could be used to extend the range of audiometric tests to high frequencies (>8 kHz) where traditional measures sometimes prove inadequate. Presently, high-frequency tests encounter problems with circumaural and supra-aural earphone calibration at high frequencies and subtle changes in ear canal resonance effects (Shaw, 1966). Insert microphones and receivers encounter problems with placement and differences between individual ears. Incident acoustic intensity has the potential to avoid these difficulties.

Similar to I_+, the outgoing acoustic intensity, I_-, does not form standing waves in the ear canal. The outgoing intensity is a combination of passive reflected sound and any active otoacoustic emissions originating in the cochlea. Separating the active component of the outgoing wave has proven useful in assessment of cochlear function and defines an important area of future investigation using traveling-wave methods.

Present results provide the instantaneous acoustic intensities I_+, I_- and I_{tot}. For periodic signals, time-average acoustic intensities, $\overline{I_+}$, $\overline{I_-}$ and $\overline{I_{tot}}$, can be computed directly from the instantaneous values (Fahy, 1995). The time-average total intensity does not exhibit standing waves like the instantaneous value and therefore could be used as a reasonable measure of the input to the ear (Neely and Gorga, 1998). The time-averaged total intensity, however, depends on both the incident wave and the outgoing wave and therefore implicitly includes the reflection coefficient. In addition, the time-averaged method does not extend directly to nonperiodic signals in the time domain. In contrast, instantaneous acoustic intensity can easily be applied in the frequency domain or the time domain. For these reasons we have focused on instantaneous acoustic intensity in the present report.

G. Current method limitations

Although the nonplanar traveling-wave approach has advantages in clearly defining the stimulus input to the ear, the output of the ear and a consistent reflection coefficient, its practical application can be challenging. The present experiments used a single probe microphone to take pressure measurements at ten different locations in the ear canal. Moving the microphone and time averaging the signals at each position proved to be quite time consuming and required at least 30 min for each subject. It was often difficult for the subjects to sit perfectly still for such a long experiment. Slight movements most certainly introduced noise. Using a single microphone also assumes reliability of the response, i.e., that statistics of the sound field within the ear canal were identical for each individual presentation of the stimulus. In future studies it may be possible to eliminate these shortcomings using multiple microphones to measure pressure simultaneously at several points. Such a method, however, would require all the microphones to be amplitude and phase calibrated to a high degree of precision. Another limitation of the method was due to transmission losses through the sides of the silicone probe tubes. This loss was the primary limiting factor that restricted our results to frequencies below 15 kHz. The spatial resolution was also limited. Results were computed on the basis of data collected over ~ 2 cm of the canal and therefore summary data represent optimized values over this interval. Finally, the present experiments did not directly control the incident intensity—we postcalculated intensity from the pressure data. It may be possible to directly control

the incident intensity in future work using a real-time feedback system or iterative stimulus presentations to tune the incident waveform. Control of the incident intensity would facilitate future exploration of the nonlinear input–output behavior of the ear.

ACKNOWLEDGMENTS

Partial support for this work was contributed by the Whitaker Foundation and the NIDCD (Grant Nos. R01 DC04928 and P01 DC01837). We would like to thank M. Simons-McCandless and the University of Utah Audiology Clinic for assistance in performing the audiometric examinations. Many thanks to Janice Kyle from the University of Utah Chemistry Department who fabricated the nonuniform horns for us. We would also like to thank Dr. J. Rosowski and Dr. D. Keefe for their insightful comments and suggestions on an earlier version of the manuscript.

APPENDIX A: WKB APPROXIMATION

The CH equation [Eq. (1)] is converted to nondimensional form by introducing $s=S/\ell$, and $p=\tilde{P}/\tilde{P}_o$ to obtain

$$\varepsilon^2\frac{d^2p}{ds^2}+\lambda^2(s)p=0, \tag{A1}$$

where S is the curved centerline coordinate, ℓ is the total length, $\tilde{P}_o$ is a nominal pressure, the parameter $\varepsilon=(c/\ell\omega)$, c is the speed of sound, and the variable coefficient is defined as

$$\lambda^2(s)=\left(\frac{A_C}{A_I}+\left(\frac{1}{2\kappa A_I}\frac{dA_I}{dS}\right)^2-\frac{1}{2\kappa^2A_I}\frac{d^2A_I}{dS^2}\right). \tag{A2}$$

The traveling wave solution of the CH equation provided in the text was prompted by results from the following WKB expansion.

1. WKB outer solution ($\lambda^2\neq 0$)

At high frequencies the characteristic wavelength is small relative to the length of the ear canal such that ε becomes a small parameter. This observation motivates seeking a solution of Eq. (A1) in the WKB form. Following Holmes (1995) we introduce the expansion

$$p=\{\varphi_0(s)+\varepsilon\varphi_1(s)+\varepsilon^2\varphi_2(s)+\cdots\}e^{i\psi(s)/\varepsilon}. \tag{A3}$$

Substitution into Eq. (A1) and taking the limit as $\varepsilon\to 0$ provides the O(1) eikonal equation

$$\left(\lambda^2-\left(\frac{d\psi}{ds}\right)^2\right)\phi_0=0, \tag{A4}$$

which requires the phase function to satisfy

$$\psi_\pm(s)=\int_0^s\pm\lambda(\xi)d\xi. \tag{A5}$$

The positive $(+)$ root corresponds to the phase of the incident wave traveling toward the TM and the negative $(-)$ root corresponds to the phase of the outgoing wave. Having satisfied Eq. (A4) by selection of the phase function, we take the limit as $\varepsilon\to 0$ a second time to obtain the O(ε) transport equation

$$\left(\lambda^2-\left(\frac{d\psi}{ds}\right)^2\right)\varphi_1=i\left(2\frac{d\psi}{ds}\frac{d\varphi_0}{ds}-\frac{d^2\psi}{ds^2}\varphi_0\right), \tag{A6}$$

which is satisfied only if the right hand side is zero. Integrating this gives

$$\varphi(s)\propto\sqrt{\frac{\lambda(s_o)}{\lambda(s)}}. \tag{A7}$$

Combining terms and returning to dimensional variables gives

$$P^{\mathrm{WKB}}(S,T)=\{P_+\Phi_+(S)e^{i(\omega T+\mathrm{Re}(\Psi_+(S)))}+P_-\Phi_-(S)e^{i(\omega T+\mathrm{Re}(\Psi_-(S)))}\}, \tag{A8}$$

where P_+ and P_- are constants giving the pressure of the incident and outgoing acoustic waves evaluated at the reference point $S=S_o$, taken to be a point near the TM. The function determining the amplitude of the waves along the length of the canal is

$$\Phi_\pm(S)=\sqrt{\frac{\lambda_o}{\lambda_\pm(S)}\frac{A_{Io}}{A_I(S)}}\exp\{-\mathrm{Im}(\Psi_\pm(S))\}, \tag{A9}$$

where $A_{Io}=A_I(S_o)$ and $\lambda_o=\lambda(S_o)$ are constants.

For the viscous case $\kappa_\pm$ becomes complex-valued with the imaginary part differing in sign for the incident and outgoing waves. For the inviscid case where $\kappa_+=\kappa_-=\kappa$, the phase functions for the incident and outgoing waves differ only in sign, $\Psi_+=-\Psi_-$.

The outer WKB solution given by Eqs. (A8) and (A9) is not valid near turning points where $\lambda=0$. In this region a transition layer is introduced as described below.

2. The transition layer ($\lambda^2\approx 0$)

The WKB expansion carried out above becomes singular near turning points where $\lambda^2=0$. This defines points in the canal where traveling and evanescent waves meet. To determine the behavior in this region of the canal we introduce a new coordinate x,

$$x=(s-s_o)/\varepsilon^n, \tag{A10}$$

where s_o is the location in the canal where $\lambda^2=0$ and $n=2$ is a scale factor introduced to balance terms in Eq. (A1). Substitution into Eq. (A1) provides

$$\frac{d^2p^{TL}}{dx^2}+\lambda^2p^{TL}=0, \tag{A11}$$

where the "TL" denotes the transformed pressure in the transition layer. In this region we assume a solution of the form

$$p_\pm^{TL}=\phi_\pm^{TL}(x)e^{i(\psi_\pm^{TL}/\varepsilon)}, \tag{A12}$$

where the phase function is from Eq. (A5) evaluated at the turning point, $\psi_\pm^{TL}=\psi_\pm(s_o)$. Expanding λ^2 in a Taylor's series around $x=0$ gives Airy's equation

$$\frac{d^2p^{TL}}{dx^2}-\gamma^3xp^{TL}=0, \tag{A13}$$

where $\gamma^3 = -\,d\lambda^2/dx\,|_{x=0}$. This assumes a first-order turning point (i.e., $\gamma^3 \neq 0$). The solution, written in dimensional form, is

$$P_{\pm}^{TL}(S,T) = \{a_{\pm}\mathrm{Ai}(\Gamma(S-S_o)) + b_{\pm}\mathrm{Bi}(\Gamma(S-S_o))\} \times \sqrt{\frac{A_{Io}}{A_I(S)}}\, e^{i(\omega T + \mathrm{Re}[\Psi_{\pm}(S_o)])}, \tag{A14}$$

where Γ is the dimensional version of γ. Ai and Bi are Airy functions of the first and second kind, respectively, and a and b are constants obtained in terms of P_+ and P_- by matching (Holmes, 1995).

APPENDIX B: CH HORN EQUATION

The CH equation [Eq. (1)] governs the lengthwise (S) dependence of the pressure distribution in curved horns. It can be derived directly from the 3D theory of linear acoustics by introduction of a curvilinear coordinate system (ξ_1,ξ_2,ξ_3) with metrics (h_1,h_2,h_3) (Agulló *et al.*, 1999; Keefe and Barjau, 1999). In the present work we denote the lengthwise coordinate in the direction of wave propagation as $S=\xi_1$ and the orthogonal equipotential coordinates as (ξ_2,ξ_3). With this, the area functions appearing in the CH equation associated with compliance and inertance are

$$A_C(S) = A_1\langle h_1\rangle \tag{B1}$$

and

$$A_I(S) = A_1\langle 1/h_1\rangle, \tag{B2}$$

where the inner product is

$$\langle f\rangle = \frac{1}{A_1(S)}\int\int_{A_1} f h_2 h_3 d\xi_2 d\xi_3. \tag{B3}$$

Specific forms of the area functions are provided for axisymmetric horns by Agulló *et al.* (1999) and Keefe and Barjau (1999).

Note that the Webster horn equation and the CH horn equation generate identical results [identical Λ in Eq. (3)] if and only if the "*Webster area function, A_W,*" is selected using

$$\left(\frac{1}{2A_W}\frac{dA_W}{dS}\right)^2 - \frac{1}{2A_W}\frac{d^2A_W}{dS^2} = \left(\frac{A_C}{A_I}-1\right)\kappa^2 + \left(\frac{1}{2A_I}\frac{dA_I}{dS}\right)^2 - \frac{1}{2A_I}\frac{d^2A_I}{dS^2}. \tag{B4}$$

In general, $A_W(S,\kappa)$ is a function of $A_I(S)$ and $A_C(S)$, varies with frequency, and is distinct from the horn cross-sectional area $A(S)$. Hence, Eq. (15) in the text provides an estimate of the generalized Webster area function A_W and not the actual cross-sectional area function of the ear canal. Equation (B3) further shows that the area function that should be used in the Webster horn equation to predict the true pressure distribution differs from the physical cross-sectional area function.

APPENDIX C: INFLUENCE OF VISCOSITY

The 1D horn equation can be extended to include the effects of viscous dissipation by using the complex-valued wave number $\hat{\kappa} = \kappa(1 \pm i\zeta)$. The $(+)$ root corresponds to the outgoing wave and the $(-)$ root corresponds to the incident wave such that each wave dissipates energy as it propagates. The WKB solution given by Eqs. (A8) and (A9) remains valid for this case provided we use the complex-valued $\hat{\kappa}$ in computing $\lambda(S)$. In this case, for $\lambda(s) \approx \lambda_o$, the amplitude function can be reduced to

$$\Phi_{\pm}(S) \approx \sqrt{\frac{A(0)}{A(S)}}\exp\left\{\frac{\mp\kappa\hat{\lambda}\mu}{\rho c^2}\int_0^S \frac{1}{A}dS\right\}, \tag{C1}$$

where μ is viscosity, ρ is density, c is the speed of sound and $\hat{\lambda}(\omega)$ is a velocity profile factor that increases with frequency. Using this result, and assuming propagating waves, the pressure reflection coefficient is approximately

$$R \approx \frac{P_+ e^{i(-\mathrm{Re}[\Psi_+(S)])}}{P_- e^{i(-\mathrm{Re}[\Psi_-(S)])}}\exp\left\{\frac{-2\kappa\hat{\lambda}\mu}{\rho c^2}\int_0^S \frac{1}{A}dS\right\}. \tag{C2}$$

Appearance of the exponential term causes the magnitude of R to vary with spatial location. Viscous dissipation is predicted to cause a slight decrease in R as the distance from the tympanic membrane is increased. The space constant of the exponential decrease in R is approximately $\rho c^2\overline{A}/2\kappa\hat{\lambda}\mu$. Experimental data collected in uniform tubes having cross-sectional area equal to the human ear canal show that this space constant is much longer than the length of the adult human ear canal. This also held for data from curved horns reported in Sec. IV of the text. Therefore, the effect of viscosity on the spatial dependence of R is small and is neglected in the current analysis.

Agulló, J., Barjau, A., and Keefe, D. H. (**1999**). "Acoustic propagation in flaring, axisymmetric horns: I. A new family of unidimensional solutions," Acustica **85**, 278–284.

Agulló, J., Barjau, A., and Martínez, J. (**1988**). "Alternatives to the impulse response $h(t)$ to describe the acoustical behavior of conical ducts," J. Acoust. Soc. Am. **84**, 1606–1612.

ASTM C384-95 (**1995**). "Standard test method for impedance and absorption of acoustical materials by the impedance tube method" (ASTM, New York), pp. 101–109.

Benade, A. H. and Jansson, E. V. (**1974**). "On plane and spherical waves in horns with non-uniform flare. I. Theory of radiation, resonance frequencies and mode conversion." Acoustica **31**, 79–98.

Berlin, C. I. (Ed.) (**1998**). *Otoacoustic Emissions: Basic Science and Clinical Applications* (Singular, San Diego).

Burns, E. M., Keefe, D. H., and Ling, R. (**1998**). "Energy reflectance in the ear canal can exceed unity near spontaneous otoacoustic emission frequencies," J. Acoust. Soc. Am. **103**, 462–474.

Fahy, F. J. (**1995**). *Sound Intensity*, 2nd ed. (E & FN Spon, London), Chap. 4, pp. 38–88.

Holmes, M. H. (**1995**). *Introduction to Perturbation Methods* (Springer-Verlag, New York), Chap. 4, pp. 161–222.

Huang, G. T., Rosowski, J. J., Puria, S., and Peake, W. T. (**2000**). "Tests of some common assumptions of ear-canal acoustics in cats," J. Acoust. Soc. Am. **108**, 1147–1161.

Huang, G. T., Rosowski, J. J., Flandermeyer, D. T., Lynch, III, T. J., and Peake, W. T. (**1997**). "The middle ear of a lion: Comparison of structure and function to domestic cat," J. Acoust. Soc. Am. **101**, 1532–1549.

Hudde, H. (**1983**). "Estimation of the area function of human ear canals by sound pressure measurements," J. Acoust. Soc. Am. **73**, 24–31.

Keefe, D. H., Ling, R., and Bulen, J. C. (**1992**). ''Method to measure acoustic impedance and reflection coefficient,'' J. Acoust. Soc. Am. **91**, 470–485.
Keefe, D. H., Bulen, J. C., Arehart, K. H., and Burns, E. M. (**1993**). ''Ear-canal impedance and reflection coefficient in human infants and adults,'' J. Acoust. Soc. Am. **94**, 2617–2638.
Keefe, D. H. (**1997**). ''Otoreflectance of the cochlea and middle ear,'' J. Acoust. Soc. Am. **102**, 2849–2859.
Keefe, D. H., and Barjau, A. (**1999**). ''Acoustic propagation in flaring, axisymmetric horns: II. Numerical results, WKB theory, and viscothermal effects,'' Acustica **85**, 285–293.
Kemp, D. T. (**1978**). ''Stimulated acoustic emissions from within the human auditory system,'' J. Acoust. Soc. Am. **64**, 1386–1391.
Khanna, S. M., and Stinson, M. R. (**1985**). ''Specification of the acoustical input to the ear at high frequencies,'' J. Acoust. Soc. Am. **77**, 577–589.
Lawton, B. W. (**1979**). ''A survey of the sound field within the occluded ear canal,'' Proc. Inst. Acoust., November.
Lawton, B. W., and Stinson, M. R. (**1986**). ''Standing wave patterns in the human ear canal used for estimation of acoustic energy reflectance at the eardrum,'' J. Acoust. Soc. Am. **79**, 1003–1009.
Lonsbury-Martin, B. L., Harris, F. P., Stagner, B. B., Hawkins, M. D., and Martin, G. K. (**1990**). ''Distortion product emissions in humans. I. Basic properties in normally hearing subjects,'' Ann. Otol. Rhinol. Laryngol. **99**, 3–42.
Lynch III, T. J., Peake, W. T., and Rosowski, J. J. (**1994**). ''Measurement of the acoustic input impedance of cat ears: 10 Hz to 20 kHz,'' J. Acoust. Soc. Am. **96**, 2184–2209.
Margolis, R. H., Paul, S., Saly, G. L., Schachern, P. A., and Keefe, D. H. (**2001**). ''Wideband reflectance tympanometry in chinchillas and humans,'' J. Acoust. Soc. Am. **110**, 1453–1464.
Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (McGraw–Hill, New York).
Neely, S. T., and Gorga, M. P. (**1998**). ''Comparison between intensity and pressure as measures of sound level in the ear canal,'' J. Acoust. Soc. Am. **104**, 2925–2934.
Pierce, A. D. (Ed.) (**1994**). *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, Woodbury, NY).
Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (**1986**). *Numerical Recipes: The Art of Scientific Computing* (Cambridge U.P., Cambridge).
Rabbitt, R. D. (**1988**). ''High-frequency plane waves in the ear canal: Application of a simple asymptotic theory,'' J. Acoust. Soc. Am. **84**, 2070–2080.
Rabbitt, R. D., and Friedrich, M. T. (**1991**). ''Ear canal cross-sectional pressure distributions: Mathematical analysis and computation,'' J. Acoust. Soc. Am. **89**, 2379–2390.
Rabbitt, R. D. and Dragicevic, J. (**1991**). ''Measurement of Acoustic Energy Reflection by the Human Ear,'' 3rd USA–China–Japan Conference on Biomechanics, Atlanta, GA, 25 August.
Rabbitt, R. D., and Holmes, M. H. (**1988**). ''Three-dimensional acoustic waves in the ear canal and their interaction with the tympanic membrane,'' J. Acoust. Soc. Am. **83**, 1064–1080.
Reinsch, C. H. (**1967**). ''Smoothing by Spline Functions,'' Numer. Math. **10**, 177–183.
Rosowski, J. J., Peake, W. T., and Lynch, III, T. J. (**1984**). ''Acoustic input-admittance of the alligator-lizard ear: Nonlinear features,'' Hear. Res. **16**, 205–223.
Salmon, V. (**1946a**). ''Generalized plane wave horn theory,'' J. Acoust. Soc. Am. **17**, 199–211.
Salmon, V. (**1946b**). ''A new family of horns,'' J. Acoust. Soc. Am. **17**, 212–218.
Shaw, E. (**1966**). ''Ear canal pressure generated by circumaural and supraaural earphones,'' J. Acoust. Soc. Am. **39**, 471–479.
Shaw, E. A. G., and Teranishi, R. (**1967**). ''Sound pressure generated in an external-ear replica and real human ears by a nearby point source,'' J. Acoust. Soc. Am. **44**, 240–249.
Stinson, M. R. (**1985**). ''The spatial distribution of sound pressure within scaled replicas of the human ear canal,'' J. Acoust. Soc. Am. **78**, 1596–1602.
Stinson, M. R. (**1990**). ''Revision of estimates of acoustic energy reflectance at the human eardrum,'' J. Acoust. Soc. Am. **88**, 1773–1778.
Stinson, M. R., and Khanna, S. M. (**1989**). ''Sound propagation in the ear canal and coupling to the eardrum, with measurements on model systems,'' J. Acoust. Soc. Am. **85**, 2481–2491.
Stinson, M. R., and Lawton, B. W. (**1989**). ''Specification of the geometry of the human ear canal for the prediction of sound-pressure level distribution,'' J. Acoust. Soc. Am. **85**, 2492–2503.
Stinson, M. R., Shaw, E. A. G., and Lawton, B. W. (**1982**). ''Estimation of acoustical energy reflectance at the eardrum from measurements of pressure distribution in the human ear canal,'' J. Acoust. Soc. Am. **72**, 766–773.
Vorländer, M. (**2000**). ''Acoustic load on the ear caused by headphones,'' J. Acoust. Soc. Am. **107**, 2082–2088.
Voss, S. E., and Allen, J. B. (**1994**). ''Measurement of acoustic impedance and reflectance in the human ear canal,'' J. Acoust. Soc. Am. **95**, 372–384.
Webster, A. G. (**1919**). ''Acoustical impedance, and the theory of horns and of the phonograph,'' Proc. Natl. Acad. Sci. U.S.A. **5**, 275–282.
Zwicker, E. (**1990**). ''On the influence of acoustical probe impedance on evoked otoacoustic emissions,'' Hear. Res. **47**, 185–190.
Zwislocki, J. (**1962**). ''Analysis of the middle-ear function. Part I: Input impedance,'' J. Acoust. Soc. Am. **34**, 1514–1523.

Temporal pitch mechanisms in acoustic and electric hearing

Robert P. Carlyon
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, England

Astrid van Wieringen
Lab. Exp. OtoRhinoLaryngology, KU Leuven, Kapucijnenvoer 33, 3000 Leuven, Belgium

Christopher J. Long and John M. Deeks
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, England

Jan Wouters
Lab. Exp. ORL, KU Leuven, Kapucijnenvoer 33, 3000 Leuven, Belgium

(Received 11 October 2001; revised 2 May 2002; accepted 3 May 2002)

Two experiments investigated pitch perception for stimuli where the place of excitation was held constant. Experiment 1 used pulse trains in which the interpulse interval alternated between 4 and 6 ms. In experiment 1a these ''4–6'' pulse trains were bandpass filtered between 3900 and 5300 Hz and presented acoustically against a noise background to normal listeners. The rate of an isochronous pulse train (in which all the interpulse intervals were equal) was adjusted so that its pitch matched that of the ''4–6'' stimulus. The pitch matches were distributed unimodally, had a mean of 5.7 ms, and never corresponded to either 4 or to 10 ms (the period of the stimulus). In experiment 1b the pulse trains were presented both acoustically to normal listeners and electrically to users of the LAURA cochlear implant, via a single channel of their device. A forced-choice procedure was used to measure psychometric functions, in which subjects judged whether the 4–6 stimulus was higher or lower in pitch than isochronous pulse trains having periods of 3, 4, 5, 6, or 7 ms. For both groups of listeners, the point of subjective equality corresponded to a period of 5.6 to 5.7 ms. Experiment 1c confirmed that these psychometric functions were monotonic over the range 4–12 ms. In experiment 2, normal listeners adjusted the rate of an isochronous filtered pulse train to match the pitch of mixtures of pulse trains having rates of $F1$ and $F2$ Hz, passed through the same bandpass filter (3900–5400 Hz). The ratio $F2/F1$ was 1.29 and $F1$ was either 70, 92, 109, or 124 Hz. Matches were always close to $F2$ Hz. It is concluded that the results of both experiments are inconsistent with models of pitch perception which rely on higher-order intervals. Together with those of other published data on purely temporal pitch perception, the data are consistent with a model in which only first-order interpulse intervals contribute to pitch, and in which, over the range 0–12 ms, longer intervals receive higher weights than short intervals. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1488660]

PACS numbers: 43.66.Ba, 43.66.Fe, 43.66.Hg, 43.66.Ts [NFV]

I. INTRODUCTION

For normally hearing listeners, the pitch of a complex sound is usually dominated by the lower-numbered harmonics (Plomp, 1967; Moore *et al.*, 1985). The frequencies of these harmonics may be encoded by their place of excitation on the basilar membrane, by the temporal pattern of auditory-nerve (AN) responses to them, or to some combination of the two. In this article, we concentrate on ''purely temporal'' pitch perception, which we define as a pitch that can only be derived from the temporal characteristics of the AN response. To do this we use two types of stimulus: acoustic pulse trains that have been bandpass filtered to remove spectral components that are resolvable by the peripheral auditory system (cf. Hoekstra, 1979; Shackleton and Carlyon, 1994; Carlyon, 1997; Kaernbach and Demany, 1998), and electric pulse trains applied to a single channel of a cochlear implant.

Although the pitch derived from these stimuli is in many ways weaker than that derived by normal listeners from resolved harmonics—for example showing larger discrimination thresholds (Hoekstra, 1979; Shackleton and Carlyon, 1994; Kaernbach and Bering, 2001) that increase markedly above 300–600 Hz (Shannon, 1983; Tong and Clark, 1985; Townshend *et al.*, 1987; McKay *et al.*, 2000; Carlyon and Deeks, 2002)—our approach is of theoretical, and, potentially, practical significance. First, by excluding place-of-excitation cues, one can perform straightforward tests of more general pitch models that rely on temporal processing (Patterson *et al.*, 1991; Meddis and O'Mard, 1997). Second, users of most modern cochlear implants rely entirely on electric pulse trains for their sense of hearing. Understanding how the auditory system derives a pitch from the temporal characteristics of pulse trains may help guide the development of new cochlear-implant signal-processing strategies. Finally, by comparing the results obtained with electric and acoustic pulse trains, one can evaluate the extent to which the acoustic stimuli provide an adequate simulation of hearing by cochlear implant users. If the simulation is found to be accurate, then it may prove possible to develop experimental procedures and/or signal-processing strategies with

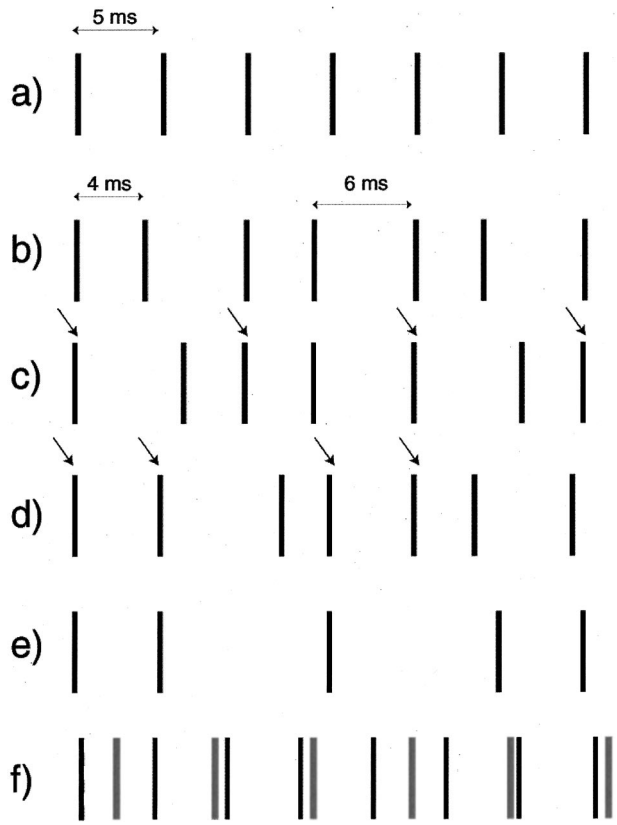


FIG. 1. Schematic representation of pulse trains used here and elsewhere. The ordinate is amplitude and the abscissa is time. (a) Periodic pulse train. (b) The ''4–6'' pulse train used in experiment 1. (c) Pulse train with regular second-order intervals and quasirandom first-order intervals. Arrows indicate the regular second-order intervals. (d) Pulse train with a proportion of regular first-order intervals, indicated by arrows. (e) Regular pulse train from which a proportion of pulses has been deleted. (f) Mixtures of two pulse trains having rates $F1$ Hz (gray lines) and $F2$ Hz (black lines).

the aid of normal listeners, before using the comparatively scarce time of cochlear implant users (cf. Blamey *et al.*, 1984).

The interpulse intervals in the pulse trains used in previous pitch studies were usually the same throughout the stimulus. The approach used here is to modify such isochronous pulse trains [Fig. 1(a)] in a way which will differentiate between three different theories of temporal pitch perception. All three theories were developed to account for data obtained with normal listeners, but make specific predictions about the pitches of both electric and acoustic pulse trains. Differences between the theories can be illustrated by the stimulus shown schematically in Fig. 1(b), in which the interval between successive pulses alternates between 4 and 6 ms. This stimulus was used in our first experiment, and was presented both acoustically to normal listeners and electrically to cochlear implant users.

Autocorrelation theories, originally applied to hearing by Licklider (1951), assume that the auditory system analyzes the intervals between each pulse and every other pulse. The largest peak in the autocorrelation of the stimulus shown in Fig. 1(b) is at 10 ms, because every second-order interval (between each pulse and the next-but-one) corresponds to this value. There are also peaks at 4 ms and 6 ms. We concentrate on an influential modern version of the autocorrelation approach, originally described by Meddis and Hewitt (1991) and subsequently modified by Meddis and O'Mard (1997). This model is similar to the original autocorrelation approach but takes into account important aspects of peripheral processing such as filtering and half-wave rectification. An implementation of this model will be described in Sec. II B., but, for the time being we note that it too produces the largest peak at 10 ms, plus two smaller peaks at 4 and 6 ms. The original version of the model (Meddis and Hewitt, 1991) assumed that the auditory system selects the largest peak in the autocorrelogram (10 ms), and, when two peaks have approximately equal size, the one corresponding to the shorter interval is selected. According to this model, then, the ''second choice'' is a period of 4 ms (pitch = 250 Hz). The later version (Meddis and O'Mard, 1997) uses a slightly more complicated decision metric, and its predictions will be discussed in more detail in Sec. II B.

In contrast to autocorrelation models, Kaernbach and Demany (1998) have suggested that the auditory system is sensitive only to *first-order* intervals between successive pulses. This conclusion was based on a set of experiments, in one of which subjects could not discriminate between a ''quasirandom'' pulse train and one whose second-order intervals were regular, but whose first-order intervals were, subject to certain constraints, random [Fig. 1(c); arrows indicate regular second-order intervals]. Subjects *could*, however, discriminate between the quasirandom train and one with a small proportion of regular first-order intervals [Fig. 1(d), arrows indicate regular first-order intervals). The predictions of their theory for the stimulus shown in Fig. 1(b) are that the pitch should correspond to a period of either 4 or 6 ms. In fact, they describe informal observations suggesting that subjects favor the longer of the two first-order intervals in pulse trains similar to those in Fig. 1(b), a finding consistent with the results of a pilot experiment described by Carlyon *et al.* (2001). Early research using filtered pulse trains (e.g., Thurlow and Small, 1955; Small and McClellan, 1963) was also consistent with pitch being based on first-order intervals, but suggested that the *shortest* interpulse intervals were dominant. However, it should be noted that those early experiments did not control for place-of-excitation cues as strictly as is possible in more modern studies, and the reports of ''temporal separation pitches'' as high as 2000 Hz (Thurlow and Small, 1955)—much higher than the few hundred Hz at which temporal pitch perception breaks down (Burns and Viemeister, 1976, 1981; Carlyon and Deeks, 2002)—suggests that place cues were in fact present in those experiments.[1]

Yet another metric was proposed by Carlyon (1996, 1997), who suggested that the pitch of a bandpass-filtered pulse train may correspond simply to the number of pulses present during a fairly long (>100 ms) segment of the stimulus. For example, he found that deleting a number of pulses at random from a pulse train [Fig. 1(e)] reduced its pitch. According to this *mean rate* hypothesis, the pitch of the pulse train shown in Fig. 1(b) should correspond to an interval of 5 ms (the average of 4 and 6 ms).

The present study consisted of two experiments. Experiment 1 presented electric pulse trains similar to those in Fig.

1(b) to a single channel of the LAURA cochlear implant, and bandpass-filtered acoustic versions of the stimulus to normal listeners. An advantage of this dual approach is that it overcomes some of the limitations inherent to purely acoustic experiments.[2] For example, the temporal response of the auditory system to the pulses illustrated in Fig. 1 will be affected by peripheral auditory processes such as ringing of auditory filters or adaptation at the inner hair cell/auditory-nerve synapse. Additionally, acoustic experiments require the presence of background noise to mask combination tones (CTs), and this noise is often not included when passing the experimental stimuli through auditory models (e.g., Carlyon, 1998; Kaernbach and Demany, 1998; Krumbholz *et al.*, 2000). Electric stimulation by-passes cochlear processing and can be performed in the absence of background noise, and so if the same results are obtained with the two types of stimulation, then a range of alternative explanations can be excluded. Experiment 2 was performed only with normally hearing listeners, and used a different stimulus, consisting of a mixture of two isochronous pulse trains of different rates, filtered into the same frequency region. The results of both experiments are consistent with pitch being derived from the first-order intervals present in the stimulus, with the greatest weight being applied to the longest first-order intervals.

II. EXPERIMENT 1

A. Method

Two different techniques were used in experiment 1. In experiment 1a, four normally hearing volunteers took part, all of whom were members of the auditory research laboratory at the University of Leuven. They adjusted the rate of an isochronous pulse train so that its pitch matched that of a ''4–6'' pulse train, whose first-order intervals alternated between 4 and 6 ms [Fig. 1(b)]. The 400-ms pulse trains were constructed at a sampling frequency of 44 100 Hz and passed through a digital bandpass filter having cutoff frequencies of 3900 and 5300 Hz. The digital filter was designed to approximate the analog eighth-order Butterworth filter used in experiment 2. The attenuation at the cutoff frequencies was no more than 3 dB, relative to the level in the center of the passband. The attenuation at half an octave below the lower cutoff, and at half an octave above the upper cutoff, was at least 24 dB relative to that at the cutoff frequencies. The rms level was set to 54 dB SPL. Stimuli were presented against a background of continuous pink noise, having a spectrum level of 9.5 dB SPL at 4 kHz. They were turned on and off with 50-ms raised-cosine ramps. A total of 29 isochronous pulse trains, with periods ranging from 2 to 14 ms in steps of 7%, with the period rounded to the nearest 0.1 ms, were generated and saved to separate waveform files. During each run of the matching test the subjects heard a triad of stimuli presented repeatedly: the 4–6 pulse train, an isochronous pulse train, and the 4–6 pulse train again. The interstimulus time was 400 ms within, and 1500 ms between, triads. Subjects were required to adjust the pitch of the second stimulus to match that of the first (and third). The particular ''starting value'' for the isochronous pulse train was chosen in a quasirandom fashion, from run to run. The subject could then increase or decrease the period of the isochronous train to be presented on the next trial, by either one, two or three 7% steps (by pressing ''---,'' ''--,'' ''-,'' or ''+,'' ''++,'' ''+++''). When s/he was satisfied with the pitch match, a separate button could be pressed which then initiated the next run. Each subject performed 30 matches. All pulse trains were played out of a portable PC (Toshiba, Satellite Pro 420CDT) via a 16-bit PCMCIA card (WAVjammer, New Media Corp), mixed (Eurorack MX 1604A) with the analog noise from a CD player (Sony, CDP-209), and presented to the right earpiece of a TDH39 headset.

Experiment 1b was performed both with normally hearing listeners and five users of the LAURA cochlear implant. In pilot experiments we found that the pitch-matching procedure was quite time-consuming, and one cochlear implant user produced quite variable results. We attributed this to the fact that the 4–6 stimulus, by virtue of its irregularity, had a different timbre to the isochronous sounds. We therefore adopted a constant-stimulus procedure in which, during each experimental block, the pitch of the isochronous sounds varied substantially from trial to trial. We reasoned that this would help subjects focus on the pitches of the sounds, and that the timbral difference would ''stick out'' less.

During each block of 50 trials in experiment 1b, the 4–6 stimulus was paired, in random order, with an isochronous pulse train having a period of 3, 4, 5, 6, or 7 ms. These values spanned the range of pitch matches made in experiment 1a. The interstimulus interval was 500 ms. After the two sounds were presented the subject indicated which had the higher pitch. No feedback was given. Another isochronous train was selected at random for the next trial. All subjects except one completed 20 blocks of trials, each time with different randomizations of the paired stimuli, so there were 200 trials for each isochronous stimulus. One normal listener (NH5) did only 50 trials per point.

Although nearly all of the matches in experiment 1a were to periods between 3 and 7 ms, we performed a supplementary experiment (1c) using the procedure of experiment 1b but with isochronous pulse trains having periods of 4, 6, 8, 10, and 12 ms. This was done as an extra check for subjects hearing a pitch corresponding to a period of 10 ms, corresponding to the largest peak in the autocorrelogram. One implant user and two normal listeners took part, both of whom had participated in experiment 1b. Two of them did 100 trials per point, and one normal listener (NH5) did 50 trials per point.

For the normally hearing listeners the stimuli were generated in the same way as in experiment 1a, and presented against the same continuous pink-noise background. For the implant users the stimuli were 400-ms-long biphasic pulse trains, with a phase duration of 40 μs, and the current was gated on and off with 50-ms linear ramps. The pulse trains were applied to the stimulation channel shown in Table I via a DSP—TIC30 board, the sequencer of the computer (Compaq Armada M700), and the external coil of the LAURA device. Table I also provides brief details of each implant user. Subjects were implanted at the university ENT Dept. of the St. Augustinus Hospital in Antwerp and the University Hospital in Leuven. Based on closed-set vowel and conso-

TABLE I. Details of the cochlear implant users who participated in experiment 1b. The table shows their age, etiology, duration of deafness prior to implantation (DD, in years), number of implanted years (CI years), percentage-correct speech recognition results reported by van Wieringen and Wouters (1999) for vowels (SRV) and for consonants (SRC), channel under test (CH) and current used for the 40 μs/phase in these experiments [C(μA)]. In the LAURA device, channel 1 is the most apical, and channel 8 is the most basal. The stimulation channel with the largest dynamic range was chosen.

Listener	Age	Aetiology	DD (years)	CI (years)	SRV (%)	SRC (%)	CH	C(μA)
CI 1	20	Meningitis	3.5	7	53	45	2	600
CI 2	44	Menière's	3	5	79	51	7	1000
CI 3	33	Unknown	6	7	73	58	3	900
CI 4	51	Congenital progressive	1	3	74	38	3	1050
CI 5	48	Congenital progressive	12	5	40	35	4	1200

nants tests (''SRV'' and ''SRC,'' respectively, Table I), they are considered moderate to good performers (van Wieringen and Wouters, 1999). They had all participated in previous psychophysical studies and were paid for their participation.

For a given subject, all electrical stimuli in experiment 1b were presented at the same current level, as shown in Table I. This level, measured using a test implant having a 2.2-kΩ resistor between electrodes, was judged to produce a comfortable loudness for all stimuli. We chose not to perform formal loudness-balancing measurements in this experiment for the different pulse rates, for two reasons. First, because this was a pitch preference task without feedback, we considered it unlikely that any small residual loudness differences would strongly and systematically influence subjects' responses. Second, pilot experiments revealed that the resulting current levels differed only slightly or not at all, even between the shortest and longest periods used here. This is similar to the measurements of McKay and McDermott (1998), who showed that loudness matches for comfortably loud pulse trains varied by less than 0.5 dB over the range of rates used in experiment 1. However, because experiment 1c used a wider range of rates, we decided to loudness balance all stimuli for the cochlear implant user (CI 2) who took part in that experiment. We did so by presenting the 4–6 stimulus (level=1000 μA) with each isochronous sound in an alternating fashion, and with an interstimulus interval of 400 ms. The subject could increase or decrease the level of the isochronous sound by 5, 20, or 50 μA by pressing one of six buttons. Two matches were obtained for each isochronous stimulus and averaged to obtain the levels used in experiment 1c. These levels were 1000, 1005, 1040, 1040, and 1050 μA for isochronous periods of 4, 6, 8, 10, and 12 ms, respectively.

All stimuli (electric and acoustic) in both parts of the experiment were presented using the APEX software/hardware package (Geurts and Wouters, 2000).

B. Results and discussion

The pitch matches made by the four normal listeners of experiment 1a were pooled, placed in 0.5-ms bins, and plotted in histogram form in Fig. 2. The geometric mean of all the matches was 5.71 ms, with a standard error across subjects of 0.23 ms. Psychometric functions describing the proportion of trials in which each isochronous pulse train in experiment 1b was judged higher than the 4–6 stimulus are shown in Fig. 3(a) for the normal listeners and Fig. 3(b) for the implant users. The point of subjective equality—at which these functions cross the 50% point, was estimated by a probit function. The geometric means of these values across subjects were 5.56 and 5.69 ms for the normal listeners and implant users, respectively. A summary of the results from both parts of the experiment is shown in Table II. It can be seen that the results are highly consistent both across different procedures (experiments 1a vs 1b) and across the two groups of listener (experiment 1b).

The results of experiment 1c are shown in Fig. 3(c). It can be seen that the psychometric functions in all three cases are monotonic, and do not show any discontinuity at 10 ms, as might be expected if subjects sometimes heard a pitch corresponding to that value. The intercepts of these psychometric functions with the 50% point were 5.9 ms for cochlear implant user CI2, and 5.7 ms for normal listener NH5. These values are reasonably close to those observed for the same subjects in experiment 1b. Subject NH1's data intercept the 50% point at a period slightly longer than 6 ms, and longer than the 5.4 ms observed in experiment 1b. Unfortunately, we could not measure the intercept precisely, because the probit functions used successfully for the other functions described here did not provide an accurate fit to NH1's functions in this condition (chi-squared, $df=3$, $p<0.001$).

The results of experiment 1 have implications for all three models of temporal pitch perception described in the Introduction. Figure 4(a) shows the output of an autocorrelogram model (Meddis and Hewitt, 1991; Meddis and

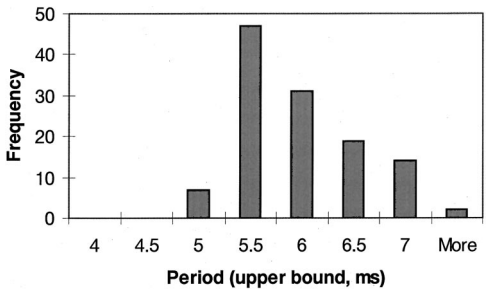


FIG. 2. Distribution of pitch matches falling into 0.5-ms bins, obtained from experiment 1a.

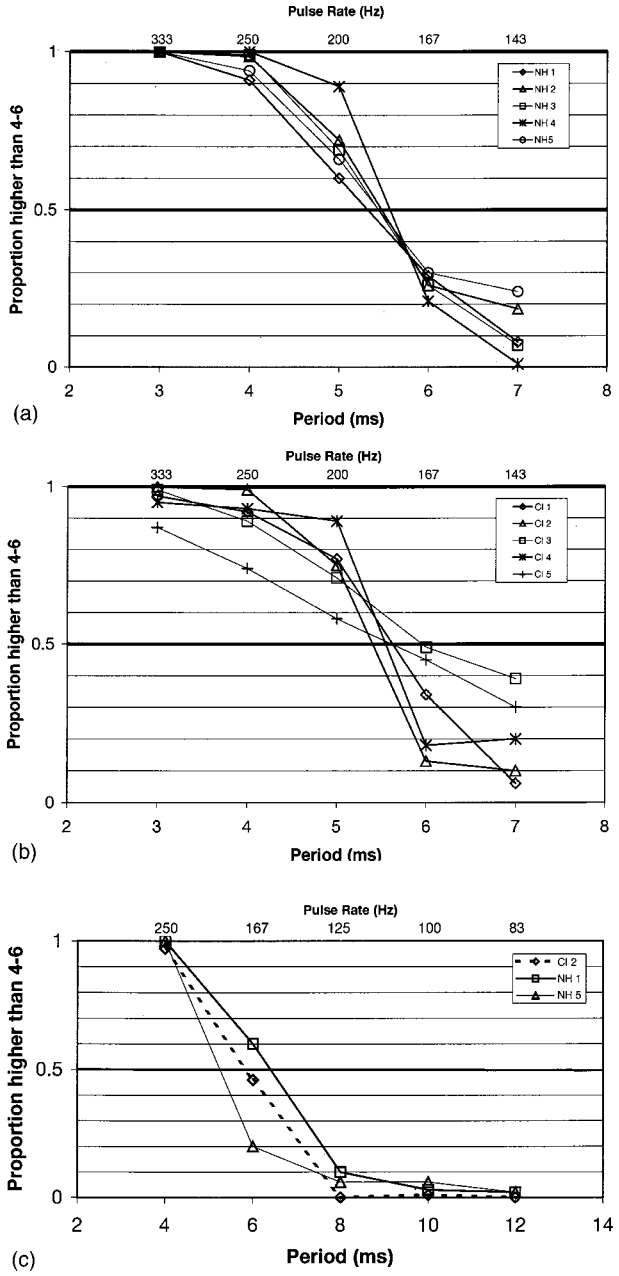


FIG. 3. (a) Proportion of trials in which isochronous stimuli, having periods shown on the abscissa, were judged as higher in pitch than the "4–6" stimulus by the normal listeners in experiment 1a. (b) As in part (a), but for the cochlear implant users (of experiment 1b). (c) Results of experiment 1c for one cochlear implant user and two normal listeners. An indication of the pulse rates corresponding to periods of the isochronous pulse trains is shown at the top of each plot.

O'Mard, 1997) in response to the 4–6 stimulus used here. (The background noise was not used in this simulation). The largest peak corresponds to a period of 10 ms, and the next largest to a period of 4 ms. As discussed in the Introduction, the original version of the model would predict matches to these values, which did not occur in the data. The revised version (Meddis and O'Mard, 1997) differs not in the autocorrelogram itself but in the decision statistic applied. Specifically, it takes the summary autocorrelograms ("SACFs") of all "matching" stimuli, and calculates the squared Euclidean distance (D^2) between each one and the SACF of the comparison stimulus. When we applied this model to the

TABLE II. Rate of an isochronous pulse train judged equal in pitch to the "4–6" pulse train of experiment 1. Results from experiment 1b ("2IFC intercept") represent the point at which the psychometric functions intercepted the 50% point, as estimated from probit fits. Means and standard errors are geometric.

	Normal listeners		Implant users	
Subject	2IFC intercept	Match	Subject	2IFC intercept
NH 1	5.40 ms	5.8 ms	CI 1	5.55 ms
NH 2	5.69 ms	6.3 ms	CI 2	5.40 ms
NH 3	5.53 ms	5.2 ms	CI 3	6.20 ms
NH 4	5.62 ms	5.6 ms	CI 4	5.58 ms
NH 5	5.7 ms		CI 5	5.65 ms
Mean	5.59 ms	5.71 ms		5.69 ms
S.E.	0.06	0.23		0.13

acoustic stimuli of experiment 1a, we observed the predictions shown in Fig. 4(b). Because a perfectly harmonic (isochronous) matching stimulus has peaks at multiples of the period, a pulse train with a period of 5 ms provides a very good (in fact, the best) match to our 4–6 stimulus. This is because it has a peak at 5 ms, which does a reasonably good job of matching the two peaks in the 4–6 SACF at 4 and 6 ms, respectively; it also has a peak at 10 ms which matches the 10-ms peak in the 4–6 SACF exactly. Accordingly, the function relating D^2 to the period of the matching sound has a minimum at 5 ms. On either side of this minimum the function rises, but does so less towards shorter periods than it does towards longer periods, because a matching sound having a period of 4 ms also produces a good match, due to the

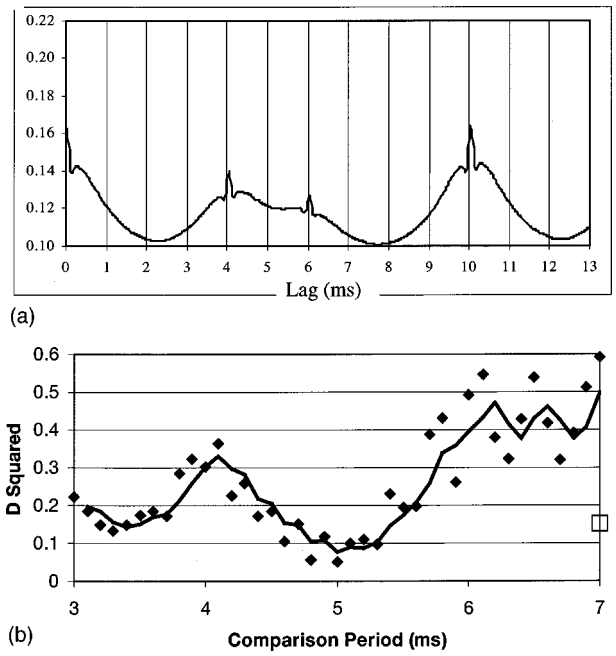


FIG. 4. (a) Summary autocorrelogram produced by Meddis and O'Mard's (1997) model in response to the "4–6" stimulus of experiment 1. (b) Squared Euclidean distances, according to Meddis and O'Mard's autocorrelogram model, between the 4–6 stimulus and various isochronous stimuli, as a function of their interpulse period. The open square on the right of the plot is the squared Euclidean distance for an isochronous stimulus with a interpulse interval of 10 ms.

large peak in the SACF of the 4–6 stimulus. Additional minima occur for matching sounds having periods of 10 ms [value shown as open square on the right of Fig. 4(b)] and at 3.33 ms, both of which produce SACFs having secondary peaks at 10 ms. Unfortunately, the revised autocorrelogram model (Meddis and O'Mard, 1997) does not specify how to deal with situations where the squared Euclidean distances indicate the possibility of multiple pitch matches, but neither of the two most obvious options is seen in the data: (i) if subjects usually match to the stimulus producing the minimum D^2, with the probability of matches to adjacent values being inversely proportional to D^2, there should be a unimodal distribution of matches with a mean lower than 5 ms but higher than 4 ms, or (ii) if subjects match to stimuli producing a low D^2, regardless of their period, there should be a bimodal distribution of matches clustered around 5 and 3.33 ms, and, in experiment 1c, there should be a ''bump'' in the psychometric function around 10 ms.

The data are also inconsistent with the ''mean rate'' metric proposed by Carlyon (1996, 1997), according to which the perceived pitch of the 4–6 stimulus should correspond to a period of 5 ms. The results could, however, be reconciled with Kaernbach and Demany's conclusion that pitch is based on first order intervals, if one additionally assumes that longer intervals receive a larger weight than shorter intervals when pitch is calculated. In Sec. IV, we describe a model in which pitch is derived from such a weighted sum of first-order intervals.

Finally, we should mention two related findings reported by other authors. First, Kaernbach and Demany (1998) mentioned the results of an informal experiment in which, for a sequence consisting of alternating 5-ms and 3-ms periods, subjects sometimes heard a pitch corresponding to 5 ms and sometimes (although more rarely) to 3 ms. In contrast, the results of experiment 1a indicate that subjects never matched to the shorter of our two intervals (4 ms). We are not sure whether this is due to a difference between the stimuli used in the two experiments, or whether a more formal investigation using Kaernbach and Demany's stimuli would yield results more similar to our own. Second, McKay and McDermott (1996) required cochlear implant users to perform a pitch-rating experiment using electrical pulse trains that consisted of ten pulses evenly distributed throughout each 10-ms period, and with two of these pulses having a higher amplitude than the others. The intervals between higher-amplitude pulses alternated between two values, which, for different stimuli, were 1 and 9 ms, 2 and 8 ms, 3 and 7 ms, and 4 and 6 ms. Consistent with our results, they found that the relative pitch varied in the same direction as the longer of the two interpulse intervals. However, the exact value of this pitch was not the primary focus of their study, and a more detailed interpretation of their results is complicated by the fact that pitch would also have been affected by the lower-amplitude pulses, in a way which varies across listeners (McKay *et al.*, 1995; McKay and Carlyon, 1999). We plan to report the results of a study of the pitch of acoustic and electric amplitude-modulated pulse trains in a future publication (van Wieringen *et al.*, 2002).

III. EXPERIMENT 2

A. Rationale and method

Carlyon (1996) presented normal listeners with pairs of inharmonically related pulse trains, presented simultaneously and passed through the same bandpass filter. When the spectral components of the pulse trains were unresolved by the peripheral auditory system, he reported that subjects heard a unitary ''crackle-like'' percept, and could not extract the pitches of the two underlying pulse trains. He further suggested that the pitch of this mixture corresponded to the average number of pulses in a fairly long (>100 ms) portion of the stimulus. This ''mean rate'' hypothesis received some support from the later finding that deleting some pulses at random from an otherwise isochronous pulse train could affect its pitch, which dropped systematically as a greater and greater proportion of pulses was deleted (Carlyon, 1997). However, the pitch of mixtures of inharmonically related pulse trains had never actually been measured, and this was the purpose of experiment 2. Only acoustic stimuli and normal listeners were used in this experiment.

Mixtures of pulse trains having rates of $F1$ and $F2$ Hz were played concurrently out of two DACs of a CED 1401 plus laboratory interface (16-bit resolution, sampling rate $=20\,000$ Hz), antialiased (Kemo VBF 25.01, cutoff 8300 Hz, attenuation rate$=100$ dB/octave), bandpass filtered between 3900 and 5400 Hz (one low-pass and one high-pass Kemo VBF 25.03 in series for each pulse train, attenuation rate 48 dB/octave), mixed with a continuous pink noise, and presented to the left earpiece of a Sennheiser HD250 headset. These mixtures had an overall level of 61 dB SPL, a duration of 400 ms, and were turned on and off abruptly prior to filtering. The continuous pink-noise background had a spectrum level of 6 dB SPL at 4000 Hz. On different pitch-matching runs the value of $F1$ was either 70, 92, 109, or 124 Hz, and $F2/F1$ was always 1.29 ($F2=90$, 119, 141, or 160 Hz). Each run started with one of the mixtures being followed, 500 ms later, by a single isochronous pulse train having a rate selected at random from a uniform distribution ranging from 50 to 400 Hz. Subjects could then adjust the rate of the single pulse train to be presented on the next trial by pressing one of four buttons, which increased or decreased the value by a factor of 1.1 or 1.4. Subjects were told that, if they did hear more than one pitch in the mixture, they were to match to the strongest pitch. The adjustment procedure continued until the subject was satisfied with the match, and s/he pressed a separate button indicating this. A total of 12 pitch matches was obtained at each value of $F1$ for each of four normally hearing subjects. These 12 values were geometrically averaged to obtain the pitch match for each subject and condition. In addition, six matches were obtained in each condition to the $F1$-alone and $F2$-alone stimuli. This was done in order to check for any systematic response bias. All subjects had absolute thresholds of less than 15 dB HL (ANSI, 1969) at all audiometric frequencies.

B. Results

The symbols in each panel of Fig. 5(a) show the pitch matches obtained for each listener in experiment 2, as a func-

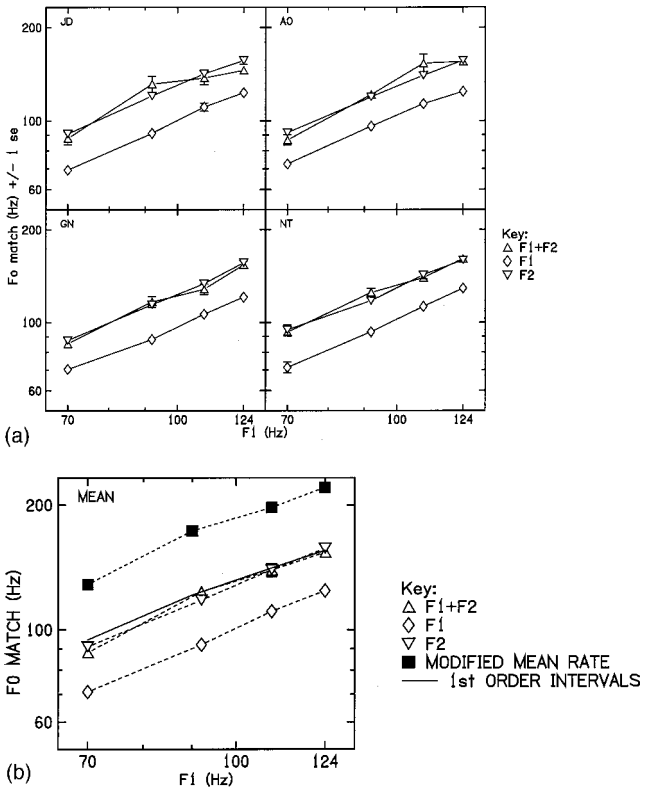


FIG. 5. (a) Pitch matches obtained by each listener in experiment 2 to pulse trains of $F1$ Hz (diamonds), $F2$ Hz (inverted triangles), and to mixtures of $F1+F2$ Hz (upright triangles). The geometric means and standard errors of the matches are plotted as a function of $F1$. (b) Geometric means of the matches obtained from the four listeners in experiment 2. Solid lines with no symbols are the predictions of the model described in Sec. IV A 2. Solid squares indicate the predictions of the revised "mean rate" model described in Sec. IV B 1.

tion of $F1$. Matches to the mixtures of two pulse trains are shown by upright triangles. In all cases, these matches are very close to the matches obtained to the $F2$ stimulus alone (inverted triangles). Mean data are shown by the dashed lines connecting open symbols in Fig. 5(b), which also contains two additional lines, whose meanings will be discussed in Sec. IV. The absence of any matches significantly greater than $F2$ argues against the mean rate hypothesis proposed by Carlyon (1996, 1997), according to which subjects would match to $F1+F2$ Hz. However, some of the shorter interpulse intervals in the stimulus will be removed by factors such as auditory filter ringing and neural refractoriness, and so it might be possible for a modified version of the mean rate model to produce matches more in line with the data. This possibility is examined further in Sec. IV.

The pattern of matches is also slightly different from the summary autocorrelogram, which has peaks of roughly equal height at the periods corresponding to $1/F1$ and $1/F2$; this is shown in Fig. 6, which plots Meddis and O'Mard's SACF to our stimulus where $F1=109$ Hz. In Sec. IV B we will present a quantitative model based on first-order intervals, which, we will argue, is consistent with the results of experiment 2, as well as those of experiment 1 and of some other results in the literature.

IV. DISCUSSION

A. Analysis based on first-order intervals

1. Overview

As discussed earlier, the results of our experiments are consistent neither with the autocorrelogram model described by Meddis and his colleagues (Meddis and Hewitt, 1991; Meddis and O'Mard, 1997), nor with Carlyon's mean rate hypothesis (Carlyon, 1996). In this section, we propose a model in which "purely temporal" pitch is derived from a weighted sum of first-order interpulse intervals, and which, we argue, can account not only for our own data but those from two other paradigms reported in the literature. Initially, we discuss the model in terms of intervals between *stimulus* pulses. This has the advantage of simplicity, and is sufficient for explaining the major features of the model. However, it is clear that any pitch mechanism must operate on the output of the auditory nerve (AN), and that this will differ from the

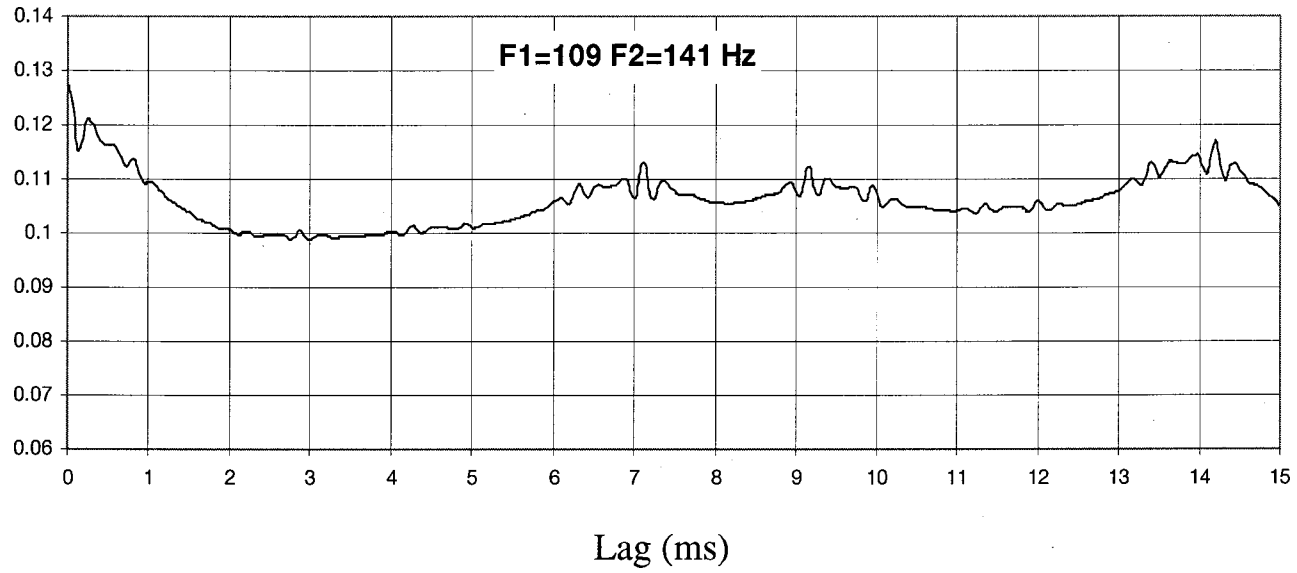


FIG. 6. Summary autocorrelogram produced by Meddis and O'Mard's (1997) model in response to the mixture of 109- and 141-Hz pulse trains used in experiment 2.

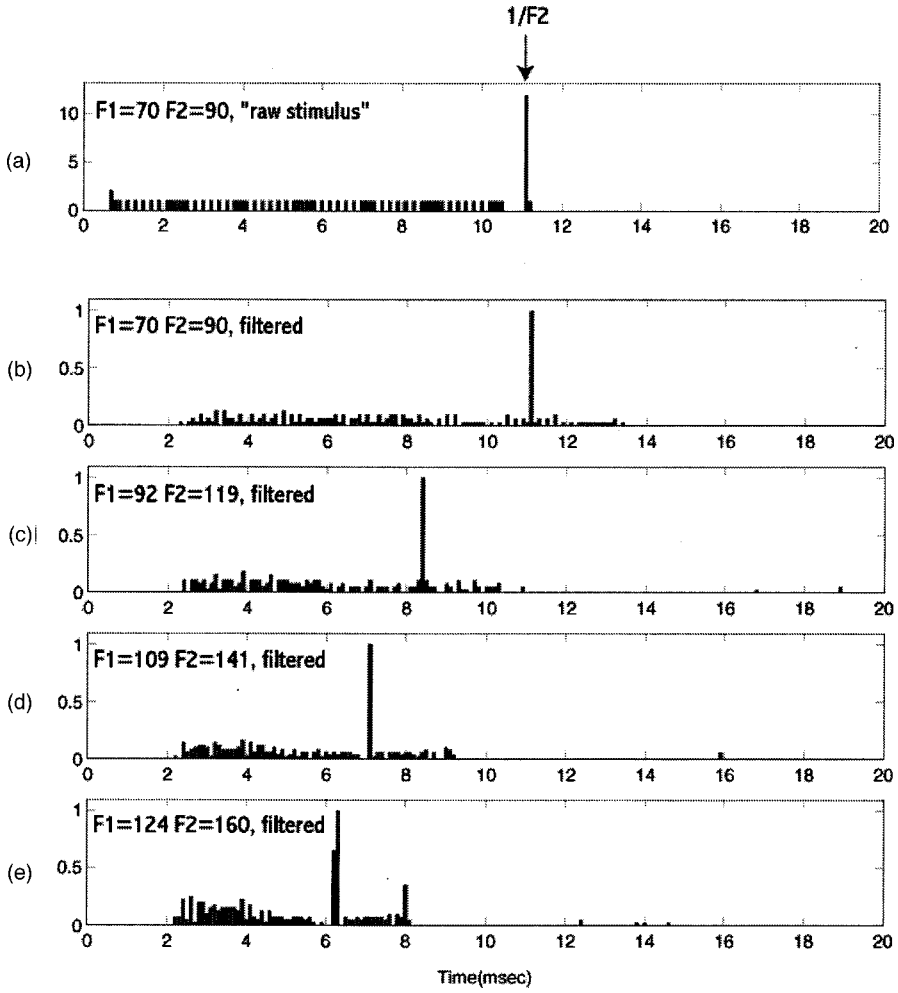


FIG. 7. (a) Distribution of first-order interpulse intervals in the mixture of 70- and 90-Hz pulse trains used in experiment 2. Parts (b)–(e) are similar, but show simulated ISIHs obtained by passing the stimulus through a simulated auditory filter and low-pass filtering to mimic properties of the inner hair cells (see the text for details). $F1$ is 70 Hz in part (b), 92 Hz in part (c), 109 Hz in part (d), and 124 Hz in part (e).

raw input due to basilar-membrane ringing (for acoustic stimuli) and to refractory effects (for both acoustic and electric stimuli). We then move on to incorporate the effects of this peripheral processing for stimuli, such as those with numerous short interpulse intervals, where they are likely to have a significant effect. Finally, for all stimuli, we consider whether the transformations imposed by the auditory system up to and including the auditory nerve are sufficient to account for the pattern of results obtained. This first subsection gives an overview of our general approach, which is then quantified in the next subsection.

At first sight, the results of experiment 2, in which subjects matched to the higher of the two rates present in the mixture, seem at odds with those of experiment 1, where they matched to the lower pitch (period close to 6 ms, rather than to 4 ms). However, this discrepancy can be resolved by considering the pattern of first-order intervals in the mixtures used in experiment 2; this is illustrated for the case where $F1=70$ Hz in Fig. 7(a). It can be seen that the most common first-order interval corresponds to $1/F2$ (1/90 Hz $=11.1$ ms), and that this is also the longest first-order interval present. The reason for the absence of first-order intervals at $1/F1$ is apparent upon inspection of Fig. 1(f). The interval between successive pulses in the $F1$-Hz pulse train (gray lines) is so long that there is always at least one intervening pulse from the $F2$-Hz stimulus (black lines). Hence, a model which operates on first-order intervals and selectively weights longer intervals will yield a match to about $1/F2$. As discussed in Sec. II B, this general scheme is also consistent with the results of experiment 1.

An additional attraction of the idea that longer first-order intervals dominate purely temporal pitch perception is that it can qualitatively account for some recent data presented by Plack and White (2000b). In one condition, they presented their subjects with a 40-ms unresolved ''standard'' harmonic complex, having a nominal $F0$ of 250 Hz, in which the fifth and all subsequent ''pitch pulses'' were advanced or delayed by either 0, 1, or 3 ms—thereby either shortening or lengthening a single interpulse interval. Subjects adjusted the $F0$ of a perfectly periodic comparison sound so that its pitch matched that of the standard. Delaying the last few pulses produced much larger pitch shifts than did advancing them by the same amount. This can be understood if one assumes that imposing a delay increases a single interpulse interval, which then receives a large weight in the pitch estimation process. Imposing an advance shortens one interpulse interval, but this shortened interval will receive a smaller weight when estimating pitch.[3]

Finally, it should be noted that a pitch analysis based on first-order intervals could, in principle, account for the data that Carlyon (1997) interpreted as support for the mean rate model. As he pointed out, deleting some pulses from an otherwise periodic pulse train will lengthen some of the first-order interpulse intervals, and it is clear that any weighted

sum of such intervals will produce a lower pitch than is the case for an uninterrupted pulse train. In the next subsection we investigate whether a simple model can account quantitatively for the reductions in pitch that resulted from deleting pulses, as well as the effects reported in the present paper and in the study described by Plack and White (2000b).

2. General model

In this subsection we describe the general format of a model which estimates pitch based on first-order intervals. We then describe some constraints imposed on that model by the data presented here and elsewhere, and consider how effectively it could account for the data. We should state at the outset that we do not have sufficient data to specify unambiguously how such a model should work. However, we believe that it is worthwhile to determine whether a model based on a reasonable set of parameters can be reconciled with data from a fairly wide range of paradigms.

As stated above in Sec. IV A 1, we will first describe the model in terms of first-order intervals between *stimulus* pulses, rather than between neural spikes. This simplification is likely to be fairly adequate for stimuli in which the first-order intervals are fairly long—such as the 4–6 pulse train of experiment 1, but will require modification in cases where numerous short interpulse intervals are present, and where the effects of basilar-membrane ringing and refractoriness are likely to be greatest. These modifications will be described after the basic model has been introduced.

In general terms, we propose that pitch (P) will be determined by the first-order intervals present (τ), by the weights applied to intervals of different lengths ($W(\tau)$), and by some function of the relative proportions of those intervals ($f(p_t)$)

$$P=\frac{\Sigma_{t=1}^{n} W(\tau_t) f(p_t)}{\Sigma_{t=1}^{n} W(\tau_t) f(p_t) \tau_t}. \quad (1)$$

For example, the standard stimulus of experiment 1 had equal proportions of 4-ms and 6-ms intervals, so $\tau_1=4$ ms, $\tau_2=6$ ms, and $p_1=p_2=0.5$. The fact that the observed pitch was much closer to 6 ms than to 4 ms reveals that $W(\tau)$ must be an expansive function, at least over the range between 4 and 6 ms. Specifically, given that the average intercept across the ten listeners of experiment 1b corresponded to an interval of 5.64 ms, the 6-ms interval should receive about 82% ((5.64−4)/(6−4)) of the total weights, and the 4-ms interval only about 18%—a ratio of 4.56:1. This relationship is approximated by the form of $W(\tau)$ shown by the thick solid line in Fig. 8, which is, as described below, also generally consistent with the data of experiment 2, of Plack and White (2000b), and of Carlyon (1997). It approximates a power law for intervals between about 1 and 7 ms, but becomes less expansive at longer intervals, and reaches a plateau between 8 and 12 ms. Given the fact that very low repetition rates do not elicit a pitch percept at all (e.g., Guttman and Pruzansky, 1962), it seems probable that $W(\tau)$ will start to drop at even longer intervals. Here, we do not attempt to model data obtained with stimuli containing such long intervals, and so have left $W(\tau)$ undefined for $\tau>12$ ms.

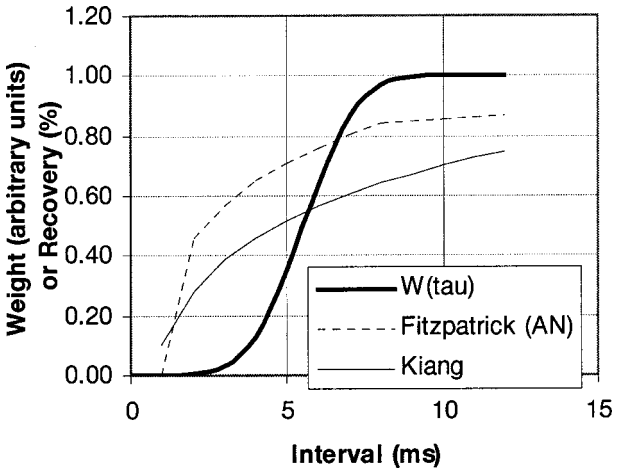


FIG. 8. Heavy solid line: a possible form of the function $W(\tau)$ in Eq. (1). Dashed line: function fitted to the mean auditory nerve data of Fitzpatrick *et al.* (1999). The function is zero for intervals (τ) below 1 ms, equal to $0.84+0.63\log_{10}(\tau/8)$ for 1 ms$<\tau<$8 ms, and to $0.9+0.15\log_{10}(\tau/20)$ for $\tau>8$ ms. Faint line: fit to Kiang's (1965) AN data, as a function of the interpulse interval (τ) of isochronous pulse trains. The function is equal to $0.70-0.6*(\log_{10}(10/\tau))$ for $\tau<10$ ms, and $0.95-0.25*(\log_{10}(100/\tau))$ for $\tau\geqslant 10$ ms.

To account for the data of Plack and White (2000b), we need to specify the form of $f(p_t)$. This was not an issue with experiment 1, because the two interpulse intervals were equiprobable. Using the form of $W(\tau)$ described in Fig. 8, we could approximate their data by assuming that $f(p_t)=p_t^2$. This "squaring" was chosen to fit the data, but has the desirable feature of minimizing the influence of very rare interpulse intervals. This could render the pitch estimate more robust to the influence of both external and physiological noise. The results obtained using the model are shown by the open diamonds in Fig. 9, along with Plack and White's data (solid line) and the predictions of Meddis and Hewitt's (1991) autocorrelogram model (dotted lines), as implemented by Plack and White (2000b, Fig. 7). It can be seen that the predictions of the present model, while underestimating the pitch shifts in some conditions, account quite well for the general form of the data. The fit can be improved slightly

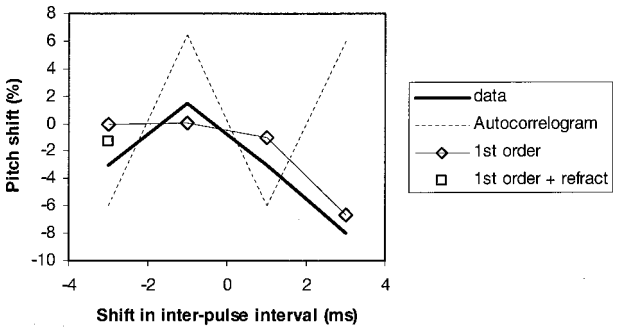


FIG. 9. The thick solid line shows the pitch shifts obtained by Plack and White (2000b) to 40-ms 250-Hz pulse trains in which one interpulse interval had been reduced (negative values on abscissa) or increased (positive values). Solid lines with diamonds show the predictions of the model described in Sec. IV B. The unconnected square shows the predictions of the model for a 3-ms reduction in interpulse interval, when refractory effects are included (see the text for details). The dashed line shows the predictions of Plack and White's implementation of Meddis and Hewitt's (1991) autocorrelogram model.

TABLE III. The middle column shows the percentage change in pitch produced by deleting an additional 10% of pulses from a pulse train in which either 15%, 30%, or 45% of pulses had already been deleted. These results are taken from Carlyon (1997). The right-hand column shows the predictions of the model described in Sec. IV B, using the function shown in Fig. 8.

	% pitch change for 10% pulse reduction	
% deleted in standard	Data	Model
15	4	3.9
30	5.2	6.0
45	3.9	5.5

if we follow Plack and White's observation that in one of their conditions, where the last few pulses were advanced by 3 ms, the resulting 1-ms interpulse interval would probably not be represented in the auditory nerve. They argued that the refractory properties of the auditory nerve (Kiang, 1965) made it likely that many fibers would not respond to the pulse marking the end of this short interval, and so, following their suggestion, we have replaced this 1-ms interpulse interval with one of 5 ms. Doing so introduces a negative pitch shift in the condition with a -3-ms pulse shift (open square), which is more in line with the data. With or without this modification, the fit, although imperfect, is substantially better than that obtained with the autocorrelogram model (dotted line), which performs very poorly, and which predicts large positive shifts for the -1-ms and $+3$-ms delays, and large negative shifts for the -3-ms and $+1$-ms delays. Similarly unsuccessful predictions were generated by another model, based on a spectral analysis of simulated neural spike trains, that was considered by Plack and White (2000b). Reasons for the failure of the autocorrelogram and ''spike spectrum'' models are discussed in detail in that article.

The present model can also account for the results of Carlyon (1997), who asked subjects to adjust the $F0$ of a pulse train from which a proportion of pulses had been deleted, so that its pitch matched that of a 200-Hz ''standard'' pulse train, which had either 15%, 30%, or 45% of pulses deleted. By manipulating the proportion of deleted pulses in the ''matching'' train, it was determined that a 10% drop in pulse probability *re* the standard produced about a 4% drop in pitch, regardless of the overall proportion of pulses deleted from the standard stimulus. The model predictions, shown in the right-hand column of Table III, produces pitch shifts which are of a similar size to those obtained (middle column), and which, like the data, do not vary markedly with the overall proportion of pulses deleted in the Carlyon (1997) paradigm.

The stimuli of experiment 2 differ from those described so far in that, as shown in Fig. 7(a), there are numerous very short first-order intervals in the stimulus. These intervals will be modified by auditory filter ringing, the time constant of the inner hair cell membrane potential, and neural refractoriness. We have simulated peripheral processes by a ''gammatone'' (Patterson *et al.*, 1988) auditory filter centered on 4500 Hz, followed by an 800-Hz low-pass filter implemented in order to simulate the receptor potential transfer function (Holton and Weiss, 1983). As can be seen in Figs. 7(b)–(e), this removes very short intervals and introduces additional intervals longer than $1/F2$. However, the details of this simulation are unlikely to be crucial, because the form of $W(\tau)$ shown in Fig. 8 greatly reduces the influence of the short intervals that the simulation removes. This is perhaps fortunate, given the preliminary report by Long *et al.* (2002) that most cochlear implant users, like normal listeners, also perceive these mixtures as having a pitch close to $F2$ Hz. Indeed, we have measured the predictions of the model with and without this additional filtering, and have observed only minor differences in the predictions obtained. The predictions of the model are shown by the solid lines with no symbols in Fig. 5(b), and show reasonable agreement with the mean data, except when $F1=70$ Hz, where the matched pitch is approximately 7% below the predictions.[4]

3. Summary

In this subsection we have shown that a fairly simple model, which selectively weights longer first-order intervals, can account reasonably well for data from four different paradigms. It is acknowledged that the parameters of the model were developed on a *post hoc* basis, in order to best fit the data, and it should also be stressed that the solution we arrived at is unlikely to be unique. Nevertheless, the model rests on only two crucial assumptions, both of which we consider reasonable. One is that the form of $W(\tau)$, although expansive for short and moderate ISIs, should not continue to expand indefinitely as τ is increased to longer and longer values. The other is that the function $f(p_t)$ should also be expansive, thereby reducing the influence of very rare interpulse intervals.

B. Possible physiological correlates and limitations on rate discrimination

1. Comparison to AN recovery data

The function $W(\tau)$ described by the solid line in Fig. 8 could be interpreted as reflecting an increased probability of neural firing in response to a stimulus pulse as the interval since the previous pulse is increased. To evaluate whether our weighting function could simply reflect refractory properties of the AN, we have plotted functions derived from two sets of AN data in Fig. 8. The first of these (thin unbroken line), described by Kiang (1965), shows the probability of a cat AN fiber firing in response to any one pulse in an isochronous acoustic pulse train, as a function of interpulse interval; the duration of the pulse train was 60 s when the interpulse interval was 5 ms or greater, and 24 s. otherwise. The second function (dashed line) is a fit to the data of Fitzpatrick *et al.* (1999), who measured the probability of a cat AN fiber firing to the second of two equal-amplitude clicks, as a function of interpulse interval.

Neither of the two AN recovery functions seems adequate to account for the results. Recall that, in order to fit the results of experiment 1, in which the pitch of the 4–6 stimulus corresponded to an average period of 5.64 ms, the 6-ms intervals had to receive more than 4.5 times the weight of the 4-ms intervals. This is reflected in the form of $W(\tau)$, but both of the AN functions are steepest between about 1

and 3 ms, and show at most a 1.2-fold increase from 4 to 6 ms. Furthermore, our attempts to model the results of experiment 2 using a revised ''mean rate'' approach, which operated not on stimulus interpulse intervals but on AN interspike intervals simulated from Fitzpatrick *et al.*'s (1999) data, were unsuccessful: The resulting pitch matches overestimated those observed by more than 40% at all values of $F1$ studied [Fig. 5(b), solid squares].[5] Finally, it is worth noting that it is unlikely that an account based purely on AN recovery functions could account for the results of experiment 1, which were very similar for electric and acoustic stimulation despite the fact that different peripheral processes were involved in the two cases. This does not mean, of course, that refractoriness effects at higher levels of the auditory system could not produce a function more similar to $W(\tau)$. In this regard, it is worth noting that Fitzpatrick *et al.* also measured recovery functions from the cochlear nucleus, inferior colliculus, and cortex. Although the form of the functions differed between these neural structures, in no case was the increase between 4 and 6 ms as great as that shown by $W(\tau)$.

2. Coding by first-order intervals

Given the above conclusion that the form of $W(\tau)$ cannot be accounted for by properties of AN fibers, a reasonable interpretation of our model is that a more central mechanism performs a weighted combination of first-order intervals between spikes in the AN response to the stimulus. One apparent problem with this approach is that representations based on first-order interspike intervals within a single neuron have the undesirable characteristic of being level dependent (e.g., McKinney and Delgutte, 1999).[6] However, the problem of level dependence can be overcome if the responses of several neurons, innervating separate hair cells, are combined *before* the pitch estimate is made. Indeed, Carlyon (1997) has argued that such a combination of responses *must* take place if the auditory system is to discriminate between a pulse train from which several pulses have been deleted, and an intact pulse train that has been reduced in level: In both cases, individual auditory-nerve fibers will ''miss'' some spikes, but with deleted pulses the same spikes will be missed by all fibers (leading to an ''irregular'' percept), whereas an attenuated intact pulse train will sound smooth but quiet, presumably because different neurons miss different pulses, and because the responses of several neurons will be summed. Note that this differs from the form of analysis used by Cariani and Delgutte (1996), who showed that, for periodic complex tones, pitches estimated from first-order intervals in the auditory-nerve firing patterns were strongly level dependent, even when the responses of multiple fibers were combined. However, they calculated an interval distribution for each and every auditory-nerve fiber, before combining each of these distributions into a summary measure. We think that it is *a priori* very unlikely that a neural pitch mechanism would actually perform this sort of calculation on a fiber-by-fiber basis.[7]

C. Implications for more general models of pitch perception

Section IV B described one way in which a weighted sum of first-order intervals can account for data obtained from a variety of paradigms. It should be stressed, however, that all of these paradigms involved purely temporal pitch perception, in that they involved either electric pulse trains applied to a single channel of a cochlear implant, or acoustic pulse trains whose spectral components were unresolvable by the peripheral auditory system. We do not wish to argue that a similar process necessarily applies to the pitch of pure tones or of complex tones containing resolved harmonics. Indeed, there is substantial evidence that pitches of resolved and unresolved harmonics are extracted by separate mechanisms: That of unresolved harmonics leads to difference limens that depend more strongly on duration (Plack and Carlyon, 1995; White and Plack, 1998), can be ''reset'' by a brief silent interval (Plack and White, 2000a), is affected by preceding and following complexes regardless of their $F0$ (Micheyl and Carlyon, 1998), and cannot be accurately compared to that of a simultaneously presented group of resolved harmonics in a different frequency region (Carlyon and Shackleton, 1994). This conclusion has recently been bolstered by evidence using a transfer-of-learning paradigm (Grimault *et al.*, 2002),[8] and by Kaernbach and Bering's (2001) finding that a given listener's $F0$ difference limen (''$F0$DL'') for a group of unresolved harmonics is a good predictor of that listener's $F0$DL for another group of unresolved harmonics, but not for a group of resolved harmonics. What our analysis *does* imply, however, is that models of pitch perception which rely on higher-order intervals (Meddis and Hewitt, 1991; Patterson *et al.*, 1995; Meddis and O'Mard, 1997) cannot account for the purely temporal form of pitch perception studied here. This is of course a problem for those models that rely on higher-order intervals *and* claim to be able to account for the pitches of both resolved and unresolved harmonics.

D. Summary and conclusions

This article describes two main findings, both of which were obtained using stimuli where place of excitation was unlikely to provide a cue, and both of which have implications for theories of temporal pitch perception. First, the pitch of a pulse train in which the interpulse intervals alternate between 4 and 6 ms corresponds to a period of about 5.7 ms, both in acoustic and electric hearing. Second, when two inharmonically related (ratio = 1.29) pulse trains, each of which was perfectly periodic, were passed through the same bandpass filter and presented acoustically, the pitch corresponded roughly to that of the higher-rate pulse train. We have argued that the results presented here and elsewhere are consistent neither with Carlyon's mean rate model (Carlyon, 1995), nor with ''autocorrelogram'' models that operate on higher-order intervals (Meddis and Hewitt, 1991; Meddis and O'Mard, 1997). Rather, these and other phenomena relating to ''purely temporal'' pitch perception can be captured by a model in which pitch is derived from a weighted sum of the first-order intervals present in the stimulus, and where the

form of this weighting favors longer intervals, at least for interval lengths up to 12 ms. The form of this weighting function is unlikely to result simply from refractory properties of the AN; rather, it is likely to be applied by more central stages to the composite pattern of firing derived from the combination of many AN fibers.

ACKNOWLEDGMENTS

We thank Peter Cariani for useful discussions concerning the limitations of using only acoustic pulse trains to investigate temporal pitch perception. Author CJL was supported by a grant from the Leverhulme Trust to RPC.

[1]In one experiment, Thurlow and Small used a regular 100-Hz pulse train in which an additional pulse had been added x ms after every one of the original pulses. In one condition they passed these stimuli through a ''narrow'' bandpass filter centered on 5 kHz, and observed that as x was increased from 0 to about 2 ms, there was a sharp increase in pitch. This would have caused the longest first-order interval to decrease from 10 to 8 ms, and so is consistent with the *longest* intervals dominating pitch. However, in a subsequent experiment they broadened the bandpass filter (by an unspecified amount) and used a stimulus in which x gradually increased from 0 to 10 ms; presumably, this was achieved by mixing two pulse trains of slightly different rates. They reported that, at small values of x, subjects heard a very high pitch of 1000–2000 Hz, which then dropped to 200 Hz as x was increased to 5 ms, and then rose to 1000–2000 Hz as x was increased further. The fact that these very high pitches were reported only for the broader filter bandwidth, for which place-of-excitation cues would have been strongest, combined with the fact that intervals corresponding to 1/2000 Hz (0.5 ms) would have been obscured by auditory filter ringing, add weight to the conclusion that their data were strongly influenced by ''place'' cues.

[2]Many of these limitations were first pointed out by Peter Cariani in postings to an online discussion group.

[3]Plack and White also studied conditions in which an additional silent interval was inserted just before the first delayed or advanced pulse. The same gap was inserted in the comparison stimulus. Generally, this produced much smaller pitch shifts than in the conditions with no gap, a finding they attributed to the gap ''resetting'' the pitch integration mechanism. They have recently provided further evidence for this ''resetting'' phenomenon (Plack and White, 2000a), which we do not attempt to capture in the present model.

[4]A second complication arises because the distribution of intervals below $1/F2$ is continuous. In order to fit within the framework of Eq. (1), we need to transform this into a discrete distribution by placing interpulse intervals that fall into a narrow range into a single bin. Because the function $f(p_t) = p_t^2$ is expansive, the width of the these bins will influence the effect those intervals have on the predicted pitch estimate: Wider bins will increase the probability within each bin, and will increase the effect of widely dispersed intervals. When plotting Fig. 7 we chose a bin width of 0.1 ms, which corresponds roughly to the change in period that normal listeners can detect in a perfectly periodic pulse train filtered in the same way as those used in experiment 2, and with an $F0$ of about 200 Hz; this is equal to a rate difference of about 4 Hz (Krumbholz *et al.*, 2000; Carlyon and Deeks, 2002). However, it is recognized that the choice of bin widths used is somewhat arbitrary, and that the deviation in predicted matches from $F2$ [the largest peak in the interspike histogram (''ISIH'')] would change as a function of bin width. It is also worth noting that the accuracy of the fits will also depend on the exact form of $f(p_t)$ at very low probabilities.

[5]The predicted mean rate was obtained by weighting the first-order intervals in the stimulus with this function, adding up the total number of intervals remaining, and then dividing by the stimulus duration. The same procedure was applied to the variable (matching) stimulus. The $F0$'s of the regular-rate stimuli matched to each of the $F1$ and $F2$ mixtures used in experiment 2, as predicted by the modified mean rate model, were then calculated.

[6]McKinney and Delgutte (1999) reported that, for a 220-Hz tone, first-order neural intervals were equally frequent at one, two, and three times the period when the tone level was 52 dB SPL, but that a single interval corresponding to the period dominated the response when the level was 67 dB SPL. This observation should be qualified by the fact that we know of no analogous measures for filtered pulse trains like those used here, and that, to the best of our knowledge, no one has performed pitch matches between pulse trains differing greatly in level. However, it seems reasonable to assume that the responses of individual nerve fibers to filtered pulse trains will be level dependent, but that the pitch of these stimuli will not vary dramatically across level.

[7]An additional advantage of combining responses across auditory-nerve fibers is that it could reduce the influence of ''spontaneous'' neural spikes. This would occur if the postsynaptic neuron (e.g., in cochlear nucleus) required more than one input neuron to fire within a narrow time window before generating a spike. Summation is not, however, essential for the influence of spontaneous spikes to be minimized, because this is also achieved by the square-law form of $f(p_t)$.

[8]Grimault *et al.* (2001) trained subjects on $F0$ discrimination of either resolved or unresolved harmonics filtered between 1375–1875 Hz, and then tested discrimination performance for resolved and unresolved harmonics filtered into different frequency regions. Compared to a pretraining baseline, those trained on one type of complex (resolved or unresolved) improved more for complexes of a similar ''resolvability.''

ANSI (**1969**). ANSI S3.6-1969, ''Specifications for audiometers'' (American National Standards Institute, New York).

Blamey, P. J., Dowell, R. C., Tong, Y. C., Brown, A. M., Luscombe, S. M., and Clark, G. M. (**1984**). ''Speech processing strategies using an acoustic model of a multiple-channel cochlear implant,'' J. Acoust. Soc. Am. **76**, 104–110.

Burns, E. M., and Viemeister, N. F. (**1976**). ''Nonspectral pitch,'' J. Acoust. Soc. Am. **60**, 863–869.

Burns, E. M., and Viemeister, N. F. (**1981**). ''Played again SAM: Further observations on the pitch of amplitude-modulated noise,'' J. Acoust. Soc. Am. **70**, 1655–1660.

Cariani, P. A., and Delgutte, B. (**1996**). ''Neural correlates of the pitch of complex tones. I. Pitch and pitch salience,'' J. Neurophysiol. **76**, 1698–1716.

Carlyon, R. P. (**1995**). ''Extracting the fundamental frequencies of two concurrent sounds,'' *A Biological Framework for Speech Perception and Production* (Kyoto, Japan).

Carlyon, R. P. (**1996**). ''Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker,'' J. Acoust. Soc. Am. **99**, 517–524.

Carlyon, R. P. (**1997**). ''The effects of two temporal cues on pitch judgments,'' J. Acoust. Soc. Am. **102**, 1097–1105.

Carlyon, R. P. (**1998**). ''Comments on 'A unitary model of pitch perception [J. Acoust. Soc. Am. **102**, 1811–1820 (1997)],' '' J. Acoust. Soc. Am. **104**, 1118–1121.

Carlyon, R. P., and Deeks, J. M. (**2002**). ''Limitations on rate discrimination,'' J. Acoust. Soc. Am. (in press).

Carlyon, R. P., Demany, L., and Deeks, J. (**2001**). ''Temporal pitch perception and the binaural system,'' J. Acoust. Soc. Am. **109**, 686–700.

Carlyon, R. P., and Shackleton, T. M. (**1994**). ''Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?,'' J. Acoust. Soc. Am. **95**, 3541–3554.

Fitzpatrick, D. C., Kuwada, S., Kim, D. O., Parham, K., and Batra, R. (**1999**). ''Responses of neurons to click-pairs as simulated echoes: Auditory nerve to auditory cortex,'' J. Acoust. Soc. Am. **106**, 3460–3472.

Geurts, L., and Wouters, J. (**2000**). ''A concept for a research tool for experiments with cochlear implant users,'' J. Acoust. Soc. Am. **108**, 2949–2956.

Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., and Collet, L. (**2002**). ''Evidence for two pitch encoding mechanisms using a selective auditory training paradigm,'' Percept. Psychophys. **64**, 189–197.

Guttman, N., and Pruzansky, S. (**1962**). ''Lower limits of pitch and musical pitch,'' J. Speech Hear. Res. **5**, 207–214.

Hoekstra, A. (**1979**). ''Frequency discrimination and frequency analysis in hearing,'' Ph.D. Institute of Audiology, University Hospital, Groningen, Netherlands.

Holton, T., and Weiss, T. F. (**1983**). ''Receptor potentials of lizard cochlear hair cells with free-standing stereocilia in response to tones,'' J. Physiol **345**, 205–240.

Kaernbach, C., and Bering, C. (**2001**). ''Exploring the temporal mechanisms involved in the pitch of unresolved harmonics,'' J. Acoust. Soc. Am. **110**, 1039–1047.

Kaernbach, C., and Demany, L. (**1998**). ''Psychophysical evidence against the autocorrelation theory of auditory temporal processing,'' J. Acoust. Soc. Am. **104**, 2298–2306.
Kiang, N. Y. S. (**1965**). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (M.I.T. Press, Cambridge, MA).
Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (**2000**). ''The lower limit of pitch as determined by rate discrimination,'' J. Acoust. Soc. Am. **108**, 1170–1180.
Licklider, J. C. R. (**1951**). ''A duplex theory of pitch perception,'' Experientia **7**, 128–133.
Long, C. J., Carlyon, R. P., McKay, C. M., and Vanat, Z. (**2002**). ''Temporal pitch perception: Examination of first-order intervals,'' Int. J. Audiol. (in press).
McKay, C. M., and Carlyon, R. P. (**1999**). ''Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains,'' J. Acoust. Soc. Am. **105**, 347–357.
McKay, C. M., and McDermott, H. J. (**1996**). ''The perception of temporal patterns for electrical stimulation presented at one or two intracochlear sites,'' J. Acoust. Soc. Am. **100**, 1081–1092.
McKay, C. M., and McDermott, H. J. (**1998**). ''Loudness perception with pulsatile electrical stimulation: The effect of interpulse intervals,'' J. Acoust. Soc. Am. **104**, 1061–1074.
McKay, C. M., McDermott, H. J., and Carlyon, R. P. (**2000**). ''Place and temporal cues in pitch perception: are they truly independent?,'' ARLO (http://ojps.org/ARLO/top.html) **1**, 25–30.
McKay, C. M., McDermott, H. J., and Clark, G. M. (**1995**). ''Pitch matching of amplitude modulated current pulse trains by cochlear implantees: The effect of modulation depth,'' J. Acoust. Soc. Am. **97**, 1777–1785.
McKinney, M. F., and Delgutte, B. (**1999**). ''A possible neurophysiological basis of the octave enlargement effect,'' J. Acoust. Soc. Am. **106**, 2679–2692.
Meddis, R., and Hewitt, M. (**1991**). ''Virtual pitch and phase sensitivity studied using a computer model of the auditory periphery: Pitch identification,'' J. Acoust. Soc. Am. **89**, 2866–2882.
Meddis, R., and O'Mard, L. (**1997**). ''A unitary model of pitch perception,'' J. Acoust. Soc. Am. **102**, 1811–1820.
Micheyl, C., and Carlyon, R. P. (**1998**). ''Effects of temporal fringes on fundamental-frequency discrimination,'' J. Acoust. Soc. Am. **104**, 3006–3018.
Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1985**). ''Relative dominance of individual partials in determining the pitch of complex tones,'' J. Acoust. Soc. Am. **77**, 1853–1860.
Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (**1991**). ''Complex sounds and auditory images,'' in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 429–446.
Patterson, R. D., Allerhand, M., and Giguère, C. (**1995**). ''Time-domain modelling of peripheral auditory processing: A modular architecture and a software platform,'' J. Acoust. Soc. Am. **98**, 1890–1894.
Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1988**). ''Spiral Vos Final Report, Part A: The Auditory Filterbank.'' Applied Psychology Unit, Cambridge, England, Contract Report: APU 2341.
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.
Plack, C. J., and White, L. J. (**2000a**). ''Perceived continuity and pitch perception,'' J. Acoust. Soc. Am. **108**, 1162–1169.
Plack, C. J., and White, L. J. (**2000b**). ''Pitch matches between unresolved complex tones differing by a single interpulse interval,'' J. Acoust. Soc. Am. **108**, 696–705.
Plomp, R. (**1967**). ''Pitch of complex tones,'' J. Acoust. Soc. Am. **41**, 1526–1533.
Shackleton, T. M., and Carlyon, R. P. (**1994**). ''The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination,'' J. Acoust. Soc. Am. **95**, 3529–3540.
Shannon, R. V. (**1983**). ''Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics,'' Hear. Res. **11**, 157–189.
Small, Jr., A. M., and McClellan, M. E. (**1963**). ''Pitch associated with time delay between two pulse trains,'' J. Acoust. Soc. Am. **35**, 1246–1255.
Thurlow, W. R., and Small, A. M. (**1955**). ''Pitch perception for certain periodic auditory stimuli,'' J. Acoust. Soc. Am. **27**, 132–137.
Tong, Y. C., and Clark, G. M. (**1985**). ''Absolute identification of electric pulse rates and electrode positions by cochlear implant listeners,'' J. Acoust. Soc. Am. **74**, 73–80.
Townshend, B., Cotter, N., Compernolle, D. v., and White, R. L. (**1987**). ''Pitch perception by cochlear implant subjects,'' J. Acoust. Soc. Am. **82**, 106–105.
van Wieringen, A., Carlyon, R. P., Long, C. J., and Wouters, J. (**2002**). ''Pitch of amplitude-modulated irregular-rate stimuli in electric and acoustic hearing,'' J. Acoust. Soc. Am. (submitted).
van Wieringen, A., and Wouters, J. (**1999**). ''Natural vowel and consonant recognition by Laura cochlear implantees,'' Ear Hear. **20**, 89–103.
White, L. J., and Plack, C. J. (**1998**). ''Temporal processing of the pitch of complex tones,'' J. Acoust. Soc. Am. **103**, 2051–2063.

Comodulation masking release in consonant recognition[a)]

Bom Jun Kwon[b)]
Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057-1922

(Received 3 January 2001; revised 19 July 2001; accepted 29 April 2002)

Comodulation masking release (CMR) refers to an improvement in the detection threshold of a signal masked by noise with coherent amplitude fluctuation across frequency, as compared to noise without the envelope coherence. The present study tested whether such an advantage for signal detection would facilitate the identification of speech phonemes. Consonant identification of bandpass speech was measured under the following three masker conditions: (1) a single band of noise in the speech band (''on-frequency'' masker); (2) two bands of noise, one in the on-frequency band and the other in the ''flanking band,'' with coherence of temporal envelope fluctuation between the two bands (comodulation); and (3) two bands of noise (on-frequency band and flanking band), without the coherence of the envelopes (noncomodulation). A pilot experiment with a small number of consonant tokens was followed by the main experiment with 12 consonants and the following masking conditions: three frequency locations of the flanking band and two masker levels. Results showed that in all conditions, the comodulation condition provided higher identification scores than the noncomodulation condition, and the difference in score was 3.5% on average. No significant difference was observed between the on-frequency only condition and the comodulation condition, i.e., an ''unmasking'' effect by the addition of a comodulated flaking band was not observed. The positive effect of CMR on consonant recognition found in the present study endorses a ''cued-listening'' theory, rather than an envelope correlation theory, as a basis of CMR in a suprathreshold task. [DOI: 10.1121/1.1490351]

PACS numbers: 43.66.Dc, 43.71.Es [SPB]

I. INTRODUCTION

Detection of a signal masked by a noise band centered at the signal frequency can be improved by additional noise at spectrally distant frequencies, provided that the temporal envelope of that noise is correlated with the envelope of the noise centered at the signal frequency. This phenomenon, known as comodulation masking release (CMR), has been typically demonstrated in two experimental paradigms: (1) extending the bandwidth of the masker surrounding the signal, similar to a traditional critical band experiment (Fletcher, 1940), and (2) the addition of one or more separate noise bands at different frequencies (flanking bands) to the noise band surrounding the signal (on-frequency band). In band-widening experiments, the detection threshold increases until the masker bandwidth reaches a certain point (roughly one critical band at the signal frequency) and starts to decrease as the bandwidth further increases if the masker is comodulated; however, the threshold increases or stays the same if the masker is uncorrelated across the frequency bands (Hall *et al.*, 1984; Haggard *et al.*, 1990; Schooneveldt and Moore, 1989). The magnitude of CMR is usually assessed by the difference in threshold between the comodulated and uncorrelated conditions. Likewise, in the experiments using one or more flanking bands, the threshold with the comodulated masker is consistently 5–10 dB lower than that with the noncomodulated masker (Hall *et al.*, 1984; Hall, 1986; Cohen and Schubert, 1987; Schooneveldt and Moore, 1987; Moore and Schooneveldt, 1990). The threshold in the comodulation condition is occasionally compared with the threshold measured with a single on-frequency masker—which may be interpreted as an ''unmasking effect,'' but the masking release in this setting does not always occur. CMR tends to increase as more flanking bands are applied, akin to the increase of CMR with the increase in the bandwidth in band-widening experiments. This indicates that as more auditory filters become involved, the effect of CMR becomes more pronounced. Thus, the mechanism of CMR can be understood as across-channel envelope processing.

The theories that have been proposed to account for CMR can be classified into the following two groups: (1) detection by the cue from the ''dip'' in the flanking masker (Buus, 1985), and (2) detection by an analysis of envelope correlation across frequency bands such as an equalization–cancellation (E–C) model (Buus, 1985; Hall, 1986; Richards, 1987, 1988; Green, 1992). While in early years a general consensus existed among researchers that ''multiple cues'' should be taken into consideration to account for CMR (e.g., Hall and Grose, 1988) and some data were better explained by an E–C model (e.g., Hall *et al.*, 1988), later studies seem to lend more support to cued-listening models than the theories based on overall envelope correlation (Grose and Hall, 1989; Moore *et al.*, 1990; Fantini, 1991; Hall and Grose, 1991; Eddins and Wright, 1994; Buss and Richards, 1996).

[a)]A portion of this study was presented at the 140th meeting of the Acoustical Society of America, Newport Beach, California, December, 2000.

[b)]Current affiliation: Cochlear Americas, 400 Inverness Drive South, Suite 400, Englewood, CO 80112. Electronic mail: bkwon@cochlear.com

Can the phenomenon of CMR, primarily a detection advantage, occur in a suprathreshold task? In one experiment (Hall and Grose, 1988), listeners were more sensitive (8 dB on average) to the level change of a narrow-band noise under the comodulated condition than under the noncomodulated condition, but most subsequent studies have found little or no evidence supporting an effect of CMR at a suprathreshold level (Grose and Hall, 1992 and Festen, 1993, for speech recognition; Hall and Grose, 1992, for gap detection; Hall and Grose, 1995, for amplitude discrimination; Hall *et al.*, 1997, for pitch ranking). However, the possibility of an effect of CMR on speech recognition is worth reconsideration. Speech recognition might be more likely to show a CMR effect than the other suprathreshold tasks. Listener's performance of intensity discrimination or pitch perception generally becomes poorer if the duration of the signal becomes too short. The dip-listening strategy does not allow the listener a long duration of ''dip;'' usually it is less than a few tens of milliseconds. On the contrary, the dip listening of a speech waveform with short dips is quite successful: in several speech recognition experiments where temporally fluctuating maskers were used, it was found that listeners do take advantage of dips in the masker as long as temporal resolution permits; it is not difficult to find instances of successful dip listening during masker dips less than 50 ms (e.g., Kwon and Turner, 2001). Thus, if listeners truly take relevant information for detection from the off-frequency cue band regarding when to listen, the possibility exists that such processing may aid suprathreshold speech recognition.

The contrast between two conditions, comodulation and noncomodulation, is clear in most traditional CMR experiments, and the data in these conditions are compared to assess the magnitude of CMR. In previous speech experiments (Grose and Hall, 1992; Festen, 1993), the assessment of CMR was rather indirect because such a contrast was not available or did not exist in their methodology. Grose and Hall's experiment primarily compared the conditions with different number of flanking bands to assess an effect of CMR. This was based on the finding that the magnitude of CMR generally increases as the number of comodulated flanking bands increases (Hall *et al.*, 1988, 1990; Schooneveldt and Moore, 1989). The speech signal consisted of several narrow spectral slits spanning a wide frequency region, and the masker was narrow bands of noise corresponding to each spectral slit plus additional narrow flanking bands. No significant improvement in speech recognition was found as more flanking bands were applied, leading Grose and Hall to conclude that there was ''no convincing effect of CMR.'' It should be noted, however, that the lack of an effect of increasing the number of flanking bands may not reflect the lack of an effect of CMR on speech recognition. This is because the initial number of bands may have provided some across-channel interactions (i.e., CMR) that may not have increased as the number of flanking bands increased (see Hall *et al.*, 1990).

In Festen's experiment (1993), both the target and masker were wideband speech. The masker was divided into several frequency bands, and, under the assumption that the original speech masker had high envelope coherence across frequency, the coherence was reduced by imposing a different time delay to each band. He argued that ''the masking release due to across-frequency signal analysis (CMR)'' was weak—a difference of 1.3 dB in the speech reception threshold (SRT) was measured between the condition with the original masker and the time-delayed masker. However, the weak effect might be attributed to the manipulation of the masker in his experiment: several spectral portions of the speech masker were replaced by a steady noise of 1/3-octave band, which would behave like ''deviant'' bands. It is well known that the insertion of dissimilar bands (deviant bands) among the flanking bands substantially reduces the effect of CMR from other comodulated bands (Hall *et al.*, 1990).

In both studies (Grose and Hall, 1992; Festen, 1993), the speech signal extended over a wide frequency region, which required a substantial modification of the method that is typically used in CMR experiments for detection (band-widening or on-signal/flanking bands). The present study attempted to set up an experimental paradigm similar to the traditional CMR experiments with a clearly defined baseline condition. It was necessary to limit the frequency region of the speech signal to incorporate the traditional CMR paradigm with a flanking band. Thus, bandlimited speech was chosen for the target and three masker conditions were compared: (1) a single masker in the on-frequency (speech) band; (2) two comodulated masker bands (on-frequency and flanking band); and (3) two uncorrelated masker bands. Consonant recognition was measured and the primary focus in the present experiment was to test whether the score in the comodulated condition would be any higher than that in the noncomodulated condition or the single masker band condition.

II. PILOT EXPERIMENT

A. Speech stimulus

As the earlier speech experiments (Grose and Hall, 1992; Festen, 1993) did not demonstrate a clear effect of CMR, it was first necessary to determine whether an effect of CMR on speech recognition even existed and, if it did, whether it was context dependent. Would the effect, if any, be observed under only certain circumstances (e.g., the identification of certain features)? To address this question, it was necessary to use a small number of test tokens that controlled those features. The following two consonant sets were constructed: (1) stop consonants (''stops,'' hereafter), /b, d, g, p, t, k/, and (2) alveolar and dental consonants (''alveolar/dentals,'' hereafter), /d, n, ð, z, t, s/. Since the feature, *stop*, is characterized by the manner of articulation and the features, *alveolar* and *dental*, are characterized by the place of articulation, the features of these two sets are complementary; all stops have the same manner but different place of articulation, whereas all alveolar/dentals have similar place but different manner of articulation. Because temporal waveforms of stops are relatively similar, cues from subtle spectral differences (which roughly correspond to place information) would have to be used for correct identification. In contrast, because the spectral characteristics of alveolar/dentals are roughly similar, subtle differences in the temporal envelope

(manner feature) would have to be used for identification. If any discrepancy were to be found in the effect of CMR between the two sets, it would indicate that the process underlying CMR for consonant recognition is more amenable to one type of feature than another. This, in turn, might provide insight into the mechanism of CMR.

The present experiment used part of a speech database developed by Shannon *et al.* (1999) as the speech material. This database was recorded by 10 speakers (5 males, 5 females) for each phoneme. A bandpass version (1200–2200 Hz) was used as the target speech signal. Ideally, it would have been preferable to use a narrow band of speech for the target, comparable to one auditory bandwidth (roughly one-third octave), to check the interaction of modulation patterns in the masker and target across auditory channels; however, such a choice of narrow band could seriously compromise the intelligibility of the target speech. A bandwidth of 1000 Hz was used to ensure that listeners could reach a score around 75% in the quiet condition.

B. Masker

There were three conditions of the noise masker as follows:

(1) OF-only: one band of masker at the signal frequency; the ''on-frequency'' band only;
(2) COMOD: two bands of masker; the on-frequency band and ''flanking'' band, with comodulation of the temporal envelopes; and
(3) NONCO: same as COMOD, but without comodulation (''noncomodulation,'' hereafter).

The on-frequency band ranged from 1200–2200 Hz (the same as the speech band). The flanking band ranged from 5000–7000 Hz. The coherence of envelopes between the flanking and on-frequency bands was generated by the method proposed by Hall *et al.* (1984). A wideband noise with a Gaussian distribution was multiplied by a low-pass noise with a cutoff frequency of 50 Hz, which produced a wideband noise with envelope fluctuations that were coherent across all frequencies. This multiplied wideband noise was further filtered to obtain the maskers for the on-frequency and flanking bands. For the comodulation condition, the on-frequency and flanking bands were obtained by filtering one source of the multiplied noise. For the noncomodulation condition, the noise bands were derived from two independent noise sources; therefore, there was no envelope correlation between them.

C. Detailed procedures

The task in the present study was identification of the intervocalic consonant in the /aCa/ context. The duration of each speech token ranged from 0.9 to 1.3 second,[1] and the level was 70 dB SPL. This bandlimited speech was presented to the subjects with the noise masker. The duration of the masker was always 0.6 s longer than the speech stimulus presented at each trial so that there was 0.3 s of noise before and after the speech signal. The spectrum level (N_0) of the on-frequency noise was 30 dB, while the rms level of the

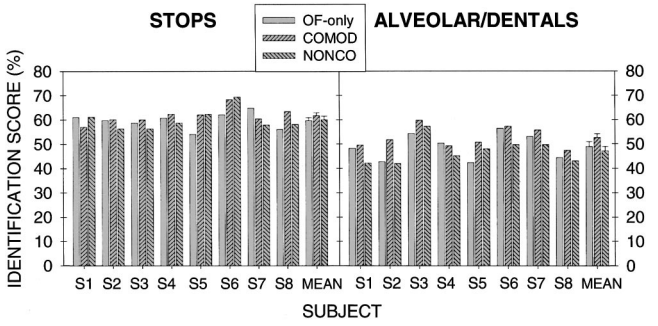


FIG. 1. Results of the pilot experiment. Percent consonant identification scores for each subject for three conditions of the masker, OF-only, COMOD, and NONCO. Left panel: stop consonants, right panel: alveolar/dental consonants.

flanking masker was adjusted to be the same as the on-frequency masker, as previous CMR studies reported that CMR was maximal when the two masker bands were the same level (Hall, 1986; McFadden, 1986; Cohen and Schubert, 1987). Upon each presentation of the stimulus, the subject responded by clicking a mouse on a response screen showing options of consonants. Trial-by-trial visual feedback was given.

Noise samples were digitally created by MATLAB routines. The sampling rate was 44.1 kHz. A low-pass noise (50 Hz) was obtained by a fourth-order elliptic filter. Bandpass filters (fifth-order elliptic filters) for the generation of the bandlimited stimuli (on-frequency and flanking band) had a slope of at least 80 dB/oct. Digitally processed speech and noise were played in a sound field via a loudspeaker located in a sound-attenuating booth. A MATLAB program administered the testing, and all the processing for the generation of the stimuli was done in real time by the program.

Eight normal-hearing subjects participated in the pilot experiment (average age: 28 years, standard deviation: 7 years). All of them had pure-tone thresholds lower than 20 dB HL. They were compensated for their participation in the study. Prior to data collection, each subject was familiarized with the stimuli and experimental settings during practice sessions without the noise masker. As described earlier, there were six experimental conditions (two types of consonants by three types of masker). All of 60 tokens (6 consonants by 10 speakers) in each consonant set were presented in a pseudorandom order in a single experimental run. A total of seven runs was completed for each condition, and the average score of the last five runs was taken as the final data for the condition.[2] One block of 60 trials normally took 4–8 min; therefore, each subject spent 3–6 h after the practice run.

D. Results and discussion

Figure 1 displays the individual scores of all subjects for the three masker conditions, as well as the mean and standard error across subjects. For stop consonants, the mean score was 59.6%, 61.7%, and 60.0% for OF-only, COMOD, and NONCO, respectively. A repeated measures analysis of variance (ANOVA) revealed that the effect of masker condition was not significant [$F(2,14)=0.949$, $p=0.41$]. For alveolar/dental consonants, in contrast, the mean score was

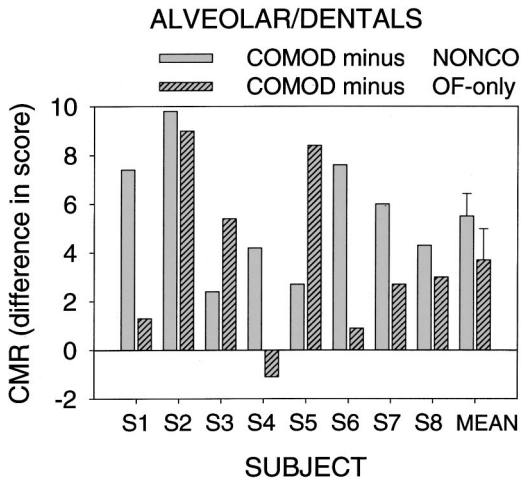


FIG. 2. The magnitude of CMR for alveolar/dental consonants, calculated in two ways; the difference in score between COMOD and NONCO (left bar), and the difference between COMOD and OF-only, an ''unmasking'' effect (right bar).

48.9%, 52.6%, and 47.1% for OF-only, COMOD, and NONCO condition, respectively, and the effect of masker condition was significant [$F(2,14)=9.697$, $p=0.002$].

In previous studies, the magnitude of CMR has often been assessed by the difference in threshold between the comodulation (COMOD) and noncomodulation (NONCO) conditions. This assessment represents the effect of comodulation in the masker bands. A comparison could alternatively be made between the comodulation and on-frequency only (OF-only) conditions. This assessment of CMR represents the ''unmasking'' nature of a comodulated flanking masker—the addition of a flanking band could lower the threshold. Figure 2 shows the magnitude of CMR (the improvement of the recognition score) assessed by these two methods for alveolar/dentals. According to a paired t-test, scores were significantly different between the COMOD and NONCO condition ($p=0.002$), and marginally significant between the COMOD and OF-only condition ($p=0.02$). The mean effect of comodulation was 5.7% (COMOD vs NONCO), and the mean effect of unmasking was 3.7% (COMOD vs OF-only).

The results indicated that a CMR effect was found only when the listeners were expected to identify the manner feature of consonants (e.g., when distinguishing /t/ from /s/), but not for the identification of place information (e.g., when distinguishing /t/ from /k/). This implies that the involvement of CMR in speech recognition is largely limited to manner distinction, and it may explain why previous experiments of speech recognition have found little effect of CMR—vowel recognition is unlikely to be affected by CMR because it does not usually require manner distinction, and consonant recognition would be affected only when the task involves manner distinction. Overall, these pilot data demonstrate a positive effect of CMR on consonant identification under some circumstances and justify the need for further experiments. The main experiment was designed for a larger set of consonants with more masking conditions.

III. MAIN EXPERIMENT

A. Stimulus conditions

The pilot experiment demonstrated that CMR has an effect on consonant identification, at least when listeners make decisions based on the manner features. The main experiment in the present study was designed to assess the effect of the location of the flanking band and the noise level, to identify the factors influencing the effect of CMR on consonant identification. Also, a larger set of consonants was employed in the main experiment to test consonant identification in a more realistic listening situation. The primary interest in the main experiment was to determine (1) whether a CMR effect with a broader choice of consonants would be significant, and (2) whether the ''selective'' involvement of CMR—strong effect for manner distinction and little effect for place—could be confirmed by feature analysis in the experiments with more consonant tokens.

To meet these goals, the following consonants were chosen to construct a consonant set for testing: /b, d, g, p, t, k, z, s, v, f, ʒ, ʃ/. This consonant set was constructed such that the articulation features were balanced across categories.[3] The spectrum level of the noise in the on-frequency band was 24 or 30 dB, and the rms level of the flanking band was equal to that of the on-frequency noise, as in the pilot experiment. Frequencies of the flanking bands were 4000–8000, 3000–6000, or 500–1000 Hz, referred to as FB1, FB2, or FB3, respectively. All the processing strategies and details were the same as in the pilot experiment.

B. Subjects and procedure

Five normal-hearing subjects participated in the experiment (average age: 26 years, standard deviation: 6 years; two of them had participated in the pilot experiment). The details of the experimental protocol were the same as those of the pilot experiment, except that one run of an experimental condition consisted of 120 trials (each of 120 tokens, 12 consonants recorded by 10 speakers, was presented once in one experimental run) and that each condition was repeated three times (and the average of the three repetitions was taken as the final score). Prior to the testing under the masker, the score in quiet was obtained. For the main data collection (with the masker), the condition of the on-frequency band only (OF-only) was initially tested, followed by conditions with the flanking band, i.e., FB1, FB2, and FB3 (the order was balanced across subjects). The 30-dB spectrum level conditions were tested first. Trial-by-trial visual feedback was given for the 30-dB noise level conditions, but not for the 24-dB conditions to avoid possible effects of acclimatization. It normally took 10–15 min for a subject to finish one block of 120 trials, and there were 45 sessions in total;[4] therefore, each listener spent 8–11 h to complete the experiments.

C. Results and discussion

1. General trends

Figure 3 shows the individual scores along with group means and standard errors for the three different masker con-

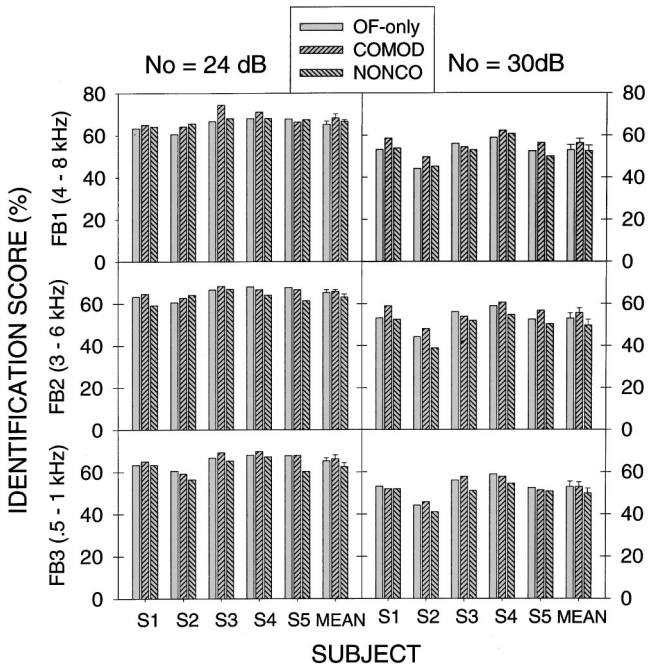


FIG. 3. Results of the main experiment. Percent consonant identification scores for individual subjects for three conditions of the masker, OF-only, COMOD, and NONCO, along with two conditions of masker level (N_0) and three conditions of flanking band frequency. The scores of OF-only at each condition of masker level (N_0) are repeated in the panels for easy comparison.

ditions (OF-only, COMOD, NONCO), the two masker level conditions ($N_0 = 24$,30 dB), and the three flanking band conditions (FB1, FB2, FB3). The OF-only scores are presented in each of the panels for easy comparisons. Figure 4 displays the magnitude of CMR, as calculated by the difference between COMOD and NONCO and also by the difference between COMOD and OF-only. A general trend is that the comodulation condition (the center bar in each condition in Fig. 3) resulted in higher identification scores than the other two masker conditions. A three-way repeated measures ANOVA was performed for identification scores of the comodulation and noncomodulation conditions, of which the factors were the masker level N_0 (2 levels), location of the flanking band (3 levels), and comodulation/noncomodulation (2 levels). The trend of higher scores under the comodulation condition was significant [$F(1,4) = 255.758$, $p < 0.001$]. The effect of location of the flanking band was not significant [$F(2,8) = 5.244$, $p = 0.036$], while the effect of the masker level on *the score* was significant [$F(1,4) = 62.43$, $p = 0.001$]. However, the effect of the masker level on *CMR* (i.e., the difference in scores) was not significant in a two-way repeated measures ANOVA for the magnitude of CMR (COMOD vs NONCO) [$F(1,4) = 1.84$, $p = 0.25$].

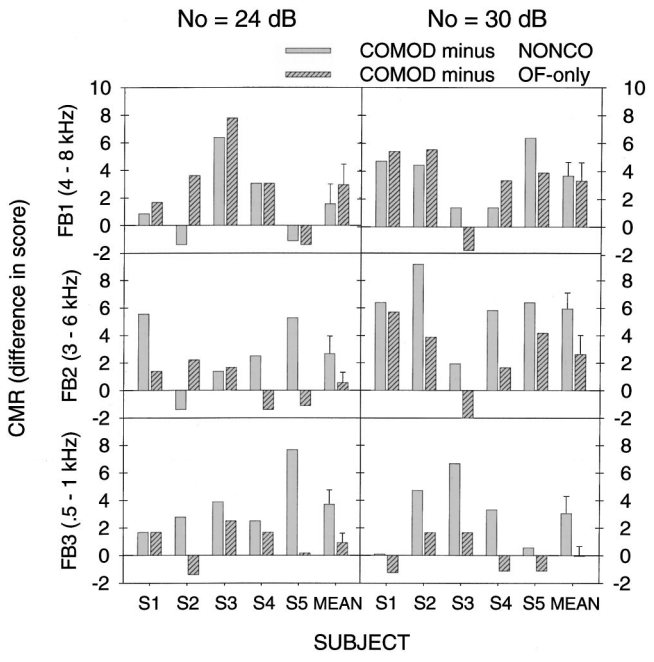


FIG. 4. The magnitude of CMR for each condition and subject, calculated by the difference in score between COMOD and NONCO (left bar), and by the difference between COMOD and OF-only, an ''unmasking'' effect (right bar).

2. The "unmasking" effect

The scores in the comodulation condition were higher for some conditions than those in the OF-only condition, implying an ''unmasking'' effect, i.e., the comodulated flanking masker could improve the scores over the OF-only scores. However, another three-way repeated measures ANOVA for the scores of the comodulation and OF-only conditions did not verify this effect [$F(1,4) = 2.54$, $p = 0.09$]. But, from the observation that the effect of location of the flanking band was significant [$F(2,8) = 22.41$, $p = 0.009$] and that the score differences between the COMOD and OF-only conditions were the least in the FB3 condition, we could suspect that this pattern of result might have been tainted by peripheral masking (or ''within-channel'' masking) from the flanking masker toward the speech target due to the spread of excitation. It should be noted that, for the FB3 condition, the flanking band was lower than (and quite close to) the signal frequencies; therefore, the addition of the flanking band masker might not be favorable regardless of comodulation. Also, due to the wide bandwidth of the speech band, across-channel masking release was likely to be in effect in the OF-only condition. Therefore, it is possible that the unmasking effect was underestimated.

3. Feature transmission analysis

Table I displays group means with standard deviations of the overall identification scores, including the score under the quiet condition, as well as the results of information transmission analysis (Miller and Nicely, 1955), which calculates the percentages of information on specific features that were correctly transmitted to (or identified by) the listeners. The overall score for the quiet condition was 77.1%, and the percentage of information transmission was 78.0%, 82.7%, and 49.7%, in the order of voicing, manner, and place, respectively. The overall scores dropped to 62.5%–68.2% with the 24-dB noise. Notice the dramatic change in manner information (from 82.7% to 39.3%–47.3%) in comparison to the small changes in place information. This indicates that listeners were less able to utilize temporal enve-

TABLE I. Percent score for consonant identification with the standard deviation and results of information transmission analysis (Miller and Nicely, 1955).

			Overall score(%)	Voicing (%)	Manner (%)	Place (%)
Quiet			77.1±3.4	78.0	82.7	49.7
$N_0=$	OF-only		65.3±3.2	57.8	43.2	44.9
24 dB	FB1	COMOD	68.2±4.4	57.2	46.6	49.6
		NONCO	66.7±1.7	57.5	41.7	49.0
	FB2	COMOD	65.8±2.1	53.8	44.9	47.4
		NONCO	63.2±3.0	52.5	39.3	47.1
	FB3	COMOD	66.2±4.3	54.8	47.3	47.1
		NONCO	62.5±4.3	49.4	39.3	45.0
$N_0=$	OF-only		53.0±5.5	42.4	25.6	34.9
30 dB	FB1	COMOD	56.3±4.7	44.1	28.5	37.2
		NONCO	52.6±5.7	42.8	24.8	33.8
	FB2	COMOD	55.6±4.9	44.3	29.2	36.2
		NONCO	49.7±6.2	39.6	20.7	29.8
	FB3	COMOD	52.9±5.0	41.7	26.2	35.3
		NONCO	49.9±5.1	39.9	23.1	30.9

lope cues in the presence of (fluctuating) noise. With the 30-dB noise, scores dropped another 10% to 15% from the 24-dB noise conditions, and the reduction in feature transmission was uniform. As for the effect of comodulation of the masker on feature transmission, in most conditions the percentages of manner information were higher in the comodulation condition than those in the corresponding noncomodulation condition.

Statistical analysis with the percentage of information transmission was also performed to identify the factors accountable for the higher scores in the comodulation condition. A three-way repeated measures ANOVA for the percentages of manner, place, and voicing information revealed that the factor comodulation/noncomodulation was significant for manner ($F=30.56$, $p=0.005$) and for place ($F=21.61$, $p=0.01$), but not for voicing ($F=9.98$, $p=0.34$). This result appears somewhat different from that of the pilot experiment: the effect of place features is now also significant. In the pilot experiment, the effect of CMR (COMOD vs NONCO) was minimal for stops (1.7%) but relatively high for alveolar/dentals (5.5%), suggesting that manner information would be primarily susceptible to CMR. While the effect of place was not always significantly observed (e.g., in the 24-dB condition), the statistical analysis in the main experiment did not provide support for the trend in the pilot experiment. Thus, it would be unwise to argue that CMR generally has more effect when the identification process involves manner than place features; these two features are susceptible to the noise masker to a different degree depending on the level of noise.

IV. GENERAL DISCUSSION

The most important finding from the present study is that an effect of CMR was consistently observed, albeit small, in a consonant recognition task. Recognition scores were generally higher in the comodulation condition than in the noncomodulation condition, and the average difference in score was 3.5% overall. One of the popular theories of CMR is based on the dip-listening (or cued-listening) strategy, which argues that the listener is cued by the flanking band masker as to ''when to listen.'' According to this concept, there should be no reason to disallow an involvement of CMR in speech recognition. Whether it is for detection or identification, useful information may fall in the region of the masker dips, and the auditory system is capable of ''listening-in-the-valleys'' to catch that information to the extent of its temporal resolution abilities. A number of studies have shown that the speech reception threshold (SRT) measured with normal-hearing listeners is 10–15 dB better in the presence of a temporally fluctuating masker compared a steady masker (e.g., Festen and Plomp, 1990; Bacon *et al.*, 1998). The idea that listeners may take advantage of temporal fluctuations in a masker for better performance of speech perception is not new; the present study rather argues that such a ''dip-listening'' activity of the auditory system for speech perception may be facilitated by coherent envelope fluctuations of the masker across frequency bands.

It should be stressed that the effect found in the present experiment is not large. We may still suspect, however, that coherent envelope fluctuations in a masker could improve speech recognition compared to a masker without the coherence. While the difference in consonant recognition score of 3.5% was measured under the conditions tested in the present study, the effect might be larger in other conditions. For example, in Festen's experiment (1993), there was about 10-dB difference in SRT between the most and the least coherent conditions.[5] This implies that the temporal characteristic of a masker needs to be carefully considered in a speech recognition experiment: we know that a temporally fluctuating masker would improve the score compared to a steady noise masker, but we also need to check if the temporal fluctuation is coherent across frequency to maximize the benefit from the dip listening.

While the cued- (or dip-) listening model could account for the effect of CMR on consonant recognition in the present study, the theory based upon envelope correlation analysis may be inappropriate to account for the effect of CMR at a suprathreshold level. This theory maintains that the detection of a signal is based upon the change in the coherent envelope pattern (when the masker bands are comodulated) which would not occur if the signal did not exist. While this explanation may be reasonable to account for CMR as a detection advantage, the process is not likely to provide a cue for a suprathreshold task, because it might ''cue'' that ''there is a signal,'' but it does not necessarily indicate ''what the signal is.''

A direct comparison of the present data to the previous data for detection CMR is complicated by a number of factors. The goals of the tasks were different, the target signal used in this study carried a different aspect of information, and we are uncertain whether (or how much) the principles found in a simple detection task could be applied to a speech identification task. However, it is of interest to compare the pattern of results observed in this study to those of the previous studies. First, in the previous CMR studies for detection the separation between the on-frequency and flanking bands was relatively small, in most studies less than one

octave, and the magnitude of CMR was high when the frequency separation was small, and tended to decrease as the separation grew (Hall *et al.*, 1984; Schooneveldt and Moore, 1987; Cohen and Schubert, 1987). This pattern of results was not observed in the present study. The FB3 condition, where the separation was the smallest, did not produce more CMR than other conditions of the flanking band. This is not difficult to interpret if the way ''within-channel'' masking contributed to CMR is considered. In CMR experiments for detection, it was argued that the perception of a beating pattern (or its change due to a signal) between the on-frequency and flanking band contributed to the detection. In the present experiment, it is not likely that such a beating cue could provide more information for better identification; therefore the within-channel effect was not apparent. It is worthwhile to note that the amount of CMR was not in general dependent upon the location of the flanking band and the amount of CMR in the FB1 condition, where the separation was the largest, was not necessarily smaller than in other conditions. It might imply that CMR for speech recognition occurs by a different mechanism from CMR for detection so that the frequency separation is not critical. If auditory grouping is more heavily involved in CMR for speech recognition than in CMR for tone detection, then the peripheral proximity might not be an important factor influencing the amount of CMR, because auditory grouping largely reflects central auditory processing.

One remaining question is ''Can the auditory system do more than just dip listening, say, can it *actively* extract information on the basis of common amplitude modulation across frequency?'' If this phenomenon is simply dip listening, then the auditory system can only catch information falling in the dip regions. On the other hand, if the auditory system can somehow separate the target from the masker by the pattern of modulation, the information critical to identification would not necessarily have to reside in the masker dip regions. For detection of a signal, studies have shown that CMR occurs only when short signal bursts exist in the dip but not in the peak regions of the masker (Grose and Hall, 1989; Moore *et al.*, 1990). The situation could be different when it concerns speech perception, however, because the speech signal intrinsically carries rich contextual information, and even the removal of some portions of a speech waveform often has little effect on the intelligibility (Warren, 1970); therefore, conceptually it would still be possible to observe a CMR-like effect even if a large chunk of speech information did not coincide with the masker dip regions. Unfortunately, the issue of ''active extraction of information'' was not addressed in the present study.

V. CONCLUSION

The present data and supporting statistical analyses verified the effect of comodulation of background noise bands on consonant identification. Significantly higher scores were obtained in the presence of the comodulated masker than the noncomodulated masker (cf. Fig. 3), and the mean difference in scores ranged from 1.6% to 5.9% for consonant recognition with 12 consonants (cf. the open bars in Fig. 4). The average of these means across the conditions was 3.5%. From the present study, we may generally expect somewhat better speech recognition in the presence of a comodulated masker than an uncorrelated, random masker. The unmasking effect, i.e., the advantage from the comodulated flanking masker compared to the on-frequency-only condition, was not found to be significant in the main experiment. Also, the result in the pilot experiment suggesting that CMR may influence manner processing more than place processing was not further supported by the data in the main experiment. The effect of CMR appears to be unbiased between manner and place processing.

ACKNOWLEDGMENTS

This work was supported by a Quota International Post-Doctoral Scholarship awarded to the author and by a grant awarded to Robert V. Shannon (NIDCD R01-01526). Helpful comments and constructive feedback during the preparation and revision of this manuscript were provided by Sid Bacon, Bob Shannon, Emily Buss, Joost Festen, Chris Turner, John Glavin III, and Monita Chatterjee.

[1]The duration was more dependent upon individual speakers rather than specific phonemes; therefore, a duration cue is unlikely to be a confounding factor.

[2]In order to reduce possible bias of individual subjects, a pair of the two masker conditions, COMOD and NONCO, was tested together in one experimental session: two 60-presentation runs of each condition were combined and randomly permutated, so that listeners could not know whether each trial was comodulated or not.

[3]These 12 consonants are balanced in the voicing (6 voiced, 6 voiceless), manner (6 stops, 6 fricatives), and place feature (a relatively even distribution from the front to the back).

[4]$45=3\times(1+2\times(1+3\times2))$, i.e. (3 repetitions)$\times$[(quiet)+(24- or 30 dB of N_0)$\times${(OF-only)+(FB1,FB2,or FB3)$\times$(COMOD or NONCO)}].

[5]Although Festen (1993) did not consider this difference as evidence of ''across-frequency masking release (CMR)'' because of the small difference after the partial replacement of steady noise with the original masker, this difference clearly demonstrates the effect of temporal coherence of the masker.

Bacon, S. P., Opie, J. M., and Montoya, D. Y. (**1998**). ''The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds,'' J. Speech Lang. Hear. Res. **41**, 549–563.

Buss, E., and Richards, V. M. (**1996**). ''The effects on comodulation masking release of systematic variations in on- and off-frequency masker modulation patterns,'' J. Acoust. Soc. Am. **99**, 3109–3118.

Buus, S. (**1985**). ''Release from masking caused by envelope fluctuations,'' J. Acoust. Soc. Am. **78**, 1958–1965.

Cohen, M. F., and Schubert, E. D. (**1987**). ''Influence of place synchrony on detection of a sinusoid,'' J. Acoust. Soc. Am. **81**, 452–458.

Eddins, D. A., and Wright, B. A. (**1994**). ''Comodulation masking release for single and multiple rates of envelope fluctuation,'' J. Acoust. Soc. Am. **96**, 3432–3442.

Fantini, D. A. (**1991**). ''The processing of envelope information in comodulation masking release (CMR) and envelope discrimination,'' J. Acoust. Soc. Am. **90**, 1876–1888.

Festen, J. M. (**1993**). ''Contributions of comodulation masking release and temporal resolution to the speech-reception threshold masked by an interfering voice,'' J. Acoust. Soc. Am. **94**, 1295–1300.

Festen, J. M., and Plomp, R. (**1990**). ''Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing,'' J. Acoust. Soc. Am. **88**, 1725–1736.

Fletcher, H. (**1940**). ''Auditory patterns,'' Rev. Mod. Phys. **12**, 47–60.

Green, D. M. (**1992**). ''On the similarity of two theories of comodulation masking release,'' J. Acoust. Soc. Am. **91**, 1769.

Grose, J. H., and Hall, J. W. (**1989**). ''Comodulation masking release using

SAM tonal complex maskers: Effects of modulation depth and signal position,'' J. Acoust. Soc. Am. **85**, 1276–1284.
Grose, J. H., and Hall, J. W. (**1992**). ''Comodulation masking release for speech stimuli,'' J. Acoust. Soc. Am. **91**, 1042–1050.
Haggard, M. P., Hall, J. W., and Grose, J. H. (**1990**). ''Comodulation masking release as a function of bandwidth and test frequency,'' J. Acoust. Soc. Am. **88**, 113–118.
Hall, J. W. (**1986**). ''The effects of across-frequency differences in masking level on spectro-temporal pattern analysis,'' J. Acoust. Soc. Am. **79**, 781–787.
Hall, J. W., and Grose, J. H. (**1988**). ''Comodulation masking release: Evidence for multiple cues,'' J. Acoust. Soc. Am. **84**, 1669–1675.
Hall, J. W., and Grose, J. H. (**1991**). ''Relative contributions of envelope maxima and minima to comodulation masking release,'' Q. J. Exp. Psychol. **43A**, 349–372.
Hall, J. W., and Grose, J. H. (**1992**). ''Masking release for gap detection,'' Philos. Trans. R. Soc. London, Ser. B **336**, 331–337.
Hall, J. W., and Grose, J. H. (**1995**). ''Amplitude discrimination in masking release paradigms,'' J. Acoust. Soc. Am. **98**, 847–852.
Hall, J. W., Haggard, M. P., and Fernandez, M. A. (**1984**). ''Comodulation masking release as a function of bandwidth and test frequency,'' J. Acoust. Soc. Am. **76**, 50–56.
Hall, J. W., Grose, J. H., and Haggard, M. P. (**1988**). ''Comodulation masking release for multicomponent signals,'' J. Acoust. Soc. Am. **83**, 677–686.
Hall, J. W., Grose, J. H., and Haggard, M. P. (**1990**). ''Effects of flanking band proximity, number, and modulation pattern on comodulation masking release,'' J. Acoust. Soc. Am. **87**, 269–283.
Hall, J. W., Grose, J. H., and Dev, M. B. (**1997**). ''Signal detection and pitch ranking in conditions of masking release,'' J. Acoust. Soc. Am. **102**, 1746–1754.
Kwon, B. J., and Turner, C. W. (**2001**). ''Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference?'' J. Acoust. Soc. Am. **110**, 1130–1140.
McFadden, D. (**1986**). ''Comodulation masking release: Effects of varying the level, duration, and time delay of the cue band,'' J. Acoust. Soc. Am. **80**, 1658–1667.
Miller, G. A., and Nicely, P. E. (**1955**). ''An analysis of perceptual confusions among some English consonants,'' J. Acoust. Soc. Am. **27**, 338–352.
Moore, B. C. J., and Schooneveldt, G. P. (**1990**). ''Comodulation masking release as a function of bandwidth and time delay between on-frequency and flanking-band maskers,'' J. Acoust. Soc. Am. **88**, 725–731.
Moore, B. C. J., Glasberg, B. R., and Schooneveldt, G. P. (**1990**). ''Across-channel masking and comodulation masking release,'' J. Acoust. Soc. Am. **87**, 1683–1694.
Richards, V. M. (**1987**). ''Monaural envelope correlation perception,'' J. Acoust. Soc. Am. **82**, 1621–1630.
Richards, V. M. (**1988**). ''Components of monaural envelope correlation perception,'' Hear. Res. **35**, 47–58.
Schooneveldt, G. P., and Moore, B. C. J. (**1987**). ''Comodulation masking release (CMR): Effects of signal frequency, flanking band frequency, masker bandwidth, flanking band level, and monotic versus dichotic presentation of flanking band,'' J. Acoust. Soc. Am. **82**, 1944–1956.
Schooneveldt, G. P., and Moore, B. C. J. (**1989**). ''Comodulation masking release (CMR) as a function of masker bandwidth, modulator bandwidth, and signal duration,'' J. Acoust. Soc. Am. **85**, 273–281.
Shannon, R. V., Jensvold, A., Padilla, M., Robert, M. E., and Wang, X. (**1999**). ''Consonant recordings for speech testing,'' J. Acoust. Soc. Am. **106**, L71–L74.
Warren, R. M. (**1970**). ''Perceptual restoration of missing speech sounds,'' Science **167**, 392–393.

The mid-level hump at 2 kHz[a)]

Lance Nizami,[b)] Jason F. Reimer,[c)] and Walt Jesteadt
Center for Hearing Research, Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 11 October 2001; revised 19 April 2002; accepted 22 April 2002)

Shortening the duration of a Gaussian-shaped 2-kHz tone-pip causes the intensity-difference limen (DL) to depart from the ''near-miss to Weber's law'' and swell into a mid-level hump [Nizami *et al.*, J. Acoust. Soc. Am. **110**, 2505–2515 (2001)]. For some subjects the size of this hump approaches or exceeds the size reported for longer tones under forward masking, suggesting that forward masking might make little difference to the DL for very brief probes. To test this hypothesis, DLs were determined over 30 to 90 dB SPL for a brief Gaussian-shaped 2-kHz tone-pip. DLs were obtained first without forward masking, then with the pip placed 10 or 100 ms after a 200-ms 2-kHz tone of 50 dB SPL, or 100 ms after a 200-ms 2-kHz tone of 70 dB SPL. DLs inflated significantly under all forward-masking conditions. DLs also enlarged under an 80 dB SPL forward masker at pip delays of 4, 10, 40, and 100 ms. The peaks of the humps obtained under forward masking clustered around a sensation level (SL) that was significantly lower than the average SL for the peaks of the humps obtained without forward masking. Overall, the results do not support the neuronal-recovery-rate model of Zeng *et al.* [Hear. Res. **55**, 223–230 (1991)], but are not incompatible with the Carlyon and Beveridge hypothesis [J. Acoust. Soc. Am. **93**, 2886–2895 (1993)] that nonsimultaneous maskers corrupt the memory trace evoked by the probe. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1485970]

PACS numbers: 43.66.Fe, 43.66.Dc, 43.66.Ba [NFV]

I. INTRODUCTION

The study of intensity-difference limens (DLs) has traditionally focused on stimuli of 50–500 ms duration, with which it has been shown (e.g., Jesteadt *et al.*, 1977) that the level-dependence of the DLs for pure tones follows the ''near-miss to Weber's law'' (McGill and Goldberg, 1968) and is uniform as a function of frequency. However, a mid-level hump in the DL, called the ''severe departure from Weber's law,'' has been observed for tones of 6.5 or 8 kHz, having steady-state durations of 26 or 16 ms (Carlyon and Moore, 1984). A hump is also evident for 6-kHz tones of 25-ms steady-state duration (Plack and Viemeister, 1992a), for 6-kHz tones having 2-ms plateaus and 2-ms onset- and offset-ramps (Plack, 1998), and for 8- and 10-kHz tones of 500 ms duration (Florentine *et al.*, 1987). Data at lower frequencies are more limited and show a greatly reduced effect. Plack and Viemeister (1993) established DLs for pure tones of 1 or 4 kHz having plateau durations of 5 ms and 2-ms on- and off-ramps. Two of the three subjects show a marginal mid-level hump for the 1-kHz tone, perhaps a transition state between a mid-level hump, and the near-miss to Weber's law that was evident for the third subject. In contrast, all three subjects showed a clear mid-level rise for the 4-kHz tone.

A mid-level hump does, however, become evident for frequencies of less than 6 kHz, when the circumstances are right. In one circumstance, the level dependence of the DL shifts from the near-miss to Weber's law, to the mid-level hump, as the equivalent rectangular duration is shortened from 10 to 2.5 ms (Nizami *et al.*, 2001). In the second circumstance, a mid-level hump appears for 1-kHz tones having an equivalent rectangular duration D (see Baer *et al.*, 1999) of about 25 ms, if the pedestal or probe is preceded by a stimulus whose frequency spectrum contains the frequency of the pedestal (Zeng *et al.*, 1991; Zeng and Turner, 1992; Turner *et al.*, 1994; Plack *et al.*, 1995). Plack and Viemeister (1992b) and Plack *et al.* (1995) also found mid-level humps for 1-kHz tones under *backward* masking, a phenomenon whose relevance will be mentioned in Sec. VII.

Different models have been proposed for the mid-level hump seen for very brief Gaussian-shaped tone-pips (Nizami *et al.*, 2001), and for the mid-level hump seen for forward-masked stimuli (Zeng *et al.*, 1991; Carlyon and Beveridge, 1993), such that the effects of shortening a tone, and the effects of forward masking it, should be complementary. Thus the imposition of a forward masker should cause further inflation in the size of the mid-level hump seen by Nizami *et al.* (2001) for the non-forward-masked, very brief Gaussian-shaped 2-kHz tone-pip. If the DLs do not inflate, all the models will have to be reexamined. Thus it is important that DLs be obtained for the Nizami *et al.* tone-pip under forward masking. These measurements would also fill an information gap, as the literature contains no reports of DLs under forward (or backward) masking for any stimuli of $D < 10$ ms.

Careful DL measurements can also resolve a problem associated with the significance of mid-level humps. Inspec-

[a)]A portion of this work was presented as ''The mid-level hump in the intensity-difference limen at 2 kHz enlarges significantly under forward masking,'' at the 25th Annual Meeting of the ARO, St. Petersburg Beach, FL, January 2002.

[b)]Electronic mail: nizami@boystown.org
[c)]Present address: Department of Psychology, California State University at San Bernardino, 550 University Parkway, San Bernardino, CA 92407-2397.

tion of the literature reveals profound across-subject differences in whether or not a hump actually appears, what its magnitude is, and where its peak is located in dB SPL (see Zeng *et al.*, 1991; Plack and Viemeister, 1992a, b, 1993; Zeng and Turner, 1992; Carlyon and Beveridge, 1993). It is thus difficult to state definitively whether humps (or differences between humps) are genuine, no matter how self-evident they may appear, because the standard deviations associated with the within-subject averages of the forward-masked DLs or of the non-forward-masked DLs sometimes approach the size of the average DLs themselves. Also, the standard deviations associated with the across-subject average DLs, the data usually taken as the definitive summaries, can vary greatly from study to study. The standard deviations for the forward-masked DLs are less than 2 dB in Zeng *et al.* (1991), for example, but may exceed 4 dB in Turner *et al.* (1994), although both studies used identical stimulus conditions, similar methods, and show a similar hump, which reaches 6 dB. These differences must surely lead to differences in whether or not the effect is statistically significant.

Such problems can be mitigated by performing analyses of variance on the data, using the greatest number of subjects possible and the greatest number of probe levels possible, in order to improve statistical power. In the present series of experiments, the level-dependence of the DL was obtained for the discrimination of a 2-kHz tone-pip having an 8-ms Gaussian-shaped envelope that represents a D of 2.51 ms. This particular tone was chosen only because of its use in a previous paper. In that paper (Nizami *et al.*, 2001), a plot of the level-dependence of the DL shows a mid-level hump that is statistically significant, whose peak is the highest of those seen for the three tone durations for which group data are available (nine subjects, $D=1.25$, 2.51, 10.03). In the present work, the level-dependence of the DL was sampled at seven probe levels for nine subjects under three conditions of forward-masker level and postmasker delay. A fourth condition was provided by the non-forward-masked DLs. In the last experiment to be described here, three other subjects provided DLs at seven probe levels for an 80 dB SPL forward masker and four postmasker delays. The results are discussed within the framework of the models that were crafted to explain the effect of the forward masker on the level-dependence of the DL, viz., the model of Zeng *et al.* (1991) and the model of Carlyon and Beveridge (1993).

II. EXPERIMENTS 1, 3, AND 5: DIFFERENCE LIMENS WITHOUT FORWARD MASKING

Intensity-difference limens were obtained for the non-forward-masked tone-pip. This procedure was then repeated twice during the course of the study in order to check whether the subjects' experience of experiment 2 (forward-masked detection thresholds) or experiment 4 (forward-masked difference limens) influenced their non-forward-masked DLs. These checks were not employed in experiment 6, which had a different emphasis and used a different subject group.

A. Subjects

The nine listeners in experiments 1–5 were students from Creighton University, four male (average age=22) and five female (average age=19). All were paid volunteers, informed as to the method but not the expectations of the study. Subject DA had some previous experience as a laboratory listener. A different set of three subjects was employed in experiment 6 (as described later). All of the subjects' quiet thresholds were $\leqslant$20 dB SPL at 500 Hz, 1 kHz, and 4 kHz in both the left and right ears for sinusoids having plateau durations of 200 ms.

B. Apparatus and stimuli

All stimuli were generated digitally at a sampling rate of 50 kHz using an array processor (TDT AP2). Signals were played through 16-bit digital-to-analog converters (TDT DD1). The pedestal or comparison tone was generated on channel 1 of the DD1, beginning at a positive-going zero crossing, while the increment was generated in quadrature phase on channel 2. The output of each channel was low-pass filtered at 10 kHz (TDT FT6) and attenuated (TDT PA4), then the outputs of the two channels were combined (TDT SM3) and presented to the listener though a headphone buffer (TDT HB6), a remote passive attenuator in the soundproof chamber, and a Sennheiser HD 250 linear II headphone. Parallel use of multiple attenuators, summers, and headphone buffers made it possible to test as many as four listeners simultaneously. The subjects were thus assigned to groups, within which they remained.

All stimuli were presented to the left ear. To obtain quiet thresholds, only the increment channel was used. The tone-pips used as pedestals and increments were created by multiplying 2-kHz carriers by Gaussian envelopes that were smoothly reduced to zero amplitude at ± 4 standard deviations (σ), so that total stimulus duration was $8\sigma = 8D/\sqrt{2\pi} \approx 3.19D$. The pip, whose total duration was 8 ms, can be thought of as consisting of a 4-ms half-Gaussian ''up'' ramp followed by a 4-ms half-Gaussian ''down'' ramp, with no steady-state portion. Stimulus levels are reported as dB SPL of the carrier.

C. Procedure

Estimates of the intensity DL were obtained using adaptive tracking of the level corresponding to 70.7% correct (Levitt, 1971). Each adaptive track consisted of 50 trials, each trial consisting of two successive observation intervals, both containing the pedestal. Within each block of trials, the increment was randomly distributed between the first and second intervals. Otherwise, the pedestal and the increment were of the same length, had the same Gaussian envelope, and appeared simultaneously whenever they occupied the same observation interval.

The subject's task was to identify which of the two listening intervals on each trial contained the more intense tone-pip. The starting level of the increment was typically set 15 dB higher than the level of the pedestal, allowing the subjects to easily identify the incremented tone on the first few trials. In each 50-trial block, 4-dB steps were followed

by 2-dB steps after the fourth reversal. The reversal points for the 2-dB steps were averaged to obtain a threshold level of the increment. The DL was defined as $\mathrm{DL}=10\log_{10}[1+(\Delta I/I)]$ where $\Delta I/I=10^{\mathrm{Threshold}/10}/10^{\mathrm{Pedestal\ level}/10}$.

The subjects observed a 16-character message window at the top of a keypad small enough to be held in the hands. At the start of each trial, a small asterisk appeared in the message window to indicate impending stimuli. The first stimulus interval started after 400 ms and lasted 50 ms, with the stimuli starting at the beginning of the interval. There followed a 600-ms period of silence, followed by another 50-ms stimulus interval. The subjects then had all the time they desired to choose the interval that they believed contained the increment, indicated by pressing the button corresponding to the first interval or the button corresponding to the second interval. Typically, no more than a few seconds was required. There was then a feedback interval of 400 ms during which the message window displayed the identity of the interval, first or second, that had contained the increment. There then followed a 500-ms delay before the next trial. When more than one subject was tested at the same time, the feedback interval started at the time of the last response.

The subjects were given three blocks of 50 trials as training, for the 90 dB SPL pedestal, the 60 dB SPL pedestal, and the 30 dB SPL pedestal, in that order. Subjects showed proficiency at the task by the end of the third block. The subjects then completed two 50-trial blocks at each pedestal level, moving from 30 to 90 dB SPL in steps of 10 dB. Between the 40 and 50 dB SPL pedestal levels, two blocks were inserted to measure quiet threshold. The entire sequence was then repeated in reverse order, and then the whole cycle was repeated once more, providing a grand total of eight measurements of the DL at each pedestal level, and a grand total of eight measurements of the quiet threshold. Arithmetic averages and standard deviations were then computed.

D. Results

The subjects' absolute detection thresholds ranged from 16.9 to 35.3 dB SPL with an arithmetic average of 22.6 dB SPL. A subject's DLs were considered valid only when the subject's quiet threshold was more than 5 dB below the level of the pedestal, a criterion found elsewhere to be reliable (Nizami *et al.*, 2001). Subjects AL, KG, and NH had quiet thresholds of 28.2, 28.5, and 35.3 dB SPL, respectively, and so the DLs were not computed for AL, KG, and NH for the 30 dB SPL pedestal, and the DL was not computed for NH for the 40 dB SPL pedestal.

The size of the DL and its pattern of level-dependence were found to vary greatly across subjects, as noted earlier (Nizami *et al.*, 2001). Meaningful patterns can nonetheless be found. Figure 1 (upper panel) presents the across-subject arithmetic average DLs and their standard deviations. The level-dependence of the DL shows a clear mid-level hump whose peak is situated somewhere between 40 and 60 dB SPL. This pattern can also be seen not only in Nizami *et al.* (2001) but in Plack (1998, 6-kHz tone, $D\approx 3$ ms).

Experiment 1 was followed by experiment 2, forward-masked detection thresholds. Experiment 1 was therefore subsequently repeated as experiment 3 in order to examine whether the experience gained in experiment 2 affected the mid-level hump that was demonstrated for the non-forward-masked tone-pip in Nizami *et al.* (2001) and in experiment 1. Experiment 3 employed the same subjects, apparatus, stimuli, and procedure as used in experiment 1, with the

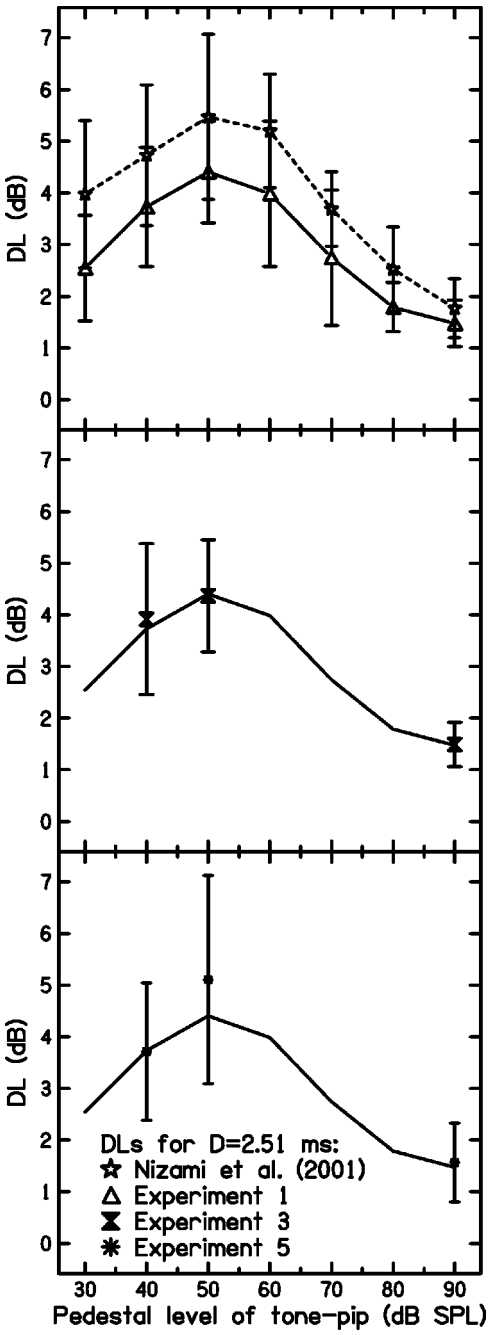


FIG. 1. Difference limens for a 2-kHz Gaussian-shaped tone-pip, averaged over nine subjects (data of experiments 1, 3, and 5). The error bars are standard deviations. Three of the subjects could not provide DLs at 30 dB SPL and one subject could not provide DLs at 40 dB SPL (see text). (Upper panel) Results of experiment 1. Provided for comparison are average DLs over nine other subjects for the same stimulus, from Nizami *et al.* (2001). (Middle panel) Results of experiment 3, the first replication of experiment 1 (same nine subjects). DLs were obtained only at 40, 50, and 90 dB SPL. The solid line connects the group DLs from the upper panel. (Lower panel) Results of experiment 5, the second replication of experiment 1 (same nine subjects). DLs were obtained only at 40, 50, and 90 dB SPL. The solid line connects the group DLs from the upper panel.

exception that, due to time constraints, the pedestal levels were restricted to 40, 50, and 90 dB SPL. The middle panel of Fig. 1 shows that there was no notable drift in the quiet thresholds, or in the DLs.

Experiment 3 was followed by experiment 4, forward-masked level discrimination. Experience gained in experiment 4 could conceivably help the subjects improve their discrimination for the non-forward-masked stimulus. Thus experiment 5 was done, in which non-forward-masked DLs were again obtained at 40, 50, and 90 dB SPL. Quiet thresholds were not obtained due to lack of time. The lower panel of Fig. 1 shows that the DLs at 40 and 90 dB SPL did not notably differ from those obtained in experiment 1. The group-average DLs at 50 dB SPL were worse than before and the standard deviations were larger, reflecting poorer performance by just two of the subjects, who produced DLs that were notably larger than before, probably due to inattention.

III. EXPERIMENT 2: FORWARD-MASKED DETECTION THRESHOLDS

In order to obtain meaningful forward-masked DLs, the level of the pedestal must be far enough above the forward-masked detection threshold that the tone-pip can be heard clearly on both intervals of any trial. In order to establish what pedestal levels could be heard under what circumstances of forward masking, forward-masked detection thresholds were obtained for the 2-kHz Gaussian-shaped tone with a variety of forward-masker levels and postmasker delays.

A. Subjects, apparatus, stimuli, and procedure

The apparatus, stimuli, and procedure were the same as used in experiments 1, 3, and 5, with the following exceptions. The 2-kHz Gaussian-shaped tone-pip was presented only on channel 2, previously used for the ''increment.'' Channel 1 now contained only a 2-kHz forward masker. The forward masker was ramped up using the 4-ms-long rising portion of the same Gaussian envelope used for the probe. The forward masker was ramped down using the 4-ms-long falling portion of the Gaussian envelope used for the probe. In between, the profile was flat, with a duration of 200 ms, for a total forward-masker duration of 208 ms. The forward masker preceded the tone-pip by either 3, 10, 30, or 100 ms, as measured from the end of the forward masker to the start of the tone-pip. This gap was fixed for any given block of trials. Each trial consisted of two stimulus intervals each lasting 500 ms, with the stimuli starting at the beginning of each interval.

The subject's task was to identify which of the two listening intervals contained the tone-pip. The starting levels for the tone-pip were high enough to allow the subjects to choose unambiguously which interval contained the tone-pip for the first few trials. A complete set of detection thresholds was established as a function of forward-masker level at the 3-ms delay before proceeding on to the 10-ms delay, then to 30 ms, then to 100 ms. At the start of each new delay condition, the subjects were given three blocks of 50 trials as training, one for the 90 dB SPL forward masker, one for the 70 dB SPL forward masker, and one for the 50 dB SPL forward masker, in that order. Subjects showed proficiency by the end of the third block. For each delay condition, the subjects completed two 50-trial blocks at each forward-masker level, moving from 50 to 90 dB SPL in steps of 10 dB. Steps of 5 dB were used for the 3-ms delay. [Forward-masker levels of <50 dB SPL were not used because they have little effect on forward-masked DLs (Zeng and Turner, 1992), at least for 100-ms narrow-band maskers centered at 1 kHz.] Once the forward-masker level reached 90 dB SPL, the data-gathering process was repeated in reverse order, and then the whole cycle was repeated once more, providing a grand total of eight forward-masked pip-detection thresholds per subject at a given delay.

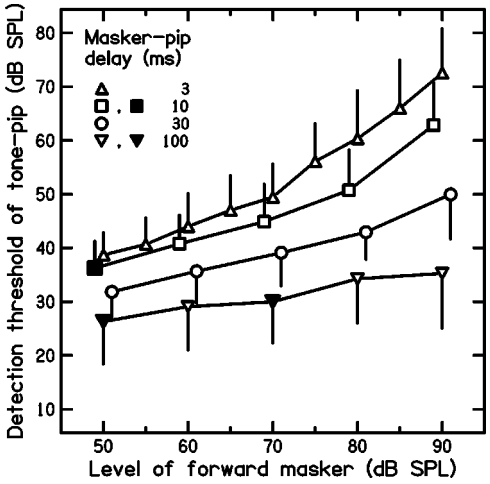


FIG. 2. Detection thresholds for a 2-kHz Gaussian-shaped tone-pip presented 3, 10, 30, or 100 ms following a 200-ms, 2-kHz tone whose level is indicated on the abscissa (experiment 2). These are the across-subject averages for the same nine subjects whose DLs appear in Fig. 1. The error bars are standard deviations. The thresholds representing the conditions chosen for experiment 4 are shown as solid symbols. For the sake of clarity, the plot for the 10-ms delay has been shifted to the left by 1 dB, and the plot for the 30-ms delay has been shifted to the right by 1 dB. The subjects' absolute detection thresholds for the Gaussian-shaped tone-pip ranged from 16.9 to 35.3 dB SPL with an arithmetic average of 22.6 dB SPL.

B. Results

For any given masker-probe delay, there was substantial variation between individuals in the shift in detection threshold caused by the forward masker's presence, and in the rates of increase of the detection thresholds with increase in forward-masker level (the slope of the growth-of-forward-masking curve). Nonetheless, common patterns emerged. Figure 2 presents the across-subject arithmetic averages and standard deviations. These curves resemble those for 4-kHz tone-pips of 10-ms total duration preceded at 0, 10, or 30 ms by 4-kHz 200-ms tones (Oxenham and Plack, 2000; Schairer *et al.*, 2001).

IV. EXPERIMENT 4: DIFFERENCE LIMENS UNDER FORWARD MASKING

The group-average detection thresholds in Fig. 2 were used to select combinations of forward-masker level and probe delay for which the tone pedestals should be clearly

audible despite the presence of the forward masker. Delays of 10 and 100 ms were chosen in order to confirm the apparent rise of the DL as the masker-probe delay increases from 12.5 to 100 ms (Plack *et al.*, 1995). At the 10-ms delay only the lowest masker level tested, 50 dB SPL, appeared to guarantee that the subjects could provide real DLs at the lowest of the desired pedestal levels, 30 and 40 dB SPL. At the 100-ms delay, masker levels of 50 and 70 dB SPL were chosen.

A. Subjects, apparatus, stimuli, and procedure

The apparatus, stimuli, and procedure were the same as used in experiment 1, with the following exceptions. Channel 1 now contained both the 200-ms 2-kHz forward-masker described under ''experiment 2,'' and the pedestal, both of which were presented at a fixed level in both intervals of every trial. Channel 2 contained the increment in quadrature phase, presented randomly in either interval 1 or interval 2. The forward masker preceded the pedestal (also in channel 1) and the increment (channel 2) by 3, 10, 30, or 100 ms, as measured from the end of the masker to the start of the tone-pip. This gap was fixed for any given block of trials. Each trial consisted of two stimulus intervals lasting 500 ms, with the stimuli starting at the beginning of each interval.

The subject's task was to identify which of the two listening intervals contained the more intense tone-pip following the forward masker. Starting levels for the increment were set high enough that the subjects could choose unambiguously which interval contained the incremented tone-pip, for the first few trials. A complete set of DLs was established as a function of pedestal level for the 70 dB SPL forward masker and the 100-ms delay. The DLs were then obtained for the 50 dB SPL forward masker and the 100-ms delay, followed by the 50 dB SPL forward masker and the 10-ms delay. At the start of each new masker-level/probe-delay condition, the subjects were given three blocks of 50 trials as training, one block for the 90 dB SPL pedestal, one block for the 60 dB SPL pedestal, and one block for the 30 dB SPL pedestal, in that order. For each masker-level/probe-delay condition, the subjects completed two 50-trial blocks at each pedestal level, moving from 30 to 90 dB SPL in steps of 10 dB. This process was then repeated in reverse order, and then the whole cycle was repeated once more, providing a grand total of eight measurements of the forward-masked difference limen for a given masker-level/probe-delay condition.

B. Results and analysis

Despite the efforts made to take detectability into account, there were still instances in which the pedestal level of the tone-pip was less than 5 dB above the forward-masked detection threshold. In those cases the pedestal tended to fade in and out of hearing range, so that a difference limen could not be resolved. Those data were not used. This problem happened for all the subjects when the 30 dB SPL tone-pip followed the 50 dB SPL forward masker at 10 ms, and it happened for all but three subjects for the 40 dB SPL tone-pip under the same forward-masking conditions. It happened for four subjects when the 30 dB SPL tone-pip followed the 50 dB SPL forward masker at 100 ms, and still happened for one of those subjects at the 40 dB SPL pedestal level. Finally, the pedestal level of the tone-pip was less than 5 dB above the forward-masked detection threshold for six subjects when the 30 dB SPL tone-pip followed the 70 dB SPL forward masker at 100 ms. One of those subjects could still not always hear the pedestal repeatedly when its level was raised to 40 dB SPL.

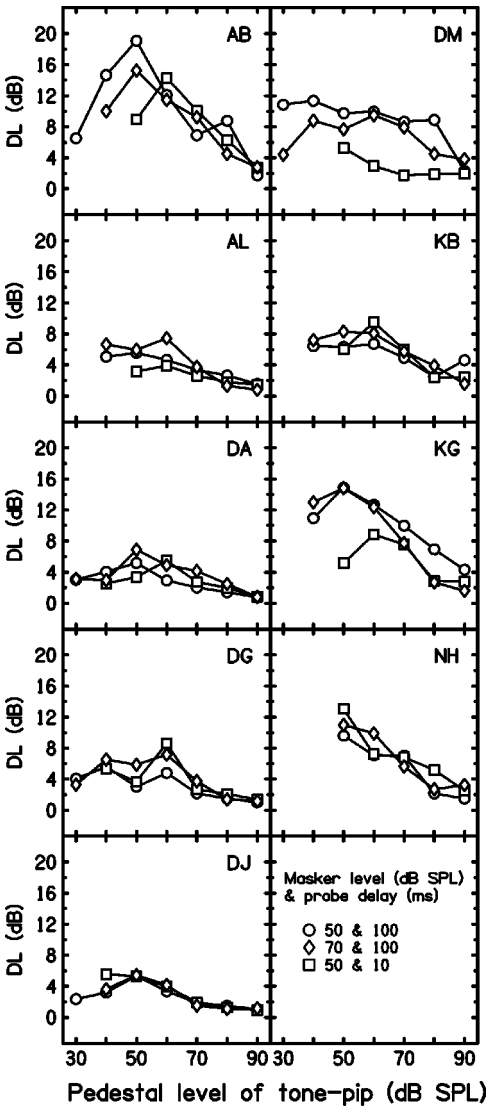


FIG. 3. Forward-masked difference limens for each of the nine subjects (experiment 4). Each frame shows one subject's forward-masked DLs for three conditions: 50 dB SPL forward-masker and 10 ms delay (open squares), 50 dB SPL forward-masker and 100 ms delay (open circles), and 70 dB SPL forward-masker and 100 ms delay (open diamonds).

Figure 3 shows DL versus pedestal level for each subject at each of the three masker-level/probe-delay conditions. Figure 4 presents across-subject arithmetic averages and standard deviations. For the 100-ms delay, the DLs appear to peak at 50 dB SPL, just as found for the non-forward-masked DLs, regardless of forward-masker level. The DLs for the 70/100 and 50/100 masker-level/probe-delay conditions do not appear to substantially diverge. The level-dependence of the DLs for the 50/10 masker-level/probe-delay condition appears to peak at a higher pedestal level.

In order to lend some rigor to these assertions, the data

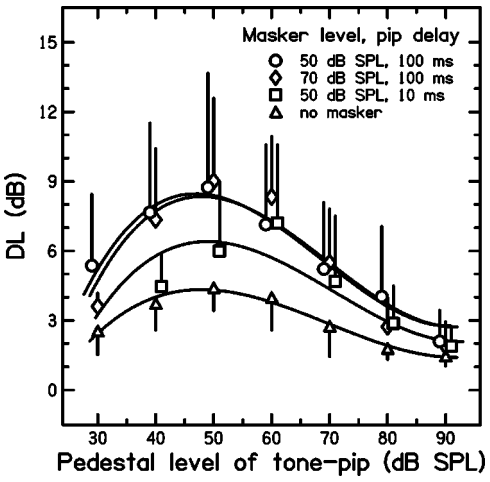


FIG. 4. The forward-masked difference limens (Fig. 3) averaged across the nine subjects. The error bars are standard deviations. For the sake of clarity, the plot for the 50 dB SPL forward masker and 100-ms delay has been shifted 1 dB to the left and the plot for the 50 dB SPL forward masker and 10-ms delay has been shifted 1 dB to the right. Also shown are the non-forward-masked DLs from the upper panel of Fig. 1. The smooth curves represent regression-fitted polynomials described in the text.

used to construct Fig. 4 were subjected to a two-way ANOVA (forward-masking condition by pedestal level), using the interaction with subjects as the error term. Four forward-masking conditions were stipulated: 50 dB SPL masker and 100-ms delay, 50 dB SPL masker and 10-ms delay, 70 dB SPL masker and 100-ms delay, and no forward masker. None of the DLs for the two lowest pedestal levels (30 and 40 dB SPL) were used, because not all subjects were able to provide DLs at those levels (see above). Significant effects were found for forward-masking condition [$F(3,24) = 6.240$, $p < 0.01$], pedestal level [$F(4,32) = 40.764$, $p < 0.001$], and the interaction of forward-masking condition with pedestal level [$F(12,96) = 7.711$, $p < 0.001$].

These significant effects were explored with further tests. Figure 4 shows large standard deviations that might render illusory any apparent differences between points on a given curve. Consequently, a separate one-way ANOVA was completed for each forward-masking condition, and the linear, quadratic, and higher-order components of the level effect were evaluated (see Keppel, 1991, pp. 141–161). These components are orthogonal and provide a means of describing the nature of a significant effect of pedestal level. The linear component represents a pattern in which the DL varies as a function of level but can be described by a straight line with nonzero slope, as in a "near miss to Weber's law," while the quadratic component provides a measure of the significance of deviation from that linearity in the form of a single peak or trough, as in a mid-level hump. There were significant effects of pedestal level, and a significant linear component, for each of the four masker-level/probe-delay conditions, 50/100 {[$F(4,32) = 14.085$, $p < 0.001$], and [$F(1,8) = 22.267$, $p < 0.01$], respectively}, 50/10 {[$F(4,32) = 13.629$, $p < 0.001$], and [$F(1,8) = 41.777$, $p < 0.001$], respectively}, 70/100 {[$F(4,32) = 36.899$, $p < 0.001$], and [$F(1,8) = 47.071$, $p < 0.001$], respectively}, and no-forward-masker {[$F(4,32) = 24.961$, $p < 0.001$], and [$F(1,8) = 22.267$, $p < 0.001$], respectively}. The consistent lack of a quadratic component presumably reflects the exclusion of DLs at 30 and 40 dB SPL.[1]

The drop in DL with increasing pedestal level, with or without forward masking, is apparently genuine. However, the overall effect could conceivably have been due to the inclusion of the non-forward-masked DLs in the initial analysis. Consequently, ANOVAs were done in which the non-forward masked DLs were excluded. The DLs for the 50 dB SPL forward masker were analyzed with a two-way ANOVA [delay (10 or 100 ms) by pedestal level], using the interaction with subjects as the error term. Pedestal level was significant [$F(4,32) = 22.015$, $p < 0.001$], but delay was not, suggesting no difference between the two plots for the 50 dB SPL forward masker for pedestal levels from 50 dB SPL onwards.

The DLs for the 100-ms delay were also analyzed with a two-way ANOVA [forward-masker level (50 or 70 dB SPL) by pedestal level], using the interaction with subjects as the error term. Pedestal level was significant [$F(4,32) = 27.070$, $p < 0.001$], but forward-masker level was not. There was a significant interaction between forward-masker level and pedestal level [$F(4,32) = 2.834$, $p < 0.05$].

The four curves in Fig. 4 appear to converge to a common DL for a pedestal level somewhere just above 90 dB SPL. This impression was confirmed by a set of ANOVAs in which pedestal level was held constant and the effect of forward-masking condition was examined. The effect of forward-masking condition was significant at 50 dB SPL [$F(3,24) = 6.353$, $p < 0.005$], at 60 dB SPL [$F(3,24) = 6.020$, $p < 0.005$], at 70 dB SPL [$F(3,24) = 4.196$, $p < 0.05$], and at 80 dB SPL [$F(3,24) = 3.066$, $p < 0.05$]. At 90 dB SPL, the effect of forward-masking condition was not significant, i.e., there was no significant difference between the DLs.[2]

The ANOVAs indicate that the DLs obtained in the presence of the forward masker for pedestal levels of 50–90 dB SPL are greater than those obtained in the masker's absence. Figure 4 also shows smooth curves, the lowest of which was obtained by fitting a third-order polynomial to DLs obtained in the absence of the forward masker. The three upper functions were obtained by multiplying the lowest function by 1.955 for the 50/100 condition, 1.927 for the 70/100 condition, and 1.478 for the 50/10 condition. With these multipliers, the three polynomials account for 96.3%, 90.6%, and 73.0% of the variance, respectively. The smaller percentage for the 50/10 condition perhaps reflects the shift of the peak to the right. Overall, however, the DLs obtained in the presence of a forward masker are proportional to those obtained in the absence of a forward masker across all pedestal levels and the DLs obtained in the presence of one forward masker are proportional to those obtained in the presence of another. This does not necessarily mean that forward-masked and non-forward-masked DLs have the same origin. The proportionality rule cannot be general, in view of the incidents in which there is no mid-level hump in the absence of the forward masker (e.g., Zeng *et al.*, 1991; Zeng and Turner, 1992). Several studies do show mid-level humps in the absence of a forward masker and apparently larger DLs in the presence of a forward masker (e.g., Carlyon and Beveridge,

1993, Fig. 1; Plack and Viemeister, 1992a, Fig. 3; Schlauch *et al.*, 1997, Figs. 3 and 6), and this data was examined in the same manner as that of Fig. 4. While simple multiplication works very well in some cases, such as S3 in Fig. 3 of Schlauch *et al.* (1997), it does not work in the majority of cases, primarily because the hump in the absence of the forward masker is too low to account for the hump seen under forward masking.

V. EXPERIMENT 6: DIFFERENCE LIMENS UNDER AN 80 dB SPL FORWARD MASKER

This experiment was actually completed before experiments 1–5, but is presented last for continuity of logic.

A. Subjects, apparatus, stimuli, and procedure

The subjects were LN (the first author, age 41), and two female students from Creighton University, SWO (age 22) and AS (age 21), paid volunteers informed as to the method but not the expectations of the study. The two students had no previous experience as psychophysical listeners.

The apparatus, stimuli, and procedure are the same as described under experiment 4, with the following exceptions. The level of the 200-ms 2-kHz forward masker was 80 dB SPL. The masker preceded both the pedestal (channel 1) and the increment (channel 2) by 0, 10, 40, or 100 ms. A complete set of DLs was first established as a function of the tone-pip's pedestal level for the 40-ms probe delay. A complete set of DLs was then obtained for the 10-ms delay, followed by the 100-ms delay, and finally the 0-ms delay. For each delay condition, the subjects performed a forward-masked detection task between doing the DLs for the 40 dB SPL pedestal condition and the DLs for the 50 dB SPL pedestal condition. In this detection task, the subjects completed two 50-trial blocks in which only the increment tone appeared (as in experiment 2). Thus for each delay the subjects provided a grand total of eight measurements of the forward-masked difference limen for each pedestal level, and a grand total of eight measurements of the forward-masked pip-detection threshold.

After forward-masked DLs and forward-masked detection thresholds were obtained, non-forward-masked DLs and quiet thresholds were obtained, as done in experiment 1.

B. Results and analysis

Figure 5 presents the within-subject average DLs found with and without forward masking. The standard deviations have been omitted for the sake of clarity, but for the forward-masked DLs, the standard deviations have the same orders of magnitude and same dependence on pedestal level as seen for the group averages of Fig. 4, and for the non-forward-masked DLs, the standard deviations have the same orders of magnitude and same dependence on pedestal level as seen for the group averages of Fig. 1. Forward-masked DLs were rejected if the forward-masked detection threshold was less than 5 dB below the level of the pedestal, just as in the analysis for experiment 4. Similarly, non-forward-masked DLs were rejected if the quiet threshold was less than 5 dB below the level of the pedestal, just as in the analysis for experiment 1. The non-forward-masked DLs for subject LN were anomalously flat compared to the previous performance of this subject for this stimulus (Nizami *et al.*, 2001), and so LN was retested with pedestals at 30, 60, and 90 dB SPL. The resulting DLs followed the peaked pattern seen in Fig. 1, reconfirming the mid-level hump for this subject.

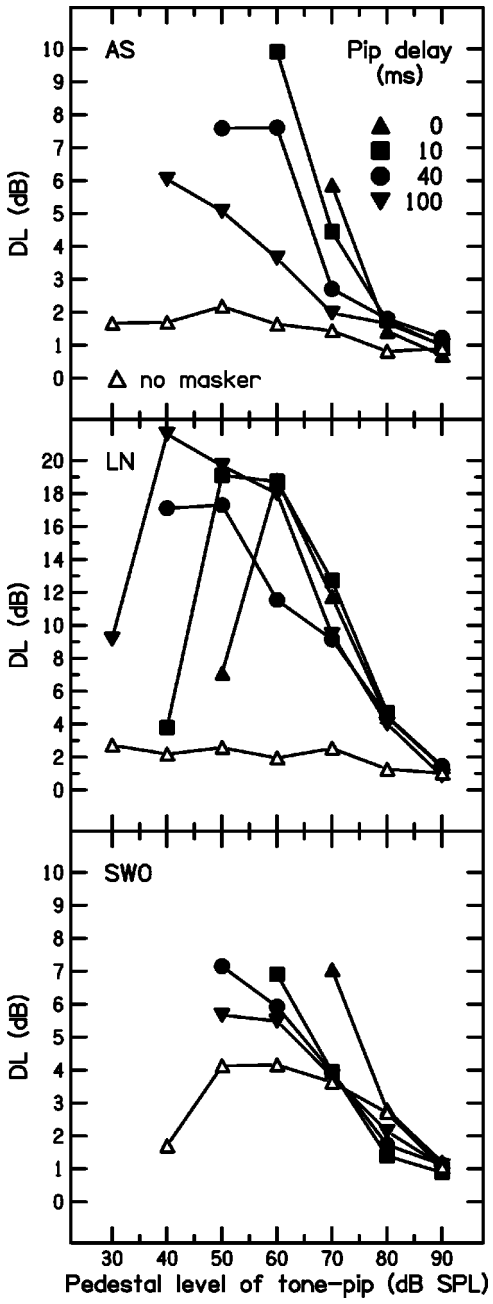


FIG. 5. Difference limens for a 2-kHz Gaussian-shaped tone-pip presented 0, 10, 40, or 100 ms following an 80 dB SPL, 200-ms, 2-kHz tone (experiment 6). The three subjects were not part of the group of nine subjects used earlier. Note the different vertical scale used for LN.

At the delays of 0, 10, 40, and 100 ms, respectively, the subjects' forward-masked detection thresholds were 43.7, 34.5, 28.1, and 23.8 dB SPL (subject LN); 62.4, 50.7, 41.8, and 36.9 dB SPL (subject SWO); and 58.7, 47.4, 37, and 28.2 dB SPL (subject AS). The subjects' quiet thresholds were 20.8 dB SPL (LN), 27.3 dB SPL (SWO), and 20.5 dB SPL (AS), numbers that are well within the range found for

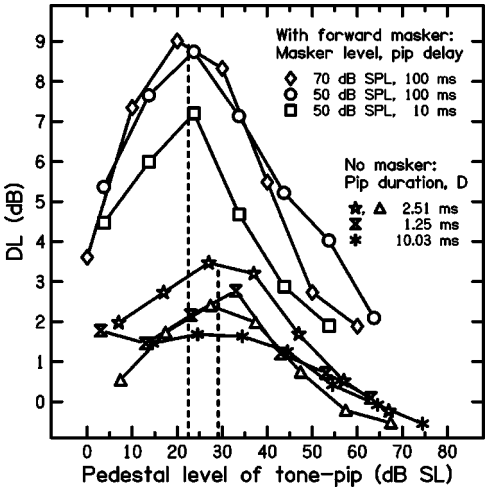


FIG. 6. The forward-masked DLs of Fig. 4, presented in SL scales that were constructed as described in the text. Error bars have been omitted for clarity. Also shown, but offset 2 dB downwards for clarity, are the non-forward-masked DLs for the Gaussian-shaped 2-kHz tone-pip of $D=2.51$ ms (open triangles, as in Fig. 1), as well as the DLs for the Gaussian-shaped 2-kHz tone-pips of $D=1.25$ ms, $D=2.51$ ms (open stars, as in Fig. 1), and $D=10.03$ ms, from Nizami *et al.* (2001). The *x*-intercept of the left-hand vertical line was determined by averaging the SLs at which the three upper curves appear to peak, and the *x*-intercept of the right-hand vertical line was determined by averaging the SLs at which the humps for the one curve for $D=1.25$ ms and the two curves for $D=2.51$ ms appear to peak.

the other nine subjects (experiment 1). Only subject LN had forward-masked detection thresholds low enough to obtain forward-masked DLs at low pedestal levels, revealing a return from the mid-level hump to the lower DLs seen at high SPLs. The other subjects' forward-masked DLs suggest a peak somewhere in the range of 40–60 dB SPL, the same range seen for the position of the peak for the forward-masked DLs at the 100-ms delay for the nine subjects of experiment 4.

VI. SL ANALYSIS OF EXPERIMENTS 1-6

Experiments 1, 3, and 5 confirmed the mid-level hump for 2-kHz Gaussian-shaped tone-pips for $D=2.51$ ms, that had first been observed by Nizami *et al.* (2001). In that study, the level-dependence of the DL had been determined for non-forward-masked 2-kHz Gaussian-shaped tone-pips of $D=1.25$, 2.51, and 10.03 ms. The resulting level-dependence curves achieved closer mutual alignment when plotted using SL scales. A common SL scale had been made for each D by subtracting the group average quiet threshold at each D from the pedestal level. For $D=2.51$ ms, this average threshold was 23.0 dB for the Nizami *et al.* (2001) DLs (nine subjects) and 22.6 dB for the present DLs (nine different subjects). Figure 6 shows all these DLs, past and present, in SL scales. Because the peaks of the two level-dependence curves for $D=2.51$ ms align in SPL scales (top panel, Fig. 1), the similarity of the group average thresholds allows them to also align in SL scales.

Comparing Fig. 6 to Fig. 4 shows that the peaks in the level-dependence plots for the forward-masked DLs are in closer alignment when plotted in an SL scale. The peaks of the non-forward-masked DLs, including those of Nizami *et al.* (2001), occur at a level that is 6.6 dB higher, a significant difference ($t=2.877$, $p<0.05$, two-tailed *t*-test, $df=4$). Thus the mid-level hump in the forward-masked DLs, and that in the non-forward-masked DLs, may have different origins.

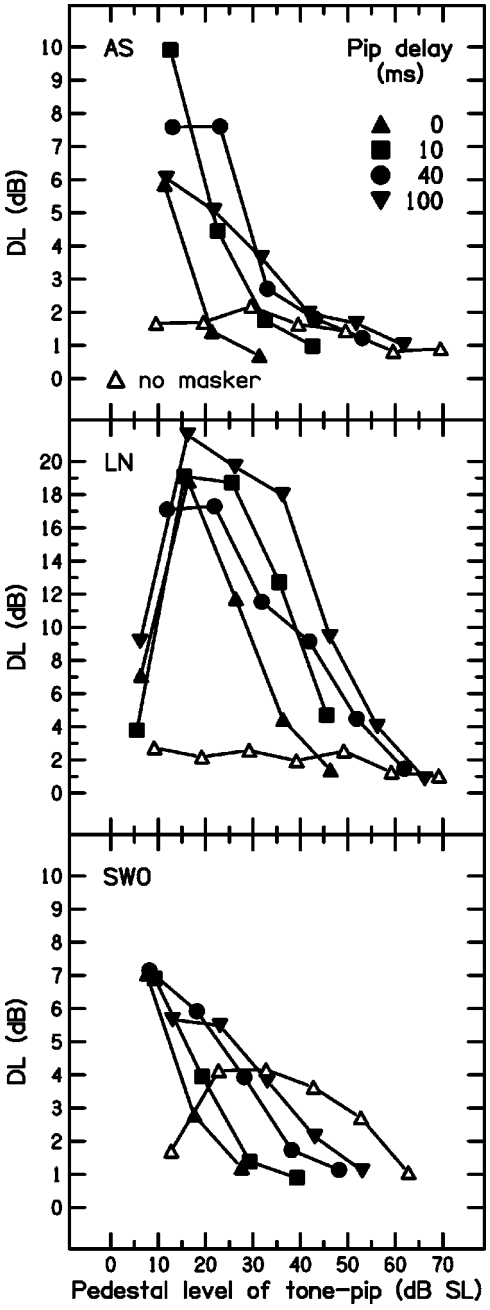


FIG. 7. Difference limens for a 2-kHz Gaussian-shaped tone-pip presented 0, 10, 40, or 100 ms following an 80 dB SPL, 200-ms, 2-kHz tone (data of Fig. 5), displayed using SL scales constructed as described in the text.

Figure 7 shows the effect of an SL transform on the DLs from experiment 6 (that was seen in Fig. 5). The subjects' individual forward-masked detection thresholds were used to make the SL scales for the forward-masked DLs, and the subjects' individual quiet thresholds were used to make the SL scales for the non-forward-masked DLs. The forward-masked DLs of LN, where peaks are well defined, come into alignment, suggesting a peak at 20 dB SL, similar to the forward-masked DLs in Fig. 6. The largest forward-masked DLs of the other two subjects also fall into respective align-

ment. These alignments suggest that each subject's level-dependence curve has a peak somewhere in the range of 10–20 dB SL.

VII. DISCUSSION

A. Absence of learning effects on the non-forward-masked mid-level hump

The group-average DLs for the non-forward-masked 2-kHz Gaussian-shaped tone-pip (Fig. 1, upper panel) did not improve with extensive experience at forward-masked detection of the probe, or with extensive experience at forward-masked discrimination of level changes in the probe (Fig. 1, middle and lower panels, respectively). It appears that the mid-level hump is not an artifact that can be removed through training.

B. The Zeng *et al.* (1991) model of the mid-level hump under forward masking

Zeng *et al.* (1991) were the first to note the emergence of a mid-level hump under forward masking. They used probe delays of 100 ms to obtain their principal data, and the literature (including the present work) has followed suit. Zeng *et al.* explained the forward-masked mid-level hump by noting that at their probe delay of 100 ms, neurons of different spontaneous firing rates (and hence, on average, different thresholds) would be in different states of recovery from the diminishing effect that the forward masker is known to have on the neuronal response to a probe. High-spontaneous-rate neurons would have recovered from the influence of the masker, but would fire at their saturation firing rates in response to a mid-level probe. In contrast, low-spontaneous-rate neurons would not be recovered at all; due to the forward masker, their dynamic ranges would still be shifted upwards in the intensity dimension. Consequently, the auditory periphery is not as sensitive to changes in probe level at moderate probe levels as at lower or higher probe levels.

Our review of the current data and of the published data suggests two problems with the Zeng *et al.* model. The first problem concerns probe delay. Zeng *et al.* (1991) obtained DLs as a function of probe delay using, for each subject, the pedestal level at which that subject had shown a maximal DL for the 100-ms delay (40, 50, or 60 dB SPL). In this manner, they found a monotonic decrease in the size of the DL with increasing delay, which they assumed reflected the relatively slow recovery of low-spontaneous-rate neurons, supporting their model. Plack *et al.* (1995) investigated the delay-dependence of the DL using a probe of the same frequency as Zeng *et al.* (1 kHz) but with longer ramps (5 vs 2 ms) and a shorter steady state (20 vs 25 ms). The level of the probe was 50 dB SPL and the level of the 1-kHz forward masker was 80 dB SPL. Plack *et al.* (1995) found that the DLs actually increased as the probe delay increased from 12.5 to 50 ms, and continued to rise at a slower rate as delay increased to 200 ms. This finding moves opposite in direction to that of Zeng *et al.* (1991).

It is at first tempting to blame this inconsistency on the differences in the stimuli used in the two studies. The forward maskers used by Plack *et al.* were shorter (30 vs 100 ms), and for nonsimultaneous masking Plack *et al.* had used 1-kHz tones whereas Zeng *et al.* (1991) had chosen a 200-Hz narrow-band noise centered at 1 kHz, the frequency of the probe. Plack *et al.* (1995) interpreted the inconsistency between their DLs and those of Zeng *et al.* (1991) as a reflection of referential coding, in which the random amplitude fluctuations of the narrow-band masker used by Zeng *et al.* would reduce its usefulness as a reference relative to the forward masker used by Plack *et al.* (1995), that had the same frequency, duration, and ramping as the probe. Indeed, Zeng *et al.* had chosen a narrow-band masker specifically because it did *not* sound like the 1-kHz probe, whereas the forward masker used by Plack *et al.* (1995) introduces a potential confound, that of ''perceptual similarity'' (Schlauch *et al.*, 1997, 1999). But if perceptual similarity is a problem, then the DLs of Plack *et al.* should have been larger than those of Zeng *et al.* (1991), not smaller as observed. Perceptual similarity is also contraindicated by the dependence of the size of the forward-masked DL on the center frequency of the narrow-band-noise masker (Zeng *et al.*, 1991).

The present data suggest an explanation for the different delay-dependences found by Zeng *et al.* (1991) and by Plack *et al.* (1995). Experiment 4 provides DLs for the probe tone at two delays following the 50 dB SPL forward masker, 100 and 10 ms. Figures 4 and 6 show that the DL at the peak of the level-dependence curve, the plotting parameter used by Zeng *et al.* (1991, Fig. 3), is lower for the 10-ms delay than for the 100-ms delay, in accord with Plack *et al.* (1995), regardless of whether plotted using SPL scales or SL scales. The DLs of experiment 6 qualify this relation. In experiment 6, DLs had been obtained at various durations following an 80 dB SPL forward masker at the probe frequency (2 kHz). Those DLs (Fig. 5) seem to become smaller with increasing delay, in accordance with Zeng *et al.* (1991). The data points cluster together at high pedestal levels, however, making interpretation difficult, and suggesting that the trend may depend on the pedestal level. Therefore the DLs of Fig. 5 were replotted as a function of delay, subject-by-subject, for pedestal levels of 50, 60, and 70 dB SPL. These levels were chosen because Fig. 5 shows that, above 70 dB SPL, there seems to be little difference in the DL with level, and, below 50 dB SPL, only one subject of the three could detect the pedestal. The DLs for subjects AS and SWO, shown in Fig. 8, generally decline with increasing delay for the 50 and 60 dB SPL pedestals, but flatten out for the 70 dB SPL pedestal. The DLs for LN, the oldest of the three subjects, actually rise over 50 to 100 ms postmasker. In sum, although the data of Fig. 8 do not allow us to unambiguously choose between the delay-dependence found by Zeng *et al.* and that found by Plack *et al.*, there is a hint nonetheless that those delay-dependences may differ according to pedestal level.

Our second issue with the Zeng *et al.* model is the effect of forward-masker level. For the 100-ms delay, the present experiments provide DLs for three levels of the forward-masker, 50, 70, and 80 dB SPL. Although the DLs obtained in the presence of the 80 dB SPL forward masker appear to be slightly larger than the others at low pedestal levels and slightly smaller at high pedestal levels, the variability asso-

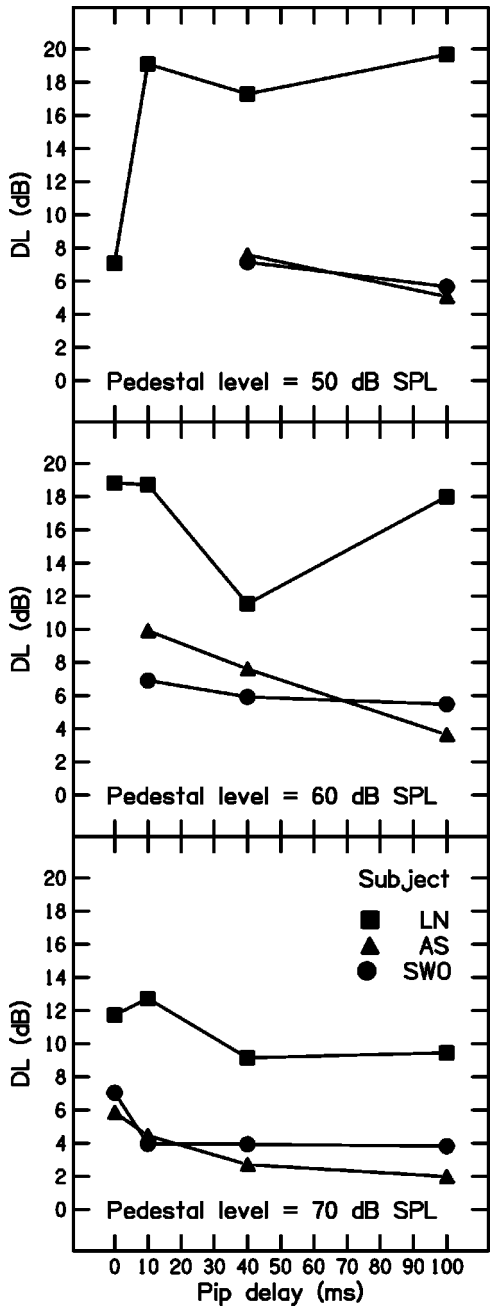


FIG. 8. The difference limens of Fig. 5, plotted as a function of tone-pip delay with the pedestal level of the tone-pip as the parameter.

ciated with these points is high and the number of subjects is lower than for the other two forward maskers. In general, masker level appears to have no effect. This result stands in stark contrast to the only published report of the effect of forward-masker level on the size of the mid-level hump. Zeng and Turner (1992) noted a general increase in the size of the average DL as a function of forward-masker level for levels 40 dB SPL and above, from about 2 dB at 40 dB SPL to about 7 dB at 90 dB SPL. In their study, the masker-probe delay was 100 ms and the stimuli were centered at 1 kHz. The units used for the DL were the same as used here. Zeng and Turner (1992) hypothesized that the elevation of the DL at moderate levels of the probe was only substantial at high levels of the forward masker because afferents of low spontaneous rate have high thresholds, so that high forward-masker levels are required to cause a significant forward-masking effect. Presumably, within the context of their model, Zeng and Turner meant that only high forward-masker levels would shift the dynamic ranges of the low-spontaneous-rate afferents enough to open up a substantial encoding gap. [Increasing a forward-masker's intensity does increase the threshold shift of the primary afferents, and lengthens the duration required for recovery (Smith, 1977; Harris and Dallos, 1979).] The present finding of a substantial mid-level hump for the 100-ms delay condition when the forward-masker level is only 50 dB SPL, and the lack of an effect of the forward-masker level on the size of the mid-level hump, run counter to the Zeng *et al.* (1991) model.

C. Context-coding models of the mid-level hump under forward masking

Plack and Viemeister (1992b) demonstrated that a profound rise in the DL is found when the masker *follows* the probe, for the stimuli and masker-probe interval originally used by Zeng *et al.* (1991). Plack *et al.* (1995) confirmed this result for ipsilateral and contralateral backward-maskers. The unlikelihood of a peripheral interaction in backward masking implies a central mechanism in which the representation of the backward masker and the pedestal or pedestal +increment overlap in a temporal window. A central interaction of some kind is also implied by the existence of contralateral effects of forward masking on intensity discrimination (e.g., Plack *et al.*, 1995; Zeng and Shannon, 1995; Schlauch *et al.*, 1999), because the left- and right-ear inputs do not have the opportunity to interact below the level of the brainstem (Irvine, 1986).

The longer integration times and contralateral effects can be accounted for in models that incorporate assumptions concerning memory. Carlyon and Beveridge (1993) provided the first such account, based on the two modes of intensity coding described by Durlach and Braida (1969). It is generally assumed that subjects in an intensity discrimination task rely on traces of the sounds in auditory memory, but Durlach and Braida described an alternative coding strategy based on context coding and verbal labeling that would generally be less accurate, but that would become the optimum strategy when the memory trace was poor. They noted that context coding was better in the proximity of perceptual anchors and that the top or bottom of the stimulus range, marked by threshold and discomfort, respectively, served as natural anchors (Braida *et al.*, 1984). Carlyon and Beveridge (1993) suggested that the forward or backward masker might be expected to interfere with memory for the tones to be discriminated, but that the effect of this might not be as observable for tones presented at low or at high levels, because the subjects would be able to fall back on context coding in those cases. At moderate levels, however, context coding would be poor, and either coding strategy would result in elevated DLs. Plack has interpreted the results of more recent studies in terms of context coding or referential coding (Plack *et al.*, 1995; Plack, 1998) and has shown that conditions that would be expected to alter referential coding have an impact on the magnitude of the mid-level hump. The occurrence of mid-level humps at comparable levels in condi-

tions with and without forward maskers presents problems for context-coding models. We review those issues in the following sections, following a brief review of the data.

The high correlation between the position of mid-level humps obtained at high frequencies with and without a forward masker (Carlyon and Beveridge, 1993; Plack and Viemeister, 1992a), together with the stronger relationship shown in Fig. 4, suggest a common mechanism underlying mid-level humps observed with and without nonsimultaneous maskers. Plack (1998) has argued that referential coding results in better performance in all cases and thus may be the common mechanism. The argument is based on the observation that notched noise presented with the pedestal reduces the mid-level hump at high frequencies (Plack, 1998), as well as in the presence of nonsimultaneous maskers (Plack and Viemeister, 1992a,b), combined with the assumption that the notched noise has a beneficial effect because it provides an intensity reference. The effect of notched noise is heavily dependent, however, on temporal parameters that would not be expected to interfere with referential coding (e.g., Oxenham and Moore, 1995). In order for the notched noise to result in better performance through referential coding, we must assume that the probe-evoked memory trace has been corrupted at mid levels and that the auditory system has fallen back on context coding. But there is no *a priori* reason to suppose that shortening the tone's duration from $D = 25$ ms to $D = 2.51$ ms or increasing its frequency from 1 to 6 kHz would weaken a memory trace. Thus we still lack an explanation for why mid-level humps occur under some conditions, but not others.

The role of context coding is debatable, but the present data do not contradict the notion that the probe evokes a memory trace that is corrupted by a forward or backward masker. The latter notion is compatible with results reported by Turner *et al.* (1994). They obtained forward-masked and non-forward-masked DLs using the stimulus conditions of Zeng *et al.* (1991) but using three different experimental methods: two-alternative forced-choice, two-alternative forced-choice with multiple looks at the stimuli, and the method of adjustment, in which subjects adjust the incremented tone to achieve equal loudness with the pedestal tone. (Each DL is then taken as the standard deviation of eight such adjustments.) Turner *et al.* found that the two-alternative forced-choice trials produced the mid-level hump, but the method of adjustment did not result in any visible divergence from the near miss to Weber's law, in the same group of subjects. Turner *et al.* then obtained non-masked, forward-masked, and backward-masked DLs using the same stimuli, with the adjustment procedure only, and another subject group. This group also showed no mid-level hump. Turner *et al.* concluded that the method of adjustment measures "a fundamentally different quantity than that measured by forced-choice procedures." It might also be thought that the method of adjustment allows the memory trace to stay uncorrupted.

Plack (1998) notes that if providing a reference or context eliminates mid-level humps of all kinds, then they cannot be assumed to have a peripheral component. If not, a two-stage model remains, in which peripheral factors of the type described by Nizami *et al.* (2001) might account for mid-level humps in the absence of forward or backward masking, and a central factor associated with the masker might lead to a further reduction in performance in some conditions. Explication of the underlying processes requires knowing more about what conditions cause a mid-level hump in the absence of nonsimultaneous masking.

VIII. SUMMARY AND CONCLUSIONS

There is a change in the level-dependence of the intensity-difference limen (DL) of a tone when that tone is forward masked by a stimulus of 50 dB SPL or higher, whose frequency is centered at the frequency of the probe. For 1-kHz probes, the near miss to Weber's law develops into a mid-level hump (Zeng *et al.*, 1991; Zeng and Turner, 1992; Plack and Viemeister, 1992b; Turner *et al.*, 1994; Plack *et al.*, 1995; Schlauch *et al.*, 1997, 1999); for higher frequencies, where the DLs show a mid-level hump even in the absence of a forward masker (e.g., Carlyon and Moore, 1984; Florentine *et al.*, 1987; Plack, 1998), the hump apparently enlarges under forward masking (Plack and Viemeister, 1992a; Carlyon and Beveridge, 1993). Although the effect observed by Zeng *et al.* (1991) is large, a hump just as big can appear for mid-frequency tones in the absence of forward maskers, if the equivalent rectangular duration of the tone is shortened to less than 10 ms (Nizami *et al.*, 2001). Because this hump could be as large as some of those seen under forward masking, it was predicted that the imposition of a forward masker would make little further difference in the size of the hump.

First, the present results confirmed the mid-level hump found by Nizami *et al.* (2001) for a 2-kHz Gaussian-shaped tone-pip having an equivalent rectangular duration D of 2.51 ms. Then a 200-ms 2-kHz sinusoid was placed in each interval of the two-interval forced-choice task in order to forward mask the Gaussian-shaped tone-pip. The masker-level/probe-delay conditions used were 50 dB SPL/10 ms, 50 dB SPL/100 ms, and 70 dB SPL/100 ms. The level-dependence of the DL showed a mid-level hump significantly larger than that found without forward masking. The DLs obtained under forward masking were found to be a constant multiple of those obtained without forward masking, so that the inflating effect of the forward masker weakened with further rises in pedestal level, the DLs converging at the highest level used (90 dB SPL). When tone-pip intensities were expressed as sensation levels, the mid-level humps found under forward masking lined up. The mid-level humps for the non-forward-masked 2-kHz Gaussian-shaped tone-pips of Nizami *et al.* (2001) also align in SL scales, but at a different SL, suggesting a different mechanism for those humps.

Zeng *et al.* (1991) were the first to note a mid-level hump under forward masking. They used masker-probe delays of 100 ms, and they explained the hump as a result of the different recovery rates of peripheral afferents from the adapting effects of the forward masker. Later, Zeng and Turner (1992) noted a dependence of each subject's maximum DL on the level of the forward masker. The present experiments provided DLs at a delay of 100 ms for forward-masker levels of 50, 70, and 80 dB SPL. Masker level made

no difference in the pedestal-level-dependence of the DLs. Further, the finding of a mid-level hump with a forward masker of only 50 dB SPL contradicts the Zeng *et al.* (1991) model.

Carlyon and Beveridge (1993) have suggested that nonsimultaneous masking acts by disrupting the memory trace evoked by the probe, causing a reliance on context coding, providing poor discrimination at mid-levels. The present data do not contradict the notion of a corruptible memory trace.

ACKNOWLEDGMENTS

The experiments were done at Boys Town National Research Hospital under support from a grant to Walt Jesteadt (R01 DC00136). We thank Dr. Carlyon, Dr. Plack, and Dr. Schlauch for providing data for mid-level humps with and without forward maskers for the analyses described in Sec. VII C. We also thank the reviewers, Dr. Christopher J. Plack and Dr. William Hellman, for their thoughtful insights.

[1]Regarding the 50 dB SPL masker and the 100-ms delay: the significant linear component reflects a nonzero slope to the curve, but does not tell us which of the DLs differ significantly from each other. For that we use pairwise comparisons. There are ten pairs of DLs to be compared, representing all possible pairs of points on the curve for pedestal levels $\geqslant$50 dB SPL. All but three differed at $p<0.05$ or better. The three nonsignificant pairings were the DLs for 50 and 60 dB SPL, the DLs for 70 and 80 dB SPL, and the DLs for 80 and 90 dB SPL. Regarding the 50 dB SPL masker and the 10-ms delay, pairwise comparisons revealed differences at $p<0.05$ or better between all but three pairs of DLs for pedestal levels $\geqslant$50 dB SPL: the DLs for 50 and 60 dB SPL, the DLs for 50 and 70 dB SPL, and the DLs for 80 and 90 dB SPL. Regarding the 70 dB SPL masker and the 100-ms delay, pairwise comparisons revealed differences at $p<0.05$ or better between all but one pair of DLs for pedestal levels $\geqslant$50 dB SPL: the DLs for 50 and 60 dB SPL. Finally, regarding the non-forward-masked DLs, pairwise comparisons revealed differences at $p<0.05$ or better between all but two pairs of DLs for pedestal levels $\geqslant$50 dB SPL: the DLs for 50 and 60 dB SPL, and the DLs for 80 and 90 dB SPL.

[2]This analysis also yielded paired comparisons, the results of which will only be mentioned for the highest and lowest SPLs used, for which the results are most telling. For the 50 dB SPL pedestal, the DL for the 50/100 masker-delay condition differed significantly from the DL for the non-forward-masked tone-pip, the DL for the 50/10 masker/delay condition differed significantly from the DL for the 70/100 masker/delay condition, and the DL for the 70/100 masker/delay condition differed significantly from the DL for the non-forward-masked tone-pip. For the 90 dB SPL pedestal, pairwise comparisons revealed no significant differences between DLs.

Baer, T., Moore, B. C. J., and Glasberg, B. R. (**1999**). ''Detection and intensity discrimination of Gaussian-shaped tone pulses as a function of duration,'' J. Acoust. Soc. Am. **106**, 1907–1916.

Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., and Purks, S. R. (**1984**). ''Intensity perception. XIII. Perceptual anchor model of context coding,'' J. Acoust. Soc. Am. **76**, 722–731.

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.

Carlyon, R. P., and Moore, B. C. J. (**1984**). ''Intensity discrimination: a severe departure from Weber's law,'' J. Acoust. Soc. Am. **76**, 1369–1376.

Durlach, N. I., and Braida, L. D. (**1969**). ''Intensity perception. I. Preliminary theory of intensity resolution,'' J. Acoust. Soc. Am. **46**, 372–383.

Florentine, M., Buus, S., and Mason, C. R. (**1987**). ''Level discrimination as a function of level for tones from 0.25 to 16 kHz,'' J. Acoust. Soc. Am. **81**, 1528–1541.

Harris, D. M., and Dallos, P. (**1979**). ''Forward-masking of auditory nerve fiber responses,'' J. Neurophysiol. **42**, 1083–1107.

Irvine, D. R. F. (**1986**). ''The auditory brainstem. Chapter 5. Superior olivary complex: anatomy and physiology,'' Prog. Sens. Physiol. **7**, 79–121.

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.

Keppel, G. (**1991**). *Design and Analysis* (Prentice Hall, Englewood Cliffs, NJ).

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.

Nizami, L., Reimer, J. F., and Jesteadt, W. (**2001**). ''The intensity-difference limen for Gaussian-enveloped stimuli as a function of level: tones and broadband noise,'' J. Acoust. Soc. Am. **110**, 2505–2515.

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.

Oxenham, A. J., and Plack, C. J. (**2000**). ''Effects of masker frequency and duration in forward masking: further evidence for the influence of peripheral nonlinearity,'' Hear. Res. **150**, 258–266.

Plack, C. J. (**1998**). ''Beneficial effects of notched noise on intensity discrimination in the region of the 'severe departure,'" J. Acoust. Soc. Am. **103**, 2530–2538.

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**, 3097–3101.

Plack, C. J., and Viemeister, N. F. (**1993**). ''Suppression and the dynamic range of hearing,'' J. Acoust. Soc. Am. **93**, 976–982.

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.

Schairer, K. S., Reimer, J. F., and Jesteadt, W. (**2001**). ''Psychometric functions for detection of tones following on- and off-frequency forward maskers,'' J. Acoust. Soc. Am. **109**, 2464.

Schlauch, R. S., Clement, B. R., Ries, D. T., and DiGiovanni, J. J. (**1999**). ''Masker laterality and cueing in forward-masked intensity discrimination,'' J. Acoust. Soc. Am. **105**, 822–828.

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.

Smith, R. L. (**1977**). ''Short-term adaptation in single auditory nerve fibers: some poststimulatory effects,'' J. Neurophysiol. **40**, 1098–1112.

Turner, C. W., Horwitz, A. R., and Souza, P. E. (**1994**). ''Forward- and backward-masked intensity discrimination measured using forced-choice and adjustment procedures,'' J. Acoust. Soc. Am. **96**, 2121–2126.

Zeng, F.-G., and Shannon, R. V. (**1995**). ''Possible origins of the nonmonotonic intensity discrimination function in forward masking,'' Hear. Res. **82**, 216–224.

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: effects upon intensity discrimination,'' Hear. Res. **55**, 223–230.

Buildup and breakdown of echo suppression for stimuli presented over headphones—the effects of interaural time and level differences

Katrin Krumbholz[a)]
Centre for the Neural Basis of Hearing, Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom

Andrea Nobbe
Institute for Human-Machine Communication, Technical University Munich, Arcisstrasse 21, D-80333 München, Germany

(Received 26 April 2001; revised 7 May 2002; accepted 9 May 2002)

The current study investigates buildup and breakdown of echo suppression for stimuli presented over headphones. The stimuli consisted of pairs of 120-μs clicks. The leading click (lead) and the lagging click (lag) in each pair were lateralized on opposite sides of the midline by means of interaural level differences (ILDs) of ± 10 dB or interaural time differences (ITDs) of ± 300 μs. Echo threshold was measured with an adaptive one-interval, two-alternative, forced-choice procedure with a subjective decision criterion, in which listeners had to report whether they heard a single, fused auditory event on one side of the midline, or two separate events on both sides. In the control conditions, referred to as the ''single'' conditions, echo threshold was measured for a single click pair, the test pair, presented in isolation. In addition to the control conditions, two kinds of test conditions were investigated, in which the test pair was preceded by 12 identical conditioning pairs: in the ''same'' conditions, the interaural configuration (ILDs or ITDs) of the conditioning pairs was identical to that of the test pair; in the ''switch'' conditions, the interaural configuration of lead and lag was reversed between the conditioning pairs and the test pair, in order to produce a switch in the lateralizations of the stimuli between the conditioning train and the test pair. No matter whether the lateralization of the clicks was produced by ILDs or by ITDs, most listeners experienced a buildup of echo suppression in the ''same'' conditions, manifested by a prolongation of echo threshold relative to the respective ''single'' conditions. However, the breakdown of echo suppression was much stronger in the ILD-switch than in the ITD-switch conditions. In five out of six listeners, the ITD switch had hardly any effect on echo threshold, although the ITDs (± 300 μs) produced roughly the same degree of lateral displacement as the ILDs (± 10 dB). These results suggest that the dynamic processes in echo suppression operate differentially in pathways responsible for the processing of interaural time and level differences. [DOI: 10.1121/1.1490594]

PACS numbers: 43.66.Pn, 43.66.Qp [LRB]

I. INTRODUCTION

In echoic environments, the sound that reaches the listener's ears often contains several reflections of the primary sound, which originate from different directions. The apparent location of the composite sound is usually dominated by the localization information in the leading, i.e., the primary sound. This phenomenon is often referred to as the ''precedence effect'' [for a review, see Blauert (1997) and Litovsky *et al.* (1999)]. While a reflection can noticeably change certain perceptual aspects of the composite sound, such as its loudness, its spaciousness, and its pitch or timbre, reflections are generally not perceived as discrete sounds. Only when the delay between the primary sound and the reflection exceeds a certain critical value, called the ''echo threshold,'' does the observer perceive the reflection as a separate auditory event localized at or near its actual physical source. In this case, the reflection is referred to as an ''echo.'' In the laboratory, the precedence effect can be investigated with two loudspeakers in an anechoic room (free-field), by presenting an acoustic stimulus (lead), simulating the primary sound, through one loudspeaker and a delayed copy of the same stimulus (lag), simulating a reflection, through the other loudspeaker (e.g., Wallach *et al.*, 1949; Freyman *et al.*, 1991).

Initially, it seemed that the precedence effect is a hard-wired, and thus presumably a fairly peripheral effect, which is necessarily driven by correlated input stimuli. More recent studies have shown, however, that the precedence effect depends not only on the properties of the stimuli themselves but also on the context in which they are presented, indicating that the precedence effect may involve central processes associated with higher cortical processing stages (Clifton, 1987; Clifton and Freyman, 1989; Freyman *et al.*, 1991; Blauert and Col, 1992; Clifton *et al.*, 1994; Yost and Guzman, 1996; Yang and Grantham, 1997). Thurlow and Parks (1961) and later Clifton and her colleagues have shown that

[a)]Current address: IME, AG Kognitive Neurologie, Forschungszentrum Jülich, 52425 Jülich, Germany. Electronic mail: k.krumbholz@fz-juelich.de

echo threshold can be elevated by exposing the listener to repeated presentations of lead and lag. If the lead–lag delay is just above the echo threshold for a single pair, the listener initially perceives lead and lag as separate auditory events at or near their actual sources. However, as presentation continues, the echo fades away and only a single event is perceived at or near the location of the lead loudspeaker, suggesting that the perceptual suppression of the lag is building up as the auditory system acquires more information about the stimuli over time. This ''buildup'' of echo suppression may extend over a period of a few seconds depending on the rate of presentation of the lead–lag pairs, and it can increase echo threshold by several milliseconds depending on the number of pairs in the train. Sudden changes in the stimuli that would be unlikely to occur in a realistic context, i.e., a switch of the locations of lead and lag loudspeakers or a change in lead–lag delay, result in a release, or ''breakdown,'' of echo suppression (Clifton, 1987; Clifton *et al.*, 1994). In a realistic context, a sudden change in lead–lag delay or in the locations of lead and lag would mean that the locations of the reflecting surface and/or the primary sound source have suddenly changed. After such an unrealistic change, the lag is perceived as a separate auditory event even if a single, fused event was perceived before the change. Sudden changes in the frequency or the intensity of *both* lead and lag, on the other hand, do not lead to a breakdown of echo suppression (Clifton *et al.*, 1994). Such changes would be encountered frequently in a realistic context, because the frequency content and the intensity of most natural sounds change widely over time. These results suggest that echo suppression is modulated by the listener's experience of realistic acoustic environments, and so probably involves input from higher cortical levels. This assumption is supported by the observation that the buildup of echo suppression is asymmetric between left and right hemi-fields (Clifton and Freyman, 1989; Blauert and Col, 1992). After buildup, echo suppression tends to be greater when the lead is presented from the right and the lag from the left than in the reverse condition. Grantham (1996) examined the left–right asymmetry of the buildup effect in a large group of listeners and found that only the right-handed listeners showed a significant asymmetry. Grantham pointed out that the left–right asymmetry of the buildup effect and its correlation to handedness is reminiscent of other asymmetries in auditory function, like the ear preference for speech and melody perception (Kimura, 1967) and the ear dominance for pitch (Efron and Yund, 1976), which are presumed to reflect hemispheric differences at higher cortical levels.

In the free field, the direction of a sound is conveyed by a combination of different acoustic cues, including interaural time differences (ITDs), which are determined by the path length differences from the sound source to the two ears, and interaural level differences (ILDs), which result primarily from the acoustic head shadowing effect. The processing of these interaural localization cues begins in the superior olivary complex, which is situated in the brainstem. It is likely that the processing of ITDs and ILDs requires different specialized neural circuits. Indeed, physiological data indicate that, to a large extent, the initial processing of ITDs and ILDs takes place in separate pathways represented by different brainstem nuclei. Converging projections from each cochlear nucleus to the medial superior olive (MSO) transmit temporal information with great fidelity, and the MSO contains cells that are sensitive to ITDs of the order of a few tens of microseconds, which makes this nucleus a likely candidate for the processing of ITDs (Oertel, 1997). The lateral superior olive (LSO), on the other hand, is assumed to process ILDs, because it contains neurons, which receive excitatory and inhibitory projections from the ipsilateral and the contralateral side, respectively, which makes these neurons sensitive to intensity differences across the ears (see, e.g., Irvine *et al.*, 2001 and references therein). Functional evidence suggests that the role of the human superior olivary complex in auditory spatial processing is similar to olivary function in other mammals (for a review, see Moore, 2000). In particular, electrophysiological studies indicate that there are two separate systems for ITD and ILD processing in the human brainstem (Pratt *et al.*, 1997). This notion is corroborated by findings of Griffiths *et al.* (1998) who reported the case of a pontine lesion in the area corresponding to the superior olive which led to an isolated deficit in the detection of ITDs, but not ILDs. Eventually, ITD and ILD information must be combined and evaluated within the framework of an integrated map of auditory space, which can be brought into register with spatial maps from the visual and the somatosensory modalities, and which would be expected to be shaped by experience. Physiological data gathered by King and coworkers (for a review, see King and Schnupp, 2000) indicate that the representation of auditory space first emerges on the level of the superior colliculus in the midbrain.

If, as suggested by the previous studies, the dynamic processes in echo suppression indeed operate on a high, possibly cortical level of processing, i.e., *after* ITD and ILD information has been combined and translated into a spatial representation, one would expect that different interaural cues (ITDs or ILDs), which lead to similar spatial percepts should have a similar effect on echo suppression. If, on the other hand, the processes involved in the buildup and the breakdown of echo suppression operate on a more peripheral level, *before* ITD and ILD information is combined, ITDs and ILDs may be expected to have differential effects on echo suppression. The experiments presented in the current study aim to test these two hypotheses. In the free field, the effects of ITDs and ILDs on echo suppression cannot be disentangled. However, echo suppression can also be observed when the stimuli are presented over headphones, where the localization cues can be manipulated independently. Headphone studies have shown that, for a certain range of lead–lag delays, the perceived lateral position of a pair of correlated stimuli is dominated by the interaural parameters of the leading stimulus (e.g., Wallach *et al.*, 1949; Zurek, 1980; Shinn-Cunningham *et al.*, 1993). Moreover, listeners' ability to detect changes in the ITD or the ILD of the lag can deteriorate markedly in the presence of the lead (Zurek, 1980; Yost and Soderquist, 1984). The buildup and the breakdown of echo suppression, however, have so far only been investigated for stimuli presented through loudspeakers in the free field. The purpose of the present study

was, first, to establish the buildup of echo suppression for stimuli presented over headphones and, second, to investigate the differential contributions of a sudden change in the ITDs or the ILDs of lead and lag to the breakdown of echo suppression.

II. EXPERIMENT 1

This experiment investigates echo threshold for pairs of clicks presented over headphones. In the control conditions, also referred to as the ''single'' conditions, echo threshold was measured for a single click pair (test pair) presented in isolation. The leading click and the lagging click in the test pair were lateralized on opposite sides of the midline by means of interaural time or level differences. In addition to the control conditions, two kinds of test conditions were investigated, in which the test pair was preceded by 12 identical ''conditioning'' pairs: in the ''same'' conditions, the interaural configuration of the conditioning pairs was identical to that of the test pair; in the ''switch'' conditions, the interaural configuration of lead and lag was reversed between the conditioning pairs and the test pair, in order to produce a switch in the lateralizations of the stimuli between the conditioning train and the test pair. From the results of the free-field studies (Clifton, 1987), we expected to observe a buildup of echo suppression in the ''same'' conditions. The ''switch'' conditions were analogous to Clifton's (1987) classical breakdown paradigm, in which the lead loudspeaker and the lag loudspeaker in an anechoic room are mounted on opposite sides of the median plane. When, during a sequence of lead–lag pairs, the locations of the lead and lag are suddenly switched between the loudspeakers, the listener experiences a breakdown of echo suppression. So, we expected to observe a breakdown of echo suppression in the ''switch'' conditions.

A. Methods

1. Stimuli and setup

All measurements were conducted in a double-walled, sound-insulated chamber. The stimuli were presented over headphones; they consisted of 120-μs clicks, which were presented at a peak-equivalent level[1] of 78 dB SPL, corresponding to a sensation level of about 35–40 dB. The clicks were lateralized on the left or right side of the midline by means of interaural time or level differences. The stimuli were generated digitally with a 16-bit amplitude resolution and a sampling rate of 50 kHz (TDT AP2). Stimuli for the left ear and the right ear were digital-to-analog converted through separate channels of the D/A-converter (TDT DD1), low-pass filtered at 10 kHz (TDT FT6-2) and passed through two separate programmable attenuators (TDT PA4). The attenuated stimuli were finally fed to the left and right channels of a headphone buffer (TDT HB6) and delivered to the headphones (AKG K 240 DF). The headphones were semi-open and diffuse-field corrected.

2. Procedure

Echo thresholds were measured with a subjective one-interval, two-alternative, forced-choice procedure with two interleaved adaptive tracks. The test pair consisted of two clicks, which were presented either with an interaural level or with an interaural time difference. In both cases, the interaural disparity of the lag (ITD or ILD) had the same absolute value but the opposite sign as the interaural disparity of the lead, so lead and lag were lateralized on opposite sides of the midline. In the following, a positive ITD or ILD means that the click at the left ear is delayed or attenuated relative to the click at the right ear, so the stimulus is lateralized on the right; conversely, a negative ITD or ILD means that the stimulus is lateralized on the left. The listener had to indicate whether he or she heard a single, fused auditory event on *one* side of the midline or two separate events on *both* sides, by pressing one of two buttons on a response box. The adaptive parameter in the measurements was the delay between the leading and the lagging click in the test pair, in the following referred to as the interclick interval (ICI). Even when the lag is not perceived as a separate event, it can noticeably change the perceptual qualities of the composite sound. In particular, a lead–lag pair with an ICI smaller than about 20 ms sounds more tonal than a single click, and the perceived pitch decreases with increasing ICI. Listeners were instructed to ignore any ''monaural'' sound quality changes introduced by the lag and use the spatial attributes of the stimuli as a criterion for their one-versus-two judgments. Some listeners found it helpful to use visual imagery to aid their decisions. They reported that, for ICIs above the echo threshold, the click pairs sounded like two marbles bouncing on a surface on both sides of the midline, whereas in subthreshold conditions, the sound was like that of a single marble bouncing on one side. Each threshold measurement was made up of two interleaved adaptive tracks, which were chosen randomly by the computer before each trial. Each track could only be presented up to two times in a row. At the beginning of each threshold measurement, the ICI was much longer than the anticipated echo threshold in one of the two tracks, and it was much shorter than the anticipated threshold in the other track. The ICI was decreased by a factor ν when the listener reported hearing two separate auditory events, and increased by the same factor when the listener reported hearing a single, fused event. In each of the two tracks, ν was equal to 1.5 up to the first reversal of the ICI, was reduced to 1.3 for the following reversal, and was set to 1.15 for the rest of a total of eight reversals. Each threshold estimate is the mean of the geometric means of the ICIs at the last six reversals in each of the two adaptive tracks in one threshold run. The thresholds presented in the figures are the mean of at least three threshold estimates and the error bars show their standard error.

In each of four control conditions, referred to as the ''single'' conditions, the test pair was presented on its own; the lead and the lag of the test pair had (1) an ILD of +10 and −10 dB (ILD-single-left), (2) an ILD of −10 and +10 dB (ILD-single-right), (3) an ITD of +300 and −300 μs (ITD-single-left), or (4) an ITD of −300 and +300 μs (ITD-single-right), respectively. So, in supra-threshold conditions, i.e., when the lag in the test pair was perceived as a separate event, it was lateralized on the right in the ''single-right'' conditions, and on the left in the ''single-left'' conditions.

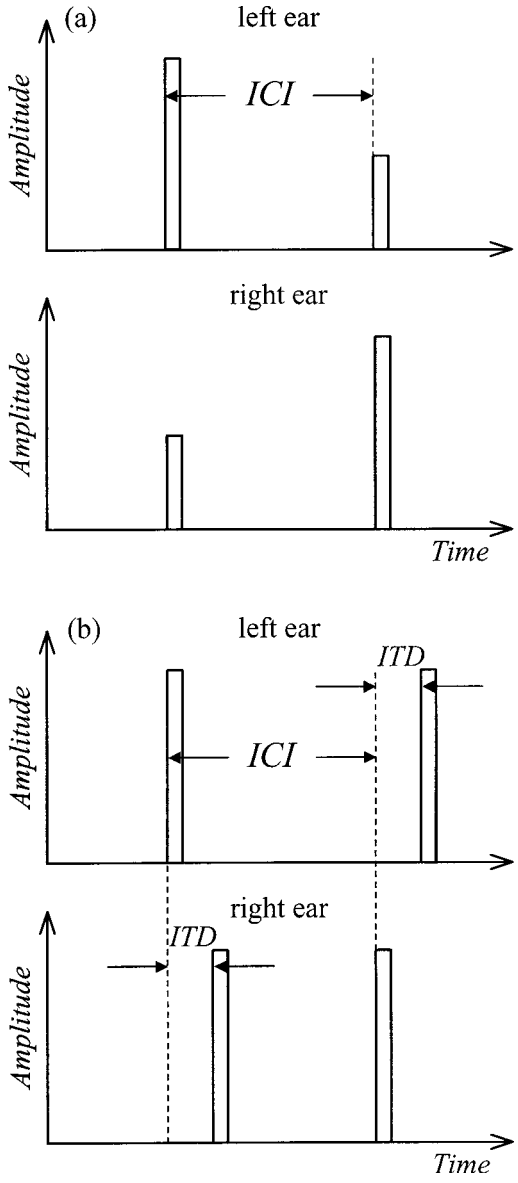


FIG. 1. Schematic diagram of the stimuli to the left ear (upper panels) and the right ear (lower panels) for the test pair in the ILD-single-right (a) and ITD-single-right (b) conditions. The interclick interval (ICI) is the delay between the leading and the lagging clicks (not to scale).

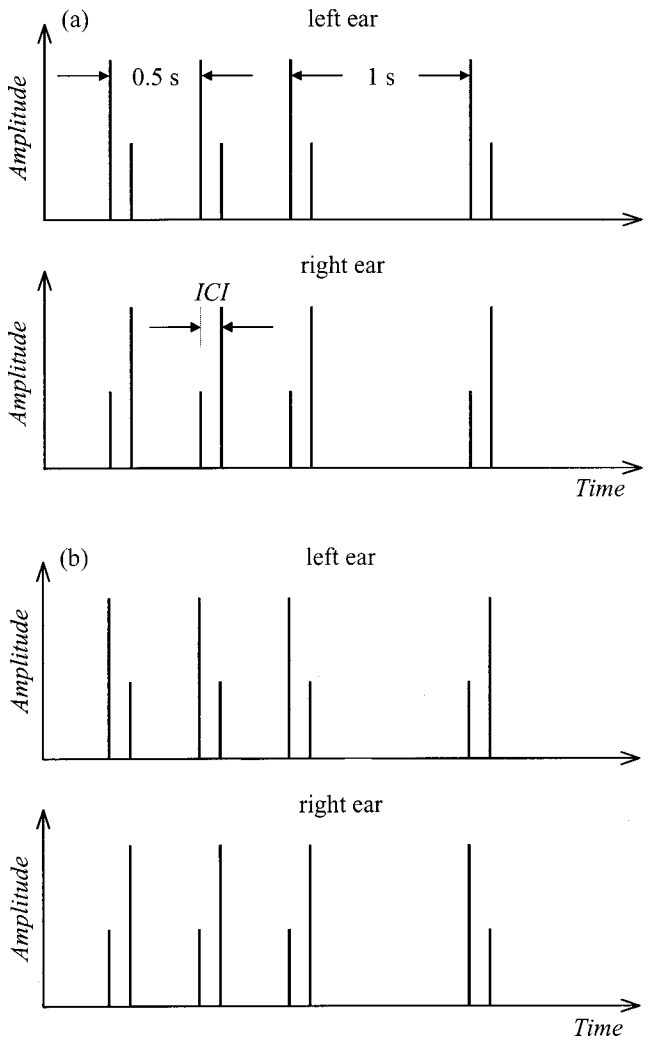


FIG. 2. Schematic diagram of the stimuli to the left ear (upper panels) and the right ear (lower panels) for the conditioning pairs and the test pair in the ILD-same-right (a) and ILD-switch-left (b) conditions. Only 3 of 12 conditioning pairs are shown. The delay between the conditioning pairs was 0.5 s and the delay between the last conditioning pair and the test pair was 1 s (not to scale).

Panels (a) and (b) of Fig. 1 show a schematic representation of the test pair in the ILD-single-right and ITD-single-right conditions, respectively.

In each of eight test conditions, the test pair was preceded by a train of 12 identical click pairs, referred to as the conditioning pairs. The ICI in the conditioning pairs was always the same as in the test pair. In four of the eight test conditions, referred to as the ''same'' conditions, the interaural configuration of the conditioning pairs was identical to that of the test pair in the four ''single'' conditions (ILD-same-right, ILD-same-left, ITD-same-right, ITD-same-left). Figure 2(a) schematically shows three click pairs of the conditioning train and the test pair in the ILD-same-right condition. In this condition, the lag in the test pair and in the conditioning pairs would be lateralized on the right, when perceived as a separate event. In the other four test conditions, referred to as the ''switch'' conditions, the interaural configuration of lead and lag in the conditioning pairs was reversed relative to the test pair in the four ''single'' conditions, so the lateralizations of the lead and, when perceived as a separate event, the lag would switch sides between the conditioning pairs and the test pair (ILD-switch-right, ILD-switch-left, ITD-switch-right, ITD-switch-left). Figure 2(b) schematically shows three click pairs of the conditioning train and the test pair in the ILD-switch-left condition. In this condition, the lag in the test pair would be lateralized on the left, when perceived separately, whereas the lag in the conditioning pairs would be lateralized on the right. The click pairs in the conditioning train were presented with a rate of 2 Hz and the test pair was presented 1 s after the beginning of the last pair in the conditioning train. Experimental runs were blocked into 1-h sessions, during which the different conditions were measured in a random order. The listeners were not informed that the experiment comprised different interaural conditions (ILD and ITD). Whereas all listeners could readily distinguish between the different test conditions (''single,'' ''same,'' and ''switch'') and between ''left'' and ''right'' conditions, none

of the listeners apart from AN and KK, who were the authors, reported to be aware of any subjective differences between ITD and ILD conditions. The same condition was never measured twice on the same day.

3. Listeners

A total of six listeners, between 20 and 31 years old, participated in the experiments. All listeners reported to be right-handed. Listeners AN and KK were the authors, the other listeners were paid for their services at an hourly rate. Listeners SB and WC were male, the other four listeners were female. None of the listeners had any known history of hearing impairment. No listeners were rejected on the basis of their performance. Listeners RB and WC required some training in order to get used to the task and reach a stable performance.

B. Results

The seven panels of Fig. 3 show the average (top) and individual echo thresholds of six listeners in the ILD conditions, in which the clicks were lateralized with ILDs of ± 10 dB. Figure 4 shows the thresholds for the ITD conditions, in which the clicks were lateralized with ITDs of ± 300 μs. Each block of three adjacent bars shows the thresholds for the ''single'' (white bars), ''same'' (hatched bars), and ''switch'' (black bars) conditions. The left block of bars shows thresholds for the ''left'' conditions, i.e., where the test lag had a negative ILD or ITD, and the right block shows thresholds for the ''right'' conditions, where the test lag had a positive ILD or ITD. For any one condition, individual echo thresholds varied by more than 10 ms among listeners. However, most of the listeners showed a similar pattern of results across the different conditions tested. Individual threshold estimates for all listeners were submitted to a four-way ANOVA with interaural disparity (ILD and ITD), condition (''single,'' ''same,'' and ''switch''), side (''left'' and ''right'') and listener's gender (male and female) as independent factors. The analysis yielded significant main effects of interaural disparity ($F=65.668$, $p<0.0001$) and condition ($F=44.513$, $p<0.0001$), and a significant interaction between these two factors ($F=8.921$, $p=0.0002$). Moreover, there was a significant interaction between the factors gender and side ($F=11.271$, $p=0.0009$).

Most listeners experienced a buildup of echo suppression in the ''same'' conditions (hatched bars), manifested by a prolongation of echo threshold relative to the respective ''single'' conditions (white bars). In some listeners, the buildup tended to be greater in the ITD than in the ILD conditions, so in the average data, the echo thresholds for the ILD-same conditions are by about 6 ms longer than for the ILD-single conditions, whereas the thresholds for the ITD-same conditions are by about 9 ms longer than the ITD-single conditions. Moreover, the average echo thresholds for the ITD-single conditions were by about 2 ms longer than the average thresholds for the ILD-single conditions.

The largest differences between ILD and ITD conditions, however, were observed for the ''switch'' conditions. The echo thresholds for the ILD-switch conditions (black bars in Fig. 3) were much smaller than for the ILD-same conditions (hatched bars in Fig. 3). For most listeners, the thresholds for the ILD-switch conditions were even smaller than for the ILD-single conditions (white bars in Fig. 3). In contrast, for five of the six listeners, the echo thresholds for the ITD-switch conditions (black bars in Fig. 4) were much larger than for the ITD-single conditions (white bars in Fig. 4), and, in most cases, the thresholds for the ITD-switch conditions were almost as large as for the ITD-same conditions (hatched bars in Fig. 4). Only listener WC experienced a strong breakdown of echo suppression in the ITD-switch conditions. In the average data, the echo thresholds for the ITD-switch conditions with test lag left and right are shorter by only 3.6 and 1.5 ms than the thresholds for the respective ITD-same conditions, compared to 9 and 7.7 ms for the ILD conditions. According to Sheffé's multiple-comparisons test applied to the ILD and the ITD data separately, the difference between ''single'' and ''same'' conditions was highly significant for both the ILD and the ITD data (both $p<0.0001$). In

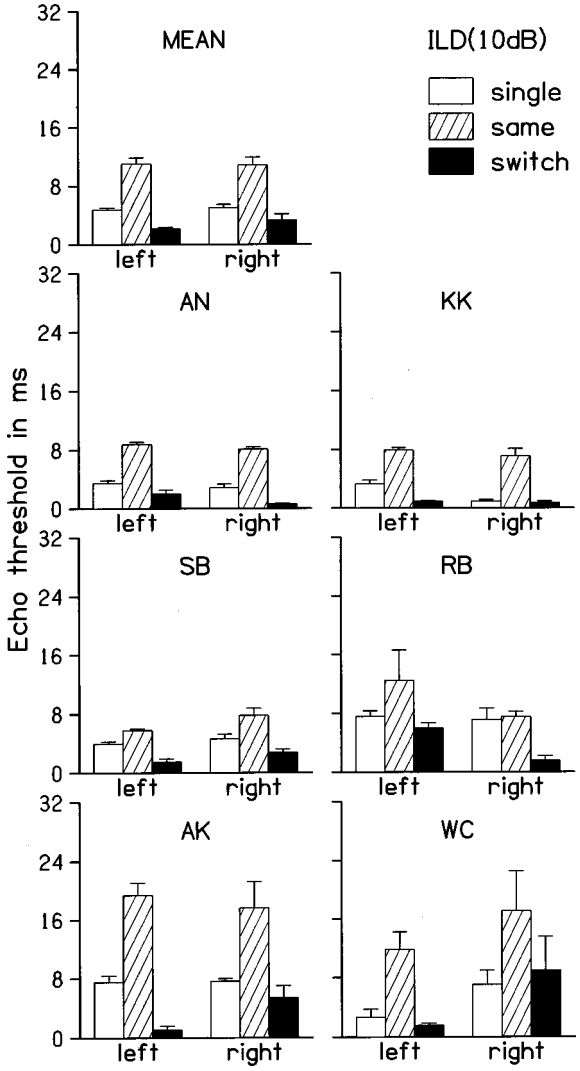


FIG. 3. Mean (upper panel) and individual echo thresholds of six listeners for the ILD conditions measured in experiment 1. Each block of three adjacent bars shows echo thresholds for the ''single'' (white bars), ''same'' (hatched bars) and ''switch'' (black bars) conditions (see legend in upper right corner of the figure). The left and right blocks of bars show the thresholds for the test-lag-left and the test-lag-right conditions, respectively. The error bars show the standard errors of the thresholds estimates.

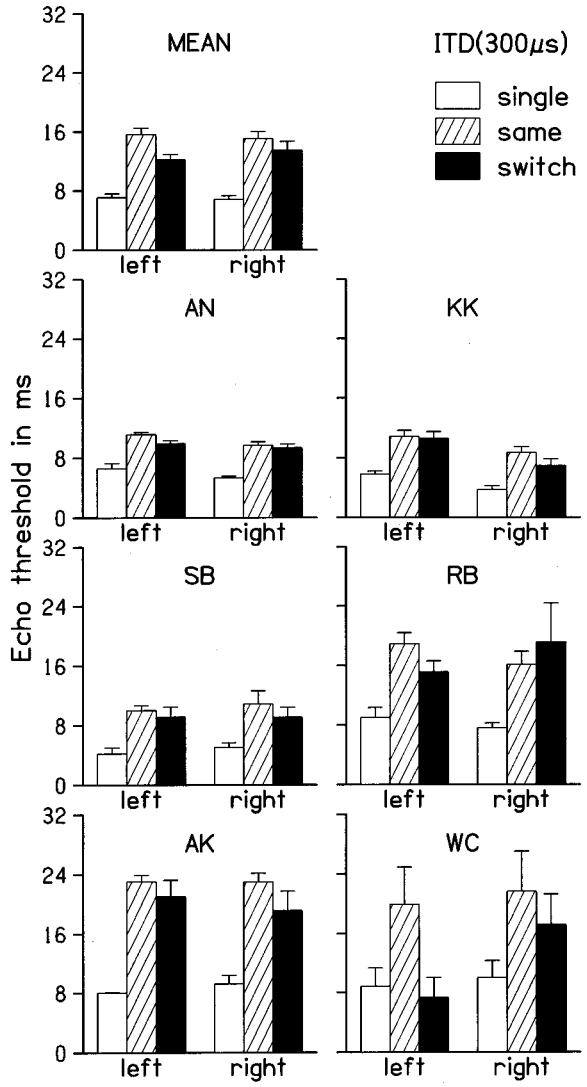


FIG. 4. Mean (upper panel) and individual echo thresholds of six listeners for the ITD conditions measured in experiment 1, plotted in the format as the thresholds for the ILD conditions in Fig. 3.

the case of the ILD data, the difference between "same" and "switch" conditions was also highly significant ($p < 0.0001$), whereas the difference between "single" and "switch" conditions was insignificant ($p = 0.127$). In the case of the ITD data, the situation was reversed, that is, the difference between "same" and "switch" conditions was insignificant, albeit with a fairly low value of p ($p = 0.0791$), whereas the difference between "single" and "switch" conditions was highly significant ($p < 0.0001$). Unpaired t-tests revealed that the differences between ILD-single and ITD-single, ILD-same and ITD-same, as well as ILD-switch and ITD-switch conditions were all significant ("single": $t = -3.653$, $p = 0.0004$; "same": $t = -3.141$, $p = 0.0023$; "switch": $t = -8.635$, $p < 0.0001$).

The interaction between gender and side revealed by the four-way ANOVA indicates that echo thresholds exhibited a significant asymmetry between "left" and "right" conditions, and that the direction of the asymmetry differed between male and female listeners. For two of the four female listeners (AN and KK), echo thresholds for the "left" conditions were slightly, but consistently, longer than thresholds for the "right" conditions. Listeners RB and AK also showed a slight left–right asymmetry for most of the conditions, but their threshold differences were less consistent. For the two male listeners SB and WC, the left–right asymmetry was reversed, i.e., echo thresholds for the "left" conditions were consistently shorter than thresholds for the "right" conditions; moreover, the threshold differences were larger than in any of the female listeners. For most listeners, the left–right differences in echo threshold were similar for "single" and "same" conditions, so there was hardly any left–right asymmetry in the extent of the buildup.

III. EXPERIMENT 2

Experiment 1 showed that a switch in the ILDs of lead and lag between the conditioning train and the test pair produced a complete breakdown of echo suppression. In some listeners, the ILD switch brought echo suppression even below its baseline level for a single lead–lag pair, suggesting that, in these listeners, the switch did not merely cause a release of the additional echo suppression that had built up during the conditioning train, but a rebound, i.e., a sensitization relative to the control condition. In contrast, the ITD switch caused no appreciable breakdown effect in most listeners. A simple explanation for the differences between the ILD-switch and the ITD-switch conditions may be that the ITD switch between ± 300 μs and the ILD switch between ± 10 dB were not equivalent in terms of spatial perception. If it were the case that the ITDs produced a smaller degree of lateral displacement than the ILDs, the perceived spatial switch associated with the ITD switch would have been smaller than the spatial switch associated with the ILD switch. So the ITD switch may simply have been too small to actually produce a strong breakdown effect. The second experiment tests this idea. For that, we measured the ITD required to produce the same degree of lateral displacement in a click stimulus as an ILD of ± 10 dB.

A. Methods

Setup and stimulus generation were the same as in experiment 1 (see Sec. I A). As in experiment 1, the stimuli were 120-μs clicks presented at a peak-equivalent level of 78 dB SPL. The ITD required to produce the same degree of lateral displacement in a click as an ILD of ± 10 dB was measured with an adaptive tracking procedure. Each trial consisted of two clicks separated by a silent interval of 700 ms. The second click was the test click. It had a fixed ILD of ± 10 dB and no ITD. The first click was the matching click. It was lateralized by means of an ITD which was varied according to the listener's responses. The stimulation was free running, i.e., trials were presented continuously with a silent gap of 1500 ms, and the same trial was repeated until the listener made a response and the ITD of the matching click was changed. At the beginning of each measurement, the matching click was lateralized on the side opposite to the test click. The listener's task was to press the left of two response buttons, if the first click was perceived to the left of the second click. Conversely, the listener was asked to press the right button if the first click was perceived to the right of the second click. If the lateral positions of matching click

TABLE I. ITDs required to produce the same degree of lateral displacement as an ILD of plus (left column) and minus (right column) 10 dB. Different rows show the matching thresholds of six listeners and their average. The numbers in brackets denote the standard errors of the threshold estimates.

ILD	−10 dB	10 dB
AN	−200 (±10.599) μs	227.78 (±10.082) μs
KK	−148.89 (±8.34) μs	122.22 (±15.805) μs
SB	−264.44 (±11.776) μs	143.33 (±5.774) μs
RB	−303.33 (±32.146) μs	360 (±50.565) μs
AK	−271.11 (±35.492) μs	218.89 (±12.239) μs
WC	−163.33 (±20.578) μs	184.44 (±20.164) μs
Mean	−225.19 (±9.185) ms	209.44 (±9.857) μs

and test click were so close that a decision was difficult, the listener was asked to keep pressing the same button until the direction of the movement between the stimuli had noticeably reversed. The ITD of the matching click was increased each time the left button was pressed, and decreased each time the right button was pressed. Increasing the ITD shifted the matching click to the right, decreasing the ITD shifted the matching click to the left. The step size for the ITD increments and decrements was 80 μs up to the first reversal, reduced to 40 μs for the following reversal, and set to 20 μs for the rest of the eight reversals that made up each matching run. Each matching threshold estimate is the mean of the last six reversals in ITD. At least three matching runs were performed for each of the two conditions (test click ILD $= \pm 10$ dB) and the results were averaged. The listeners were the same as in experiment 1.

B. Results and interim discussion

The individual matching thresholds for six listeners and their average are listed in Table I. The left and right columns show the ITDs which produce the same degree of lateral displacement as an ILD of minus or plus 10 dB, respectively. On average, an ILD of +10 dB was matched with an ITD of about +209 μs, and an ILD of −10 dB was matched with an ITD of −225 μs. In both cases, the absolute values of the matching ITDs are smaller than 300 μs, the value used in experiment 1. This indicates that the degree of lateral displacement produced by the ITDs (±300 μs) was, if anything, greater than the degree of lateral displacement produced by the ILDs (±10 dB).

The current matching results are in general agreement with the results from previous matching experiments, in which listeners adjusted the ILD of a pointer stimulus to match the lateral position of target stimuli with varying ITDs (Feddersen *et al.*, 1957; Buell *et al.*, 1991). The results are also in accordance with the fact that listeners were generally unaware of any subjective differences between the stimuli in the ILD and the ITD conditions.

The matching results rule out the idea that the differences between the ILD and ITD conditions in experiment 1 are attributable to differences in the degree of lateral displacement produced by the ITDs and the ILDs used in that experiment. If anything, the perceived spatial switch associated with the ITD switch was larger than the spatial switch associated with the ILD switch. So, if the breakdown of echo suppression were determined by the degree to which the lateral positions of lead and lag are changed, the effect of the ITD switch on echo threshold should have been at least as large as the effect of the ILD switch. The third experiment examines whether a larger ITD switch would eventually produce an appreciable breakdown of echo suppression or whether it is principally impossible to produce a strong breakdown effect with an ITD switch, however large.

IV. EXPERIMENT 3

The third experiment is a repetition of the ITD conditions of experiment 1, with the difference that, in this case, the clicks were lateralized with ITDs of ±600 μs, instead of ±300 μs used in experiment 1. An ITD of 600 μs is expected to produce an intracranial image that is almost fully lateralized towards one ear. As in experiment 1, echo threshold was measured in the "single," "same," and "switch" conditions, and for a test lag with a negative ("left") and a positive ("right") ITD (see Sec. I A 2). Setup, stimulus generation, procedure, and listeners were the same as in experiment 1.

A. Results

The seven panels of Fig. 5 show the average (top) and the individual echo thresholds of six listeners for the ITD conditions of experiment 3 (600 μs). The echo thresholds in Fig. 5 were plotted in the same format as in Figs. 3 and 4. In the following, the ITD conditions from experiments 1 and 3 will be referred to as "ITD(300)" and "ITD(600)," respectively. For most listeners, the echo thresholds for the ITD(600)-single and the ITD(600)-same conditions (white and hatched bars in Fig. 5) tended to be longer by a few milliseconds than the thresholds for the respective ITD(300)-single and ITD(300)-same conditions (white and hatched bars in Fig. 4). Moreover, the threshold difference between "same" and "switch" conditions tended to be slightly larger for the ITD(600) than for the ITD(300) conditions, indicating that there was indeed a stronger breakdown in the ITD(600) conditions. In the average data, the same-switch threshold difference for the ITD(600) conditions amounts to 4.1 and 4.6 ms for test-lag-left and test-lag-right conditions, compared to 3.6 and 1.5 ms for the respective ITD(300) conditions.

The differences between the ITD(300) and the ITD(600) data were statistically analyzed by submitting the data to a two-way ANOVA with interaural disparity [ITD(300) and ITD(600)] and condition ("single," "same," and "switch") as independent factors. Whereas the effect of condition was highly significant ($F = 49.730$, $p < 0.0001$), neither the effect of interaural disparity ($F = 2.612$, $p = 0.1073$) nor the interaction between the two factors ($F = 0.526$, $p = 0.5914$) were significant. However, Sheffé's multiple-comparisons test applied to the ITD(600) data separately showed that the difference between the ITD(600)-same and the ITD(600)-switch conditions was significant ($p = 0.0101$), which had not been the case for the respective ITD(300) conditions (see Sec. I B). However, the difference between the ITD(600)-single and ITD(600)-switch conditions was also significant ($p = 0.0006$), indicating that, although there was a significant

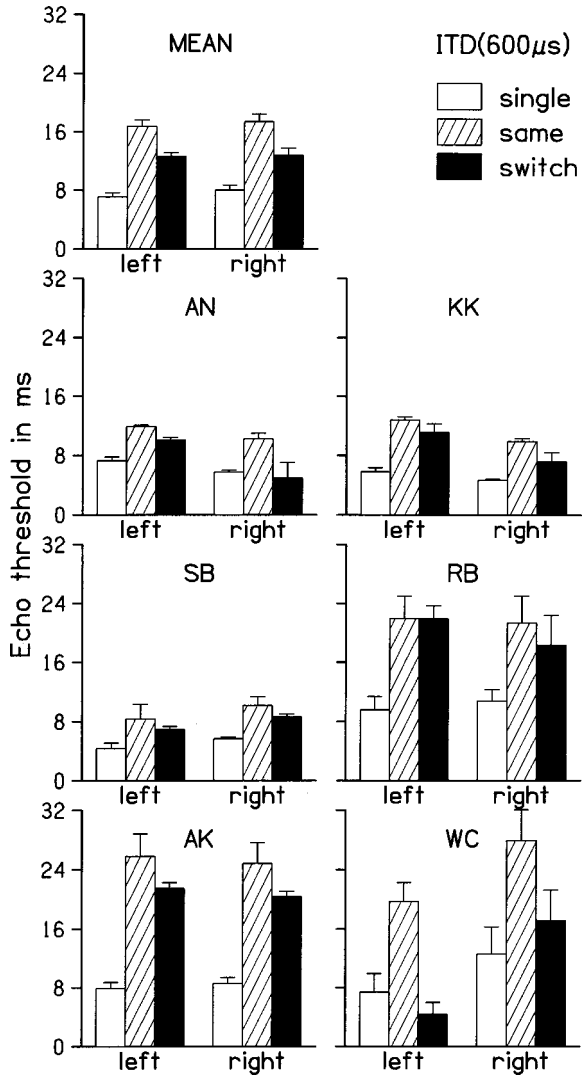


FIG. 5. Mean (upper panel) and individual echo thresholds of six listeners for the ITD(600) conditions measured in experiment 3. The thresholds for the ITD(600) conditions were plotted in the same format as the thresholds for the ILD and the ITD(300) conditions in Figs. 3 and 4.

reduction of echo threshold in the ITD(600)-switch conditions relative to the ITD(600)-same conditions, the ITD(600)-switch thresholds were still significantly larger than the ITD(600)-single thresholds.

V. DISCUSSION

The results from the current experiments showed that ITDs and ILDs may have different effects on echo suppression, even if they produce similar spatial percepts. Figure 6 shows a summary of the present data, expressed in terms of echo threshold differences representing the buildup and the breakdown of echo suppression. The open boxes labeled ''L'' and ''R'' show the buildup threshold shifts (differences between echo thresholds for ''same'' and ''single'' conditions) for test lag left and test lag right, respectively. The cross-hatched boxes show the respective breakdown threshold shifts (differences between echo thresholds for ''same'' and ''switch'' conditions). The data are represented in a box plot, which shows the minimum and the maximum, as well as the 25% and 75% percentiles, and the median of the individual

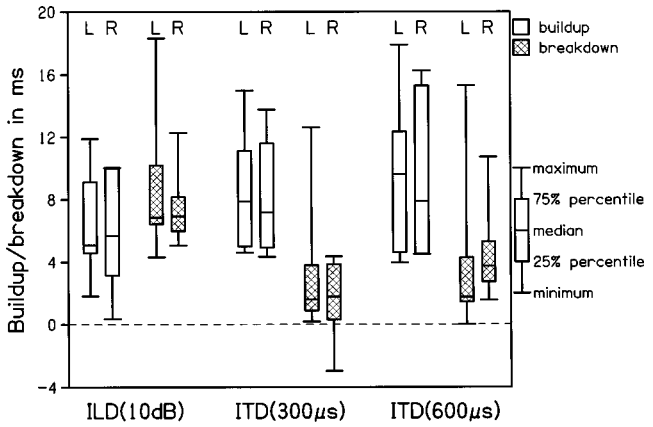


FIG. 6. Threshold shifts for buildup (open bars) and breakdown (hatched bars) for the ILD, ITD(300), and ITD(600) conditions measured in experiments 1 and 3. The data of six listeners are represented in a box plot, which shows minimum and maximum, as well as 25% and 75% percentiles, and median of the individual results of six listeners (see legend at the right of the figure). Results for test-lag-left and test-lag-right conditions are labeled ''L'' and ''R'' at the top of the figure.

results of six listeners. The median buildup threshold shifts for the ILD-left/right conditions (about 5 ms) are similar to the threshold shifts reported by Freyman *et al.* (1991), who presented clicks through loudspeakers in the free field and measured echo threshold with a subjective paradigm similar to that used in the current study. The median breakdown threshold shifts for the ILD conditions (about 7 ms) are even slightly larger than the buildup threshold shifts, indicating that the ILD switch caused a complete breakdown or even an undershoot of echo suppression below the baseline level for a single lead–lag pair. It appears that the buildup threshold shifts tended to be larger in the ITD(300) than in the ILD conditions, and that they were larger still in the ITD(600) conditions. In the ITD(300) and ITD(600) conditions, the median buildup threshold shifts ranged between 8 and 9 ms, compared to about 5 ms for the ILD conditions. The median breakdown threshold shifts for the ITD(300) conditions, on the other hand, amounted to only about 2 ms, indicating that there was hardly any breakdown of echo suppression in the ITD(300)-switch conditions. In the ITD(600) conditions, the median breakdown threshold shifts ranged between 2.5 and 3.6 ms. Thus, even in the ITD(600) conditions, the median breakdown threshold shifts were still much smaller than the ''buildup'' threshold shifts, indicating that the breakdown of echo suppression in the ITD(600)-switch conditions was still incomplete for most of the six listeners. Only one listener experienced a comparable breakdown of echo suppression in the ILD and the ITD conditions.

A. Theoretical implications of the current results

The differences between the effects of ITDs and ILDs on the breakdown and, to a lesser extent, on the buildup of echo suppression suggest that the dynamic processes in echo suppression are not associated with a single, central mechanism, which acts in a pathway that is common to both kinds of interaural cues. Rather, it appears that the processes involved in the breakdown effect operate differentially in pathways responsible for the processing of ITDs and ILDs. This

may mean that, at least for stimuli presented over headphones, the breakdown of echo suppression is determined by the neural information which is present at a level *before* information from ITDs and ILDs is combined and transformed into a spatial representation, suggesting that the breakdown effect operates on a more peripheral processing level than previously assumed. Alternatively, it might be the case that differential information about ITDs and ILDs is retained up to the highest stages of processing, and that, at central stages, the differential representation exists in parallel to an integrated, spatial representation of the localization cues. Either the differential or the integrated representation may be dominant in different situations. King and Carlile (1995) found that the spatial selectivity of neurons in the mammalian superior colliculus, which are assumed to constitute the first integrated representation of auditory space, is not only based on interaural but also on monaural spectral localization cues. It may be the case that the auditory system is unable to evaluate stimuli presented over headphones in an integrated spatial framework, because these stimuli have unrealistic monaural spectral profiles and therefore produce intracranial images.

B. Left–right asymmetry in echo suppression

The present data exhibited a significant interaction between the left–right asymmetry of echo threshold and listeners' gender (see Sec. II B). For female listeners, echo threshold tended to be larger in the test-lag-left than in the test-lag-right conditions; for male listeners, the situation was reversed. In most listeners, the amount of left–right asymmetry was similar for ''single,'' ''same,'' and ''switch'' conditions, so there was only little asymmetry in the extent of buildup or breakdown. In contrast, Grantham (1996) observed a significant left–right asymmetry only in his buildup conditions. Moreover, it appears that Grantham's right-handed listeners showed the same direction of asymmetry as our female listeners, irrespective of gender. However, Grantham did not explicitly test for an interaction with gender. Moreover, neither in the present nor in Grantham's study were the listeners submitted to a standard test battery for handedness. Rather, classification of handedness was merely based on listeners' own reports. This may have been insufficient in some cases, so it cannot be ruled out that the group of listeners in the current study was inhomogeneous with respect to handedness. Part of the discrepancies between Grantham's and the present results may also be due to differences in the methods and the procedure; in contrast to the present study, Grantham presented the stimuli through loudspeakers in the free field, and used an objective localization discrimination paradigm.

C. Comparison to results from simultaneous grouping

The relative ineffectiveness of ITDs with respect to the breakdown of echo suppression is reminiscent of the lack of an influence of ITDs on the grouping of simultaneous frequency components (see, e.g., Culling and Summerfield, 1995; Hukin and Darwin, 1995). Culling and Summerfield (1995) measured listeners' ability to identify two concurrent synthetic vowels, each of which was represented by two discrete bands of noise, centered at the formant frequencies. They showed that listeners were unable to identify the vowels when they were presented with different ITDs, indicating that listeners could not segregate the respective formant noise bands on the basis of different ITDs. On the other hand, the vowels were identified accurately, when each was presented to each ear separately. Similarly, Hukin and Darwin (1995) showed that the contribution of a single harmonic component to the quality of a vowel can be significantly reduced by presenting that component to the opposite ear than the rest of the vowel, but not by displacing the component by means of an ITD. ITDs do seem to have an effect on grouping in tasks, where the stimuli to be grouped or segregated are spread over time. Hartman and Johnson (1991) found that two interleaved melodies can be segregated on the basis of different ITDs. However, the segregation was less effective than when each melody was played to each ear separately. Darwin and Hukin (1999) observed that ITDs can help to track one spoken message in the presence of another. However, at the same time, the authors showed that listeners were unable to explicitly track individual frequency components in a sentence on the basis of a common ITD. Darwin and Hukin concluded that ITDs play little or no role in determining the constituents of auditory objects and that simultaneous grouping must be mainly based on other cues, such as onset and offset asynchronies, harmonic frequency relations, and ear of presentation. The difference between the effects of ITDs on simultaneous and on sequential grouping has been interpreted as an indication that simultaneous grouping may be mainly based on more peripheral processes whereas sequential grouping may involve fairly central processes (Carlyon *et al.*, 2001).

As far as it would be admissible to extrapolate the effects of ear of presentation on simultaneous grouping to ILDs in general, one could suggest an admittedly tentative interpretation of the current results as a case of auditory source segregation: For a certain range of ICIs, the leading click and the lagging click in a single click pair or at the beginning of the conditioning train in the ''same'' and ''switch'' conditions are perceptually segregated, i.e., they are perceived as separate auditory events lateralized on opposite sides of the midline. Presumably, segregation in the ITD conditions is mainly based on the onset asynchrony between lead and lag, whereas segregation in the ILD conditions may be based on both onset asynchrony and the ILD difference between lead and lag. This might explain why echo threshold for the ''single'' and ''same'' conditions tended to be longer in the ITD than in the ILD conditions. During the conditioning train, echo suppression builds up and the auditory scene is reanalyzed; eventually, lead and lag are perceived as a single, fused auditory event despite the cues (i.e., onset synchrony and, possibly, ILD difference) which indicate that they are actually two separate sounds. The perceptual grouping of lead and lag breaks up, if the ILDs of the stimuli are suddenly switched. If our interpretation is appropriate, this would indicate that a sudden ILD switch triggers a reset of the auditory grouping cues. Simi-

larly, a sudden change in the onset asynchrony between lead and lag, i.e., a change in the ICI, has been shown to produce a break-up of their perceptual fusion (Clifton *et al.*, 1994). In contrast, when the ITDs of lead and lag are switched, grouping persists for a much larger range of ICIs. Unless the ITD switch is very large *and* the ICI is very long, the fused auditory event simply hops to the opposite side of the midline. This indicates that, unlike an ILD switch, an ITD switch causes no or only a partial reset of the auditory grouping cues.

ACKNOWLEDGMENTS

This research was supported by the Medical Research Council (G9901257). We thank Hedwig Gockel for bringing up the issue of auditory grouping and for useful discussions.

[1]The peak-equivalent level is defined as the level of a 1000-Hz tone with the same amplitude as the click.

Blauert, J. (**1997**). *Spatial Hearing*, revised edition (MIT, Cambridge, MA).
Blauert, J., and Col, J.-P. (**1992**). ''A study of temporal effects in spatial hearing,'' in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 531–538.
Buell, T. N., Trahiotis, C., and Bernstein, L. R. (**1991**). ''Lateralization of low-frequency tones: Relative potency of gating and ongoing interaural delays,'' J. Acoust. Soc. Am. **90**, 3077–3085.
Carlyon, R. P., Cusack, R., Foxton, J. M., and Robertson, I. H. (**2001**). ''Effects of attention and unilateral neglect on auditory stream segregation,'' J. Exp. Psychol. Hum. Percept. Perform. **27**, 115–127.
Clifton, R. K. (**1987**). ''Breakdown of echo suppression in the precedence effect,'' J. Acoust. Soc. Am. **82**, 1834–1835.
Clifton, R. K., and Freyman, R. L. (**1989**). ''Effect of click rate and delay on breakdown of the precedence effect,'' Percept. Psychophys. **46**, 139–145.
Clifton, R. K., Freyman, R. L., Litovsky, R. Y., and McCall, D. (**1994**). ''Listeners' expectations about echoes can raise or lower echo threshold,'' J. Acoust. Soc. Am. **95**, 1525–1533.
Culling, J. F., and Summerfield, Q. (**1995**). ''Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay,'' J. Acoust. Soc. Am. **98**, 785–797.
Darwin, C. J., and Hukin, R. W. (**1999**). ''Auditory objects of attention: the role of interaural time differences,'' J. Exp. Psychol. Hum. Percept. Perform. **25**, 617–629.
Efron, R., and Yund, E. W. (**1976**). ''Ear dominance and intensity independence in the perception of dichotic chords,'' J. Acoust. Soc. Am. **59**, 889–898.
Feddersen, W. E., Sandel, T. T., Teas, D. C., and Jeffress, L. A. (**1957**). ''Localization of high-frequency tones,'' J. Acoust. Soc. Am. **29**, 988–991.
Freyman, R. L., Clifton, R. K., and Litovsky, R. Y. (**1991**). ''Dynamic processes in the precedence effect,'' J. Acoust. Soc. Am. **90**, 874–884.
Grantham, D. W. (**1996**). ''Left-right asymmetry in the buildup of echo suppression in normal-hearing adults,'' J. Acoust. Soc. Am. **99**, 1118–1123.
Griffiths, T. D., Elliott, C., Coulthard, A., Cartlidge, N. E. F., and Green, G. G. R. (**1998**). ''A distinct low-level mechanism for interaural timing analysis in human hearing,'' NeuroReport **9**, 3383–3386.
Hartmann, W. M., and Johnson, D. (**1991**). ''Stream segregation and peripheral channeling,'' Music Percept. **9**, 155–183.
Hukin, R. W., and Darwin, C. J. (**1995**). ''Effects of contralateral presentation and of interaural time differences in segregating a harmonic from a vowel,'' J. Acoust. Soc. Am. **98**, 1380–1387.
Irvine, D. R. F., Park, V. N., and McCormick, L. (**2001**). ''Mechanisms underlying the sensitivity of neurons in the lateral superior olive to interaural intensity differences,'' J. Neurophysiol. **86**, 2647–2666.
Kimura, D. (**1967**). ''Functional asymmetry of the brain in dichotic listening,'' Cortex **3**, 163–178.
King, A. J., and Carlile, S. (**1995**). ''Neural coding for auditory space,'' in *The Cognitive Neurosciences*, edited by M. S. Gazzaniga (MIT, Cambridge, MA), pp. 279–293.
King, A. J., and Schnupp, J. W. H. (**2000**). ''Sensory Convergence in Neural Function and Development,'' in *The New Cognitive Neurosciences*, edited by M. S. Gazzaniga (MIT, Cambridge, MA), pp. 437–450.
Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). ''The precedence effect,'' J. Acoust. Soc. Am. **106**, 1633–1654.
Moore, J. K. (**2000**). ''Organization of the human superior olivary complex,'' Microsc. Res. Tech. **51**, 403–412.
Oertel, D. (**1997**). ''Encoding of timing in the brain stem auditory nuclei of vertebrates,'' Neuron **19**, 959–962.
Pratt, H., Polyakov, A., and Kontorovich, L. (**1997**). ''Evidence for separate processing in the human brainstem of interaural intensity and temporal disparities for sound lateralization,'' Hear. Res. **108**, 1–8.
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.
Thurlow, W. R., and Parks, T. E. (**1961**). ''Precedence-suppression effects for two click sources,'' Percept. Mot. Skills **13**, 7–12.
Wallach, H., Newman, E. B., and Rosenzweig, M. R. (**1949**). ''The precedence effect in sound localization,'' Am. J. Psychol. **62**, 315–336.
Yang, X., and Grantham, D. W. (**1997**). ''Echo suppression and discrimination suppression aspects of the precedence effect,'' Percept. Psychophys. **59**, 1108–1117.
Yost, W. A., and Guzman, S. J. (**1996**). ''Auditory processing of sound sources: Is there an echo in here?'' Curr. Dir. Psychol. Sci. **5**, 125–131.
Yost, W. A., and Soderquist, D. R. (**1984**). ''The precedence effect: Revisited,'' J. Acoust. Soc. Am. **76**, 1377–1383.
Zurek, P. M. (**1980**). ''The precedence effect and its possible role in the avoidance of interaural ambiguities,'' J. Acoust. Soc. Am. **67**, 952–964.

The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal

Douglas S. Brungart[a)]
Air Force Research Laboratory, 2610 Seventh Street, Wright-Patterson AFB, Ohio 45433-7901

Brian D. Simpson
Veridian, 5200 Springfield Pike, Suite 200, Dayton, Ohio 45431

(Received 24 October 2000; revised 10 May 2002; accepted 10 May 2002)

Although many studies have shown that intelligibility improves when a speech signal and an interfering sound source are spatially separated in azimuth, little is known about the effect that spatial separation in distance has on the perception of competing sound sources near the head. In this experiment, head-related transfer functions (HRTFs) were used to process stimuli in order to simulate a target talker and a masking sound located at different distances along the listener's interaural axis. One of the signals was always presented at a distance of 1 m, and the other signal was presented 1 m, 25 cm, or 12 cm from the center of the listener's head. The results show that distance separation has very different effects on speech segregation for different types of maskers. When speech-shaped noise was used as the masker, most of the intelligibility advantages of spatial separation could be accounted for by spectral differences in the target and masking signals at the ear with the higher signal-to-noise ratio (SNR). When a same-sex talker was used as the masker, the intelligibility advantages of spatial separation in distance were dominated by binaural effects that produced the same performance improvements as a 4–5-dB increase in the SNR of a diotic stimulus. These results suggest that distance-dependent changes in the interaural difference cues of nearby sources play a much larger role in the reduction of the informational masking produced by an interfering speech signal than in the reduction of the energetic masking produced by an interfering noise source. [DOI: 10.1121/1.1490592]

PACS numbers: 43.66.Pn, 43.66.Qp, 43.66.Rq [LRB]

I. INTRODUCTION

In multitalker speech-perception tasks, performance is much better when the target speech signal and the interfering sound sources are located at different azimuth positions in the horizontal plane than when both the target and masking sounds originate from the same location in space. This so-called ''cocktail-party'' phenomenon has been studied extensively with speech maskers (Drullman and Bronkhorst, 2000; Duquesnoy, 1983; Freyman *et al.*, 1999; Hawley *et al.*, 1999; Festen and Plomp, 1990; Peissig and Kollmeier, 1997; Plomp, 1976) and speechlike noise maskers (Bronkhorst and Plomp, 1988, 1992; Plomp and Mimpen, 1979), and these studies have consistently shown that the intelligibility of the target speech increases systematically with the angular separation between the target and the masker. The release from masking can exceed 10 dB when the target is presented directly in front of the listener and the masker is presented near 90 degrees in azimuth (Bronkhorst, 2000).

Several different mechanisms contribute to this improvement in intelligibility. Perhaps the most important is the increase in signal-to-noise ratio (SNR) that inevitably occurs at one of the two ears when the target and masker signals originate from different directions in the horizontal plane. When two competing sources are located at different angles in the horizontal plane, differences in the head-shadowing effects for the two sources will cause one source to have a higher SNR in the left ear than it does in the right ear and the other source to have a higher SNR in the right ear than it does in the left ear. By selectively attending to the ear with the higher SNR (the ''better'' ear), the listener is able to effectively increase the SNR of either of the two sources. Differences in the spectral shapes of the target and masker signals at the better ear, which are determined by the head-related transfer functions (HRTFs) associated with the target and masker locations, can also influence performance (Zurek, 1993). These differences in the relative levels and spectral shapes of the target and masker signals at the better ear can account for most, but not all, of the intelligibility improvement afforded by spatial separation. Spatial unmasking is also influenced by a binaural interaction effect that is based on differences between the low-frequency interaural time delays (ITDs) and interaural level differences (ILDs) of the target and masker signals (Zurek, 1993; Levitt and Rabiner, 1967). Bronkhorst and Plomp (1988) found that the ITD portion of this binaural interaction effect could account for as much as a 5-dB release from masking for a speech source at 0 degrees and a noise masker near 90 degrees when the head-shadow was removed from the stimulus, but that it contributed only about 2.5 dB to the overall spatial release from masking in natural listening where the head-shadow cues were also available. More recently, Zurek (1993) integrated the better-ear and binaural interaction effects into a single model capable of predicting intelligibility with a spatially separated speech target and noise masker. For a more detailed review of the effects of angular separation in the

a)Electronic mail: douglas.brungart@wpafb.af.mil

 0001-4966/2002/112(2)/664/13/$19.00

''cocktail-party'' effect, see the recent reviews by Ericson and McKinley (1997) and Bronkhorst (2000).

One aspect of the ''cocktail-party'' phenomenon that has received almost no attention in the literature is the role that spatial separation in distance plays in the perception of multiple competing talkers in the region near the listener's head. Virtually all previous multitalker experiments have focused on relatively distant sound sources, located 1 m or more from the listener. Because the anechoic HRTF is independent of distance in this region (Brungart and Rabinowitz, 1999), differences in the distances of a target and a masker should have no impact on speech intelligibility when their overall levels are similar at the location of the listener. However, when the source is located within 1 m of the head, the HRTF is highly dependent on distance. Specifically, the ILD increases dramatically with decreasing distance in this region, while the ITD increases only modestly (Brungart and Rabinowitz, 1999). There are also substantial distance-dependent spectral changes in the HRTFs of nearby sound sources. Experiments have shown that listeners are able to use these distance-dependent changes in the HRTF to make reasonably accurate judgments about the distances of nearby sound sources in free-field environments (Brungart *et al.*, 1999; Brungart, 1999a). Until recently, however, almost nothing was known about the impact of these distance-dependent changes in the HRTF on the segregation of sound sources near the listener's head.

In order to examine the effects of distance on the segregation of nearby sources, Shinn-Cunningham and her colleagues (Shinn-Cunningham *et al.*, 2001) have recently adapted Zurek's model (1993) to account for the effects of spatial separation in distance on the intelligibility of a nearby speech signal masked by a nearby speech-shaped noise source. Their results have shown that virtually all of the effects of spatial separation with a noise masker can be explained by spectral differences in the target and masker signals at the ear with the better SNR, and that binaural factors can explain only 1–2 dB of the release from masking obtained by spatially separating the signal and masker in distance. However, there is some reason to believe that these results may underestimate the advantages of spatially separating multiple speech signals near the head. Recent studies with sound sources at distances greater than 1 m have shown that the binaural cues play a much larger role in the segregation of speech from a competing speech signal at a different location in azimuth than in the segregation of speech from a competing noise signal at a different azimuth (Freyman *et al.*, 1999; Hawley *et al.*, 2000). This difference seems to occur because interfering speech signals and interfering noise signals produce different kinds of masking: interfering noise signals produce only ''energetic'' masking, while interfering speech signals may produce both ''energetic'' and ''informational'' masking (Brungart, 2001b; Freyman *et al.*, 1999; Kidd *et al.*, 1998). In this context, energetic masking refers to the traditional concept of masking where the interfering signal overlaps in time and frequency with the target signal in such a way that portions of the target signal are rendered inaudible. Informational masking refers to the interference that occurs when the target and masker signals do not overlap in time and frequency but the listener is still unable to segregate the acoustic elements of the target signal from the acoustic elements of a similar-sounding masker. Freyman and his colleagues have suggested that listeners derive a greater benefit when two speech signals are spatially separated in azimuth than when a speech signal and a noise signal are spatially separated in azimuth because the listeners are able to use differences in the apparent locations of the two sounds to reduce the informational component of speech-on-speech masking (Freyman *et al.*, 1999; Freyman *et al.*, 2001). If this hypothesis is true, then there is reason to believe that spatial separations in distance that cause a difference in the apparent locations of the target and masking sounds will also produce a greater benefit when a target speech signal is masked by another speech signal than when it is masked by a noise signal. In this experiment, stimuli from a speech corpus that produces primarily informational masking in two-talker listening were used to determine whether the larger binaural advantages that have been reported in the segregation of competing speech signals that are spatially separated in azimuth also occur in the segregation of competing speech signals that are spatially separated in distance near the listener's head.

II. METHODS

A. Listeners

A total of nine paid listeners, five male and four female, participated in the experiment. All had normal hearing ($<$15 dB HL from 500 Hz to 8 kHz), and their ages ranged from 21 to 55 years. All of the listeners had participated in previous experiments that utilized the speech materials used in this study.

B. Stimuli

1. Speech materials

The speech stimuli were taken from the publicly available Coordinate Response Measure (CRM) speech corpus for multitalker communications research (Bolia *et al.*, 2000). This corpus consists of phrases of the form ''Ready (call sign) go to (color) (number) now'' spoken with all possible combinations of eight call signs (''arrow,'' ''baron,'' ''charlie,'' ''eagle,'' ''hopper,'' ''laker,'' ''ringo,'' ''tiger''), four colors (''blue,'' ''green,'' ''red,'' ''white''), and eight numbers (1–8). Thus, a typical utterance in the corpus would be ''Ready baron go to blue five now.'' Eight talkers (four male, four female) were used to record each of the 256 possible phrases, so a total of 2048 phrases are available in the corpus. Variations in speaking rate were minimized by instructing the talkers to match the pace of an example CRM phrase that was played prior to each recording. The sentences in the corpus, which are band-limited to 8 kHz, were resampled from the original 40-kHz sampling rate to 25 kHz to reduce computation time in the processing of the stimuli. The phrases were time aligned to ensure that the word ''ready'' started at the same time in all the speech signals in the stimulus, but no additional efforts were made to synchronize the call signs, colors, and numbers in the competing CRM phrases.

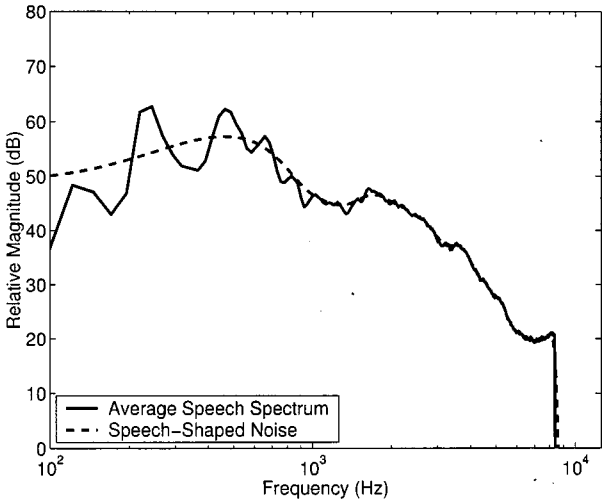


FIG. 1. Average spectrum of the speech utterances in the CRM corpus and frequency response of the filter used to shape the speech-shaped noise maskers. Note that the speech signals in the CRM corpus have been low-pass filtered with an 8-kHz cutoff frequency.

The CRM corpus was selected for this experiment for two reasons. First, the presence of the call sign provides a convenient way to instruct the listener which phrase to attend to in the speech-on-speech masking conditions of the experiment. Second, the small response set of the CRM corpus makes it easy to determine the correct color and number in the target phrase in the presence of relatively high levels of energetic masking (Brungart, 2001a). Previous experiments have shown that this insensitivity to energetic masking causes informational masking to dominate in speech-on-speech masking with the CRM corpus—in two-talker stimuli, listeners are generally able to hear the colors and numbers spoken by both the target and masking talkers, but are unable to correctly determine which color and number were spoken by the target talker (Brungart, 2001b). Thus, the CRM corpus is well suited to experiments such as this one that are designed to concentrate on the informational component of speech-on-speech masking rather than on the energetic component of speech-on-speech masking. Note that one would expect to find smaller differences between speech-on-speech masking and speech-on-noise masking with a measure of speech intelligibility that is more sensitive to the effects of energetic masking, such as the identification of nonsense syllables.

2. Speech-shaped noise

In some trials, a speech-shaped noise signal was used as the masker. The spectrum of this noise masker was determined by averaging the log-magnitude spectra of all of the phrases in the CRM corpus. This average spectrum was used to construct a 129-point finite impulse response (FIR) filter that was used to shape Gaussian noise to match the average spectrum of the speech signals (Fig. 1).

3. Spatial processing

The stimuli in the experiment were processed with HRTFs in order to simulate sound sources at different distances along the listener's interaural axis (Wightman and Kistler, 1989a, b). The HRTFs used for this spatial processing were derived from an earlier set of HRTFs measured for nearby source locations with a Knowles Electronics Manikin for Acoustic Research (KEMAR). These HRTFs, which are described in detail elsewhere (Brungart and Rabinowitz, 1999), were measured in a large anechoic chamber with an acoustic point source located directly to the left of the manikin (90 degrees azimuth) at distances of 12 cm, 25 cm, and 1.0 m from the center of the manikin's head. The overall level effects of distance and the frequency characteristics of the point source were removed from these HRTFs by subtracting the free-field spectrum of the sound source (as measured by a single microphone placed at a location corresponding to the center of the manikin's head) from the HRTFs measured at the manikin's left and right ears. The HRTF measurements were made in the frequency domain and consisted of 600-point transfer functions with 32-Hz resolution from 100 Hz to 19.2 kHz.

The filters used to spatially process the stimuli in this experiment were derived directly from these HRTFs using the following procedure. First, the headphones used in the experiment (Sennheiser HD540) were placed on the KEMAR manikin and the same frequency-domain method used to measure the original HRTFs was used to measure the 600-point left- and right-ear transfer functions of the headphones. These transfer functions were subtracted from the raw HRTFs for the left and right ears in order to determine the desired transfer functions of the headphone-corrected HRTFs for each stimulus location. Then the MATLAB FIR2 command was used to generate 251-point, linear-phase FIR filters matching the magnitudes of the frequency responses of the desired transfer functions over the frequency range from 100 Hz to 15 kHz at a 44.1-kHz sampling rate. These linear-phase filters were up-sampled to a 1-MHz sampling rate in order to delay the contralateral-ear HRTF by the interaural time delay, which was determined from the average slope of the unwrapped phase of the original interaural HRTF over the frequency range from 160 to 1700 Hz.[1] Finally, the HRTFs were down-sampled to a 25-kHz sampling rate to efficiently accommodate the 8-kHz band-limited speech corpus used in this experiment. The resulting HRTFs were stored in a MATLAB file and directly convolved with the target and masker signals immediately prior to each stimulus presentation. Figure 2 shows the frequency responses of the HRTFs used for each source location in this experiment (without headphone correction), and the ILDs and ITDs corresponding to each set of HRTFs.

Although these KEMAR HRTFs do not capture the high-frequency, listener-specific detail that would be present in individualized HRTFs, they do produce distance- and direction-dependent cues that are similar to the ones that would occur with a nearby sound source in the free field. They are therefore able to provide listeners with some information about the directions and distances of virtual sounds. An earlier experiment that required listeners to localize noise bursts that were processed with the same HRTFs used in this experiment has shown that listeners are able to localize both the distances and directions of nearby virtual sounds processed with the KEMAR HRTFs (Brungart and Simpson,

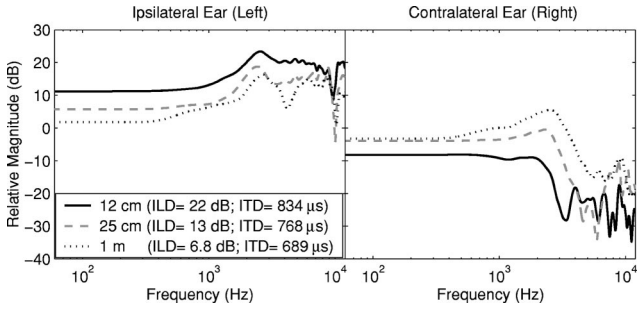


FIG. 2. These curves show the frequency responses of the HRTF filters used to spatially process the stimuli used in the experiment. The headphone response corrections described in the text have been removed from these plots, so they represent the frequency responses of the raw HRTFs measured directly from the KEMAR manikin (as described in Brungart and Rabinowitz, 1999). The numbers in the legend show the average interaural level difference (ILD) (measured from overall rms power for a speech-shaped noise stimulus) and the interaural time delay (ITD) (implemented with a linear phase delay in the HRTF for the contralateral ear) for each stimulus distance used in the experiment. Note that in each case the HRTF has been normalized to the sound pressure level that would occur at the location of the center of the head if the manikin's head were removed.

2001). The polar plot in Fig. 3 shows the median response locations in that experiment for three virtual sound locations along the listener's interaural axis. Although the localization judgments of the listeners in the virtual experiment were generally not as accurate as those of a different group of listeners who were asked to localize a nearby acoustic point source in the free field[2] (Brungart, 1999b), the median response locations shown in the figure indicate that the HRTF processing techniques used in this experiment can be used to generate virtual sounds along the interaural axis that are perceived at systematically increasing distances in roughly the same direction relative to the listener. It is not possible to know exactly what effect the nonindividualized HRTFs used in this experiment had on performance, but it should be noted that previous researchers who have compared the effect of spatial separation in azimuth on multitalker speech perception with virtual sources generated with nonindividualized HRTFs to the effects of spatial separation in azimuth on multitalker speech perception with free-field sources (Nelson *et al.*, 1999; Abouchacra *et al.*, 1997; Hawley *et al.*, 1999) or virtual sources generated with individualized HRTFs (Drullman and Bronkhorst, 2000) have reported no significant differences between the generic virtual presentations and the more realistic free-field and individualized virtual presentations.

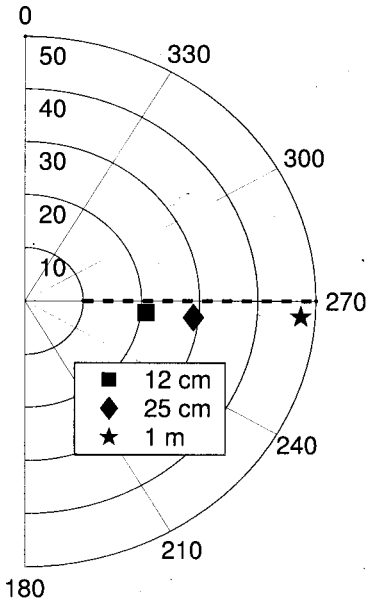


FIG. 3. Median direction and distance judgments for nearby virtual noise bursts. The results have been adapted from an earlier experiment that asked listeners to move an electromagnetic position sensor to the perceived location of a random-amplitude noise burst that was processed with the same set of HRTFs used in this experiment and presented over headphones (Brungart and Simpson, 2001). Each point represents the median location of 162 trials collected with seven normal-hearing listeners: the radius on the polar plot represents the median response distance (in cm), and the angle on the polar plot represents the median response azimuth (in degrees). Although the distance judgments were somewhat compressed, the results clearly show that the stimuli were perceived at systematically increasing distances at approximately the same angle in azimuth.

4. Stimulus configurations

All of the target and masker stimuli were presented along the interaural axis directly to the left of the listener. A total of five different target and masker configurations were tested (as shown in the first two columns of Fig. 4). In the 1 m–1 m configuration, both the target and the masker were presented at the same distance. In the 12 cm–1 m and 25 cm–1 m configurations, the target was presented at a closer distance than the masker. In the 1 m –12 cm and the 1 m–25 cm configurations, the masker was presented at a closer distance than the target. The target and masker locations were selected randomly in each trial in a process that resulted in roughly twice as many trials with the target and masker colocated at 1 m than in the other possible configurations.

5. Normalization

In real-world environments, the overall intensity of a stimulus varies with the distance of the source. Thus, if two equally intense speech signals were separated in distance, one would expect the closer speech signal to be substantially easier to comprehend simply because it would be more intense at the location of the listener; the contribution of binaural cues to the release from masking would be minimal relative to these distance-dependent intensity cues. Therefore, in order to examine the contribution of binaural cues and control for these distance-based intensity variations, the relative levels of the target and masker signals were adjusted in two different ways. In the *center-of-the-head* normalization condition (COH), the overall rms levels of the target and masker signals were equalized before they were convolved with the HRTFs, and the SNRs in the left and right ears were determined by the relative levels of the HRTFs shown in Fig. 2. This effectively normalized the rms levels of the target and masking sounds at the center of the listener's head (with the head removed from the sound field). In contrast, in the *better-ear* normalization condition (BE), the rms levels of target and masker were normalized after they were convolved with the appropriate HRTFs. The SNRs of the spatially processed target and masking signals were computed

	Configuration	BE Normalization Ipsi Ear SNR (dB)	BE Normalization Contra Ear SNR (dB)	COH Normalization Ipsi Ear SNR (dB)	COH Normalization Contra Ear SNR (dB)	Trials
Speech Masker	1m-1m	-0.1	-0.1	0.0	-0.1	2442
	1m-25cm	-5.1	0.0	-3.0	2.0	1305
	1m-12cm	-15.0	0.0	-7.7	7.0	1130
	25cm-1m	0.0	-5.4	3.3	-1.9	1222
	12cm-1m	0.0	-14.8	8.5	-6.6	1325
Noise Masker	1m-1m	-9.4	-9.0	-9.9	-9.5	1795
	1m-25cm	-15.0	-9.0	-13.1	-7.0	781
	1m-12cm	-25.0	-9.0	-18.1	-2.2	720
	25cm-1m	-9.0	-13.9	-6.7	-11.6	776
	12cm-1m	-9.0	-23.7	-1.5	-16.1	841

FIG. 4. Target and masker configurations used in the experiment. Column 2 shows a graphical representation of the target and masker locations in each configuration indicated in column 1. Columns 3–6 show the average SNRs (measured from rms power) of the spatially processed target and masker signals in the listener's left (ipsi) and right (contra) ears for better-ear (BE) and center-of-head (COH) normalization. Column 7 shows the number of trials completed in each condition.

from their rms levels at each ear, and the filtered target speech signal was scaled (by an equal amount in both ears) to make the SNR at the ear with the greater SNR (the *better ear*) equal to 0 dB.

The middle columns of Fig. 4 show the average SNRs at the ipsilateral and contralateral ears with BE and COH normalization for the HRTF-processed stimuli in each of the target-masker configurations tested in the experiment. In BE normalization, the SNR at the ear with the higher SNR is forced to be 0 dB, and the SNR in the other ear is determined by the ILDs in the HRTFs. Note that the location of the better ear depends on the relative distances of the target and masker. When the target is closer, the ipsilateral ear is the better ear. When the masker is closer, the contralateral ear is the better ear. In COH normalization, the SNRs at the two ears are determined directly by the normalized levels of the HRTFs shown in Fig. 2. In both BE and COH normalization, the absolute difference between the SNR at the ipsilateral ear and the SNR at the contralateral ear is approximately the same for each target and masker configuration. This difference is approximately equal to the difference in ILD between the HRTF of the target position and the HRTF of the masker position. For example, in the 12 cm–1 m configurations where the ILD is approximately 22 dB for the 12-cm source and 7 dB for the 1-m source (see Fig. 2), the difference between the ipsilateral ear SNR and contralateral ear SNR is approximately 15 dB. Small variations in these average SNR levels occurred because of differences in the spectral content of the target and masking signals and differences in the spectral shapes of the HRTFs at the two ears.

In the speech-shaped noise masker conditions, the masker level was increased by 9 dB after the normalization process in order to produce an SNR of −9 dB at the normalization point. This was done because previous speech-perception experiments in our laboratory have shown that performance with the CRM is most sensitive to changes in the relative level of a speech-shaped noise masker when the SNR of the target phrase is approximately −9 dB (Brungart, 2001b).[3] Note that in the noise-masker conditions shown in Fig. 4, the SNRs in the ipsilateral and contralateral ears are 9 dB lower than in the corresponding configurations with a speech masker.

The signals were presented at a comfortable listening level (approximately 65 dB SPL on average) as measured at the output of the headphones, and the overall level of each stimulus presentation was randomly roved over a 6-dB range (in 1-dB steps). This roving ensured that the listeners were not able to use absolute level to identify the target and masking phrases.

C. Procedure

In each trial, the target phrase was selected randomly from the 256 phrases in the speech corpus with the call sign ''baron,'' with the restriction that each talker was used the same number of times in each listening session. In the trials with a speech masker, the masking phrase was selected randomly from the 1176 phrases in the speech corpus with a different call sign, a different color coordinate, and a different number coordinate than the target phrase. Note that the random selection of the phrases resulted in same-sex target and masking talkers in 50% of the trials and different-sex target and masking talkers in 50% of the trials. In the trials with a noise masker, a random Gaussian noise was filtered with the speech-shaped noise filter and gated rectangularly to the beginning and end of each phrase. The normalization scheme (COH or BE) was also randomly chosen on each trial.

The data were collected with the listeners seated in front of the CRT of a Windows-based control computer in a quiet, sound-treated listening room. The stimuli for each trial were generated by an interactive MATLAB script, which selected the stimulus signals, processed the signals with the appropriate HRTFs, and presented the signals over headphones (Sennheiser HD540) through a Soundblaster AWE-64 sound card. The listeners were instructed to listen for the target phrase, which was always addressed to the call sign ''baron,'' and use the mouse to select the color and number contained in the target phrase from an array of colored digits displayed on the screen of the control computer. Each listener first participated in a total of 1560 trials with a speech masker. These trials were collected in 13 blocks of 120 trials each, with each block taking approximately 15 min to complete. Each listener then heard a total of 1000 trials with a speech-shaped noise masker. These trials were collected in five blocks of 200 trials each, with each block taking approximately 20 min to complete. One or two blocks were run per day for each listener over a period of several weeks. Note that some of the data were collected with normalization schemes or target-masker distance configurations that are not discussed in this paper, and that these points were excluded from the data analysis. Thus, the results that follow represent a total of 7435 trials collected with the speech masker and 4913 trials collected with the noise masker.

The distribution of these trials across the different target-masker configurations is shown in the last column of Fig. 4. Within each configuration, approximately half of the trials were conducted with BE normalization and approximately

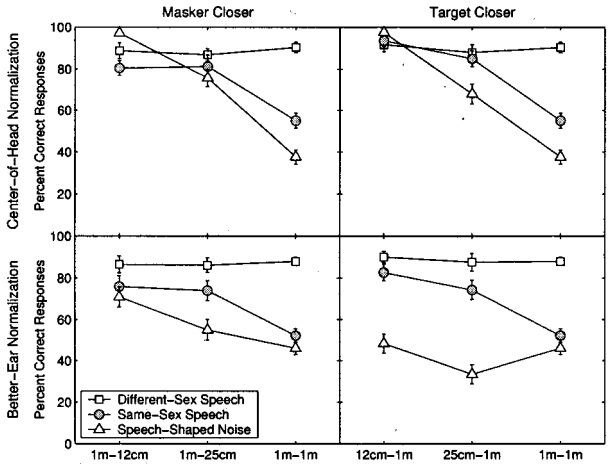


FIG. 5. Percentage of correct color and number identifications for two competing sound sources directly to the left of the listener (90 degrees azimuth). The left panels show performance for each target-masker configuration when the masker is closer than the target. The right panels show performance when the target is closer than the masker. The two rows represent the two different types of normalization used in the experiment. The symbols represent different kinds of target and masking signals (as indicated by the legend). Note that the 1m–1m condition is shown in both columns of the figure. The error bars show 95% confidence intervals calculated from the raw data for each data point.

half of the trials were conducted with COH normalization. Because the number of trials in each possible condition varied across the listeners, all of the mean performance values in the results that follow were calculated by first finding the mean performance values of each listener in that condition and then averaging across these nine individual means to determine overall performance. The standard errors bars shown in each condition represent an rms combination of the nine standard error values calculated for the individual listeners.

III. RESULTS AND DISCUSSION

A. Overall results

The overall results of the experiment are shown in Fig. 5. The data from each target-masker configuration have been plotted separately for BE normalization (bottom row) and COH normalization (top row), and the speech-masking data have been plotted separately for same-sex masking speech (circles), different-sex masking speech (squares), and speech-shaped masking noise (triangles). The left column shows the results for configurations where the masker was closer than the target talker, and the right column shows results for configurations where the target talker was closer than the masker. Each data point in the figure represents the percentage of trials in which the listeners correctly identified both the color and the number contained in the target phrase containing the call sign ''baron,'' and the error bars represent the 95% confidence intervals of each data point. The results indicate that the effects of distance separation on speech intelligibility are different for different types of maskers. When the target speech was masked by a different-sex talker (squares in Fig. 5), spatial separation in distance had little or no impact on performance. The listeners correctly identified both the color and number coordinates in the target phrase in approximately 85%–90% of the trials in all of the conditions tested. Apparently, the monaural cues that allow listeners to segregate different-sex talkers are so effective that no additional intelligibility advantage can be obtained by presenting the target and masking utterances at different distances.

When the target speech was masked by a same-sex talker (circles in Fig. 5), substantial improvements in performance occurred when the target and masking signals were spatially separated in distance. The overall percentage of correct identifications was 30–40 percentage points greater in the 12 cm–1 m configuration than in the 1 m–1 m configuration and about 25 percentage points greater in the 1 m–12 cm configuration than in the 1 m–1 m configuration. There was not much difference between the conditions where the closer talker was at 12 cm and the conditions where the closer talker was at 25 cm: performance was only about 10% better in the 12 cm–1 m configuration than in the 25 cm–1 m configuration (right panels of the figure), and was essentially identical in the 1 m–12 cm and 1 m–25 cm configurations (left panels of the figure). Apparently most of the benefits of spatial separation in distance for same-sex competing talkers can be obtained by moving one of the talkers within 25 cm of the listener's head, even though the 6-dB increase in ILD associated with a decrease in distance from 1 m to 25 cm is much smaller than the 15-dB increase in ILD associated with a decrease in distance from 1 m to 12 cm (Fig. 2).

When the target speech was masked by a speech-shaped noise masker (triangles in Fig. 5), the effects of spatial separation were substantially different for the different normalization conditions. The benefits of spatial separation in distance were greatest with the COH normalization (top panels of the figure), where the percentage of correct identifications systematically increased from approximately 40% in the 1 m–1 m configuration to approximately 75% in the 1 m–25 cm and 25 cm–1 m configurations, and to near 100% in the 1 m–12 cm and 12 cm–1 m configurations. When BE normalization was used (bottom panels of the figure), distance separation had a much smaller effect on performance. When the masker was closer than the target, the percentages of correct responses with BE normalization were 20–25 percentage points lower than in the corresponding configurations with COH normalization, and when the target was closer than the masker, separation in distance essentially had no effect on the intelligibility of the target phrase with BE normalization.

B. Target proximity

The results in Fig. 5 suggest that there are some important differences in performance between conditions where the target was closer than the masker and conditions where the target was farther away. These differences were examined by conducting a three-factor repeated-measures ANOVA for the factors of target-masker configuration (12 cm–1 m or 25 cm–1 m), relative target proximity (target closer or masker closer), and masker type (same-sex speech or speech-shaped noise) with the percentages of correct responses for each listener. The arcsine transform was applied to normalize the percentage data prior to conducting the ANOVA. The results of this ANOVA confirm that there was a significant

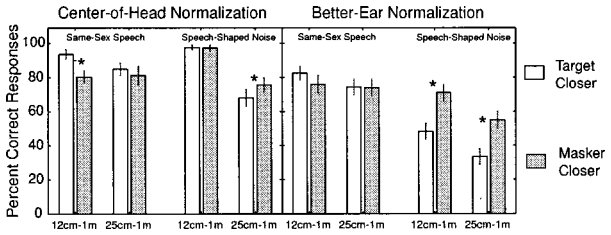


FIG. 6. Comparison of performance when the target was located closer than the masker and when the masker was located closer than the target. The asterisks indicate differences that were significant at the $p<0.05$ level (two-tailed *t*-tests). The error bars show the 95% confidence intervals calculated from the raw data in each configuration.

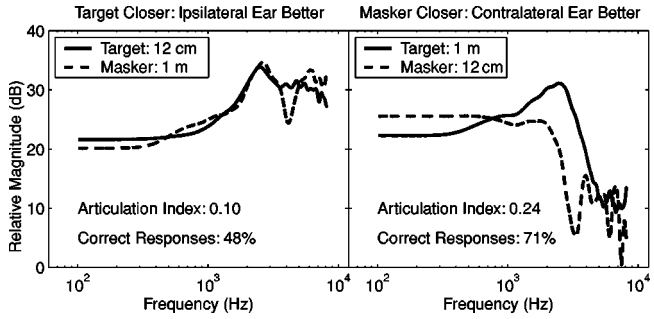


FIG. 7. Effects of spectral shape on the intelligibility of speech at the ear with the better SNR in the 12cm–1m (left panel) and the 1m–12cm (right panel) configurations of the experiment. These are the same 12 cm and 1 m HRTFs shown in Fig. 2, but their relative levels have been adjusted to equalize the overall power of speech-shaped target and masker signals (with the same frequency spectrum shown in Fig. 1) at the location of the better ear with the same method used in the BE normalization conditions of the experiment. In each case, the 20-band method developed by Kryter (1962) has been used to calculate the articulation index (AI) of the target signal when the target is presented at 56 dB SPL and the masker is presented at 65 dB SPL. This difference in AI explains the substantially larger number of correct responses that occurred in the masker-closer conditions of the experiment.

interaction between the relative distance of the target and the masker type ($F_{(1,8)}=6.573$, $p=0.033$). This interaction is shown in more detail by Fig. 6, which directly compares performance in the target-closer and masker-closer conditions for each target-masker configuration and each masker type. In the same-sex speech-masking conditions, the relative locations of the target and masker signals had only a modest impact on performance. The only significant difference occurred in the COH condition (left panel of Fig. 6), where performance in the 12 cm–1 m configuration (target closer) was approximately 10% better than in the 1 m–12 cm condition (masker closer). In part, this effect can be explained by the higher SNR that occurred at the better ear in the 12 cm–1 m condition with COH normalization. [Figure 4 shows that the SNR ratio in the better ear was 1.5 dB higher in the 12 cm–1 m configuration than in the 1 m–12 cm configuration (8.5 dB vs 7 dB).] This effect may also reflect a bias on the part of the listeners to direct their attention to the closer talker, who was located only a few centimeters from the ear.

In the noise-masking conditions, the relative distances of the target and masker had the opposite effect on performance. In three of the four noise-masking configurations shown in Fig. 6, performance was significantly worse when the target was closer than the masker than when it was farther away than the masker. This performance differential was particularly large with BE normalization (right panel of the figure), where the percentage of correct responses was more than 20 percentage points higher in the 1 m–12 cm and 1 m–25 cm configurations than in the 12 cm–1 m and 25 cm–1 m configurations. The only noise-masking configuration where performance was not significantly better with a more distant target was the 1 m–12 cm COH configuration, where performance was already near 100% in the 12 cm–1 m configuration and no measurable effect of source proximity was found.

The somewhat counterintuitive effect that relative distance had on performance with the noise masker can be explained by spectral differences in target and masker HRTFs at the ear with the more advantageous SNR in each listening configuration. Figure 7 shows the transfer functions of the HRTFs of the target and masker at the ''better ear'' for a target at 12 cm and the masker at 1 m (left panel) and for a masker at 12 cm and a target at 1 m (right panel). In both cases, the relative levels of the transfer functions have been normalized with BE normalization to make the overall rms power of a speech-spectrum-shaped noise the same when processed by either HRTF. Note that the spectral shapes of the 12 cm and 1 m HRTFs are substantially more similar at the ipsilateral ear than at the contralateral ear, and that BE normalization produces an SNR in the 2–4-kHz range that is 5–20 dB higher in the masker-closer condition than in the target-closer condition.

In order to analyze the effects of these spectral differences quantitatively, the articulation index (AI) was calculated at the better ear for a speech target presented at 56 dB SPL masked by a speech-shaped noise masker presented at 65 dB SPL with the 20-band method described by Kryter (1962).[4] This calculation indicates that the AI at the better ear was 0.24 when the target was at 1 m and 0.10 when the target was at 12 cm. This difference may explain why performance in the 1 m–12 cm configuration was substantially better than performance in the 12 cm–1 m configuration in the noise-masking conditions with BE normalization. A previous diotic experiment that measured performance in the CRM task as a function of AI with a speech-shaped noise masker (Brungart, 2001a) found that the percentage of correct color-number identifications was approximately 90% when the AI was 0.24 (compared to 71% in the 1 m–12 cm configuration of this experiment) and approximately 50% when the AI was 0.1 (compared to 48% in this 12 cm–1 m configuration of this experiment).

It is interesting to note that the spectral advantages of a more distant target talker did not have any meaningful effect on performance with the speech masker. This can be explained by the fact that speech-on-speech masking with the CRM speech task is dependent primarily on informational masking rather than on energetic masking. Previous experiments have shown that listeners are able to hear both the target and masking phrases with the CRM corpus at SNRs near 0 dB, and that most of their incorrect responses occur because the listeners are unable to segregate the content of the target phrase from the content of the masking phrase (Brungart, 2001b). Thus it is not surprising that relative per-

formance in this task does not correspond to the predictions of the articulation index, which was designed specifically to characterize the effects of energetic masking in speech.

C. Characterizing the advantages of spatial separation in distance

To this point, the performance advantages that occur when two talkers are spatially separated in distance have been described in terms of a difference in the percentage of correct identifications in the CRM task. Although this method of measuring ''spatial advantage'' is appropriate for comparing performance across different spatialization conditions with the same masking signal (as in the different-sex, same-sex, and speech-shaped noise curves in Fig. 5), it is generally not appropriate for comparing the relative advantages of spatial separation across different speech stimuli or different masking signals. These comparisons require methods of measuring spatial advantage that do not depend on the particular characteristics of the speech intelligibility test used to make the measurements.

One measure of spatial advantage that can be generalized across different speech perception tests is the change in the speech reception threshold (SRT) that occurs when the target and masking signals are spatially separated. The SRT is defined as the minimum presentation level of the target speech required to produce a predetermined threshold level of performance in the speech perception task. It is usually measured by adaptively adjusting the level of the target speech until the desired threshold level of performance is reached. A decibel measure of spatial advantage can be obtained by subtracting the SRT measured in the spatially separated condition from the SRT measured in the nonspatialized condition. This method of measuring spatial advantage does not depend on the difficulty of the particular speech intelligibility task used in the experiment, so it provides a better means of comparing the advantages of spatial separation across different stimulus and masker types than the change in the percentage of correct responses. It also allows direct comparison to the large number of experiments in the literature that have measured spatial advantage this way with noise maskers (Bronkhorst and Plomp, 1988; Festen and Plomp, 1990; Hawley *et al.*, 2000; Shinn-Cunningham *et al.*, 2001; Peissig and Kollmeier, 1997) and with speech maskers (Duquesnoy, 1983; Hawley *et al.*, 2000; Peissig and Kollmeier, 1997).

In this experiment, performance in each condition was measured only at the two SNR values determined by the BE conditions and the COH conditions shown in Fig. 4. Because these SNR values resulted in different levels of performance in each condition tested, it is not possible to directly determine a decibel measure of spatial advantage from these results. It is, however, possible to derive a decibel estimate of spatial advantage by comparing the results of this experiment to the results of a previous experiment that used the same panel of listeners and the same CRM stimuli to measure performance as a function of SNR for nonspatialized (diotic) stimuli (Brungart, 2001b). This method of estimating spatial advantage is illustrated in Fig. 8. The figure separates the data into four separate panels, with each row corresponding

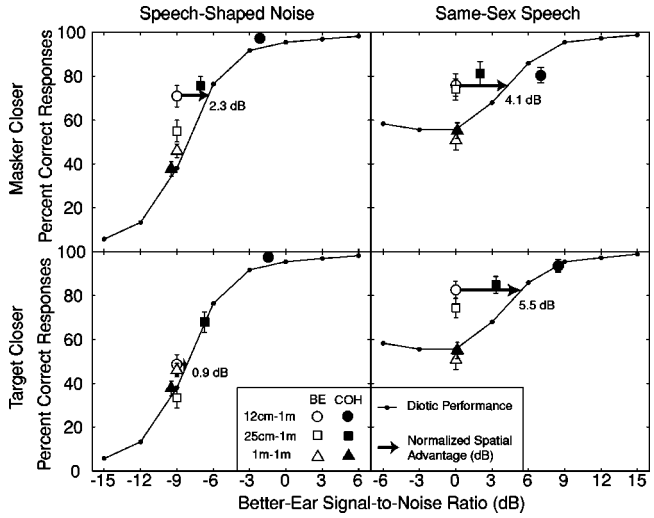


FIG. 8. Percentage of correct color and number identifications for two competing sound sources directly to the left of the listener (90 degrees azimuth) as a function of the SNR at the better ear. The left column shows results for the speech-shaped noise masker, and the right column shows results for the same-sex speech masker. The top row shows results when the masker is closer than the target, and the bottom row shows results when the target is closer than the masker. The results from the 1m–1m configuration are shown in both rows. The symbols show results for the different target-masker configurations in each panel. The open symbols represent conditions with BE normalization, and the shaded symbols represent conditions with COH normalization. The solid line shows performance as a function of signal-to-noise ratio in diotic (nonspatialized) presentations of the same target and masker signals to the same panel of nine listeners (Brungart, 2000b). The arrows in each panel show a decibel estimate of spatial advantage that has been calculated from the difference between the better-ear SNR in the BE 12cm–1m conditions (open circles) and the SNR required to achieve the same level of performance in the corresponding diotic conditions (solid lines in each panel). The error bars show 95% confidence intervals calculated from the raw data for each condition.

to a different target-masker configuration and each column corresponding to a different type of masking sound. Within each panel, the symbols show performance in each target-masker configuration as a function of the SNR at the better ear. The open symbols show performance in the BE conditions (where the better-ear SNR was forced to 0 dB), and the filled symbols show performance in the COH conditions (where the better-ear SNR was determined by the HRTFs). The error bars on each symbol represent the 95% confidence intervals for each data point. The lines in each panel of the figure show performance as a function of SNR from the previous nonspatialized (diotic) experiment that used the same speech-in-speech and speech-in-noise stimuli used in this experiment (Brungart, 2001b).

In the 1 m–1 m listening configurations, shown by the triangles in Fig. 8, overall performance was approximately the same as in the corresponding diotic configuration with the same SNR value at the better ear. This is not a surprising result, because no binaural or spectral difference cues were available to help segregate the target and masking signals in the 1 m–1 m configurations. It does, however, indicate that the 1 m HRTFs that were used to spatially process the stimuli had little impact on the overall intelligibility of the target speech.

In the spatially-separated listening configurations, per-

formance was generally better than in the corresponding diotic configurations with the same SNR values at the better ear. The right-facing arrows in the figure show a decibel estimate of this spatial advantage for the 12 cm–1 m and 1 m–12 cm listening configurations with BE normalization (open circles in the figure). In each case, the spatial advantage was estimated from the difference between the better-ear SNR in the spatially separated condition and the minimum SNR value required to achieve the same percentage of correct identifications in the corresponding diotic condition. For example, with the same-sex speech masker, correct identifications occurred in approximately 75% of the trials in the 12 cm–1 m configuration when the better-ear SNR was 0 dB (open circle in the top-right panel of the figure). In order to obtain comparable performance in the diotic condition, a better-ear SNR of approximately 4.1 dB would be required (arrow in the figure). Thus, the ''normalized'' spatial advantage in the 12 cm–1 m configuration is about 4.1 dB. Note that we refer to this estimate of spatial advantage in the BE condition as ''normalized'' spatial advantage because it compares performance in nonspatialized and spatialized listening configurations that produce the same SNR at the better ear: this is in contrast to other measures of spatial advantage that include the effects of any increase in better-ear SNR in the spatially separated condition.

These estimates of normalized spatial advantage clearly illustrate the differences that can occur between the percentage estimates of spatial advantage and decibel estimates of spatial advantage. Spatial separation produced the largest percentage point increase in performance ($\approx$30%) in the 12 cm–1 m speech-shaped noise condition with a closer masker (upper left corner of Fig. 8). However, because performance in the CRM test increases much faster with SNR with a noise masker than with a speech masker, the decibel spatial advantage was substantially larger for the speech-masking conditions than for the noise-masking conditions.

The larger normalized spatial advantages that occurred in the speech-masking conditions of the experiment also suggest that binaural difference cues play a greater role in distance-based speech segregation with a speech masker than with a noise masker. The overall decibel measure of normalized spatial advantage shown in Fig. 8 includes the effects of two different types of spatial segregation cues. The first is a monaural spectral cue based on differences in the spectral characteristics of the target and masking signals at the better ear. The second is a binaural cue, often referred to as binaural advantage or binaural interaction (Zurek, 1993; Hawley *et al.*, 2000), which allows listeners to segregate sounds on the basis of variations in the interaural difference cues produced by the target and masking signals. In this experiment, there is reason to believe the monaural spectral cue was responsible for most of the 2.3-dB normalized spatial advantage found in the 1 m–12 cm noise configuration shown in the top left panel of Fig. 8. As discussed in the previous section, differences in the spectral shapes of the 12 cm and 1 m HRTFs produce an SNR advantage at the better ear that substantially improves performance with the noise masker in the masker-closer configurations but has little effect on performance in the target-closer configurations. These monaural spectral cues could explain most of the 2.3-dB spatial advantage found in the masker-closer configurations, and the absence of these spectral cues may explain why the spatial advantage was smaller (0.9 dB) in the target-closer configurations. At the same time, there is little evidence that monaural spectral cues had much influence on performance in the speech-masker conditions. Indeed, the spatial advantage was actually lower in the masker-closer conditions where the monaural spectral segregation cues should have provided the most benefit. Overall, though, the spatial advantage was roughly comparable in the target-closer and masker-closer configurations with the speech masker, suggesting that the segregation was based primarily on binaural difference cues that were symmetric across the two configurations. Thus, it appears that binaural cues were responsible for a large portion of the relatively large spatial advantages found for the same-sex speech masker, but only for a small portion of the relatively small spatial advantages found with the speech-shaped noise masker.

Further evidence that the segregation of speech from a noise masker is dominated by monaural cues is provided by the relatively large influence that better-ear SNR had on performance in the noise-masking conditions of the experiment. In all the spatially separated configurations tested with the noise masker, large increases in performance occurred with the larger better-ear SNR values in the COH normalization conditions (comparing the open and black symbols in the left column of Fig. 8). This is in direct contrast to the spatially separated configurations with the speech masker, where performance increased only modestly in the COH conditions (comparing the open and black symbols in the right column of Fig. 8). Clearly the SNR at the more advantageous ear was a more important factor in determining performance with the noise masker than it was with the speech masker.

Overall, these data suggest that listeners who are attempting to segregate a speech signal from a speech masker perform much better when they have access to the acoustic signals at both ears than when they only have access to the acoustic signal at the ear with the higher SNR. In contrast, listeners who are attempting to extract information from a speech signal masked by noise receive little benefit from having access to the signals at both ears. This supports the hypothesis that the binaural difference cues associated with spatial separation in distance contribute more to spatial unmasking with an informational speech masker than with an energetic noise masker. It does not, however, make it clear whether the difference is due to low-level binaural signal processing or if it is a higher-level process related to a difference in the apparent locations of the sounds. The next section describes a second experiment that was designed to explore this issue in more detail.

IV. DISTANCE PERCEPTION IN TWO-TALKER SPEECH STIMULI

The results of the first experiment suggest that listeners are able to use the distance-dependent changes that occur in the HRTFs of nearby sound sources to segregate speech signals that originate from different locations along the interaural axis. They do not, however, provide any information

about where the listeners perceived those speech signals. In order to draw any definitive conclusions about the effect that spatial separation in distance has on multitalker speech segregation, it is necessary to verify that two conditions were met by the virtual stimuli presented in this experiment: (1) that the target and masking signals appeared to be located at the same angle in azimuth and (2) that the target and masking signals appeared to be located at different distances.

The first of these conditions is required to ensure that the listeners were not performing the segregation task on the basis of differences in the apparent directions of the target and masking signals. The results of a previous experiment that measured localization judgments for noise bursts that were processed with the same set of HRTFs used in experiment 1 (Brungart and Simpson, 2001) provide some evidence that the two-talker stimuli used in this experiment met this ''equal apparent azimuth'' requirement. These results, which are illustrated in Fig. 3, show that 12 cm, 25 cm, and 1 m HRTFs measured at 270 degrees in azimuth with the same techniques used to produce the HRTFs used in this experiment all resulted in median location judgments near 270 degrees. Although there was some variability in the responses, these results show that there was no systematic tendency to perceive the stimuli at different locations in azimuth. Even if there were small differences in the apparent azimuth locations of the signals, the impact of these differences would be limited somewhat by the relative insensitivity of human listeners to changes in the directions of sounds near the interaural axis. Experiments that have measured the minimum audible angle (MAA) for lateral source positions have found that the MAA is roughly 10 degrees for sound sources at 75 degrees azimuth and roughly 20 degrees for sound sources at 90 degrees azimuth (Chandler and Grantham, 1992; McKinley *et al.*, 1994). Thus, the 3–6-degree variations in the median azimuth judgments for the different HRTF distances shown in Fig. 3 are small relative to the MAA in this region. Consequently, we do not believe the advantages of spatial separation in distance found in the first experiment can be explained by differences in the apparent azimuth locations of the competing signals.

The results shown in Fig. 3 also provide evidence that the ''different apparent distance'' requirement was met by the stimuli used in the first experiment. Although the responses were compressed relative to the range of simulated distances, the median response distances increased systematically with the simulated distances of the noise bursts. However, this experiment did not measure perceived distance with virtual speech sounds, and it did not measure the listener's ability to localize the distances of two simultaneous virtual sounds. This makes it difficult to know for certain whether the listeners in the first experiment were actually perceiving the target and masking signals at different distances. In order to address this issue, a second experiment was conducted that examined how well listeners were able to judge the relative distances of the target and masking talkers in the two-talker stimuli used in the first experiment.

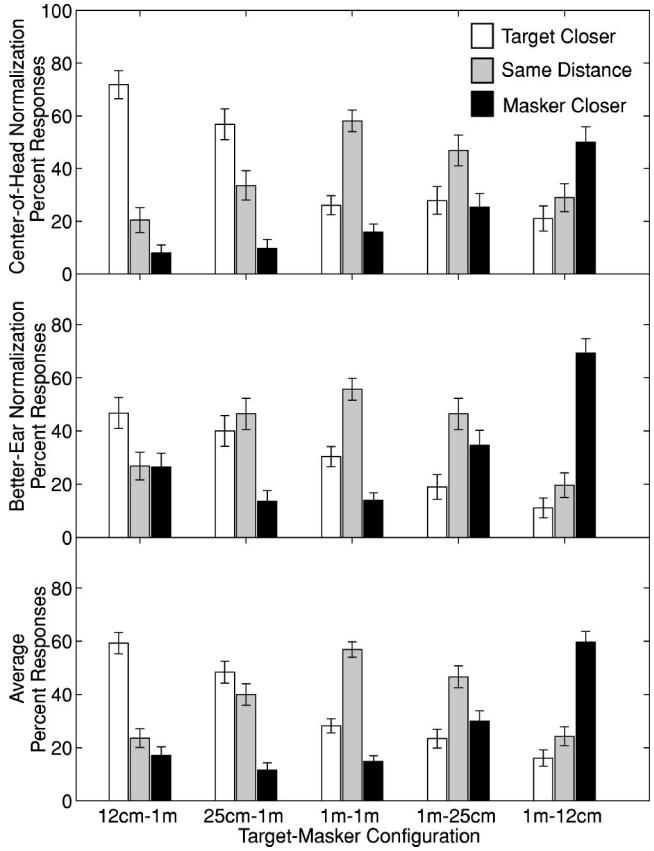


FIG. 9. Distribution of responses in experiment 2. Each group of three bars shows the percentages of target-closer, same-distance, and masker-closer responses for a different target-masker configuration. Note that the leftmost bar in each group represents correct responses in the 12 cm–1 m and 25 cm–1 m configurations, the middle bar represents correct responses in the 1 m–1 m configurations, and the rightmost bar represents correct responses in the 1 m–25 cm and 1 m–12 cm configurations. The top panel shows the results for COH normalization, the middle panel shows the results for BE normalization, and the bottom panel shows the results averaged across both normalization conditions. The error bars show 95% confidence intervals calculated from the raw data in each configuration.

A. Methods

The stimuli used in experiment 2 were almost identical to those used in the speech-masking conditions of experiment 1. They consisted of pairs of randomly selected phrases from the CRM corpus that were processed with KEMAR HRTFs measured at different distances (12 cm, 25 cm, or 1 m) along the listener's interaural axis. One of the phrases (the target phrase) always contained the call sign ''baron,'' and the other phrase (the masking phrase) always contained the call sign ''ringo.'' The HRTF-processed stimuli were normalized to have an SNR equal to 0 dB either at the center of the head (COH normalization) or at the ear with a higher SNR (BE normalization). Then they were mixed together digitally and played back to the listener over stereo headphones (Sennheiser HD540) in a quiet, sound-treated listening room. The overall level of each stimulus presentation was randomly roved over a 6-dB range in 1-dB steps.

Although the stimuli were similar to those used in the first experiment, the task was quite different. After each stimulus presentation, the listeners were asked to determine whether the target phrase was closer than the masking

phrase, the same distance as the masking phrase, or farther away than the masking phrase. They responded by using the computer mouse to select ''Baron Closer,'' ''Same Distance,'' or ''Baron Farther'' on the screen of the control computer. No feedback was provided about the actual locations of the stimuli.

A total of seven listeners participated in the second experiment, including five who also participated in the first experiment. Each listener participated in four blocks of 120 trials, with each block consisting of five repetitions of all combinations of two target-masker voice configurations (same-sex or different-sex), two normalization schemes (BE or COH), and five target-masker distance configurations (12 cm–1 m; 25 cm–1 m; 1 m–1 m; 1 m–25 cm; and 1 m–12 cm). Note that, as in the first experiment, twice as many trials were collected in the 1 m–1 m configuration than in the other configurations.

B. Results and discussion

Figure 9 shows the distribution of responses for each target-masker configuration and normalization scheme used in the experiment. The top panel shows the results for COH normalization, the middle panel shows results for BE normalization, and the bottom panel shows the results averaged across these two normalization schemes. Within each target-masker configuration, the left bar shows the percentage of target-closer responses, the middle bar shows the percentage of same-distance responses, and the right bar shows the percentage of target-farther responses (as indicated in the legend). The error bars represent the 95% confidence intervals of each data point.

The averaged data shown in the bottom panel of the figure indicate that the listeners were able to make reasonably accurate judgments about the relative distances of the target and masking phrases. In the 12 cm–1 m, 1 m–12 cm, and 1 m –1 m configurations, the listeners correctly identified the relative location of the target phrase in approximately 60% of the trials. While this is far from perfect performance, it is well above chance and is perhaps remarkably good when one considers that the task required the listener to correctly identify the phrase containing ''baron'' while simultaneously determining the location of that phrase relative to the masker. It is also important to note that these results cannot be explained by distance-dependent differences in the overall levels of the competing talkers. Although there was a systematic relation between overall intensity and distance in the COH normalization conditions (where the closer talker was always more intense in the left ear and less intense in the right ear), there were no consistent level-based distance cues in the BE normalization conditions. In the 12 cm–1 m condition with BE normalization, for example, the closer target talker was presented at the same level as the more distant masking talker in the left ear, and at a level almost 15 dB lower than the more distant masking talker in the right ear (see Fig. 4). Thus, depending on how the listeners integrated the intensity of the stimuli across the two ears, the overall level cue in the 12 cm–1 m condition with BE normalization was either nonexistent or in opposition to the usual inverse relationship between intensity and distance. Despite this misleading intensity cue, the listeners in this condition performed well above chance in identifying the 12 cm target talker as the closer talker (left-hand side of the middle panel in Fig. 9). The ability of the listeners to correctly identify the relative distances of the target and masking talkers without feedback and in the presence of misleading intensity cues confirms that the KEMAR HRTFs used in these experiments provided sufficiently realistic acoustic cues for the listeners to perceive the competing talkers at different distances.

It is, however, apparent that the listeners were much better at segregating the two speech messages in experiment 1 than they were at determining the relative distances of the two speech signals in experiment 2. In the 12 cm–1 m configuration with BE normalization, for example, the listeners correctly identified both the color and number in the target phrase approximately 85% of the time in experiment 1 (lower right panel of Fig. 4), but correctly identified the relative distance of the target talker in less than 50% of the trials of experiment 2. Most of this reduction in performance can likely be attributed to the increased complexity of the task in experiment 2, where the listeners had to both identify the target phrase and determine its relative distance at the same time. In contrast, listeners were required only to identify the target phrase in experiment 1.

One final interesting aspect of the data from the second experiment is that there were no indications of the large differences in performance that occurred in the same-sex and different-sex masking conditions of experiment 1. In fact, overall performance was identical with the same- and different-sex masking voices in experiment 2 (52% correct responses). Thus, although both experiments required the listeners to segregate the target phrase from the masking phrase, it is apparent that differences in the vocal characteristics of the two talkers provided a much larger benefit in the speech intelligibility task in experiment 1 than in the distance localization task in experiment 2.

V. DISCUSSION AND CONCLUSIONS

The results of these experiments provide insights into the role that spatial separation in distance plays in determining the intelligibility of a nearby talker masked by a competing nearby sound. When the listening task is relatively easy to perform with spatially co-located signals, spatial separation of the target and masker in distance does not improve the intelligibility of the talker. This is apparent from the lack of any discernible differences between the conditions where the target and masker phrases were presented at different distances and those where they were presented at the same distance when two phrases were spoken by different-sex talkers.

When the target phrase is masked by noise, spatially separating the target and masker can produce a tremendous improvement in speech intelligibility (from 40% to near 100%). However, nearly all of this benefit is derived from spectral differences in the target and masker signals at the better ear. The binaural difference cues that appear to dominate the perception of distance for nearby sound sources (Brungart, 1999a) contribute little or nothing to our ability to segregate a nearby talker from a masking noise—the overall

spatial advantage was not much larger than 2 dB even for the largest spatial separations used in this experiment, and most of this overall spatial advantage was the result of monaural spectral cues at the listener's better ear. This is consistent with the findings of Shinn-Cunningham *et al.* (2001), who showed that spatial unmasking effects with a noise masker are dominated by spectral ''better ear'' advantages, and that binaural interaction effects account for less than 2 dB of a total threshold shift of 25 dB or more for sound sources located at 15 cm and 1 m along the interaural axis of the listener. They also found that their predictions overestimated performance when the target and masker were at 90 degrees, suggesting that even 2 dB is a generous estimate of the actual binaural advantage that can be obtained by separating the distances of a nearby speech signal and noise masker.

When the target phrase is masked by same-sex speech, however, spatially separating the target and masker can produce improvements in intelligibility that substantially exceed those that would be predicted from spectral differences at the better ear. The overall spatial advantage of separating the target and masker in distance was as large as 5.5 dB when the SNR at the better ear was normalized to 0 dB, and monaural spectral cues seemed to contribute very little to this overall spatial advantage. In fact, in contrast to the relatively minor role that binaural cues play in spatial unmasking with a noise masker, binaural cues appear to account for most of the spatial unmasking that occurs with a speech masker. The results in Fig. 5 show that spatial separation in distance improved performance with a same-sex speech masker by about 28 percentage points with BE normalization and by about 33 percentage points with COH normalization. Thus, about 85% of the intelligibility improvement afforded by spatially separating the talkers in distance was maintained when the SNR advantage at the better ear was eliminated. With the noise masker, spatial separation in distance improved performance by about 60 percentage points with COH normalization and only by about 15 percentage points with BE normalization (averaging the target-closer and masker-closer conditions). Thus, only about 25% of the intelligibility improvement was maintained when the better-ear SNR advantage was eliminated with a noise masker. Clearly binaural difference cues play a much larger role in the spatial unmasking of sound sources that are spatially separated in distance when the masker is same-sex speech than when the masker is speech-shaped noise.

These results are consistent with other recent multitalker experiments that have found similar differences between the binaural advantages of spatial separation in azimuth with a distant noise masker or a distant speech masker. Hawley *et al.* (2000), for example, also found that the binaural advantages of spatial separation in azimuth were 3–4 dB larger when two or more speech or time-reversed speech signals were used as the maskers than when two or more noise or modulated noise signals were used as the maskers. Freyman *et al.* (1999) used the precedence effect to manipulate the apparent locations of a target and a masker without affecting the SNR at either ear, and found that apparent location had a substantial effect on speech intelligibility with a speech masker but essentially no effect with a noise masker. Thus it appears to be generally true that the binaural difference cues associated with spatial separation of the target and masking sounds have a much greater impact on speech intelligibility with speech maskers than with noise maskers. Freyman and his colleagues have suggested that this occurs because differences in the apparent locations of sounds can dramatically reduce informational masking by enhancing the listener's ability to selectively attend to the target speech and avoid being distracted by the contents of the interfering speech signal. If this is the case, then the results of this experiment can be explained in the same way: the informational masking that occurred with the same-sex speech masker was reduced when the target and masker were spatially separated in distance because they appeared to originate at different locations in space. The results of the second experiment support this hypothesis, because they show that the listeners generally perceived the talkers at different distances along the interaural axis. However, further research is necessary to conclusively determine whether the intelligibility advantages of spatial separation found in experiment 1 were caused by differences in the apparent locations of the competing sounds or if they were the result of some other kind of binaural processing related to the enlarged ILDs that occur for sound sources near the head.

Although the mechanisms involved are not yet fully understood, it is clear from the results of these experiments that the spatial unmasking of speech associated with the traditional ''cocktail-party'' effect can be achieved by spatial separation in distance as well as spatial separation in direction when the target and masking sounds are located near the listener. Additional experiments are now needed to examine the effects that spatial separation in distance has at locations off of the listener's interaural axis in order to form a more complete picture of the interactions that occur between apparent distance and apparent direction in the segregation of competing sound sources near the head.

ACKNOWLEDGMENTS

The authors would like to thank Rich Freyman and Larry Feth for their insightful comments during the preparation of the manuscript. Portions of this research were presented at the 2001 International Conference on Auditory Display in Espoo, Finland. This research was funded in part by AFOSR LRIR 01HE01COR.

[1]Although this does not preserve the exact phase information of the original HRTFs, previous research has shown that linear-phase HRTFs that maintain the correct low-frequency phase information are indistinguishable from HRTFs that preserve the original phase information (Kulkarni *et al.*, 1999).

[2]The absolute azimuth errors were approximately 16 degrees with the virtual sounds and approximately 9 degrees with the free-field sounds. The stimulus-response correlation coefficient in distance for sources along the interaural axis was approximately 0.6 for the virtual sounds and approximately 0.8 for the free-field sounds.

[3]If the experiment were conducted with a 0 dB SNR in the better ear with a noise masker, performance in the CRM task would asymptote to near 100%. Conversely, if the experiment were conducted with a −9 dB SNR in the better ear with a speech masker, performance would be in a region where SNR is known to have relatively little effect on performance. In order to avoid these problems, the noise masker was normalized to a level 9 dB higher than the level of the speech masker in each corresponding stimulus configuration.

[4]Note that this analysis assumed a speech-shaped noise signal that exactly matched the spectrum of the talker. This differs slightly from the actual noise masking conditions of the experiment, where the speech-shaped noise signal matched the average spectrum across all the talkers in the corpus and not the average spectrum of any individual talker. Presumably the AI would be slightly higher with a noise masker that did not exactly match the spectrum of the target talker. This is likely to be a small effect, however. Festen and Plomp (1990) found only a small difference between speech-shaped noise that matched the spectrum of a same-sex talker and noise that matched the spectrum of a different-sex talker.

Abouchacra, K., Tran, T., Besing, J., and Koehnke, J. (**1997**). ''Performance on a selective attention task as a function of stimulus presentation mode,'' in *Proceedings of the Midwinter Meeting of the Association for Research in Otolaryngology*, St. Petersburg Beach, Florida.

Bolia, R., Nelson, W., Ericson, M., and Simpson, B. (**2000**). ''A speech corpus for multitalker communications research,'' J. Acoust. Soc. Am. **107**, 1065–1066.

Bronkhorst, A. (**2000**). ''The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions,'' Acustica **86**, 117–128.

Bronkhorst, A., and Plomp, R. (**1988**). ''The effect of head-induced interaural time and level difference on speech intelligibility in noise,'' J. Acoust. Soc. Am. **83**, 1508–1516.

Bronkhorst, A., and Plomp, R. (**1992**). ''Effects of multiple speechlike maskers on binaural speech recognition in normal and impaired listening,'' J. Acoust. Soc. Am. **92**, 3132–3139.

Brungart, D. (**1999a**). ''Auditory localization of nearby sources. III: Stimulus effects,'' J. Acoust. Soc. Am. **106**, 3589–3602.

Brungart, D. (**1999b**). ''Auditory Parallax Effects in the HRTF for Nearby Sources,'' in *Proceedings of 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, 17–20 October 1999, pp. 171–174.

Brungart, D. (**2001a**). ''Evaluation of speech intelligibility with the coordinate response measure,'' J. Acoust. Soc. Am. **109**, 2276–2279.

Brungart, D. (**2001b**). ''Informational and energetic masking effects in the perception of two simultaneous talkers,'' J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D., and Rabinowitz, W. (**1999**). ''Auditory localization of nearby sources. I: Head-related transfer functions,'' J. Acoust. Soc. Am. **106**, 1465–1479.

Brungart, D., and Simpson, B. (**2001**). ''Auditory Localization of Nearby Sources in a Virtual Audio Display,'' in *Proceedings of 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, 21–24 October 2001, pp. 107– 110.

Brungart, D., Durlach, N., and Rabinowitz, W. (**1999**). ''Auditory localization of nearby sources. II: Localization of a broadband source,'' J. Acoust. Soc. Am. **106**, 1956–1968.

Chandler, D. W., and Grantham, D. W. (**1992**). ''Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth, and velocity,'' J. Acoust. Soc. Am. **91**, 1624–1636.

Drullman, R., and Bronkhorst, A. (**2000**). ''Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation,'' J. Acoust. Soc. Am. **107**, 2224–2235.

Duquesnoy, A. (**1983**). ''Effect of a single interfering noise or speech source on the binaural sentence intelligibility of aged persons,'' J. Acoust. Soc. Am. **74**, 739–943.

Ericson, M., and McKinley, R. (**1997**). ''The intelligibility of multiple talkers spatially separated in noise,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Hillsdale, NJ), pp. 701–724.

Festen, J., and Plomp, R. (**1990**). ''Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing,'' J. Acoust. Soc. Am. **88**, 1725–1736.

Freyman, R., Balakrishnan, U., and Helfer, K. (**2001**). ''Spatial release from informational masking in speech recognition,'' J. Acoust. Soc. Am. **109**, 2112–2122.

Freyman, R., Helfer, K., McCall, D., and Clifton, R. (**1999**). ''The role of perceived spatial separation in the unmasking of speech,'' J. Acoust. Soc. Am. **106**, 3578–3587.

Hawley, M., Litovsky, R., and Colburn, H. (**1999**). ''Speech intelligibility and localization in a multi-source environment,'' J. Acoust. Soc. Am. **105**, 3436–3448.

Hawley, M., Litovsky, R., and Culling, J. (**2000**). ''The 'cocktail party' effect with four kinds of maskers: Speech, time-reversed speech, speech-shaped noise, or modulated speech-shaped noise,'' in *Proceedings of the Midwinter Meeting of the Association for Research in Otolaryngology*, p. 31.

Kidd, G. J., Mason, C., Rohtla, T., and Deliwala, P. (**1998**). ''Release from informational masking due to the spatial separation of sources in the identification of nonspeech auditory patterns,'' J. Acoust. Soc. Am. **104**, 422–431.

Kryter, K. (**1962**). ''Methods for calculation and use of the articulation index,'' J. Acoust. Soc. Am. **34**, 1689–1697.

Kulkarni, A., Isabelle, S., and Colburn, H. (**1999**). ''Sensitivity of human subjects to head-related transfer function phase spectra,'' J. Acoust. Soc. Am. **105**, 2821–2840.

Levitt, H., and Rabiner, L. (**1967**). ''Predicting binaural gain in intelligibility and release from masking of speech,'' J. Acoust. Soc. Am. **42**, 820–829.

McKinley, R., Ericson, M., and D'Angelo, W. (**1994**). ''Three dimensional audio displays: development, applications, and performance,'' Aviat. Space Environ. Med. **65**, A31–38.

Nelson, W. T., Bolia, R. S., Ericson, M. A., and McKinley, R. L. (**1999**). ''Spatial audio displays for speech communication. A comparison of free-field and virtual sources,'' in *Proceedings of the 43rd Meeting of the Human Factors and Ergonomics Society*, pp. 1202–1205.

Peissig, J., and Kollmeier, B. (**1997**). ''Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners,'' J. Acoust. Soc. Am. **35**, 1660–1670.

Plomp, R. (**1976**). ''Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of the azimuth of a single competing sound source (speech or noise),'' Acustica **34**, 325–328.

Plomp, R., and Mimpen, A. (**1979**). ''Improving the reliability of testing the speech reception threshold,'' Audiology **18**, 43–52.

Shinn-Cunningham, B., Schickler, J., Kopco, N., and Litovsky, R. (**2001**). ''Spatial unmasking of nearby speech sources in a simulated anechoic environment,'' J. Acoust. Soc. Am. **110**, 1118–1129.

Wightman, F., and Kistler, D. (**1989a**). ''Headphone simulation of free-field listening. I: Stimulus synthesis,'' J. Acoust. Soc. Am. **85**, 858–867.

Wightman, F., and Kistler, D. (**1989b**). ''Headphone simulation of free-field listening. II: Psychological validation,'' J. Acoust. Soc. Am. **85**, 868–878.

Zurek, P. M. (**1993**). ''Binaural advantages and directional effects in speech intelligibility,'' in *Acoustical Factors Affecting Hearing Aid Performance*, 2nd ed., edited by G. Studebaker and I. Hochberg (Allyn and Bacon, Portland).

Broadband sound generation by confined turbulent jets

Zhaoyan Zhang, Luc Mongeau,[a] and Steven H. Frankel
School of Mechanical Engineering, Purdue University, West Lafayette, Indiana 47907-1077

(Received 9 November 2001; revised 14 May 2002; accepted 17 May 2002)

Sound generation by confined stationary jets is of interest to the study of voice and speech production, among other applications. The generation of sound by low Mach number, confined, stationary circular jets was investigated. Experiments were performed using a quiet flow supply, muffler-terminated rigid uniform tubes, and acrylic orifice plates. A spectral decomposition method based on a linear source-filter model was used to decompose radiated nondimensional sound pressure spectra measured for various gas mixtures and mean flow velocities into the product of (1) a source spectral distribution function; (2) a function accounting for near field effects and radiation efficiency; and (3) an acoustic frequency response function. The acoustic frequency response function agreed, as expected, with the transfer function between the radiated acoustic pressure at one fixed location and the strength of an equivalent velocity source located at the orifice. The radiation efficiency function indicated a radiation efficiency of the order $(kD)^2$ over the planar wave frequency range and $(kD)^4$ at higher frequencies, where k is the wavenumber and D is the tube cross sectional dimension. This is consistent with theoretical predictions for the planar wave radiation efficiency of quadrupole sources in uniform rigid anechoic tubes. The effects of the Reynolds number, Re, on the source spectral distribution function were found to be insignificant over the range $2000<\mathrm{Re}<20\,000$. The source spectral distribution function approximately obeyed a St^{-3} power law for Strouhal number values $\mathrm{St}<0.9$, and a St^{-5} power law for $\mathrm{St}>2.5$. The influence of a reflective open tube termination on the source function spectral distribution was found to be insignificant, confirming the absence of a feedback mechanism. [DOI: 10.1121/1.1492817]

PACS numbers: 43.70.Aj, 43.28.Ra [MSH]

LIST OF SYMBOLS

a	sphere diameter (m)
c	speed of sound (m/s)
d	orifice diameter (m)
D	tube cross section dimension (m)
f	frequency (Hz)
E	nondimensional sound pressure spectrum
F	source spectral distribution function
g	Green's function
G	acoustic frequency response function
G_{exp}	measured system transfer function
He	Helmholtz number, fD/c
k	wave number (m^{-1})
L	distance from the orifice (m)
M	radiation efficiency function
P	sound power
p	total pressure (Pa)
p'	acoustic pressure (Pa)
r	distance between source and observer (m)
R	reflection coefficient
Re	Reynolds number, Ud/v
S_{pp}	sound pressure spectral density
St	Strouhal number, fD/U
T_{ij}	Lighthill stress tensor
U	jet centerline velocity (m/s)
v_c	velocity of the sphere center (m/s)
x	observer position
y	source position
Δp	mean pressure drop across the orifice (Pa)
ρ_0	ambient density ($\mathrm{kg/m}^3$)
ρ'	acoustic density ($\mathrm{kg/m}^3$)
ω	angular frequency (rad/s)
ν	kinematic viscosity ($\mathrm{m/s}^2$)
σ_{ij}	viscous stress tensor (Pa)
τ	source time coordinate
σ	standard deviation
ϕ	sound source strength

I. INTRODUCTION

Sound generation by jet flows has been the object of many previous investigations (Blake, 1986). Most of these studies have considered the problem of sound production by free high-speed jets because of its importance for aircraft engine noise (Hubbard, 1994). There has been relatively less interest in noise from low-speed confined jets. The motivation for the study described in this paper is the potential significance of this problem for speech production. Sound generation by confined low-speed jet flows is always involved in the generation of fricative consonants, an important component of speech. It may also contribute to voice production, since the jet plume downstream of the glottis is

[a]Author to whom correspondence should be addressed. Electronic mail: mongeau@ecn.purdue.edu

most likely turbulent over a large fraction of one glottis oscillation. Turbulence may be responsible for a random component in the sound produced by the air flow through the larynx, which may play a significant role in defining voice quality. A better understanding of this contribution could help develop better models for speech synthesis, and assist in the diagnosis of pathological voice conditions.

Jet noise theory is usually based on the work of Lighthill (1952), and subsequent refinements by many investigators, in which the basic equations for the fluid motion are formulated in the form of an inhomogeneous wave equation. This approach, termed the acoustic analogy, allows presumed sound source terms to be identified. In particular, for nonheated subsonic free jet flows, the source term is the double divergence of Lighthill's stress tensor, a quantity dominated by the instantaneous, time varying Reynolds stress tensor. In a free field, the source distribution is of a quadrupole-type, known to be an inefficient sound generation mechanism relative to monopole or dipole contributions. The radiated sound pressure may be obtained from the convolution of the source term and a free space Green's function over the entire region where the instantaneous Reynolds stress is significant.

The confinement of the source in a channel modifies the generation and propagation of sound waves significantly. The characteristics of the turbulent jet flow are also different due to the effects of confinement. For example, the strength of the recirculation region is often increased. Additional sources may arise from the impingement of the jet plume on the rigid walls. Finally, the one-dimensional nature of the radiated sound waves at low frequencies changes the basic character of the sound sources from quadrupole- to dipole-, or monopole-type sources. This generally enhances sound production efficiency.

The theory of sound production by turbulence in a rigid pipe has been discussed by Davies and Ffowcs Williams (1968). Lighthill's acoustic analogy was used, with a theoretical expression for the Green's function satisfying rigid boundary conditions on the walls of a rectangular pipe. The radiated sound pressure spectral density was expressed as the summation of acoustic modes within the pipe. A dimensional analysis of the predicted radiated acoustic power was performed. Two limiting cases were considered: (1) low frequencies, where the planar wave mode of propagation dominates; and (2) high frequencies, where higher order duct modes are excited. It was found that the sound power from large-scale turbulence scales with the sixth power of the jet centerline velocity, regardless of frequency. The sound power from small-scale turbulence at low frequencies scales with the sixth power law of the jet velocity. At high frequency (above the cut-on frequency of the first high acoustic mode in the tube), the sound power scales with the eighth power of the jet velocity, similar to the power scaling law of free subsonic jets.

A similar approach was followed by Nelson *et al.* (1981) to study sound production by sound absorbing splitters in a flow duct. The sound source term was modeled as a dipole. The differences in behavior at low and high frequencies were again highlighted. The sound power within the low- and high-frequency regions obeyed a U^4 and a U^6 power law, respectively. Measured sound pressure spectra were cast in nondimensional form. It was shown that one-third octave band sound power spectra measured for various flow speeds were in satisfactory agreement with the postulated scaling laws.

The production of broadband sound from air flow through the larynx, the vocal tract, and the airway has been sparsely investigated (Harper, 2000). In early experiments by Meyer-Eppler (1953), a critical Reynolds number was identified below which there was no significant turbulent sound generation. Meyer-Eppler postulated that the sound pressure increased with the square of the Reynolds number. The contribution of flow impingement on obstructions in the vocal tract was investigated using a hierarchy of physical models by Shadle (1986). The sound pressure generated by flow impingement was found to increase with the third power of the volume flow rate, and decrease with the 2.5th power of the cross-sectional area of the obstruction (Stevens, 1998). Stevens describes the sound spectrum associated with flow impingement as having a broad peak distributed around a certain Strouhal number value, based on the local flow velocity and the duct diameter near the obstruction. Landau and Lifshitz (1997) have estimated that for turbulence in a tube, the roll-on frequency of the sound spectrum is of the order of u/D, and the cutoff frequency is of the order of $u\mathrm{Re}/D$, where u is the flow velocity and D is the diameter of the tube. The recent work of Harper (2000), aimed at a better understanding of breathing sounds, is again based on a Reynolds number power law. Such models neglect the possible influence of frequency on sound production (i.e., Strouhal number effects). The influence of the acoustic response of the cavities and the duct on the power law exponents is based on an inverse filtering approach, which may not yield accurate results when the sound source region is spatially distributed and incoherent.

In confined jet flows, acoustic effects such as sound reflection and resonance are often present. At low frequencies, longitudinal plane acoustic waves may be reflected at the tube termination or at any location where the cross-sectional area changes, setting up standing waves and causing resonance. At high frequencies, where higher order modes are excited, the acoustic field is three-dimensional. The acoustic response of the system then becomes more complicated, with acoustic modes depending on boundary conditions at the tube walls and the termination. These phenomena significantly affect the level and the spectral content of the radiated sound, which makes it difficult to isolate and study the source characteristics directly from sound pressure measurements.

It is desirable to separate the spectral characteristics of the sound source from the acoustic response of the ducting system in order to develop source-filter models that can be utilized over a range of flow conditions, and different working fluids. Weidemann (1971) developed a spectral decomposition method for the separation of source characteristics and system response from experimental data based on acoustic similarity laws for fan noise. The similarity laws express nondimensional radiated sound pressure spectra as the product of a source function, which is presumed to be a function

of only the Strouhal number, and an acoustic frequency response function, presumed to be a unique function of the Helmholtz number. The Mach and Reynolds numbers also appear in the form of power laws. Neise (1975) verified the acoustic similarity laws for the tonal sound emitted by industrial centrifugal fans. Bartenwerfer *et al.* (1976) argued that the source spectral characteristics should be a function of both the Strouhal number and the Reynolds number, and questioned the relevance of power law terms. Mongeau *et al.* (1995) applied the spectral decomposition method to study the aerodynamic sound generating mechanisms in a turbomachinery rotor. The usefulness of the source-filter based method was demonstrated by varying the source characteristics and the acoustic loading independently.

The purpose of the present study was to investigate the characteristics of aerodynamic broadband sound generation by stationary confined jet flows. The ultimate goal is to develop semi-empirical models (scaling laws) that could allow the amplitude and the spectrum of the random component of synthesized voice signals to be established. To accomplish this, the spectral decomposition method was used to isolate source characteristics from other sound propagation effects. Circular orifices were used, as a first step, to investigate the validity and the usefulness of the approach. The effects of orifice geometry are considered in an ongoing, parallel study.

II. SOUND GENERATION BY STATIONARY CONFINED TURBULENT JETS

Consider sound generation by a finite region of turbulence in a long rigid rectangular tube. The tube may or may not be anechoically terminated. Lighthill's equation, which is derived from the continuity and momentum equations, is

$$\frac{\partial^2 \rho'}{\partial \tau^2} - c_0^2 \nabla^2 \rho' = \frac{\partial^2 T_{ij}}{\partial y_i \partial y_j}, \tag{1}$$

where $\rho' = \rho - \rho_0$ is the fluctuating density, and $T_{ij} = \rho u_i u_j + \delta_{ij}(p - c_0^2 \rho') - \sigma_{ij}$ is Lighthill's stress tensor. The corresponding inhomogeneous Helmholtz equation is, with $p' = c_0^2 \rho'$,

$$\left(\frac{\partial^2}{\partial y_i^2} + k_0^2 \right) p'(y,\omega) = -\phi(y,\omega), \tag{2}$$

where $k_0 = \omega/c$ is the wave number, and $\phi(y,\omega) = \partial^2 T_{ij}(y,\omega)/\partial y_i \partial y_j$ is the Fourier transform of the sound source term. Variables with ω as a parameter are the Fourier transform of their time domain counterparts.

To solve Eq. (2), the normal velocity is assumed to vanish on the walls. From the momentum equation, neglecting viscous drag forces on the walls, this boundary condition corresponds to the vanishing of the normal pressure gradient. Equation (2) may be solved by means of a Green's function defined as the solution of

$$\left(\frac{\partial^2}{\partial y_i^2} + k_0^2 \right) g(x,y;\omega) = -\delta(x-y) \tag{3}$$

with the boundary condition $\partial g/\partial n|_{\text{wall}} = 0$ on the wall, as well as additional boundary conditions at the tube ends for finite length tubes. The coordinates x and y are the observer and source location, respectively. Note that the Green's function defined here accounts for both the impedance and the geometry of the boundary. Multiplying Eq. (2) by $g(x,y;\omega)$, subtracting the product of Eq. (3) and $p'(y,\omega)$, and integrating with respect to y over the whole flow region yields, after applying the divergence theorem and the boundary conditions,

$$p'(x,\omega) = \int_V g(x,y;\omega)\phi(y,\omega)dV(y). \tag{4}$$

Note that the neglection of the viscous drag on the orifice walls is probably justified. A parallel computational study of a similar flow in a rigid duct (Zhao *et al.*, 2001) showed that the viscous drag component of the dipole sound is very small compared with the pressure drag component.

In the case of an observer far from the source region, the distance between the observer and the source location is much larger than the extent of the source region. Then time retardation effects between the source and observer due to source motion can be neglected. Therefore, the Green's function can be expressed as the product of two terms as follows:

$$g(x,y;\omega) = g_x(x,\omega) g_y(y,\omega). \tag{5}$$

The far-field term, g_x, accounts for sound propagation from immediately outside the source region to the observer, therefore representing the acoustic response of the system. The near field part of the Green's function, g_y, reflects the radiation efficiency of the sound generation processes described by the source term $\phi(y,\omega)$.

The sound pressure spectral density S_{pp} and the cross-spectral density of the source term $S_{\phi\phi}$ are defined here as

$$S_{pp}(x,\omega) = \lim_{T\to\infty} \frac{2\pi}{T} \{p'(x,\omega) p'^*(x,\omega)\}, \tag{6}$$

$$S_{\phi\phi}(y_1,y_2;\omega) = \lim_{T\to\infty} \frac{2\pi}{T} \{\phi(y_1,\omega)\phi^*(y_2,\omega)\}. \tag{7}$$

From Eqs. (4) and (5), the sound pressure spectral density is related to the source cross-spectral density through

$$S_{pp}(x,\omega) = |g_x(x,\omega)|^2 \int_V \int_V S_{\phi\phi}(y_1,y_2;\omega) g_y(y_1,\omega) \times g_y(y_2,\omega) dV(y_1) dV(y_2). \tag{8}$$

According to the principle of superposition in linear acoustics, the sound pressure radiated by a distribution of sound sources is the sum of the sound pressures generated by each elemental source within the distributed source region. Radiation enhancement or cancellation occurs depending on time or phase retardation effects. These depend on the distance between the source and observer, as well as the spatial and phase distribution of the distributed sources. These interferences are described by the convolution integral in Eq. (4). For a far-field observer, the effect of varying the source-observer distance on sound waves' superposition can be separated from retarded time effects. In other words, the far-

field part of the Green's function may be moved outside the integral, as shown in Eq. (8).

For a compact source distribution, the turbulence length scale of the source is much smaller than the wavelength typical of the sound radiation so that the effect of the spatial and phase distribution of the source is insignificant and can be ignored. The parameter kD is much smaller than unity, where $k=\omega/c$ is the wave number and D is the typical length scale of the source distribution. Large scale of turbulence contributes the most to the radiated sound field, and so the length scale D is comparable to a typical tube dimension (width or height). The convolution then becomes a simple summation, instead of a summation at the retarded time. For a multipole source of order n, this integral is of the order of $(kD)^{2n}F^2$ (Howe, 1998), where F is the typical value of the source strength. The $(kD)^{2n}$ term is a measure of the so-called radiation efficiency, which is defined as the ratio of the radiated sound power to the mechanical power driving the sound generation phenomena. The radiation efficiency decreases with ascending multipole order.

The far-field sound spectra in Eq. (8) can be reorganized as the product of a source spectral function, a radiation efficiency function, and a far-field term. As discussed before, the far-field term describes the effects of sound propagation phenomena from the source region to the observer. It may be construed as the frequency response function between the Thevenin equivalent source strength (see, for example, Flanagan, 1965) and the sound pressure radiated at the observer position. The radiation efficiency, on the other hand, is a measure of sound production. It is straightforward, then, to express the nondimensional sound pressure spectrum, $E=S_{pp}(x,f)/(\frac{1}{2}\rho U^2)^2 d/U$, as

$$E=F(d,D,f,U,\upsilon)M(kD)G(x,D,f,c,R). \tag{9}$$

The sound pressure spectral density, S_{pp}, has the dimensions of pressure^2/frequency and is made nondimensional using the scaling factor $(\frac{1}{2}\rho U^2)^2 d/U$. The function F is the source spectral distribution function, the function G is the system transfer function or response function term, and M is a radiation efficiency function which depends on (kD).

The source function F depends on flow parameters such as the centerline velocity of the jet, U, the tube cross-sectional dimension, D, the orifice diameter, d, the fluid viscosity υ, as well as frequency, f. The system frequency response, G, depends on the observer position, x, the tube dimension, D, the speed of sound, c, and the acoustic boundary conditions, R. Defining the dimensionless groupings $\mathrm{Re}=Ud/\nu$, $\mathrm{St}=fD/U$, $\mathrm{He}=fD/c$, the above equation may be rewritten as

$$E=F\left(\mathrm{St},\mathrm{Re},\frac{d}{D}\right)M(\mathrm{He})G\left(\mathrm{He},\frac{x}{D},R\right). \tag{10}$$

The M function and the G function can be combined to form a new function G_1:

$$G_1\left(\mathrm{He},\frac{x}{D},R\right)=M(\mathrm{He})\cdot G\left(\mathrm{He},\frac{x}{D},R\right). \tag{11}$$

Substituting back into Eq. (10) yields the simple nondimensional source-filter form:

$$E=F\left(\mathrm{St},\mathrm{Re},\frac{d}{D}\right)G_1\left(\mathrm{He},\frac{x}{D},R\right). \tag{12}$$

III. EXPERIMENTAL APPARATUS

The experimental apparatus was a slightly modified version of an apparatus used in a parallel voice production study (Zhang *et al.*, 2002). A circular orifice plate was installed between two rectangular tubes, as illustrated in Fig. 1(a). Each tube had a square cross-section, 2.25×2.25 cm, and was 22.5 cm long. The geometry of the orifice is shown in Fig. 2. The circular orifices featured a converging nozzle at the entrance and a sharp trailing edge. Two orifice diameters, 0.381 and 0.635 cm, were used. Acoustically treated terminations were connected to each test section to reduce possible sound reflections. The terminations consisted of two corrugated, rubber hoses with different inner diameters. The smaller hose, with a 2.54-cm i.d., was connected directly to the rectangular tube at one end. At the other end, the smaller hose was perforated and wrapped using Fiberglass over a 1-m-long section. This end section was then inserted into a larger rubber hose with a 5.08-cm i.d. The junction was sealed to avoid any flow leakage.

The flow supply allowed for the use of different gas mixtures [Fig. 1(b)] in the experiments. Use of different working fluids makes it possible to study the possible dependence of sound generation on fluid properties. A helium–CO_2 mixture was used because it allowed the speed of sound and viscosity to be varied over a large range. The two gases were fed into a mixing tank, and then into the test sections. A mass flow controller (MKS Type 1559A) on each branch of the inlet gases was used to maintain the desired composition as well as the total flow rate. A third port allowed pressurized air to be delivered to the mixing tank through a switch valve. This made it possible to use either helium–CO_2 mixtures or air. A thermocouple was installed in the outlet to measure the mixture temperature, from which the density of the mixture was calculated. To reduce possible noise from the flow supply, the mixing tank walls were lined with a 2.54-cm-thick layer of Fiberglass.

Radiated sound pressure spectra were measured using two microphones (6.35-mm diameter, B&K 4938) located 14 cm upstream and downstream from the orifice plate. The volumetric flow rate through the orifice was measured using a precision mass-flow meter (Baratron type 558A) at the inlet of the test section. The time-averaged pressured gradient across the orifice was measured using a pressure gauge (Baratron type 220C). The output signals from the microphones were acquired at a sampling rate of 32 768 Hz using an HP356XA data acquisition system.

IV. EXPERIMENTAL RESULTS

The sound pressure spectra radiated downstream were measured for different flow rates, and two different orifice diameters. The jet centerline velocity was calculated using Bernoulli's equation:

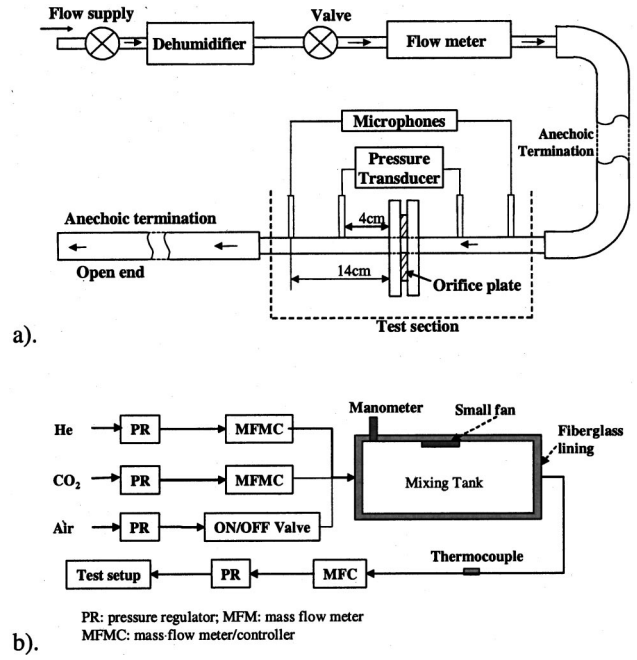


FIG. 1. (a) Schematic of the experimental apparatus. (b) Flow supply system.

$$U=\sqrt{2\Delta p/\rho_0}, \tag{13}$$

where Δp is the pressure drop across the orifice and ρ_0 is the medium density. Three different helium–CO_2 mixtures, with helium mole ratios of 0, 0.33, and 0.57, were used, as well as air.

Figure 3 shows radiated sound spectral densities, S_{pp}, measured downstream for air, several flow rates, and for the 0.381-cm-diam orifice. The spectral levels, as expected, increase as the flow rate, or jet velocity, increases. The high

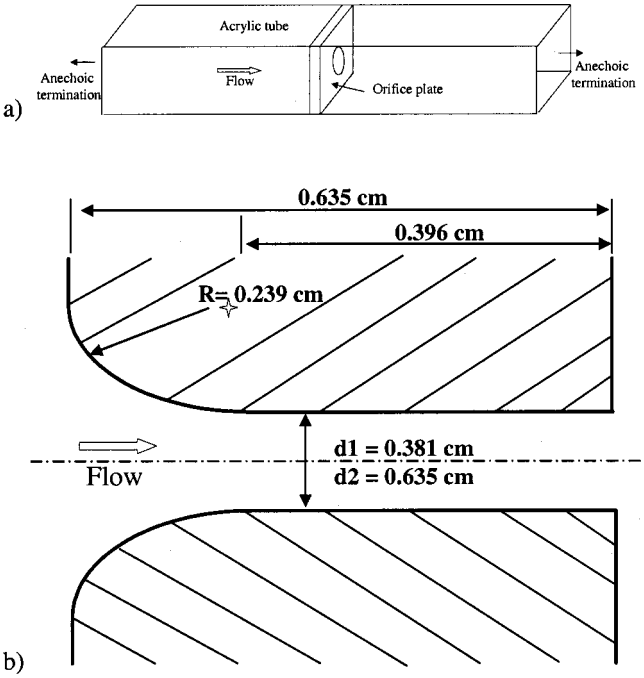


FIG. 2. (a) Test section. (b) Cross section of the circular orifices. Dimensions in cm.

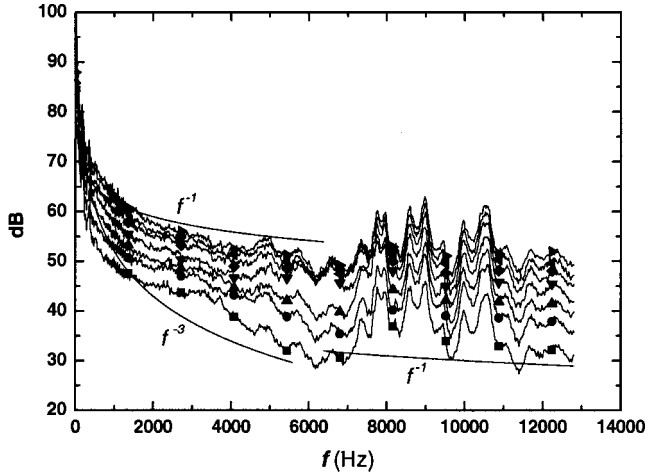


FIG. 3. Downstream sound pressure spectra for different orifice static pressure drops, $d=0.381$ cm, air is the medium. ■: $\Delta p=6$ cm H_2O; ●: $\Delta p=8$ cm H_2O; ▲: $\Delta p=10$ cm H_2O; ▼: $\Delta p=12$ cm H_2O; ♦: $\Delta p=14$ cm H_2O; ◀: $\Delta p=16$ cm H_2O; ▶: $\Delta p=18$ cm H_2O.

sound pressure levels at very low frequencies are believed to be due to flow supply noise. The effectiveness of the anechoic terminations is reduced at low frequencies, causing resonance (below 500 Hz).

Efforts were first made to establish power laws based on the jet velocity to collapse the sound spectral data. Different scaling laws work best at low frequencies and at high frequencies. At frequencies lower than the high mode cut-on frequency (approximately 7000 Hz for the rectangular tube), the spectra collapse when scaled using U^6, while at frequencies higher than the cutoff frequency, the spectra collapse using a U^8 law. Similar scaling laws for jet velocity were obtained for other mixtures, although the optimal value of the exponents varied slightly for different flow conditions.

The sound spectral level was observed to increase with the orifice diameter, as shown in Fig. 4. Different power laws for the dependence on jet diameter were also found for low frequencies and high frequencies. The low frequency components collapse very well when scaled with d^2. The high frequency components collapse better when scaled with d,

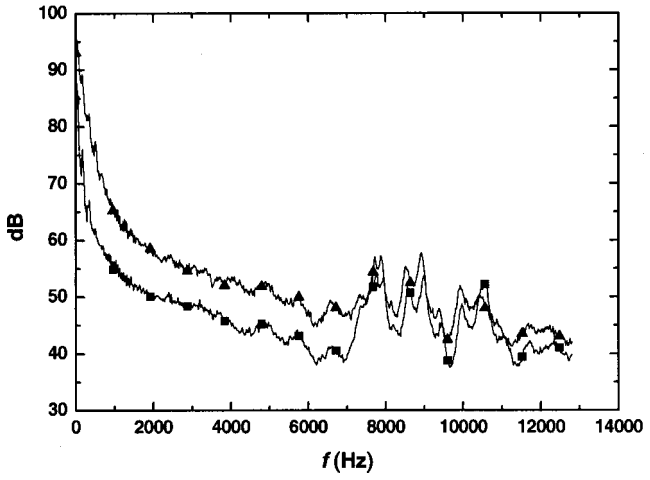


FIG. 4. Downstream sound pressure spectra for different orifice diameters, $\Delta p=10$ cm H_2O, air is the medium. ■: $d=0.381$ cm diameter; ▲: $d=0.635$ cm diameter.

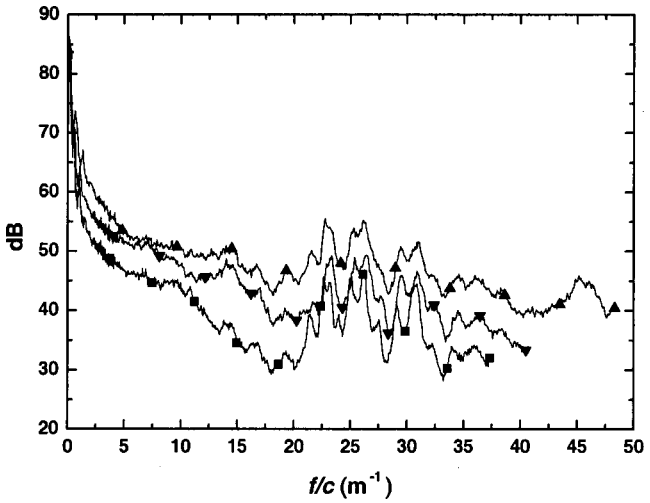


FIG. 5. Downstream sound pressure spectra for different media, and speed of sound, $d=0.381$ cm, $U=31.3$ m/s. ■: $c=343$ m/s; ▼: $c=316$ m/s; ▲: $c=268$ m/s.

with the optimal value of the exponent lying between 1 and 2. The diameter exponent also varied slightly with the flow conditions.

Figure 5 shows radiated sound spectra for different gas mixtures. Generally, the sound level is observed to increase as the speed of sound is decreased. The sound spectra were found to collapse best when scaled using a $c^{-2}\rho^2$ power law. As for the previous cases, such power scaling law does not work well at all frequencies.

To find the optimal value of the power law exponents (assuming the sound power is proportional to $U^m d^n$) as a function of frequency, the spectra were integrated from a starting frequency f_2 up to the upper limit frequency of 12.8 kHz. Figure 6 shows the exponents m, n as a function of the integration starting frequency f_s. For $f_s<300$ Hz, m decreases towards 3 due to the presence of strong flow supply noise at very low frequency. Above 8 kHz, m tends towards a value of 8. Between 300 and 8000 Hz, m is around 7. These values are consistent with the observations of Davies and Ffowcs Williams (1968). The diameter exponent, n, decreases with frequency until it approaches a value near unity above 5 kHz. It is clear that a simple power law formulation

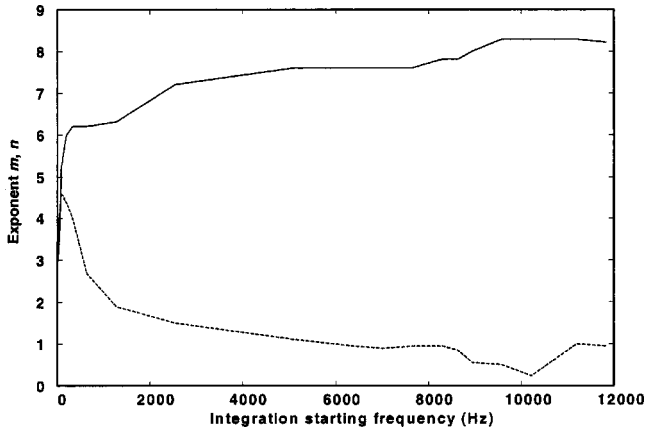


FIG. 6. Frequency dependence of the exponents in the sound power laws $P\sim U^m d^n$. ———: m; - - -: n.

is not sufficient to accurately predict the effects of flow velocity or orifice diameter. The spectral distribution of the sound spectra must be considered in the analysis.

V. SPECTRAL DECOMPOSITION PROCEDURE

The spectral decomposition method assumes a linear source-filter system, with the sound generated by the turbulent jet through the orifice acting as the source, and the filter related to the acoustic response of the system. The output signal is the measured radiated sound pressure. It is assumed that there is little or no feedback, or back reaction of the flow to the radiated sound field.

Expressing the nondimensional spectrum using a logarithmic scale, Eq. (12) is rewritten as

$$10\log E = 10\log F\left(\mathrm{St},\mathrm{Re},\frac{d}{D}\right)+10\log G_1\left(\mathrm{He},\frac{x}{D},R\right). \tag{14}$$

For a fixed microphone position, constant acoustic boundary conditions and orifice diameter, G_1 is a function of only the He number. Removing the d/D, x/D, and R dependence, Eq. (14) becomes

$$10\log F(\mathrm{St},\mathrm{Re}) = 10\log E - 10\log G_1(\mathrm{He}). \tag{15}$$

Nondimensional sound spectra may be plotted against the St number and the Re number for values of f and c forming a constant He number. Since no coupling is assumed between the F and G_1 functions, surfaces of constant He number should be parallel to each other. They would have the same functional shape and only differ by a nearly constant level difference over the St-Re plane. This level difference, according to Eq. (15), is the value of $10\log G_1$. The spectral surfaces He=const can be interpolated from the available data points. By calculating the level difference between the spectral surfaces He=const, using a least squares method, the relative amplitude of G_1 values can be determined. This process can be tentatively formulated:

$$10\log G_1(\mathrm{He}_2)-10\log G_1(\mathrm{He}_1) = \overline{10\log[F\times G_1](\mathrm{He}_2)-10\log[F\times G_1](\mathrm{He}_2)}, \tag{16}$$

where the overbar means an ensemble average of all data available for the discrete He_1 and He_2 values. Arbitrarily imposing $G_1(\mathrm{He}_{\min})=0$, where $\mathrm{He}_{\min}$ is minimum value of the He number in the experimental data, the G_1 function can then be uniquely constructed by repeating the same procedure over the entire range of He. This allows $10\log F^k$ functions to be obtained for each combination of the Strouhal number and the Reynolds number using Eq. (15). Finally, the function $10\log F$ is obtained by ensemble-averaging all $10\log F^k$ functions.

Two test cases were used to verify the decomposition method, as described below.

A. Accuracy of the spectral decomposition method

Explicit analytical expressions for the F and G_1 functions were used to generate calculated ''data'' in a simulated ''experiment,'' this to study the accuracy of the method. The F and G_1 function were arbitrarily chosen as

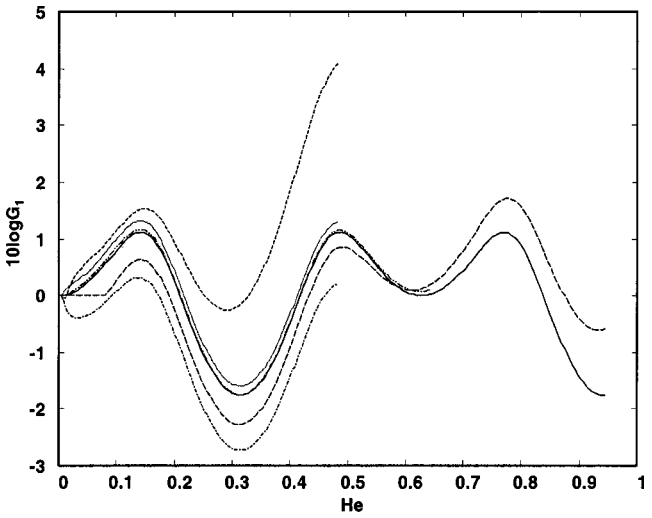


FIG. 7. Comparison between the correct and the estimated values of the G function, and constructed values obtained using spectral decomposition. The accuracy of the decomposition increases as the input spectra overlap in the St-Re plane is increased. ———: exact solution; – –: case Ia; - - -: case Ib; —·—·—·—: case II; –··–: case III; · · · ·: case IV.

$$F=\mathrm{Re}^{0.2}(\mathrm{St}^2+10), \quad G_1=\left|\frac{e^{5j\mathrm{He}}+0.2e^{-5j\mathrm{He}}}{e^{10j\mathrm{He}}+0.2e^{-10j\mathrm{He}}}\right|. \tag{17}$$

The nondimensional spectra E were then constructed based on Eq. (14). To simulate the experiments, the sound spectra S_{pp} were constructed for six different values of velocity and four values of speed of sound (or four different media), with a frequency resolution bandwidth of 16 Hz.

To test the accuracy of the decomposition method, four different spectral data sets were used as input to the decomposition program. Spectra constructed from 1, 2, 3, and 4 different values of speed of sound were used in cases I, II, III, and IV, respectively. It was found that the accuracy of the reconstructed functions depends highly on the distribution and the size of overlap region of the data sets over the St-Re plane. For a constant He number, the data set constructed for one discrete value of speed of sound maps to a hyperbolic curve in the St-Re plane. One curve in the St-Re plane is certainly not enough for accurate interpolation of the He =const surface for the 10log$E$ function over the whole St-Re plane, and errors in the interpolation process could lead to errors in calculation of the 10logG_1 function. Figure 7 shows the reconstructed G_1 functions for all four cases along with the exact solution, Eq. (17). Case I, using spectra constructed for two discrete values of the speed of sound (cases Ia and Ib), resulted in a biased estimate, with an accumulated error in excess of 3 dB over a reasonable range of He numbers. This bias gradually disappeared as more and more data sets were used, increasing the number of curves available for surface interpolation as well as the overlap (or redundancy) in the data set. For case III, in which spectra were constructed using three different values of speed of sound, the reconstructed G_1 function matched very well the exact solution. Case IV also yielded a very good agreement, except for a constant offset due to the arbitrary selection of the value of G_1 at $\mathrm{He}_{\min}$.

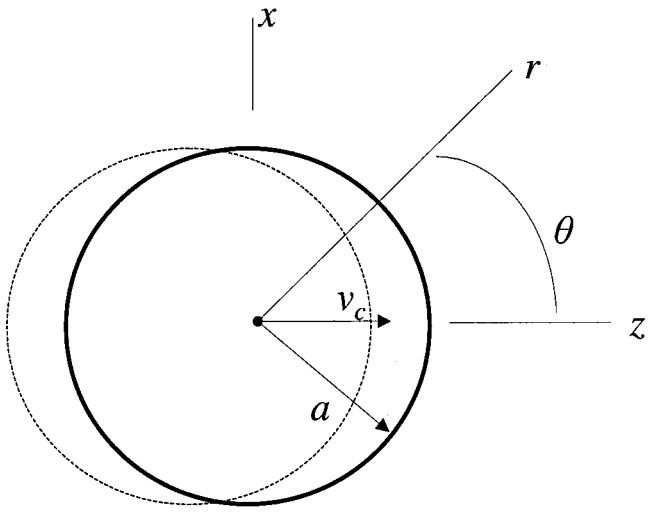


FIG. 8. Transversely oscillating rigid sphere in free space.

B. Transversely oscillating rigid sphere in free space

Next, the simple problem of sound radiation from a transversely oscillating rigid sphere in free space was used to illustrate the application of the similarity laws formulation. Consider, as illustrated in Fig. 8, the center of a rigid sphere with radius a oscillating along the z-axis about the origin with a velocity $v_c(t)=\mathrm{Real}(\hat{v}_c e^{-iwt})$, where $\hat{v}_c$ is the complex velocity. This is a classical example of a dipole sound source. The complex sound pressure at a point at a distance r from the origin along the z-axis is (Pierce, 1989)

$$\hat{p}=\frac{-i\omega\rho}{r^2}\frac{\hat{v}_c a^3 e^{-ika}}{2-(ka)^2-2ika}e^{ikr}(ikr-1). \tag{18}$$

Assuming a compact source, and when the observer is in the far field,

$$\hat{p}=(ka)^2\frac{a}{r}\frac{\rho c\hat{v}_c}{2}e^{ik(r-a)}. \tag{19}$$

In the hydrodynamic near field ($kr\ll 1, r\sim a$), the sound pressure is approximately

$$\hat{p}\approx -i(ka)\frac{\rho c\hat{v}_c}{2}e^{ik(r-a)}. \tag{20}$$

The radiation efficiency, which is the ratio of the far-field sound pressure and the sound pressure in the hydrodynamic near field, is

$$\left|\frac{\hat{p}_{\mathrm{far}}}{\hat{p}_{\mathrm{near}}}\right|^2\approx(ka)^2. \tag{21}$$

Now assume the sphere center velocity time history has a spectrum with spectral components at St numbers which are multiples of a fundamental St_0. The magnitude of each component is assumed to be proportional to the St number:

$$v_c(f)=U\sum_{n=1}^{f_n<f_s/2}\mathrm{St}\cdot\delta(f-f_n), \tag{22}$$

where $f_n=n\cdot\mathrm{St}_0U/a$, f_s is the sampling rate, and U is a constant proportionality factor playing a role similar to that

of velocity in a flow-excited sound problem. The Strouhal number may then be chosen as $\mathrm{St}=fa/U$. The complex sound pressure in the far field then becomes

$$\hat{p}(f)=U\sum_{n=1}^{f_n<f_s/2}\mathrm{St}^2\frac{a}{r}\frac{\rho c}{2}e^{ik(r-a)}\delta(f-f_n), \tag{23}$$

where the wave number $k=f/c$.

Sound power spectra can be constructed from Eq. (23) for different values of U and the speed of sound. Because the complex pressure in Eq. (23) has the dimension of ''pressure'' instead of ''pressure/frequency,'' the nondimensional spectrum is

$$E=\frac{S_{pp}(x,f)}{(\frac{1}{2}\rho U^2)^2}. \tag{24}$$

In this case study, six values of ''velocity'' U and three values of speed of sound were used to construct the sound spectra. The value of St_0 was chosen to be very small so that the spectra were broadbandlike and the spectral data set had significant overlap in the St-Re plane.

For this particular case, the Green's function is the three-dimensional free-space Green's function. For a fixed observer position in the far field of a compact source, this indicates a uniform acoustic response function at all frequencies. The radiation efficiency function M, shown in Eq. (21), is of the order of $(ka)^2$. Therefore, the G_1 function in this case, as defined in Eq. (11), is simply

$$G_1(\mathrm{He})\approx(\mathrm{He})^2. \tag{25}$$

Figure 9(a) shows a comparison between the reconstructed G_1 function and Eq. (25). The two G_1 functions are in very good agreement. The reconstructed and exact source spectral distribution functions F are shown in Fig. 9(b). As prescribed in Eq. (22), the source strength depends linearly on the St number, and thus the spectra should follow a St^2 law. A St^4 dependence is obtained instead, as shown in Fig. 9(b). The additional St^2 dependence is a consequence of the similarity laws formulation. In Eq. (24), the spectra were nondimensionalized by a factor $\frac{1}{2}\rho U^2$, which is the hydrodynamic energy in the near field. Not all of this energy is converted into acoustic power. Only the energy associated with the dilatation component of the velocity field is radiated to the far field. This conversion efficiency is St^2 for a compact source, which can be shown as follows:

$$\left(\frac{|\hat{p}_{\mathrm{near}}|}{1/2\rho U^2}\right)^2\approx\left(\frac{1/2ka\rho cU}{1/2\rho U^2}\right)^2=\left(\frac{kac}{U}\right)^2=\mathrm{St}^2. \tag{26}$$

Taking this into account then yields the observed St^4 trend of Fig. 9(b).

For a confined environment, the near-field flow-acoustic interaction is much more complicated due to the presence of solid surfaces. However, it is expected that the same formulation should also be valid for far-field measurements and compact sources.

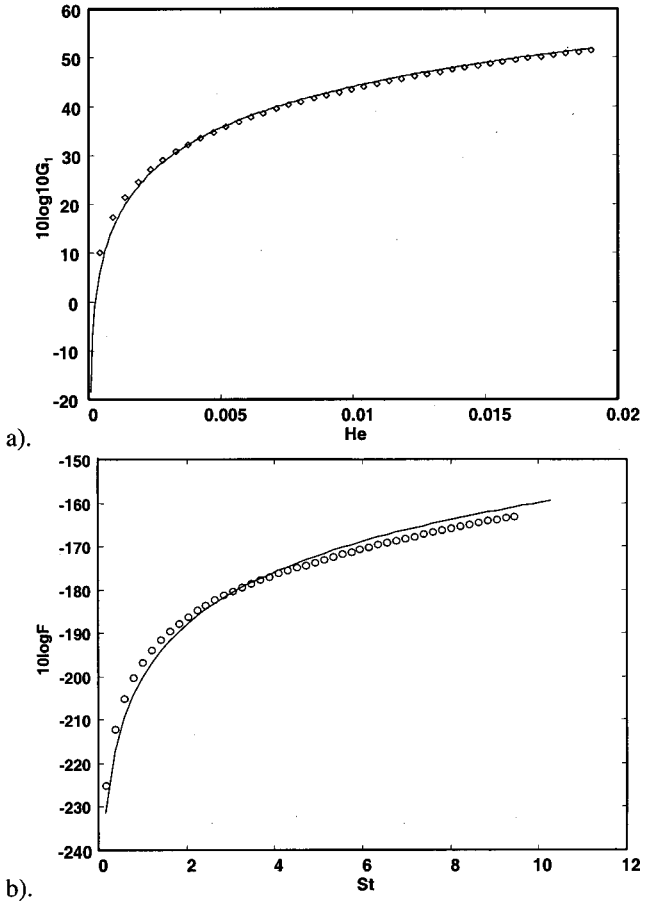


FIG. 9. (a) Comparison between the correct G_1 function, and estimated values obtained using spectral decomposition procedures. ———: 20logHe; ◇: $10\log G_1$. (b) Comparison between the correct F function, and estimated values obtained using spectral decomposition procedures. ———: 40logSt: ○: $10\log F$.

VI. RESULTS, ANALYSIS AND DISCUSSION

The sound spectra measured for varying gas mixtures, orifice diameters, and flow rates, shown in Figs. 3–5, were used to attempt the reconstruction of the source function for radiation by subsonic confined jets. Figure 10 shows nondimensional spectra $10\log E$ for two discrete values of the He

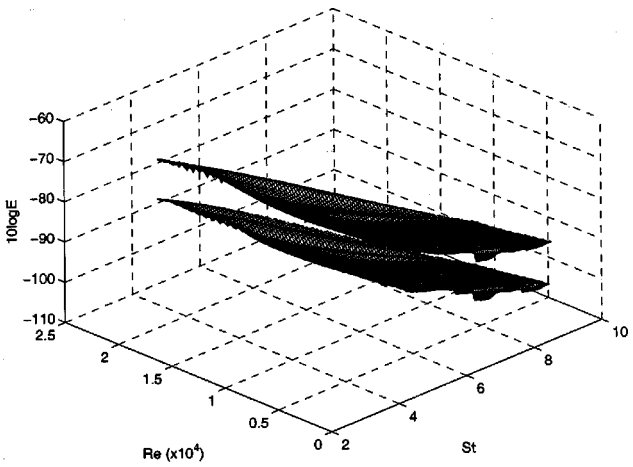


FIG. 10. The nondimensional spectra E for two consecutive He numbers. The two spectral surfaces are parallel to each other. $d=0.635$ cm.

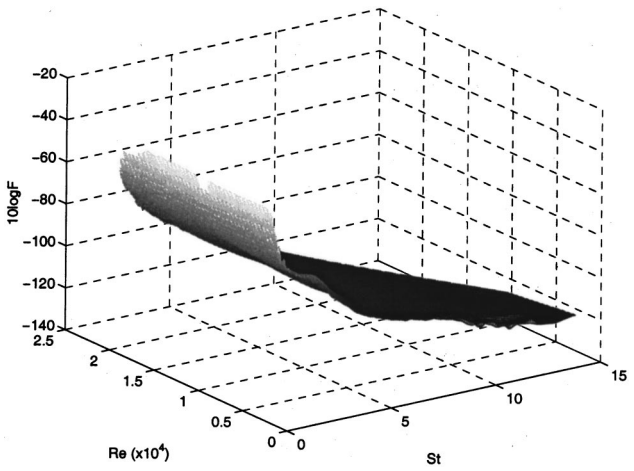


FIG. 11. Averaged source spectral distribution function F. $d=0.635$ cm.

number. For visual convenience, one of the spectral surfaces was offset by 10 dB. The two spectral surfaces are parallel to each other, which supports the validity of the similarity law formulation.

The $10\log G_1$ function was obtained by calculating the difference between neighboring spectral surfaces. The $10\log F$ functions were then obtained by subtracting the known $10\log G_1$ function from the nondimensional spectra $10\log E$. The resulting $10\log F^k$ functions collapsed very well, which is another indication of the validity of the similarity formulation. The ensemble-averaged $10\log F$ function is shown in Fig. 11. The corresponding standard deviation of $10\log F$, shown in Fig. 12, is less than 2 dB. This confirmed the validity of the product law formulation within approximately 1 dB.

Note that the sound pressure spectra were measured at one fixed location along the tube wall. At very low frequency, the microphone may have been in the acoustic near field, as the acoustic wavelength becomes large. Vortices shed from the orifice and/or turbulence may not have been dissipated, and may have caused ''pseudo-sound'' recorded by the microphone. The influence of the near field may have

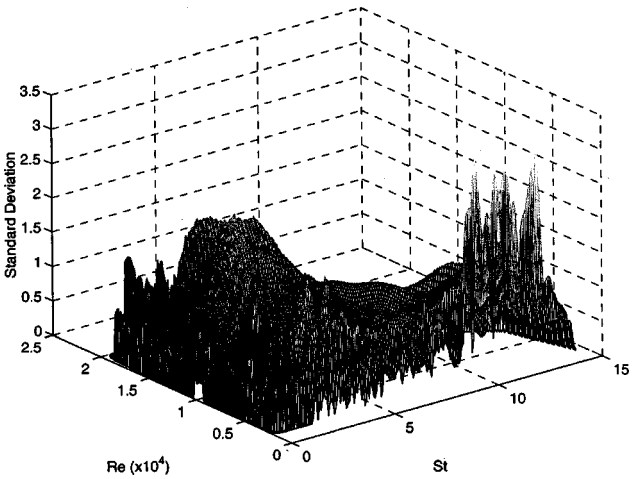


FIG. 12. Standard deviation of the estimated source functions. $d=0.635$ cm.

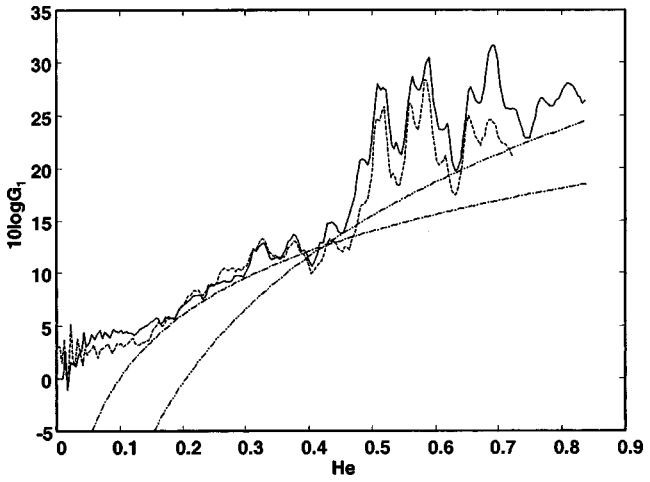


FIG. 13. The G_1 functions obtained for two orifice diameters. ———: $d=0.381$ cm; — — —: $d=0.635$ cm; —·—·—·—: 20logHe; –··–: 40logHe.

colored the radiation efficiency function (or M function) at low frequencies.

A. System response function and radiation efficiency

The acoustic frequency response functions, G_1, obtained for two different orifice diameters are shown in Fig. 13. They are nearly identical. Additional maps showing He^2 and He^4 power laws are also shown in the figure for comparison. These describe approximately the behavior of the G_1 functions at low He numbers and high He numbers, respectively.

As indicated in Eq. (11), the G_1 function is composed of the system frequency response function, G, and the radiation efficiency function, M. The acoustic response function G, as discussed before, can be calculated as the ratio of the sound pressure measured at the microphone position and that which would be measured at the orifice plate location if the tube was not there. Assume planar wave propagation in the tube, as illustrated in Fig. 14. The sound source is located on the left side of the tube, and the tube is terminated by acoustic impedance Z on the right end. With an incidence sound wave of frequency f, the complex sound pressure at the microphone location is

$$\hat{p}_{\mathrm{mic}}(l,f)=A(e^{-ikl}+R\cdot e^{ikl}), \tag{27}$$

where A is the amplitude of the incident wave, $k=2\pi f/c$ is the wave number, and R is the reflection factor which is

$$R=\frac{Z-\rho_0 c_0}{Z+\rho_0 c_0}. \tag{28}$$

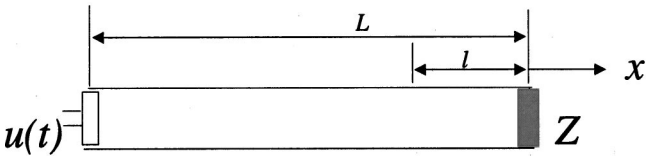


FIG. 14. Schematic of a tube with a velocity type source at one end and terminated with an acoustic impedance.

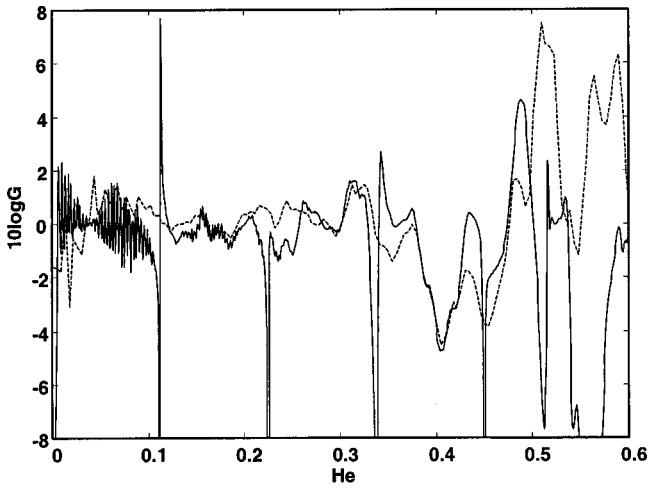


FIG. 15. Comparison between the system acoustic transfer function and the G function. ———: system acoustic transfer function; — — —: 10log$G$.

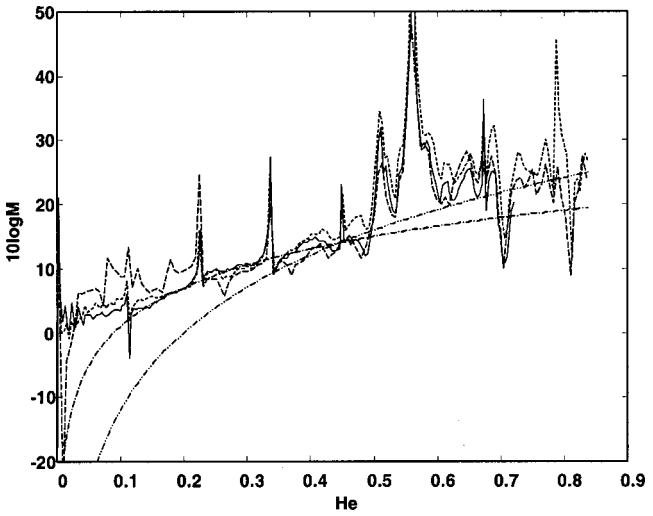


FIG. 16. The radiation efficiency functions obtained for different orifice diameters and downstream tube terminations. ———: $d=0.635$ cm, with anechoic termination; - - -: $d=0.381$ cm, with anechoic termination; — —; $d=0.381$ cm, with finite length tube downstream: —·—·—·—: 20logHe; –···–: 40logHe.

Assuming a velocity type of sound source, the complex sound pressure that would be measured at the source location if the tube was not there is

$$\hat{p}_{\text{source}}(L,f)=A(e^{-ikL}-R\cdot e^{ikL}). \tag{29}$$

The G function is then the ratio of the two complex pressures:

$$G_{\text{exp}}=\left|\frac{A(e^{-ikl}+R\cdot e^{ikl})}{A(e^{-ikL}-R\cdot e^{ikL})}\right|^2=\left|\frac{e^{-ikl}+R\cdot e^{ikl}}{e^{-ikL}-R\cdot e^{ikL}}\right|^2. \tag{30}$$

The notation G_{exp} is used here instead of G to indicate that this response function is obtained experimentally.

The termination impedance was measured using the two-microphone method (Seybert, 1977). The G_{exp} function was then calculated using Eq. (30) and shown in Fig. 15. It is possible to compare functions G_{exp} and G, the latter a component of G_1. The radiation efficiency function was extracted from the G_1 function by matching G with G_{exp}. The result is shown in Fig. 15 together with the G_{exp} function for comparison. There is generally a good agreement, except at frequencies where a pressure node coincides with the microphone location, because of flow noise. Flow noise tends to limit the dynamic range of the G function. The sharp peaks in G_{exp} are most probably due to numerical errors in the calculation of the reflection factor. Discrepancies at high frequencies are likely caused by the excitation of higher order acoustic modes.

To a good approximation, the M function can be determined by regression, as shown in Fig. 13. The mismatch at very low frequencies is, as discussed above, due to the sound pressure measurement in the acoustic near field, although there could be other sources of errors. The M function can also be obtained by subtracting the G_{exp} function from the G_1 function. This is plotted in Fig. 16. The M function approaches He^2 and He^4 for low He numbers and high He numbers, respectively. These trends are in good agreement with the theoretical radiation efficiency of quadrupole sources in tubes. The value of the He number corresponding to the cut-on frequency of higher modes in the test section was about 0.45. At greater He numbers, again, higher order modes are excited and the calculated G_{exp} function is not accurate. The agreement between the He^4 regression and the reconstructed M function is not as good.

There are two major sound sources in sound radiation from confined stationary jets: quadrupole sources due to the turbulence region downstream the orifice and dipole sources due to the unsteady force applied to the fluid by the downstream wall of the orifice (Zhang *et al.*, 2002). The quadrupolelike radiation efficiency obtained in this study indicates that the dipole sources contributions may have been negligible in the present case. The flow within the orifice may have been laminar or very stable, so that the fluctuating axial force induced by the interaction between the orifice wall and the flow was very small.

B. Source spectral distribution function

Figure 17 shows the source spectral function (also shown in Fig. 11) versus Reynolds number for constant

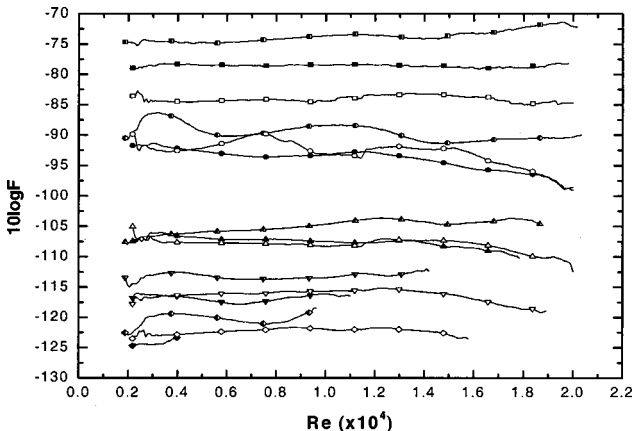


FIG. 17. The source distribution function F as a function of Reynolds number for constant Strouhal number. ■: St=1; ●: St=3; ▲: St=6; ▼: St=9; ♦: St=12. Solid symbols: $d=0.381$ cm, with anechoic termination; open symbols: $d=0.381$ cm, finite-length tube; half open symbols: $d=0.635$ cm, with anechoic termination.

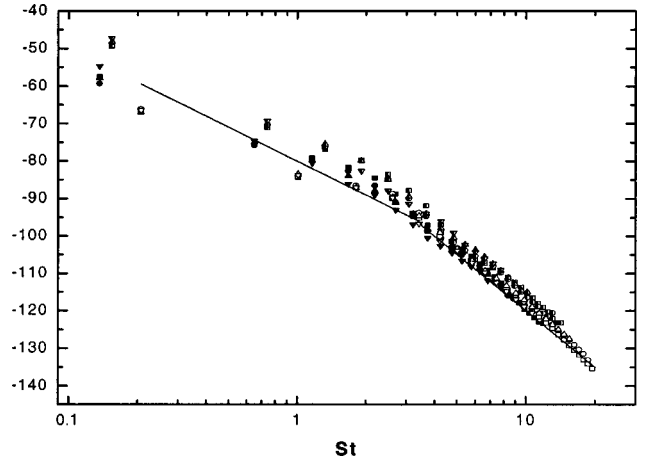


FIG. 18. The source distribution function F as a function of Strouhal number for constant Reynolds numbers. ■: $\mathrm{Re}=0.4\times10^4$; ●: $\mathrm{Re}=0.8\times10^4$; ▲: $\mathrm{Re}=1.2\times10^4R$; ▼: $\mathrm{Re}=1.6\times10^4$. Solid symbols: $d=0.381$ cm, with anechoic termination; open symbols: $d=0.381$ cm, finite-length tube; half open symbols: $d=0.635$ cm, with anechoic termination. ———: two-segment curve.

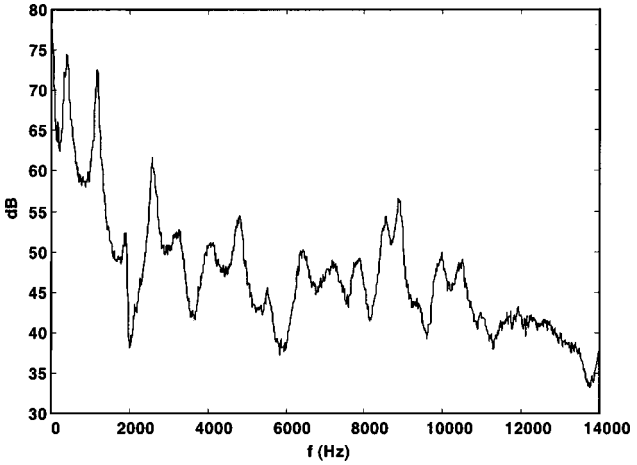


FIG. 19. Downstream sound pressure spectrum for the case with finite length tube downstream of the orifice plate. $d=0.381$ cm, $\Delta p=12$ H_2O, and air.

Strouhal number. The Reynolds number influence was nearly negligible over the range considered (2000 to 20 000), though a very progressively slow increase in the source amplitude with the Reynolds number was observed. The F function obtained for the larger orifice diameter (0.635 cm) was approximately 2 or 3 dB higher in level than the F function obtained for the smaller orifice diameter (0.318 cm). This difference decreased slightly as the St number increased, indicating a dependence on d at low values of St, which is consistent with the d^2 low-frequency scaling law discussed in Sec. IV.

Note that the tube dimensions remained unchanged, while orifices with different diameters were used. Therefore, the two cases studies are not strictly geometrically similar, since the area ratio between the orifice and tube cross section was different. One of the main consequences is that the recirculation regions in these two cases are different. This could contribute to the differences between the two cases observed in Fig. 17.

Figure 18 shows the source spectral function versus St number for constant values of the Reynolds number for both orifice diameters. As mentioned already, the influence of Re number is small, and the source function levels for the larger orifice diameter are greater than for the smaller orifice diameter. A regression, shown in the figure, suggests a St^{-3} dependence of the source spectral function on the St number for $\mathrm{St}<0.9$, and a St^{-5} dependence for $\mathrm{St}>2.5$. Considering the $(\rho U^2)^2 d/U$ factor, these correspond to U^6 and U^8 dependences on velocity for the sound pressure spectra, respectively. The source function transitioned from the St^{-3} dependence to a St^{-5} dependence over Strouhal number values between 1 and 2.5.

The frequency dependence of the sound pressure spectra can be deduced from the basic trends of the source and radiation functions. At low frequencies, the M function exhibits a f^2 dependence. If the jet velocity is very small, the F function exhibits a f^{-5} dependence, the Strouhal number being large. The sound pressure spectra therefore follows a f^{-3} dependence, as shown in Fig. 3. As the jet velocity increases, the Strouhal number decreases. When the Strouhal number is small enough the F function then follows a f^{-3} law, and the sound pressure spectra a f^{-1} law. At higher frequencies, the radiation efficiency becomes approximately He^4. Accordingly, the sound pressure spectra exhibit a f^{-1} dependence.

The source function was compared with the nondimensional spectra obtained by Nelson *et al.* (1981). In their study, different formulations were used for the planar wave frequency range and the multimode frequency range. Over the planar wave range, the sound power spectra were scaled by $(1/\rho c)\rho^2U^4$, while over the multimode range the spectra were scaled by $(1/\rho c)\rho^2U^4\mathrm{He}^2$. These are similar to the $\rho^2U^4\mathrm{He}^2$ and $\rho^2U^4\mathrm{He}^4$ scaling laws for sound pressure spectral density in this study, except they considered a dipole source instead of a quadrupole source. They therefore obtained a nondimensional $\frac{1}{3}$-oct spectrum which showed a 20 dB decay per decade of St number, or a St^{-2} dependence, for St numbers above 0.2. This corresponds to a St^{-3} dependence of the spectral density. This result is in contrast with the St^{-5} dependence of the F function for a quadrupole source (for St numbers above 2.5) observed in the present study.

It should be noted that the source spectral distribution function, as discussed before, describes the frequency distribution of the propagating component of the turbulence energy. The energy in the turbulence associated with zero dilatation velocity can only produce near-field pressure fluctuations, and cannot propagate as sound in the far field.

C. Sound radiation in a finite length tube

The experiment was repeated with the downstream anechoic termination removed. The orifice had a 0.381-cm diameter. The jets discharged directly into open space from the end of the downstream rectangular tube (22.5 cm in length). This significantly changed the acoustic loading of the system, as well as the system acoustic response function.

Sound pressure spectra were measured downstream of the orifice, for different jet velocities and speed of sound. Figure 19 shows one typical sound spectrum, for $\Delta p=12$ cm

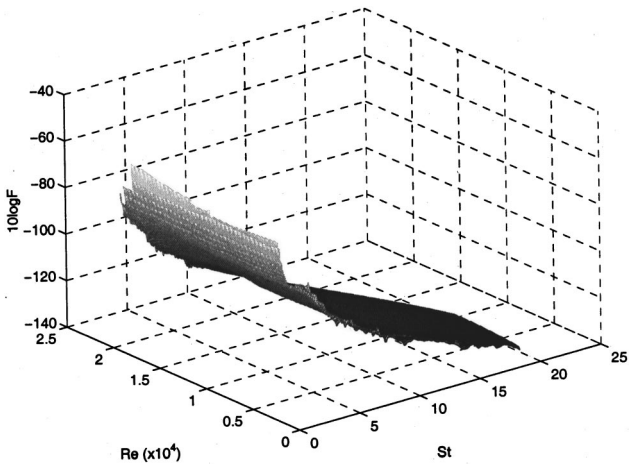


FIG. 20. Source spectral distribution for the case of a finite length tube downstream of the orifice plate.

H_2O and air as the medium. The high frequency part of the spectrum was similar to that obtained for the case with anechoic termination. However, the low frequency part was quite different. The peaks and resonance in the spectrum suggest a different system acoustic response function.

The spectral decomposition method was then used to separate the three different functions. The source distribution function F, shown in Fig. 20, is similar to that for the case with anechoic termination. The fluctuations at low Reynolds number are believed to be due to numerical errors. Figure 21 shows the function $10\log G_1$. Also shown in the figure are two curves He^2 for low He numbers ($\mathrm{He}<0.45$) and He^4 for high He numbers ($\mathrm{He}>0.45$). The good match of these two curves with the shape of the G_1 function suggests a radiation efficiency of He^2 and He^4 at low He numbers and high He numbers, respectively.

The G_1 function frequency distribution was similar to that of the measured sound pressure spectrum. For comparison, the system acoustic response function, G_{exp}, was calculated using the same method as that for the case with anechoic termination. The G function, obtained by removing the He^2 trend from the G_1 function, was compared to the G_{exp} function, which is shown in Fig. 22. A very good agreement was obtained for the He number lower than the critical value corresponding to the cut-on frequency of the acoustic higher modes. The discrepancies at high He numbers are expected since the plane wave assumption is no longer valid. This good agreement at low He numbers also indicates a radiation efficiency function M in the form of He^2.

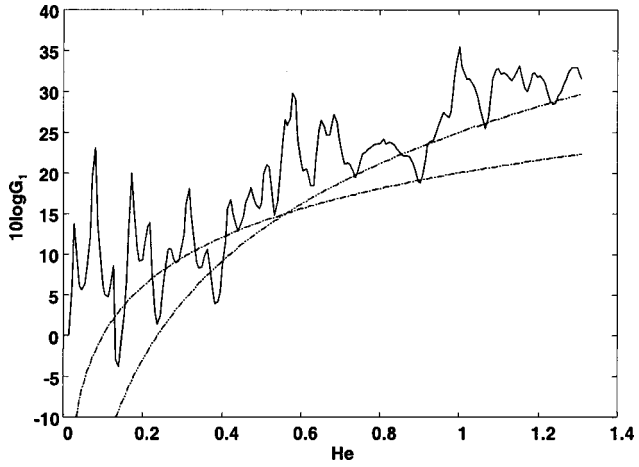


FIG. 21. G_1 function obtained for the case of a finite-length tube. ———: $10\log G_1$, $d=0.381$ cm; —·—·—·—: 20logHe; –··–: 40logHe.

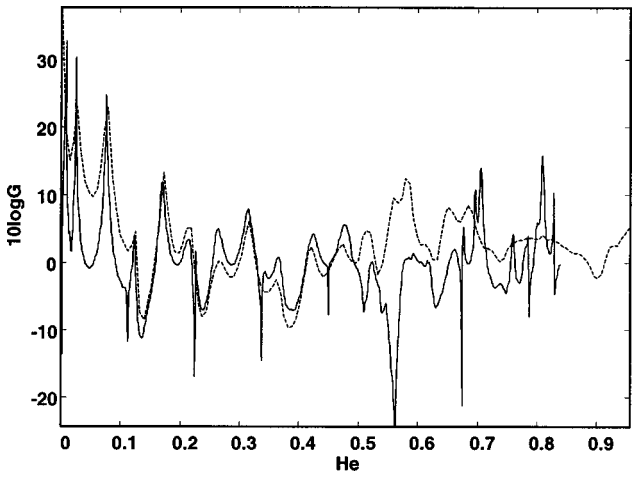


FIG. 22. Comparison between the G function obtained by removing the He^2 trend from the G_1 function, and the G_{exp} function obtained experimentally. The orifice plate was connected to a finite length tube downstream. ———: system acoustic transfer function; — — —: $10\log G$.

The M function was also obtained by subtracting the G_{exp} function from the G_1 function, as shown in Fig. 16. The He^2 curve matches the M function well at low He numbers. Since the G_{exp} function is not accurate at high He number, the agreement between the He^4 curve and the M function is poor, as expected.

The variation of the F function with the Re number for constant St numbers and St number for constant Re number are respectively shown in Figs. 17 and 18 in comparison with the F function obtained for the same orifice diameter with anechoic termination. These two F functions matched very well except at very low St number, over which the F function for finite length tube had a lower value than that for the case with anechoic termination. This difference may be due to the numerical error, i.e., there is not enough overlapping at very low St numbers. The two F functions showed a similar dependence on the St number and Re number. The effects of possible interaction between the flow field and acoustic loading were thus negligible.

D. Sound power

Sound power can be determined by integrating Eq. (10) over the entire frequency range and dividing by the characteristic impedance, ρc. At low frequencies where only plane waves propagate, $M\sim\mathrm{He}^2$. With $\mathrm{He}=Ma\cdot\mathrm{St}$, this He^2 term can be converted into a Ma^2 term. Ignoring the Re dependence, the integration then leads to a sound power of the order $\rho U^3 Ma^3$. For the high frequency component, $M\sim\mathrm{He}^4$. Following a similar approach, the sound power is of the order $\rho U^3 Ma^5$. These observations are consistent with the derivation by Davies *et al.* (1968).

VII. CONCLUSIONS

Experiments were performed to study the characteristics of sound generated by confined stationary jets through circular orifices. Nondimensional sound spectra were formulated as the product of a source spectral distribution function, a radiation efficiency function, and a system response function which describes the acoustic effects present in the tube. The source function is a function of both St and Re numbers, and the other two are functions of He number only. A spectral decomposition method based on a source-filter model was used to separate these three functions. Experiments were conducted for two downstream tube configurations and two orifice-tube dimension ratios. The results show that the system response function agrees well with the acoustic transfer function between the sound pressure spectra in the source and observer locations in the tube. The tube configuration has little effect on the source function and radiation efficiency function, which confirms the negligible interaction between the flow and acoustic loadings and the source-filter model. For a straight downstream tube configuration in this study, the radiation efficiency is similar to that of sound generation by a quadrupole source in tubes. This indicates that the dipole source contribution, which is due to the fluctuating axial force applied by the orifice wall, is negligible, and that the flow coming out of the straight orifice is very stable. This radiation efficiency also leads to a $\rho U^3 Ma^3$ and $\rho U^3 Ma^5$ scaling for the sound power at low and high frequencies, respectively. The source function obtained for confined jets having two different orifice diameters matched very well, within 2–3 dB difference. Analysis of the source spectral distribution function showed that Reynolds number effects were nearly negligible over the range 2000 to 20 000, except for a very slow increase of the source strength with the Reynolds number. The source distribution function was found to decrease with the Strouhal number as approximately St^{-3}, at very low Strouhal number and St^{-5} at high Strouhal number.

ACKNOWLEDGMENTS

This study was supported by Grant No. 5 RO1 DC03577-02 from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health.

Bartenwerfer, M., and Agnon, R. (**1976**). ''On the influence of viscosity of the working fluid on the sound production of centrifugal fans,'' DFVLR Report, DLR-FB, 76-30.

Bartenwerfer, M., Agnon, R., and Gikadi, T. (**1976**). ''An experimental investigation on noise production and noise propagation in centrifugal fans,'' DFVLR Report, DLR-FB, 76-14.

Blake, W. K. (**1986**). *Mechanics of Flow-induced Sound and Vibration* (Academic, New York).

Davies, H. G., and Ffowcs Williams, J. E. (**1968**). ''Aerodynamic sound generation in pipe,'' J. Fluid Mech. **32**(4), 765–778.

Flanagan, J. L. (**1965**). *Speech Analysis, Synthesis and Perception* (Academic, New York).

Harper, V. H. (**2000**). ''Respiratory tract acoustical modeling and measurements,'' Ph.D. thesis, Purdue University, West Lafayette, IN.

Howe, M. S. (**1998**). *Acoustics of Fluid-structure Interactions* (Cambridge U.P., Cambridge).

Hubbard, H. H. (**1994**). *Aeroacoustics of Flight Vehicles: Theory and Practice. Vol. 1: Noise Sources* (Acoustical Society of America, New York).

Landau, L. D., and Lifshitz, E. M. (**1997**). *Fluid Mechanics* (Butterworth Heinemann, Boston).

Lighthill, M. J. (**1952**). ''On sound generated aerodynamically—I. General theory,'' Proc. R. Soc. London, Ser. A **211**, 564–587.

Meyer-Eppler, W. (**1953**). ''Zum Erzeugungsmechanismus der Geräuschlaute,'' Z. Phonet. all. Sprachwissensch. **7**, 196–212.

Mongeau, L., Thompson, D. E., and McLaughlin, D. K. (**1995**). ''A method for characterizing aerodynamic sound sources in turbomachines,'' J. Sound Vib. **181**(3), 369–389.

Neise, W. (**1975**). ''Application of similarity laws to the blade passage sound of centrifugal fans,'' J. Sound Vib. **43**(1), 61–75.

Nelson, P. A., and Morfey, C. L. (**1981**). ''Aerodynamic sound production in low speed flow ducts,'' J. Sound Vib. **79**(2), 263–289.

Pierce, A. D. (**1989**). *Acoustics: An Introduction to its Physical Principles and Applications* (Acoustical Society of America, New York).

Seybert, A. F., and Ross, D. F. (**1977**). ''Experimental determination of acoustic properties using a two-microphone random-excitation technique,'' J. Acoust. Soc. Am. **61**, 1362–1370.

Shadle, C. H. (**1986**). ''The acoustics of Fricative Consonants,'' MIT, Cambridge, MA, RLE Technical Report 506, March 1986.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (The MIT Press, Cambridge, Massachusetts).

Weidemann, J. (**1971**). ''Analysis of the relations between acoustic and aerodynamic parameters for a series of dimensionally similar centrifugal fan rotors,'' NASA Technical Translation TT F-13, 798.

Zhang, Z., Mongeau, L., and Frankel, S. H. (**2002**). ''Experimental verification of the quasi-steady assumption for aerodynamic sound generation by pulsating jets in tubes,'' J. Acoust. Soc. Am., in press.

Zhao, Z., Frankel, S. H., and Mongeau, L. (**2001**). ''Numerical Simulations of Sound from Confined Pulsating axisymmetric Jets,'' AIAA J. **39**(10), 1868–1874.

Modulation frequency and modulation level owing to vocal microtremor

Jean Schoentgen[a)]
Laboratory of Experimental Phonetics, Université Libre de Bruxelles, 50, Av. F.-D. Roosevelt, B-1050 Brussels, Belgium

(Received 12 October 2001; revised 12 April 2002; accepted 21 May 2002)

Vocal microtremor designates a normal slow modulation of the vocal cycle lengths of speakers who do not suffer from pathological tremor of the limbs and whose voices are not perceived as tremulous. Vocal microtremor is therefore distinct from pathological vocal tremor. The objective is to report data about the modulation frequency and modulation level owing to vocal microtremor. The modulation data have been obtained for vowels [a], [i], and [u] sustained by normophonic and mildly dysphonic male and female speakers. The results are the following. First, modulation frequencies and relative modulation levels do not differ significantly for male and female speakers, normophonic and mildly dysphonic speakers, as well as for vowel timbres [a], [i], and [u]. Second, the typical interquartile intervals of the modulation frequency and modulation level are equal to 2.0–4.7 Hz and 0.4%–1.3%, respectively. Third, dissimilarities between data reported by different studies are due to different cutoff frequencies below which spectral peaks are considered not to contribute to vocal microtremor. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1492820]

PACS numbers: 43.70.Gr, 43.70.Jt [AL]

I. INTRODUCTION

The objective of this article is to report data about the modulation frequency and modulation level owing to vocal microtremor. Vocal microtremor designates a normal low-frequency modulation of the vocal cycle lengths of speakers who do not suffer from pathological tremor of the limbs and whose voices are not perceived as tremulous. Vocal microtremor is therefore distinct from pathological vocal tremor. The modulation data have been obtained for vowels [a], [i], and [u] sustained by normophonic and mildly dysphonic male and female speakers.

Analyses of voiced speech sounds sustained by normophonic speakers show that the vocal cycle lengths vary stochastically. Conventionally, one distinguishes perturbations of the cycle lengths owing to vocal jitter from perturbations owing to vocal microtremor (Schoentgen, 2001). This distinction can be illustrated by the time series of the vocal cycle lengths of two normophonic male and female speakers. Figures 1(a) and (c) show the cycle length in number of samples on the vertical axis and the cycle number on the horizontal axis. One sees that the length perturbations are formed by stochastic cycle-to-cycle disturbances that coexist with slower supra-cycle oscillations. The former are designated as vocal jitter and the latter as vocal microtremor.

Generally speaking, vocal jitter and vocal microtremor are thought to be distinct phenomena, because they are observed on different frequency scales. Jitter is a cycle-to-cycle disturbance, that is, the upper bound of its rhythm is equal to half the phonatory frequency, whereas vocal tremor is characterized by frequencies in an interval of 1–15 Hz (Titze, 1995a). The phonatory frequency is the frequency of the voice source signal, which is generated by the vibrating vocal folds.

Tremor of the limbs has been defined as involuntary, rhythmic movements that are seen better distally as in the fingers or hand but may sometimes be seen in the eyelids, tongue, face, or other body parts (Fucci and Petrosino, 1984). Freund (1987) distinguishes between physiological and pathological tremor. According to Freund, tremor is an inevitable side effect of any muscle activity, although the physiological tremor accompanying any movement or muscle contraction is usually not perceived. Factors that have been recognized as the major determinants of physiological tremor are rhythmic changes due to pulsatile blood flow, breathing, and mechanically or neurally determined oscillations. Pathological tremors, on the contrary, are characterized by strong motor unit synchronization. This leads to large-amplitude tremor that becomes disturbing to the patient and interferes with movement.

The previous description refers to tremor of the limbs or other body parts. Vocal microtremor, however, is a property of the cycle lengths of voiced speech sounds. Voiced speech sounds involve the vibration of the vocal folds, which is the result of the interaction between the mass and tension of the vocal folds, the shape of the glottis, the speed of the airflow through the glottis, and subglottal pressure. Most of these factors are controlled via the intrinsic and extrinsic muscles of the larynx, and the respiratory muscles that maintain subglottal pressure (Titze, 1995b). The vibrations of the vocal folds determine the cyclic evolution of the glottal volume that rhythmically connects the infra- and supra-laryngeal cavities, giving rise to pulsatile airflow. The time derivative of the airflow rate is heuristically related to the voice source signal at the glottis. Finally, the speech signal is emitted after the voice source signal has been filtered by the vocal tract

a)Electronic mail: jschoent@ulb.ac.be

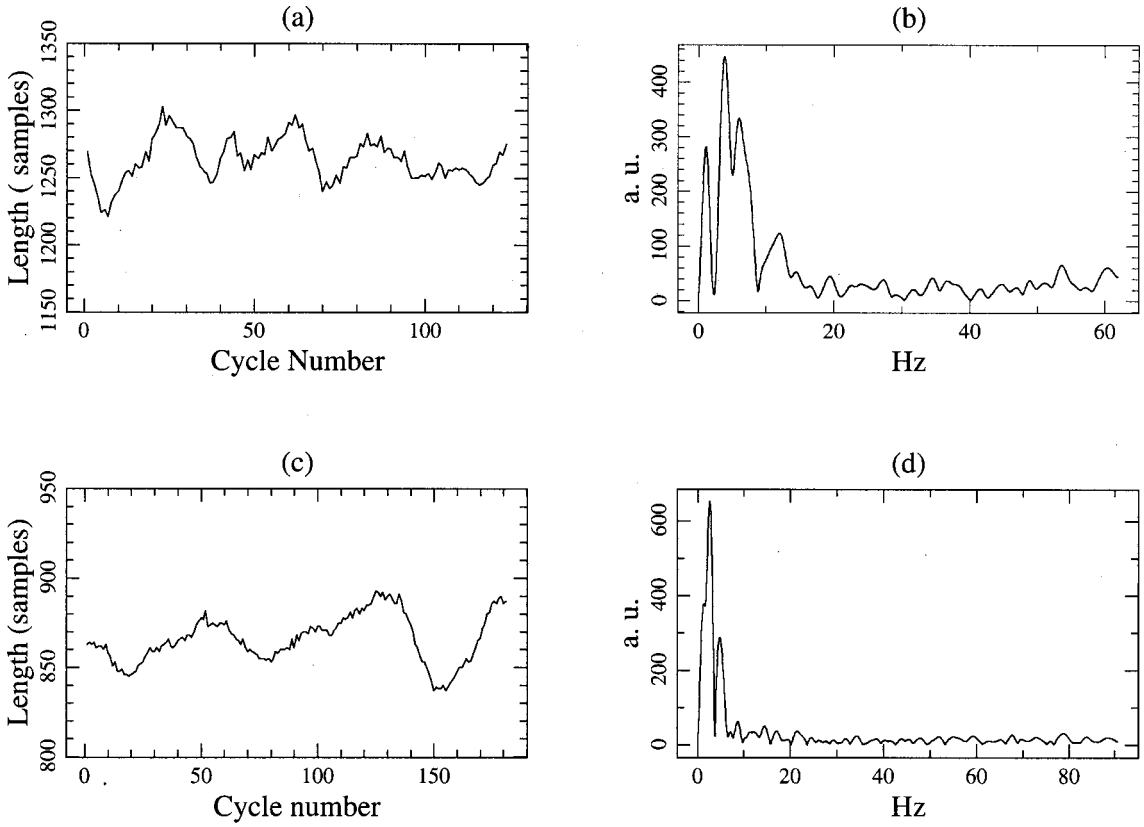


FIG. 1. Two cycle length time series and their Fourier magnitude spectra, which illustrate the distinction between perturbations of the cycle lengths owing to vocal jitter and perturbations owing to vocal microtremor. Time series (a) and its spectrum (b), and time series (c) and its spectrum (d) have been produced by two normophonic male and female speakers, respectively. The vertical axes of (a) and (c) show the cycle length in number of samples and the horizontal axes the cycle number. The sampling frequency is 160 kHz. The vertical axes of (b) and (d) show the spectral magnitude in arbitrary units and the horizontal axes the spectral frequency in Hz.

transfer function and radiated at the lips (Stevens, 1998).

One may therefore assume that vocal microtremor is caused by the physiological tremor of the laryngeal muscles, vocal folds, and respiratory muscles and relayed via the laryngeal cartilages and the subglottal pressure. The length of the chain of causality that is involved and the spatial and temporal distribution of the causes of vocal microtremor, however, entail that the properties of physiological tremor and vocal microtremor are not identical and that the properties of the one cannot be inferred directly from the other.

Vocal tremor has therefore been defined as a modulation of the phonatory frequency or speech cycle amplitude (Titze, 1995a; Sataloff, 1997, p. 825). Commonly, a modulation designates the change of a parameter of an oscillating signal by means of another time-varying signal. In the framework of our study, the modulated quantity is the cycle length of the voice source signal and speech signal. Because any modulating signal is characterized by its amplitude and instantaneous frequency, the modulation of the vocal cycle lengths is summarized by two descriptors, which are the modulation level and modulation frequency.

This article is focused on perturbations of the vocal cycle lengths rather than on perturbations of the speech signal amplitude. The modulation of the amplitude of the speech signal has not been considered for the following reasons. Vocal jitter, vibrato, microtremor, or pathological tremor of the phonatory frequency, as well as intonation, declination, glissandi, and accentuation-related and segment-related perturbations of the phonatory frequency, can be interpreted as modulations of the frequency of the voice source signal. Perturbations of the amplitude of the speech signal, however, do not have a straightforward interpretation in terms of an amplitude modulation of the voice source signal. Unplanned amplitude perturbations of the speech signal may be caused by the modulation conversion owing to the vocal tract transfer function (Panter, 1965).

With simulations, one can indeed show that a modulation of the order of 0.1% of the frequency of the voice source signal may be converted into a modulation of the order of 1% of the amplitude of the speech signal (Michaelis *et al.*, 1998). Also, a modulation of the amplitude of the speech signal may be caused by microtremor of the vocal tract walls (Jiang *et al.*, 2000). Consequently, the modulation of the amplitude of the speech signal cannot be inferred from the modulation of the amplitude of the voice source signal and vice versa.

Some authors have examined the voices of tremor patients together with normal control speakers, but they have not described the modulation level or frequency. Instead they have focused on acoustic cues such as vocal jitter and shimmy (also known as vocal shimmer) in percent, harmonics-to-noise ratio, maximum phonation time, average phonatory frequency, and so on (Schultz-Coulon *et al.*, 1979; Ludlow *et al.*, 1983; Kotby *et al.*, 1995; Gamboa *et al.*,

1997, 1998). Hirose *et al.* (1995) have estimated, in addition, the low- and high-frequency spectral energies of the time series of the phonatory frequency and speech cycle amplitude. Other authors have relied on the visual examination of spectrograms of the speech signal (Lebrun *et al.*, 1982; Kiritani *et al.*, 1995).

Few authors have attempted to characterize the modulation level or modulation frequency. Warwick has determined the modulation level of the phonatory frequency tremor of one patient in the framework of a longitudinal study over an interval of 16 weeks (Warwick *et al.*, 2000). The modulation level, expressed as a percentage of the phonatory frequency, has evolved in an interval from 1% to 3%.

Ludlow *et al.* (1986) have presented some vocal tremor characteristics of normophonic control speakers and tremor patients. The authors have expressed the modulation level relative to vocal jitter. They also have estimated heuristically the typical modulation frequency for each speaker. The range for the normophonic speakers has been equal to 3.8–20 Hz.

Yair and Gath (1988) have presented a vocal tremor analysis method that is based on the spectra of trains of synthetic unit pulses placed at the positions of the prominent peaks of the cycles of sustained vowels. Nine Parkinson patients and three normophonic speakers have been analyzed. The conclusion has been that for normal speech, vocal microtremor is characterized by rhythmic variations of the phonatory frequency that are slower than 3 Hz.

Winholtz and Ramig (1992) have studied vocal tremor with a purpose-built hardware demodulator. The study has included 12 normophonic control speakers. The normal modulation frequency and modulation level intervals have been equal to 4.9–8 Hz and 0.8%–1.8%, respectively.

Gallasch *et al.* (1999) have presented a numerical simulation of the demodulator of Winholtz and Ramig, which is based on the Hilbert transform of low-pass filtered speech. Twelve control speakers and twelve semi-professional singers have sustained vowels [a]. The authors have reported maxima at about 3 Hz in the spectra of the time series of the instantaneous phonatory frequency.

One should note that the heuristic of Ludlow *et al.*, as well as the demodulator of Winholtz and Ramig, and its simulation, include a cutoff frequency below which spectral peaks owing to frequency modulation are not taken into account.

In this article, modulation level and modulation frequency data owing to vocal microtremor in normophonic and mildly dysphonic male and female speakers are examined. The reasons are the following.

First, data on normal vocal microtremor are scarce, as far as we know.

Second, jitter and microtremor together form the perturbations of the vocal cycle lengths. Knowledge of the statistical properties of these perturbations is important, especially in the framework of clinical applications, because vocal jitter and vocal microtremor are base-line phenomena that are present all the time, even in the speech of healthy speakers. From these base-line phenomena, other aperiodicities of the vocal cycles must be distinguished, either in size or kind.

Third, frequency modulation data owing to vocal tremor have been obtained by analyzing point processes (Yair and Gath, 1988), by demodulation (Winholtz and Ramig, 1992), and by inspecting oscillograms or spectrograms (Lebrun *et al.*, 1982; Kiritani *et al.*, 1995). Here an alternative is preferred that is based on the time series of the vocal cycle lengths. The reason is that many articles have addressed the problem how to measure the vocal cycle lengths with a precision that is high enough to describe vocal jitter and vocal microtremor (Hess, 1983; Hess and Indefrey, 1987; Schoentgen and De Guchteneere, 1991).

Fourth, in addition to healthy speakers, mildly dysphonic speakers have been included in the study for the following reason. Simulations with a model have suggested that modulation frequencies and modulation levels owing to microtremor are not altered for mildly disordered voices as long as the perturbations of the instantaneous frequencies of vibration of the glottal walls are normal (Schoentgen, 2001). Mildly disordered voices therefore enable testing predictions of a vocal perturbation model, as well as obtaining microtremor data for a wider range of signals than would have been possible otherwise. Mildly dysphonic speakers are patients whose vocal cycle lengths series satisfy the following criteria. The voice source signal is monocycle periodic, the disturbances owing to microtremor of the prominent peaks of the speech cycles are not masked by additive acoustic noise, no transients owing to register breaks or voice arrests are observed, and vocal jitter is not so strong as to mask vocal microtremor.

Data about five acoustic cues are presented; two describe the modulation frequency, two the modulation level, and the fifth the average phonatory frequency. The data have been obtained for 246 vowel signals sustained by 82 speakers. Data analyses have been based on multi-factor variance analyses. The objectives have been the following.

First, test whether the modulation frequency or level differ significantly between vowel timbres. In theory, no differences are expected. In practice, comparing different vowel timbres enables checking whether the unavoidable amplitude and shape modulation owing to the vocal tract transfer function have biased significantly the speech signal markers on which the measurements of the cycle lengths have been based.

Second, test the prediction that, in the absence of pathological tremor, modulation frequency and modulation level data are comparable for normophonic and mildly dysphonic speakers (Schoentgen, 2001).

Third, test whether the modulation frequency or level data differ significantly for male and female speakers. Published modulation data have not shown any significant differences between six male and six female normophonic speakers (Winholtz and Ramig, 1992).

II. METHODS

This section presents the corpora, the statistical data processing, and the signal processing that obtains the characteristic vocal cycle length, the individual cycle lengths, and the modulation frequencies and levels.

A. Corpora

The utterances are vowel segments [a], [i], and [u] sustained by 38 French-speaking subjects (22 males and 16 females) who have not reported any laryngeal problems, and 51 speakers (19 males and 32 females) who have been patients of the department of otology, rhinology, and laryngology of the St. Pierre University Hospital in Brussels, Belgium. The normal speakers have been from the university staff or have been recruited by advertisement. The average ages are 46 and 49 years for the normophonic male and female speakers, and 48 and 47 years for the dysphonic male and female speakers. The pathologies diagnosed by the department's laryngologists fall under the following categories: oedema of the vocal folds, nodules, hypotonia and asthenia, pseudo-cysts, granuloma, congestion of the vocal folds, polyps, chronic laryngitis, hyperkinesia, and paralysis (De Guchteneere, 1996).

Segments [a], [i], and [u] have been repeated in that order at least three times. Each speaker has been instructed to sustain the vowels at a comfortable pitch and loudness level, and avoid getting out of breath. The recordings have been made in an isolated booth with a Sennheiser microphone (HME 1410 K) mounted on a headset, to keep the mouth-to-microphone distance constant (approximately 2 cm). The microphone has been positioned in front of the right corner of the mouth. Signals have been digitized at 48 kHz on a PCM Sony audio processor and stored on videotape.

Later, a stable segment of each vowel has been selected and redigitized at 20 kHz at a resolution of 12 bits and stored on computer disk for further processing. Before resampling, the signals have been filtered with a low-pass filter at 10 kHz. The stored segments are 1-s-long, central, steady segments of the vowels.

B. Estimation of the characteristic vocal cycle length

The characteristic cycle length for each vowel segment has been obtained with the empirical auto-covariance function of the two-level vowel signal. The two-level signal is obtained by assigning $+1$ to positive and -1 to negative signal samples. The purpose is to enhance the pseudo-periodicity of the signal by repressing distracting features of the original (Rabiner and Schafer, 1978). Among the three highest local maxima of the auto-covariance function, the maximum whose position (in number of positive lags) agrees within 5% with the positions of the other two, modulo an integer number, is kept. The position in number of lags of this local maximum is the characteristic vocal cycle length in number of samples of the vowel signal. When this rule fails to provide a length estimate, the program halts, outputs all possible maxima, and invites the experimenter to select one by hand.

C. Extraction of the vocal cycle length time series

In a first stage, each vowel signal is up-sampled by repeating each original sample eight times. The result is a staircase function sampled at 160 kHz, which is then low-pass filtered at 800 Hz. Indeed, one has shown that speech signals that have been low-pass filtered in the vicinity of 1000 Hz are adequate for extracting cycle lengths for characterizing vocal jitter (Titze *et al.*, 1987). The reason for up-sampling is that, typically, vocal jitter is less than 1%, and the maximum excursion owing to vocal microtremor is a few percent at most of the average cycle length (Winholtz and Ramig, 1992; Schoentgen and De Guchteneere, 1995). The low-pass filter is a symmetric finite-impulse-response filter with a length equal to 1000 samples (Stearns and David, 1993). This method for increasing the sampling frequency, together with its validation, has been presented elsewhere (Schoentgen and De Guchteneere, 1991, 1995).

Because the focus of the study is vocal microtremor rather than vocal jitter, the cycle lengths have been obtained by means of the positions of the prominent peaks of the speech cycles, because peak positions are less likely than zero-crossing positions to be biased by low-frequency additive noise. The prominent peaks occur in the vicinity of the instants of the main acoustic excitation of the vocal tract. The recursive search for the prominent peaks is similar to the one proposed by Davis (1979). In a first stage, all positive peaks and their positions are recorded. In a second stage, the prominent peaks are selected among the recorded peaks and the cycle lengths are defined as the distances between the positions of adjacent prominent peaks. The starting position is the position of the highest peak. The distance between the starting position and the next peak position to the left is divided by the characteristic vocal cycle length, which has been obtained previously. When the distance in number of samples is not within 10% modulo one of the characteristic length, the peak is dropped and the next peak to the left is selected. When the distance is within 10% modulo one of the characteristic length, the distance is stored as a valid cycle length, the starting position is assigned to the most recent prominent peak, and the search is carried out again. When no more valid cycle lengths can be found by leftward shifts, the starting position is switched back to the highest peak and the search is carried out to the right. The polarity of the vowel signal is then reversed and the search is carried out again. Consequently, two length time series are obtained for each vowel segment.

Each time series has been visually inspected for the following defects. First, the time series is too short because the search has stopped prematurely owing to isolated perturbations that exceed 10%. Second, the time series exhibits outliers, that is, the raw length perturbations are greater than four times the standard deviation. Third, the cycle length time series displays multi-cycle periodicity or phonatory frequency glissando.

When both time series of a vowel are free of defects, the one is kept that has been obtained via the signal peaks that are nearer to the acoustic excitation epochs. When only one time series is free of defects, it is kept by default. When both time series exhibit defects, one or several of the following operations are carried out in due order. First, the search interval is increased from 10% to 15% of the typical cycle length. Second, the up-sampled signal is low-pass filtered with a cutoff frequency of 500 Hz, instead of 800 Hz. Third, the signal is discarded and replaced by another vowel of the same timbre produced by the same speaker. Fourth, the

speaker is removed from the corpus. The latter has been the case for dysphonic speakers only. The final number of male and female dysphonic speakers has therefore been equal to 16 and 28.

Finally, a straight line is fitted to the length time series in the least squares sense. The fitted line is then subtracted to remove any linear trends owing to phonatory frequency drift, declination, or glissando.

Clearly, the purpose of the study is not to distinguish acoustically between normophonic and dysphonic speakers, or to obtain the vocal cycle lengths for any vowel signal produced by any speaker. The purpose is to obtain cycle length time series that are as free as possible from defects that may prevent the analysis of vocal microtremor to be carried out meaningfully.

D. Estimation of the modulation frequency

The time series of the vocal cycle lengths are stochastic. A conventional nonparametric analysis of a stochastic signal consists in the calculation of the spectral density function, which is the Fourier transform of the auto-covariance function of the signal. In practice, the formal spectral density function and auto-covariance function are replaced by the magnitude of the discrete Fourier transform of the empirical auto-covariance function, which can be computed for discrete time series of finite length. The empirical auto-covariance function is calculated for positive and negative lags. When the data length equals N, the maximum lag equals $N-1$ and the length of the empirical auto-covariance function equals $2N-1$. For high values of the lag, the auto-covariance coefficients are unreliable because few time series samples are involved in the calculation of the coefficient. The auto-covariance function is therefore windowed by a Hamming window before the discrete Fourier transform is computed (Chatfield, 1984).

An alternative consists in calculating the magnitude spectrum of the Hamming windowed time series itself. Generally speaking, the magnitude spectrum of a stochastic time series does not converge to the smooth contour of the theoretical spectral density function when the length of the time series increases. One therefore recommends smoothing the magnitude spectrum with a running average of $N/40$ spectral samples to approximate the spectral contour better (Chatfield, 1984). Each spectral sample has therefore been replaced by the average of itself and its left and right neighbors, because the number of cycles, N, has been comprised between 100 and 200.

If the microperturbations of the vocal cycle lengths were correlation-free jitter only, the contours of the magnitude spectra would be flat. Vocal microtremor, however, gives rise to strong spectral components at or near the modulation frequencies, which can therefore be estimated via the positions of the prominent spectral peaks of the magnitude spectrum [Figs. 1(b) and (d)]. Prominent peaks are those whose sizes are statistically different from the sizes of the spectral components owing to white noise. The prominent peaks have been found by means of an empirical statistical method, which consists in shuffling the original time series 100 times to create 100 artificial white noise time series that have the same average and empirical first-order moments as the original (Noreen, 1989). The spectral peaks of the unshuffled time series that are kept are those whose sizes always exceed the sizes of the matching spectral components of the shuffled time series. In other words, spectral peaks are assigned to vocal microtremor when the empirical probability is less than 1% that they are due to white noise. These peaks and their positions are stored for further processing. When no statistically significant spectral peaks can so be found, a modulation frequency equal to zero is assigned to the time series. This value flags the inability to detect spectral peaks owing to microtremor.

The following is a summary of the computer-intensive statistical processing of the spectral peaks:

(a) windowing of the original time series and subtraction of the average;
(b) calculation of the spectrum either directly or via the autocorrelation function;
(c) shuffling of the original time series;
(d) windowing of the shuffled time series and subtraction of the average;
(e) calculation of the spectrum either directly or via the autocorrelation function;
(f) comparison of the sizes of the spectral peaks of the original time series to the sizes of the matching spectral components of the shuffled time series;
(g) counter-incrementation for each ''original'' peak whose size exceeds the size of the matching ''shuffled'' spectral component;
(h) go to (c) if the number of shuffles is less than 100, otherwise go to (i);
(i) flagging as ''prominent'' of every peak whose counter equals 100.

The modulation frequency in Hz can be inferred from the positions of the prominent spectral peaks. N cycles in the time or lag domain are transformed via the discrete Fourier transform in N samples in the frequency domain. Consequently, when the cycle lengths are sampled at a constant frequency, f_0, the frequency spacing in Hz equals f_0/N if the cycle perturbation time series is Fourier transformed, and $f_0/(2N-1)$ if the empirical auto-covariance function is transformed. In practice, frequency f_0 equals the average phonatory frequency.

The assumption that the perturbations have been sampled at a constant frequency, f_0, is equivalent to making a small error while measuring the perturbations. The reason is that the length perturbations have been recorded cycle-by-cycle and not at the rhythm of a constant time interval. But, one can show by linear interpolation that when the perturbations are of the order of 0.1% of the average cycle length, the equivalent measurement error is of the order of one-tenth of the quantization noise, which is due to the sampling of the speech signal at 160 kHz.

The typical spectral resolution is the following. The analysis interval T over which the cycle lengths have been recorded equals 1 s. This means that the number of cycles equals Tf_0, rounded to the next smallest integer. Consequently, the frequency spacing approximately equals $1/T$,

when the spectral density function is estimated via the cycle perturbation time series, and half as much when it is estimated via the empirical auto-covariance function. Accordingly, the modulation frequency values are obtained with a precision of ± 0.5 Hz or ± 0.25 Hz. In short, for voices that are subject to normal microtremor, the equivalent measurement error that has been referred to in the previous paragraph is negligible compared to the quantization noise and the uncertainty owing to the finite spectral resolution.

For many time series, more than one statistically significant spectral peak has been detected. The peak positions in Hz are estimates of the modulation frequencies. They have been replaced by a weighted average frequency (Winholtz and Ramig, 1992). The reason is that single frequencies are easier to compare statistically and easier to compare to the microtremor data published by others. Weighting the individual frequencies by the sizes of the corresponding spectral peaks and dividing by the total weight, one obtains the average. The upper limit above which spectral peaks have not been taken into account equals 25 Hz (Winholtz and Ramig, 1992). The lower limit has been 0.5 or 1 Hz, i.e., the smallest possible frequency strictly greater than zero, taking into account the finite spectral resolution.

In the text, the modulation frequency estimates obtained via the empirical auto-covariance function are labeled f_{m1} and the modulation frequencies directly obtained via the perturbation time series are labeled f_{m2}.

E. Estimation of the modulation level

Formally, the modulation level is a cue of the size, relative to the average cycle length, of the perturbations owing to microtremor (Winholtz and Ramig, 1992). The distinction between perturbations owing to vocal jitter and perturbations owing to vocal microtremor is based on their typical frequency intervals, which is half the phonatory frequency for vocal jitter and of the order of 1 Hz for vocal microtremor. Smoothing the length time series can therefore decrease the influence of vocal jitter on the size of the measured perturbations owing to microtremor.

No hard rules are available that would dictate the choice of the smoother. We therefore require that, first, frequency components below typically 25 Hz be kept. Second, the variance of the time series that is removed plus the variance of the time series that is kept equals the variance of the original time series. Third, the analyzing function is as short as possible. Fourth, the smoothed time series is free of artefacts owing to the boundaries of the time series. These four requirements are fulfilled by the Haar wavelet transform, which is a multi-resolution analysis method.

Because the Haar wavelet transform is carried out on numbers of samples that are a power of 2, the original time series is padded to the right with zeros till the total number of samples equals a power of 2. Then all wavelets above the sixth scale are zeroed and the inverse Haar transform is carried out to obtain the smooth time series. The reason is the following. The sixth scale comprises 32 wavelets that are shared out equally over the length of the analysis interval. According to an empirical rule, the frequency typical of a wavelet scale equals the number of wavelets in the scale divided by the length of the analysis interval (Newland, 1993). Therefore, when the phonatory frequency equals 100 Hz, for instance, the extended analysis interval equals 1.28 s and the frequency typical of the highest wavelet scale of the smoothed time series equals 25 Hz, which is the desired upper boundary.

A modulation level in percent has been calculated as follows (Winholtz and Ramig, 1992). Length $L_{\max}$ is the absolute maximum of the smoothed time series and length L_0 the average:

$$m_1 = \frac{L_{\max} - L_0}{L_0} \times 100.$$

A possible problem with level m_1 is that its value depends on the maximum length only, which may ill represent the typical length. A second cue has therefore been computed that is based on the mean square root of the N cycle perturbations. It combines all cycle perturbations, $L_i - L_0$, and can be interpreted as the standard deviation of the excursions from the average, L_0, owing to vocal microtremor.

$$m_2 = \frac{\sqrt{(1/N)\Sigma_{i=1}^{N}(L_i - L_0)^2}}{L_0} \times 100.$$

F. Data analyses

At this stage, the phonatory frequency, as well as the weighted vocal modulation frequencies and vocal modulation levels, have been estimated by two methods each for 82 speakers sustaining three vowel timbres. The objectives of the statistical processing are to investigate whether statistically significant differences exist between male and female speakers, normophonic and dysphonic speakers, and vowel timbres [a], [i], and [u]. Five (one per vocal cue) two-factor analyses of variances with repeated measures have therefore been carried out. The factors are gender (male, female) and laryngeal status (normophonic, dysphonic). Each speaker has produced three vowel timbres ([a],[i],[u]). The null hypotheses are that the averages are the same for the vowel timbres, male and female, and normophonic and dysphonic speakers.

III. RESULTS

In the results section, the vocal cue data are summarized by the quartiles, which provide a five-number summary. The typical quartile intervals that are given in the text are obtained by citing the minimum in the first quartile row and the maximum in the third quartile row of the matching tables. The quartiles are tabulated separately for vowel timbres [a], [i], [u], for the male and female speakers, and for the normophonic and dysphonic speakers. Tables I–V, in that order, show the quartiles for modulation levels m_1 and m_2 (in %), modulation frequencies f_{m1} and f_{m2} (in Hz), and average phonatory frequency f_0 (in Hz).

Attributes m_1 and m_2 are two modulation level cues; the first takes into account the maximum perturbation, and the second all the perturbations relative to the average cycle length. For cues m_1 and m_2, the typical interquartile intervals are equal to 1.0%–3.0% and 0.4%–1.3%, respectively. Predictably, the values of the first are greater than the values

TABLE I. Modulation level m_1 in percent. Level m_1 is the maximum perturbation divided by the average cycle length.

	Male						Female					
	Normophonic			Dysphonic			Normophonic			Dysphonic		
	[a]	[i]	[u]	[a]	[i]	[u]	[a]	[i]	[u]	[a]	[i]	[u]
Min	0.6	0.8	0.7	0.9	0.6	0.7	0.7	0.6	1.0	0.9	0.7	0.6
First quartile	1.2	1.1	1.1	1.4	1.1	1.4	1.2	1.0	1.5	1.6	1.5	1.4
Second quartile	1.8	1.4	1.7	1.8	1.4	1.7	1.9	1.5	1.7	1.9	2.0	1.9
Third quartile	2.4	2.0	2.5	3.0	2.5	2.2	2.9	1.8	1.9	2.5	2.8	2.3
Max	4.0	4.6	6.0	8.5	12.0	2.6	5.2	3.5	5.3	6.2	6.9	14.2

TABLE II. Modulation level m_2 in percent. Level m_2 is the standard deviation of the cycle length perturbations divided by the average cycle length.

	Male						Female					
	Normophonic			Dysphonic			Normophonic			Dysphonic		
	[a]	[i]	[u]	[a]	[i]	[u]	[a]	[i]	[u]	[a]	[i]	[u]
Min	0.3	0.3	0.3	0.4	0.3	0.3	0.4	0.2	0.4	0.4	0.3	0.2
First quartile	0.5	0.5	0.5	0.6	0.4	0.6	0.6	0.4	0.6	0.7	0.6	0.6
Second quartile	0.8	0.7	0.7	0.8	0.6	0.7	0.9	0.6	0.8	0.8	0.9	0.7
Third quartile	1.1	0.8	0.9	1.1	1.0	0.8	1.3	0.8	0.8	1.1	1.1	0.9
Max	2.2	2.0	2.3	3.1	3.9	1.5	1.6	1.4	1.9	2.4	1.7	4.5

TABLE III. Modulation frequency f_{m1} in Hz. Frequency f_{m1} has been obtained via the magnitude spectrum of the windowed empirical auto-covariance function of the cycle length time series.

	Male						Female					
	Normophonic			Dysphonic			Normophonic			Dysphonic		
	[a]	[i]	[u]	[a]	[i]	[u]	[a]	[i]	[u]	[a]	[i]	[u]
Min	2.1	1.4	1.0	1.7	1.0	1.6	1.8	2.1	1.6	1.5	1.5	1.0
First quartile	2.6	2.7	2.2	3.1	2.5	2.5	2.5	2.7	2.2	2.7	2.6	2.5
Second quartile	3.5	3.3	3.1	3.8	3.2	3.1	2.9	3.2	2.8	3.5	3.0	3.3
Third quartile	3.8	4.5	3.8	4.3	3.7	4.1	3.6	4.1	3.4	3.9	4.0	4.4
Max	5.4	8.9	8.3	6.9	5.9	8.9	4.7	7.9	5.0	6.3	7.6	6.2

TABLE IV. Modulation frequency f_{m2} in Hz. Frequency f_{m2} has been obtained via the smoothed magnitude spectrum of the windowed cycle length time series.

	Male						Female					
	Normophonic			Dysphonic			Normophonic			Dysphonic		
	[a]	[i]	[u]	[a]	[i]	[u]	[a]	[i]	[u]	[a]	[i]	[u]
Min	2.0	2.0	1.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
First quartile	3.0	2.1	2.0	2.2	2.0	3.0	2.1	3.1	2.1	2.3	2.0	3.0
Second quartile	3.2	3.1	3.1	3.2	3.4	3.3	3.1	3.4	3.1	3.7	3.2	3.1
Third quartile	4.1	4.1	4.2	4.4	4.1	4.1	4.1	4.4	4.0	4.1	4.0	4.7
Max	5.2	9.3	8.3	6.7	13.3	8.9	5.6	7.8	6.2	8.2	7.0	7.8

TABLE V. Characteristic phonatory frequency f_0 in Hz.

	Male						Female					
	Normophonic			Dysphonic			Normophonic			Dysphonic		
	[a]	[i]	[u]	[a]	[i]	[u]	[a]	[i]	[u]	[a]	[i]	[u]
Min	74	91	88	97	100	98	133	158	158	124	124	128
First quartile	107	115	112	113	121	121	175	181	173	181	194	182
Second quartile	113	121	125	123	129	136	191	210	200	203	210	213
Third quartile	127	143	145	146	147	158	204	234	225	220	223	224
Max	179	195	189	205	221	217	325	353	322	260	284	285

of the second. Both are of the same order, however. A visual inspection of Tables I and II does not disclose any qualitative differences between normophonic and dysphonic speakers, male and female speakers, or low and high vowels.

Attributes f_{m1} and f_{m2} are two modulation frequency cues; the first is based on the spectra of the empirical autocovariance function and the second on the spectra of the time series itself. For frequencies f_{m1} and f_{m2} the interquartile intervals are equal to 2.2–4.5 Hz and 2.0–4.7 Hz, respectively. A comparison of Tables III and IV shows that matching quartiles agree better than the spectral resolution (± 0.5 Hz) in 56 out of 60 table cells. A visual inspection of Tables III and IV does not disclose any qualitative differences regarding vowel timbre, gender, or laryngeal status.

The estimation of the modulation frequency is based on the detection of spectral peaks whose sizes differ significantly from spectral components owing to white noise. When no statistically significant spectral peaks are found, the time series is flagged as microtremor-free. For a total of 246 time series, no statistically significant microtremor has been observed for four time series in the case of modulation frequency cue f_{m1} and for five time series in the case of f_{m2}. The matching percentages are 1.6% and 2.0%.

Finally, Table V summarizes the phonatory frequencies for each speaker group. A visual inspection predictably shows that the phonatory frequencies are greater for the female than for the male speakers, and greater for high vowels [i] and [u] than for low vowel [a].

Five analyses of variances have been carried out, one for each vocal cue. Table VI summarizes the results via the *F*-statistic and the corresponding probability *p* that the null hypothesis is true given the observed *F*-value. Table VI confirms the following.

(a) The modulation levels and modulation frequencies do not differ significantly (i.e., $p > 0.05$) for vowel timbres [a], [i], [u], as well as for male and female speakers, and for normophonic and dysphonic speakers.
(b) The phonatory frequencies differ significantly for male and female speakers, as well as for vowel timbres [a], [i], and [u].

IV. DISCUSSION

The topics in this section are the following: the dependence of phonatory frequency on vowel timbre and gender, and the predicted consequences for the modulation level; the lack of dependence of the modulation level on gender, vowel timbre, and laryngeal status; the lack of dependence of the modulation frequency on gender, vocal timbre, and laryngeal status; and the differences between data reported here and elsewhere (Ludlow *et al.*, 1986; Winholtz and Ramig, 1992).

A. Modulation level and phonatory frequency

The phonatory frequencies are greater for female than for male speakers, and greater for high than for low vowels (Tables V and VI). These are common observations that are explained by anatomical differences between males and females (Baken, 1987), and by an interaction between the tongue and larynx, respectively (Honda, 1995).

Differences in phonatory frequency must be discussed because modulation levels m_1 and m_2 are ratios of the typical perturbation and of the average vocal cycle length. The reason the typical size of the perturbations is expressed relative to the average cycle length is the following. One knows that vocal microperturbations increase and decrease with increasing and decreasing cycle length. This is a robust phe-

TABLE VI. *F*-Statistics (left) and probabilities *p* (right) obtained by means of analyses of variance of the modulation frequencies, modulation levels, and phonatory frequency. ''Laryngeal status'' refers to whether speakers are normophonic or dysphonic; ''timbre'' refers to vowel timbres [a], [i] and [u]. Letters G, T and S refer to (G)ender, (T)imbre and laryngeal (S)tatus, respectively.

	f_{m1}		f_{m2}		m_1		m_2		f_0	
	F	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
Laryngeal status	0.28	0.60	1.06	0.31	2.44	0.12	1.31	0.26	0.14	0.71
Gender	0.79	0.38	0.60	0.44	10^{-4}	0.99	0.01	0.93	93.7	10^{-4}
Timbre	0.70	0.50	0.51	0.60	0.52	0.60	2.34	0.10	18.8	10^{-4}
Gender×status	0.49	0.49	0.26	0.61	0.01	0.91	0.01	0.93	1.13	0.29
Timbre×gender	0.52	0.60	0.06	0.94	1.23	0.29	1.01	0.36	0.95	0.39
Timbre×status	2.67	0.07	0.72	0.49	1.22	0.30	1.21	0.30	2.34	0.10
G×T×S	0.23	0.80	0.52	0.60	1.82	0.17	0.79	0.46	0.73	0.48

nomenon that has been explained in the framework of a simulation study as follows (Schoentgen, 2001). The observed perturbations of the cycle lengths are the outcome of many small spatially and temporally distributed perturbations of the instantaneous frequencies of vibration of the glottal walls. When the glottal cycle lengths are long, more perturbations occur within one cycle. Their cumulative effects therefore increase. The division of the typical size of the perturbations by the average cycle length consequently compensates for the dependence of the perturbations on the average cycle length.

B. Modulation level, gender, vowel timbre, and laryngeal status

Modulation levels m_1 and m_2 do not differ significantly for male and female speakers (Tables VI). These observations agree with previously published data (Winholtz and Ramig, 1992). The likely explanation is that the division of the size of the perturbations by the average cycle length has successfully compensated for the dependence of the perturbations on the average vocal cycle length, which differs for males and females. This suggests that the perturbations owing to microtremor evolve proportionally to the vocal cycle length and that other experimental factors are not important.

The same tables show that modulation levels m_1 and m_2 do not differ significantly for vowel timbres [a], [i], and [u]. Again, this suggests that dividing the perturbation size by the average cycle length successfully compensates for any differences due to segment-specific effects on phonatory frequency. High vowels are indeed characterized by a higher phonatory frequency than low vowels (Tables V and VI).

The main reason data for different vowel timbres have been recorded is that the vocal tract transfer function converts any frequency or amplitude modulation of the voice source signal into an amplitude and shape modulation of the speech signal (Panter, 1965). Modulation conversion may change the size of the length perturbations spuriously because the measurements of the vocal cycle lengths are based on features of the radiated speech signal shape. But, the lack of significant differences between vowel timbres suggests that modulation conversion has not biased significantly the measured cycle lengths.

The modulation levels do not differ significantly for normophonic and dysphonic speakers (Table VI). Vocal jitter and frequency microtremor are thought to be consequences of many small perturbations of the instantaneous frequencies of vibration of the glottal walls. A model of the vocal perturbations suggests that if, first, the instantaneous frequencies of the left and right glottal walls influence each other equally, second, the voice source signal is monocycle periodic and, third, precautions are taken so that additive noise or experimental artefacts do not mask genuine microperturbations, then vocal jitter and vocal microtremor level are not increased by pathological glottal conditions, as long as the perturbations of the instantaneous frequencies of vibration of the glottal walls are normal (Schoentgen, 2001). This is confirmed by our data, which have been obtained for healthy and mildly dysphonic speakers, and selected so that the above conditions are satisfied as far as possible.

C. Modulation frequency, gender, vowel timbre, and laryngeal status

The vocal modulation frequency data do not differ significantly for male and female speakers (Table VI). These observations agree with previously published data (Winholtz and Ramig, 1992). Also, the vocal perturbation model that has been referred to earlier suggests that the vocal modulation frequency does not depend on parameters of the glottal shape that evolve with phonatory frequency (Schoentgen, 2001).

The modulation frequencies are the same for vowel timbres [a], [i], and [u] (Table VI). This agrees with what has been discussed in the previous paragraph. If the vocal modulation frequency does not depend on gender and therefore on phonatory frequency, vowel timbre is not expected to influence the modulation frequency via a segment-specific effect on phonatory frequency either. Also vowel timbre is not expected to influence the vocal modulation frequency via the vocal tract transfer function, because a constant linear filter does not shift the frequencies of the inputted spectral components.

Also the modulation frequencies do not differ significantly for normophonic and dysphonic speakers (Table VI). The explanation is similar to the one given above for the modulation level. Indeed, the vocal perturbation model suggests that if the voice source signal is monocycle periodic, and precautions are taken so that additive noise or experimental artefacts do not mask genuine microperturbations, then the vocal microtremor frequency is not influenced by pathological glottal conditions, as long as the perturbations of the instantaneous frequencies of vibration of the glottal walls are normal (Schoentgen, 2001). This is the reason mildly dysphonic speakers have been included in the first place.

D. Former studies

The vocal microtremor frequency data that have been obtained in the framework of this study agree with the six vocal microtremor spectra published by Yair and Gath (1988), Winholtz and Ramig (1992), and Titze (1994). However, differences between the data reported here and elsewhere exist (Ludlow *et al.*, 1986; Winholtz and Ramig, 1992). These differences can be explained.

Winholtz and Ramig (1992) have reported modulation frequencies and levels for six male and six female normal control speakers sustaining vowels [a]. The corresponding interquartile intervals have been equal to 5.5–8 Hz, 4.9–6.1 Hz, 0.9%–1.8%, and 0.8%–1.3%, respectively. Ludlow *et al.* (1986) have reported total tremor frequency ranges equal to 8–20 Hz for one half and equal to 3.8–7.5 Hz for the other half of 20 normal control speakers sustaining vowels [a]. Ludlow *et al.* have not reported the corresponding modulation levels. In the framework of our study, the matching interquartile intervals are equal to 3.0–4.1 Hz, 2.1–4.1 Hz, 1.2%–2.4%, and 1.2%–2.9% for male and female normophonic speakers sustaining vowel [a] (Tables I and IV).

A possible explanation for the observed discrepancies between frequency and level ranges is the differences be-

tween the frequency analysis intervals. Indeed, in our study statistically significant spectral peaks have been taken into account whose frequencies are strictly greater than zero and less than 25 Hz. Winholtz and Ramig have taken into account all spectral peaks in an interval between 2.5 and 25 Hz. Ludlow *et al.* have not reported their analysis intervals. But, one may tentatively infer from the description of their heuristic that the maximum tremor cycle length has been equal to 18 vocal cycles for the males and 26 vocal cycles for the females. One may therefore hypothesize that the corresponding cutoff frequencies for vocal microtremor have been equal to 5.5 and 7.7 Hz, respectively, assuming phonatory frequencies of 100 Hz for the males and 200 Hz for the females.

At first sight, one may therefore conclude that similar microtremor frequencies would have been reported if in the framework of the three studies the same frequency analysis intervals were used. This applies also to the modulation level intervals reported by Winholtz and Ramig, which are smaller than the ones reported here. This is to be expected because Winholtz and Ramig have omitted spectral components below 2.5 Hz, which are prominent in normal circumstances.

The choice of the analysis interval therefore appears to be a major issue. But, no final agreement seems to have been reached in the literature regarding this choice. The definition of vocal tremor whose frequency range includes the frequency ranges expected by most authors is the following. Vocal tremor is a 1–15 Hz modulation of phonatory frequency or amplitude (Titze, 1995a). According to Titze, tremor has a neurological origin or is due to an interaction between neurological and biomechanical properties of the vocal folds. Freund (1987) includes, in addition, pulsatile blood flow and breath, and Fucci and Petrosino (1984) state that the origins of tremor are wide and varied. In 1989, Orlikoff and Baken (1989) have reported circumstantial evidence that the phonatory frequency in humans is modulated by the heartbeat. These data suggest that the lower boundary of the microtremor frequency range should indeed be fixed at 0.5 Hz or lower.

V. CONCLUSION

An analysis of the vocal cycle lengths has been carried out to determine the modulation frequency and modulation level owing to frequency microtremor in normophonic and mildly dysphonic male and female speakers. The results are the following.

(1) Modulation frequency and relative modulation level do not differ significantly for male and female speakers (Table VI).
(2) Modulation frequency and relative modulation level do not differ significantly for normophonic and mildly dysphonic speakers (Table VI).
(3) Modulation frequency and relative modulation level do not differ significantly for vowel timbres [a], [i], and [u] (Table VI).
(4) The typical interquartile intervals of the modulation frequency and the relative modulation level are equal to 2.0–4.7 Hz and 0.4%–1.3%, respectively (Tables II and IV).
(5) The dissimilarities between modulation data reported by different studies are possibly due to different cutoff frequencies below which spectral peaks are considered not to contribute to vocal microtremor.

ACKNOWLEDGMENTS

The author is grateful to Professor Denis Hennebert and Dr. Raoul De Guchteneere for the recordings of the normal and disordered voices, and to Renaud Beeckmans for the statistical analysis of the data. This work has been supported by the National Fund for Scientific Research, Belgium, and has been carried out in the framework of an *Action de Recherche Concertée (98-02; 226), Communauté Française de Belgique*.

Baken, R. J. (**1987**). *Clinical Measurement of Speech and Voice* (College Hill, Boston), pp. 125–196.
Chatfield, C. (**1984**). *The Analysis of Time Series* (Chapman and Hall, London), p. 44.
Davis, St. B. (**1979**). ''Acoustic characteristics of normal and pathologic voices,'' in *Speech and Language, Advances in Basic Research and Practice, Vol. 1*, edited by N. J. Lass (Academic, New York), pp. 273–331.
De Guchteneere, R. (**1996**). ''Mesure et analyse chronologique des fluctuations de la durée des cycles glottiques'' (''Measurement and time series analysis of the fluctuations of the glottal cycle lengths''), unpublished doctoral thesis, Université Libre de Bruxelles, Brussels, Belgium.
Freund, H.-J. (**1987**). ''Central rhythmicities in motor control and its perturbances,'' in *Temporal Disorder in Human Oscillatory Systems*, edited by L. Rensing, U. an der Heiden, and M. C. Mackey (Springer, Berlin), pp. 79–82.
Fucci, D., and Petrosino, L. (**1984**). ''The practical applications of neuroanatomy for the speech-language pathologist,'' in *Speech and Language, Advances in Basic Research and Practice*, edited by N. J. Lass (Academic, New York), Vol. 11, pp. 249–317.
Gallasch, E., Girardi, L., and Kenner, T. (**1999**). ''Characterization of vocal microtremor,'' Abstract 2nd International Conference on Voice Physiology and Biomechanics, Berlin.
Gamboa, J., Jimenez, F., Nieto, A., Montojo, J., Orti-Pareja, M., Molina, J. A., Garcia-Albea, E., and Cobeta, I. (**1997**). ''Acoustic voice analysis in patients with Parkinson's disease treated with dopaminergic drugs,'' J. Voice **11**(3), 314–320.
Gamboa, J., Jimenez, F., Nieto, A., Cobeta, I., Vegas, A., Orti-Pareja, M., Gasalla, T., Molina, J. A., and Garcia-Albea, E. (**1998**). ''Acoustic voice analysis in patients with essential tremor,'' J. Voice **12**(4), 444–452.
Hess, W. (**1983**). *Pitch Determination of Speech Signals—Algorithms and Devices* (Springer, Berlin).
Hess, W., and Indefrey, H. (**1987**). ''Accurate time-domain pitch determination of speech signals by means of a laryngograph,'' Speech Commun. **6**, 55–68.
Hirose, H., Imaizumi, S., and Yamori, M. (**1995**). ''Voice quality in patients with neurological disorders,'' in *Vocal Fold Physiology*, edited by O. Fujimura and M. Hirano (Singular, San Diego), pp. 235–248.
Honda, K. (**1995**). ''Laryngeal and extra-laryngeal mechanisms of F_0 control,'' in *Producing Speech: Contemporary Issues*, edited by F. Bell-Berti and L. J. Raphael (AIP, New York), pp. 215–232.
Jiang, J., Lin, E., Wu, J., Gener, C., and Hanson, D. (**2000**). ''Effects of simulated source of tremor on acoustic and airflow voice measures,'' J. Voice **14**, 47–57.
Kiritani, S., Niimi, S., Imagawa, H., and Hirose, H. (**1995**). ''Vocal fold vibrations associated with involuntary voice changes in certain pathological cases,'' in *Vocal Fold Physiology*, edited by O. Fujimura and M. Hirano (Singular, San Diego), pp. 269–281.
Kotby, M. N., El-Sady, S., Khidr, A. A., Alloush, T., Abdel Nasser, N. H., Gamal, N., Mahmoud H., El-Sharkawi A. E., and Kamal, A. (**1995**). ''Voice quality in neurological disorders of the larynx: pathophysiological correlates,'' in *Vocal Fold Physiology*, edited by O. Fujimura and M. Hirano (Singular, San Diego), pp. 249–267.
Lebrun, Y., Devreux, F., Rousseau, J.-J., and Darimont, Ph. (**1982**). ''Tremulous speech,'' Folia Phoniatr. **34**, 134–142.

Ludlow, C., Coulter, D., and Gentges, F. (**1983**). ''The differential sensitivity of frequency perturbation to laryngeal neoplasms and neuropathologies,'' in *Vocal Fold Physiology*, edited by D. M. Bless and J. H. Abbs (College Hill, San Diego), pp. 381–392.
Ludlow, C., Bassich, C. J., Connor, N. P., and Coulter, C. (**1986**). ''Phonatory characteristics of vocal fold tremor,'' J. Phonetics **14**, 509–515.
Michaelis, D., Fröhlich, M., Strube, H., Kruse, E., Story, B., and Titze, I. (**1998**). ''Some simulations concerning jitter and shimmer measurement,'' Proceedings 3rd International Workshop Advances in Quantitative Laryngoscopy, Voice and Speech Research, Aachen, pp. 71–80.
Newland, D. E. (**1993**). *An Introduction to Random Vibrations, Spectral & Wavelet Analysis* (Longman Scientific and Technical, Harlow, UK), p. 344.
Noreen, E. W. (**1989**). *Computer Intensive Methods for Testing Hypotheses* (Wiley, New York), pp. 125–158.
Orlikoff, R. F., and Baken, R. J. (**1989**). ''Fundamental frequency modulation of the human voice by the heartbeat: preliminary results and possible mechanisms,'' J. Acoust. Soc. Am. **85**, 888–893.
Panter, P. (**1965**). *Modulation, Noise and Spectral Analysis* (McGraw–Hill, New York), pp. 273–300.
Rabiner, L. R., and Schafer, R. W. (**1978**). *Digital Processing of Speech Signals* (Prentice–Hall, Englewood Cliffs, NJ), p. 150.
Sataloff, R. T. (**1997**). *Professional Voice* (Singular, San Diego).
Schoentgen, J. (**2001**). ''Stochastic models of jitter,'' J. Acoust. Soc. Am. **109**, 1631–1650.
Schoentgen, J., and De Guchteneere, R. (**1991**). ''An algorithm for the measurement of jitter,'' Speech Commun. **10**, 533–538.
Schoentgen, J., and De Guchteneere, R. (**1995**). ''Time series analysis of jitter,'' J. Phonetics **23**, 189–201.
Schultz-Coulon, H.-J., Battmer, R.-D., and Fedders, B. (**1979**). ''Zur quantitativen Bewertung der Tonhöheschwankungen im Rahmen der Stimmfunktionsprüfung'' (''On the quantitative evaluation of fluctuations of the tone height in the framework of the functional examination of voice''), Folia Phoniatr. **31**, 56–69.
Stearns, S. D., and David, R. A. (**1993**). *Signal Processing Algorithms in Fortran and C* (Prentice–Hall, Englewood Cliffs, NJ), pp. 162–181.
Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT, Cambridge, MA), pp. 55–126.
Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice–Hall, Englewood Cliffs, NJ).
Titze, I. R. (**1995a**). ''Definitions and nomenclature related to voice quality,'' in *Vocal Fold Physiology*, edited by O. Fujimura and M. Hirano (Singular, San Diego), pp. 335–342.
Titze, I. R. (**1995b**). ''Motor and sensory components of a feedback control model of fundamental frequency,'' in *Producing Speech: Contemporary Issues*, edited by F. Bell-Berti and L. J. Raphael (AIP, New York), pp. 309–320.
Titze, I. R., Horii, Y., and Scherer, R. C. (**1987**). ''Some technical considerations in voice perturbation measurements,'' J. Speech Hear. Res. **30**, 252–260.
Warwick, P., Dromey, C., Irish, J., and Durkin, L. (**2000**). ''The Treatment of essential voice tremor with Botulinum toxin A: a longitudinal case report,'' J. Voice **14**, 410–421.
Winholtz, W. S., and Ramig, L. O. (**1992**). ''Vocal tremor analysis with the vocal demodulator,'' J. Speech Hear. Res. **35**, 562–573.
Yair, E., and Gath, I. (**1988**). ''On the use of pitch power spectrum in the evaluation of vocal tremor,'' Proc. IEEE **76**, 1166–1175.

Normalized amplitude quotient for parametrization of the glottal flow[a)]

Paavo Alku[b)] and Tom Bäckström
Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, Box 3000, Fin-02015 TKK, Finland

Erkki Vilkman
University of Oulu, Department of Otolaryngology and Phoniatrics, Oulu, Finland and Helsinki University Central Hospital, Box 220, Fin-00029 HUS, Finland

(Received 23 October 2001; accepted for publication 2 May 2002)

Normalized amplitude quotient (NAQ) is presented as a method to parametrize the glottal closing phase using two amplitude-domain measurements from waveforms estimated by inverse filtering. In this technique, the ratio between the amplitude of the ac flow and the negative peak amplitude of the flow derivative is first computed using the concept of equivalent rectangular pulse, a hypothetical signal located at the instant of the main excitation of the vocal tract. This ratio is then normalized with respect to the length of the fundamental period. Comparison between NAQ and its counterpart among the conventional time-domain parameters, the closing quotient, shows that the proposed parameter is more robust against distortion such as measurement noise that make the extraction of conventional time-based parameters of the glottal flow problematic. Experiments with breathy, normal, and pressed vowels indicate that NAQ is also able to separate the type of phonation effectively. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1490365]

PACS numbers: 43.70.Jt, 43.70.Bk, 43.70.Gr, 43.72.Ar [AL]

I. INTRODUCTION

Inverse filtering provides a noninvasive method to estimate the excitation of voiced speech, the glottal volume velocity waveform. Inverse filtering studies typically involve two stages. In the first one, an estimate of the glottal flow waveform is computed either from the oral flow using the pneumotachographic mask (also called the Rothenberg's mask) (Rothenberg, 1973) or from the speech pressure waveform (e.g., Wong *et al.*, 1979). In the second stage, the obtained estimates of the glottal flow or its derivative are parametrized by expressing their most important features using few numerical values.

Parametrization of the voice source has been the target of intensive research during the past few decades. This, in turn, has resulted in a large variety of methods to quantify the waveforms given by inverse filtering. One of the most widely used approaches to parametrize the voice source is to apply time-based parameters. This corresponds to quantifying the glottal flow using certain quotients between the closed phase, the opening phase, and the closing phase of the glottal volume velocity waveform (e.g., Holmberg *et al.*, 1988). Time-based measures have also been computed using the first derivative of the glottal flow by applying, for example, the time difference between the beginning of the closing phase and the instant of the maximal negative peak (Sundberg *et al.*, 1993). If inverse filtering is based on recording the oral flow using the properly calibrated Rothenberg mask, it is possible to parametrize voice production by measuring absolute amplitude domain values (e.g., ac flow, minimum flow, negative peak amplitude of the differentiated flow) (Holmberg *et al.*, 1988; Hertegård *et al.*, 1992; Sundberg *et al.*, 1999). Moreover, methods have been developed to quantify the voice source in the frequency domain. These techniques are typically based on measuring the decay of the voice source spectrum either from the spectral harmonics (Howell and Williams, 1992; Childers and Lee, 1991) or from the pitch-synchronously computed spectrum (Alku *et al.*, 1997). Finally, one category of voice source parametrization methods is represented by techniques that fit certain predefined mathematical functions using, for example, the Liljencrants–Fant (LF) model (Fant *et al.*, 1985) to the glottal waveform obtained by inverse filtering. In these methods, quantification of the voice source corresponds to determining the optimal parameter values of the underlying mathematical functions so that the glottal waveform given by inverse filtering is matched by its synthetic model (Carlson *et al.*, 1989; Strik and Boves, 1992).

Among the quantification methods listed above, the use of time-based parameters is one of the most prevalent. The three most commonly used time-based parameters are: (1) open quotient (OQ), which is the ratio between the open phase of the glottal pulse and the length of the fundamental period; (2) speed quotient (SQ), which is the ratio between the glottal opening and closing phases; and (3) closing quotient (CQ), which is defined as the ratio between the glottal closing phase and the length of the fundamental period. The prevalence of these time-based parameters as a method to quantify the voice source comes from the fact that they can be defined without knowing the absolute flow values of the glottal volume velocity waveform. In other words, applying

[a)]Portions of this work were presented in "Normalized amplitude quotient for parameterization of the glottal flow," Proceedings of the 5th International Workshop, Advances in Quantitative Laryngoscopy, Voice and Speech Research, Groningen, The Netherlands, April 2001.

[b)]Electronic mail: paavo.alku@hut.fi

the time-based parameters does not require the use of the Rothenberg mask in the inverse filtering stage. Moreover, computation of the time-based parameters is, at least in principle, straightforward because their values can be determined from simple time-length measurements computed directly from the waveforms given by inverse filtering. However, application of the time-based parameters in practice has shown, unfortunately, that accurate computation of their values is problematic (Holmberg *et al.*, 1988; Dromey *et al.*, 1992). This is due to the fact that time instants of glottal opening and closure are sometimes difficult to extract exactly due to formant ripple and noise that is present in the glottal waveforms given by inverse filtering. Even in the absence of formant ripple, measuring OQ and SQ is difficult because of the gradual opening of the vocal folds. Due to these problems, computation of the time-based parameters is sometimes performed by replacing the true time instants of glottal opening and closure by the time instants when the glottal flow crosses a level which is set to a certain ratio (e.g., 50%) of the difference between the maximum and minimum amplitude of the glottal cycle (Dromey *et al.*, 1992).

Among the three time-based parameters mentioned above, the use of CQ is justified for two reasons. First, CQ constitutes a measure that is affected by changes of the glottal pulse during its closing phase. Closing phase, in turn, corresponds to the portion of the glottal cycle during which the main excitation of the vocal tract occurs (Fant, 1993). Therefore, the value of CQ reflects changes that occur in the glottal source when vocal intensity or phonation type is altered: the glottal closing phase typically decreases when intensity is increased or when phonation is changed from soft to pressed (Monsen and Engebretson, 1977; Alku and Vilkman, 1996a; Sulter and Wit, 1996). Second, computation of CQ does not require determining the time instant of glottal opening, which is typically much more difficult to extract accurately than the time instant of glottal closure. Hence, when parametrization of voice production is computed from glottal flows distorted by formant ripple or noise, CQ typically yields more robust results than OQ or SQ.

In studying quantification of the glottal flow with the time-based parameters, it has become evident that these are computed by measuring the *time lengths* between corresponding events (i.e., glottal opening and closure as well as the instant of the maximal flow). However, it is also possible to measure time-domain features of the glottal closing phase using the *amplitude-domain* values extracted from the glottal flow and its first derivative. This is based on the voice source parametrization schemes developed independently and in parallel by Fant (Fant and Lin, 1988; Fant *et al.*, 1994; Fant, 1995, 1997) and Alku and Vilkman (1996a, 1996b). In these studies, the application of the ratio between the amplitude of the ac flow and the negative peak amplitude of the flow derivative has been analyzed in parametrization of the glottal source. This ratio was shown by Fant and his co-authors to yield ''a measure of effective decay time of the glottal flow pulse'' and it was used as a method to reduce the number of the LF parameters in modeling of the glottal source (Fant *et al.*, 1994; Fant, 1995, 1997).

Based on these previous studies, the current survey introduces a time-domain voice source parameter, the normalized amplitude quotient (NAQ), which is closely related to CQ. The goal of the study is to analyze, first, whether NAQ provides a more robust method for parametrizing the time-domain features of the glottal flow than CQ. Second, our aim is to analyze how NAQ behaves in the parametrization of glottal flows of different phonation types.

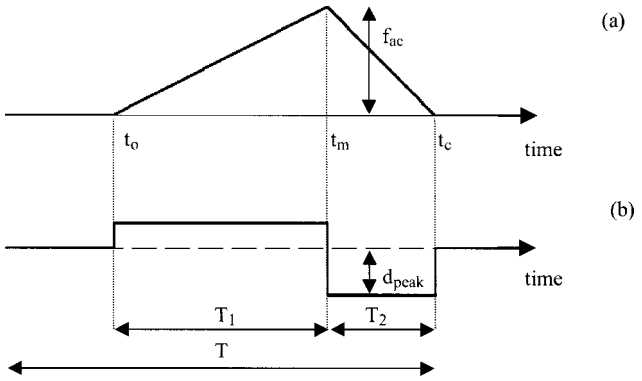


FIG. 1. A triangular-shaped glottal flow pulse (a) and its first derivative (b). Amplitude values shown in the graphs: ac flow (f_{ac}), negative peak amplitude of the differentiated flow (d_{peak}). Time values shown in the graphs: lengths of the fundamental period (T), glottal opening phase (T_1), and the glottal closing phase (T_2), instants of glottal opening (t_o), maximal glottal flow (t_m), and glottal closure (t_c).

II. MATERIALS AND METHODS

A. Normalized amplitude quotient

In order to derive the normalized amplitude quotient, let us start from a simplified model of the glottal flow represented by a triangular pulse during the glottal open phase and a zero flow during the closed phase [Fig. 1(a)]. The length of the opening phase, the closing phase, and the fundamental period is denoted by T_1, T_2, and T, respectively. The only amplitude domain value needed to define this simplified glottal pulse is the maximum value of the flow, which is denoted by f_{ac}. The first derivative of the triangular-shaped glottal pulse is given by two rectangular pulses shown in Fig. 1(b). The first of these pulses is positive and it lies between the time instant of glottal opening (t_o) and the instant of the maximum flow (t_m). The second rectangular pulse is negative; it starts at t_m and ends at the instant of glottal closure (t_c). It is worth noticing that the areas of both of the two rectangular pulses are equal to f_{ac}. This derives from the fact that the integral of the signal shown in Fig. 1(b) is the triangular pulse shown in Fig. 1(a), which starts from the zero level at glottal opening, reaches its maximum value f_{ac} at time instant t_m, and returns to the zero level at glottal closure. Due to the rectangular shape of the flow derivative, the following equation holds true during the glottal closing phase:

$$A_2 = d_{peak} \cdot T_2 = f_{ac} \rightarrow T_2 = \frac{f_{ac}}{d_{peak}} = \mathrm{AQ}, \tag{1}$$

where A_2 denotes the area of the rectangular pulse between t_m and t_c in Fig. 1(b) and AQ denotes the ratio between f_{ac} and d_{peak}. Equation (1) yields, for the simplified triangular

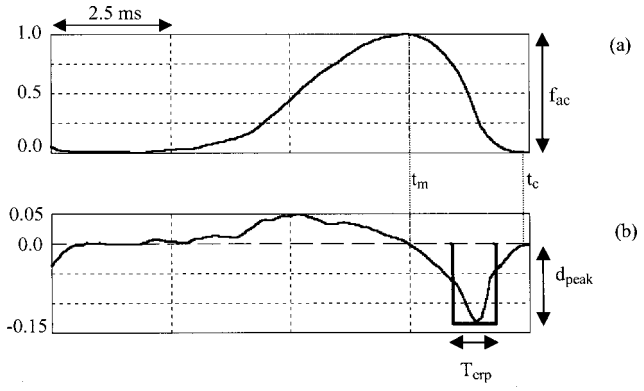


FIG. 2. A glottal flow pulse (a) and its first derivative (b) computed from natural speech by inverse filtering. Amplitude values shown in the graphs: ac flow (f_{ac}), negative peak amplitude of the differentiated flow (d_{peak}). Time values shown in the graphs: length of the equivalent rectangular pulse (T_{erp}), instants of maximal glottal flow (t_m), and glottal closure (t_c).

glottal pulse, the exact time length of the glottal closing phase as a ratio of two amplitude values, the first of which is the flow maximum and the second of which is the negative peak amplitude of the flow derivative. The value of the closing quotient can now be obtained for the triangular glottal pulse using Eq. (1) as follows:

$$\mathrm{CQ}=\frac{T_2}{T}=\frac{f_{\mathrm{ac}}}{d_{\mathrm{peak}}\cdot T}=\frac{\mathrm{AQ}}{T}. \tag{2}$$

Obviously, glottal pulses of a triangular shape do not exist in real human speech production. This can be demonstrated by comparing the glottal flow [Fig. 2(a)] of a natural vowel, obtained by inverse filtering, to its triangular-shaped artificial counterpart [Fig. 1(a)]. In particular, the shape of the differentiated glottal flow computed from the real utterance [Fig. 2(b)] is considerably different from its rectangular-shaped counterpart shown in Fig. 1(b). However, it is still possible to use the amplitude-based quotient presented above to obtain a CQ-related measure that needs no extraction of the time instant of glottal closure. Computation of the ratio $f_{\mathrm{ac}}/d_{\mathrm{peak}}$, which yields an exact time length of the glottal closing phase only for the triangular flow pulse, can be considered in the case of a natural glottal pulse as follows [see Fig. 2(b)]. An *equivalent rectangular pulse*, the height of which equals d_{peak}, is set at the instant of the negative peak of the flow derivative. The time length of this pulse is initially infinitesimal. The length of the rectangular pulse is increased until its area becomes equal to f_{ac}. (Similar to the case of the triangular glottal pulse, the value of f_{ac} equals the area computed from the flow derivative between time instants of the maximal flow and glottal closure for the natural glottal pulse as well.) When the area of the equivalent rectangular pulse is equal to f_{ac}, the length of the pulse is denoted by T_{erp} as shown in Fig. 2(b). In this case, the following equation holds:[1]

$$A_{\mathrm{erp}}=T_{\mathrm{erp}}\cdot d_{\mathrm{peak}}=f_{\mathrm{ac}}\rightarrow T_{\mathrm{erp}}=\frac{f_{\mathrm{ac}}}{d_{\mathrm{peak}}}=\mathrm{AQ}, \tag{3}$$

where A_{erp} denotes the area of the equivalent rectangular pulse and AQ denotes the ratio between f_{ac} and d_{peak}. Hence, in the case of the natural glottal pulse, computation

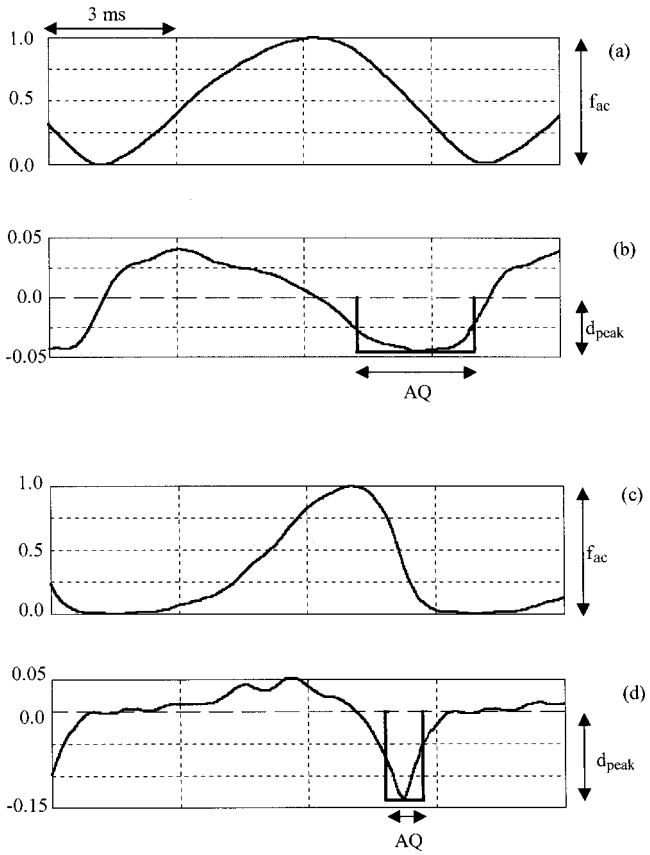


FIG. 3. Glottal flow (a) and its first derivative (b) estimated by inverse filtering in breathy phonation, glottal flow (c) and its first derivative (d) estimated by inverse filtering in pressed phonation. Amplitude values shown in the graphs: ac flow (f_{ac}), negative peak amplitude of the differentiated flow (d_{peak}). According to Eq. (3), the length of the equivalent rectangular pulse equals AQ, i.e., the ratio between f_{ac} and d_{peak}, shown in the graphs by arrows.

of the amplitude ratio $f_{\mathrm{ac}}/d_{\mathrm{peak}}$ corresponds to adjusting the time length of a rectangular pulse, whose height equals d_{peak} and whose center point[2] is located at the instant of the negative peak of the flow derivative, until the area of the pulse becomes equal to f_{ac}.

Equation (3) yields an amplitude-domain quotient to quantify the closing phase of the glottal flow that differs in two ways from measurements used in the conventional time-based parameters of the glottal flow. First, the value of AQ does not require extraction of the time instant of glottal closure. Second, AQ yields a time length, which does not correspond simply to the entire length of the glottal closing phase; rather, it represents a measure that reflects characteristics of the flow derivative in the vicinity of its negative peak. This is demonstrated in Fig. 3, which shows how the length of the equivalent rectangular pulse, i.e., AQ defined in Eq. (3), is large when the glottal flow derivative is smooth [Fig. 3(a)], which occurs in breathy phonation. However, the time length of the equivalent rectangular pulse is small in the case of a rapidly fluctuating derivative [Fig. 3(b)] corresponding to the pressed phonation type. It is worth noticing that Eq. (3) yields a value that equals the true length of the glottal closing phase only in the case of the triangular flow pulse. For natural glottal flows, however, the value given by

Eq. (3) is always smaller than the length of the glottal closing phase. The equivalent rectangular pulse has been applied previously in different forms in other fields of science. In telecommunications, frequency characteristics of filters are quantified using the concept of noise-equivalent bandwidth (Carlson, 1986). This corresponds to defining an ideal rectangular pulse for a given filter in the frequency domain that would pass as much white-noise power as the filter in question. A similar approach has also been used in psychoacoustics in quantifying auditory filters using the approach of the equivalent rectangular bandwidth (ERB) (Moore, 1982).

The time-length measure given in Eq. (3) can be normalized with respect to the length of the fundamental period in a similar manner as is done in the computation of CQ. This finally yields the following equation for normalized amplitude quotient (NAQ):

$$\mathrm{NAQ}=\frac{\mathrm{AQ}}{T}=\frac{f_{\mathrm{ac}}}{d_{\mathrm{peak}}\cdot T}. \tag{4}$$

B. Speech material

The performance of the proposed voice source parametrization method was evaluated by collecting speech data from five female and five male speakers. None of the subjects had a history of voice or hearing disorders. The voices were also perceptually within normal limits as judged by a phoniatrician. The age of the subjects varied between 29 and 52 years for females and between 32 and 47 years for males. The speakers were asked to produce a sustained /a/ vowel using breathy, normal, and pressed phonation types. The pitch was kept as constant as possible throughout the recording. The length of the pronunciation was 2 s. Subjects were allowed to use their natural fundamental frequency and intensity level during the recording. All the speakers were first trained to produce the vowel with the three different phonation types by mimicking a qualified instructor. During the recording, voice quality was assessed by a phoniatrician who asked the subject to repeat the speaking task until phonation was satisfactory. Recording of the utterances was performed in an anechoic chamber using a condenser microphone (Brüel & Kjær 4133 together with preamplifier Brüel & Kjær 2636), held 40 cm from the lips of the speaker. Speech data were saved onto a digital tape (DAT recorder TEAC RD-200T) using a sampling frequency of 22 050 Hz and a resolution of 16 bits.

After recording the speech signals an informal listening test was made in order to verify that the voices belonged to the three different phonation types. Three phonation types of each speaker were played to a panel in random order. Each panelist was asked to mark the order of the phonation type. The panel consisted of five members, all experienced in voice research. The experiment proved conclusive: all the members of the panel sorted the phonation types of each speaker correctly. The judgments of the panel were thus in line with those of the phoniatrician. Hence, we can conclude that all the subjects succeeded in producing the phonation types as required.

C. Inverse filtering

The glottal volume velocity waveforms were estimated using an inverse filtering technique that is described in detail in Alku and Vilkman (1994). This inverse filtering technique applies the acoustic speech pressure waveform that has been recorded in a free field for estimation of the voice source, i.e., no flow mask is required. The method developed is based on modeling the vocal tract transfer function with an all-pole filter which is determined using a sophisticated algorithm called discrete all-pole modeling (DAP) (El-Jaroudi and Makhoul, 1991). In comparison to the conventional linear predictive coding (LPC), which is usually applied in automatic inverse filtering, the DAP technique has been shown to yield more accurate estimates of formants, especially for high-pitched voices (El-Jaroudi and Makhoul, 1991). Hence, the estimated glottal airflow waveforms are less distorted by formant ripples (Alku and Vilkman, 1994).

The sampling frequency of the signals was first decreased from the original value of 22.050 kHz to 8.0 kHz. In order to avoid aliasing, all the signals were low-pass filtered before the downsampling with a linear phase FIR filter, that had its cutoff frequency at 4.0 kHz. Signals were then high-pass filtered with a linear phase FIR filter using a cutoff frequency of 50.0 Hz in order to remove any possible low-frequency air-pressure variations picked up during the recordings. Inverse filtering was computed by modeling the vocal tract transfer function with an all-pole filter, the order[3] of which was varied between 8 and 12. Glottal flows were estimated using a block length of 32 ms together with Hamming windowing. The position of the analysis window was initially set to the middle of the utterance. However, if the estimated glottal waveform showed evidence of formant ripple (i.e., there was an oscillating component present in the closed phase), the position of the analysis window was varied in order to find a setting that yielded a waveform with a smaller amplitude of such distortion.

III. RESULTS

The behavior of NAQ was analyzed in two parts. In the first part, robustness of the parameter was tested and compared to that of CQ. In this experiment, problematic conditions in extracting data values from glottal flows were simulated by degrading the waveforms given by inverse filtering with additive noise. The goal of the second part was to find out what kind of values NAQ typically yields when parametrizing glottal flows of three different phonation types.

A. Robustness of NAQ

One of the most important reasons behind the development of NAQ was the fact that extraction of time-based measures from glottal waveforms estimated by inverse filtering is often problematic due to noise and formant ripple that is present in the waveforms computed from natural speech. In order to evaluate the performance of NAQ in problematic conditions, we generated test material using the glottal flows inverse filtered from the voices of the ten speakers. A single cycle was cut from each glottal flow waveform. By concatenating ten such cycles, a pulse form with no jitter in the

length of the fundamental period was computed for each speaker and each phonation type, yielding a total of 30 regular glottal pulse forms. [In real speech, even when this is produced using sustained phonation, the characteristics of a single glottal pulse vary between consecutive cycles. Therefore, it is common practice to average measurements extracted from a single cycle over four to six periods (e.g., Holmberg *et al.*, 1988). However, in the current experiment a larger number of cycles was acquired for statistical analyses and, hence, ten periods were used in the averaging. The glottal excitation was synthesized by concatenating identical single pulses so that the parametrization showed no cycle-to-cycle variation.] Glottal pulse forms were then distorted by additive noise (zero mean, Gaussian distribution) in order to simulate suboptimal conditions in the parametrization of the waveforms. The amount of noise was quantified by measuring the signal-to-noise ratio (SNR) from the degraded glottal pulse forms. Eight different noise conditions were created corresponding to the following SNR values: infinity, i.e., no noise added, 60, 55, 50, 45, 40, 35, and 30 dB. Hence, the total number of glottal waveforms used in this first experiment was 240 (10 speakers, 3 phonation types, and 8 SNR categories).

Both CQ and NAQ were determined automatically for each individual glottal cycle using the following procedure. The time instant of the negative peak (t_{peak}) of the flow derivative was first identified during the glottal cycle. The length of the glottal closing phase was then determined as the sum of time spans of consecutive negative samples of the derivative before and after t_{peak}. The value of CQ was obtained as the ratio between this closing phase and the length of the fundamental period. In computing the value of NAQ using Eq. (4), f_{ac} was determined as the largest ac-flow value during the fundamental period. The value of d_{peak} was obtained by taking the amplitude of the derivative at t_{peak}. Finally, CQ and NAQ extracted from individual glottal cycles were averaged over the ten periods for each of the 240 glottal pulse forms. In order to analyze cycle-to-cycle variation in the quotient values extracted in noisy conditions, we also computed standard deviation of both CQ and NAQ over the ten glottal cycles in all 240 cases.

Two examples describing behavior of CQ and NAQ as a function of SNR are shown in Fig. 4 and Fig. 5 based on single-subject data from a female and a male subject, respectively. Both results were obtained from speech samples produced in normal phonation. Two characteristic differences between the quotients can be observed from the examples shown. First, the value of NAQ is clearly smaller (approximately 0.15 when SNR$=\infty$) than that of CQ (approximately 0.30 when SNR$=\infty$). Second, extraction of NAQ is more robust against noise, because its value changes less when SNR decreases.

In order to analyze the behavior of CQ and NAQ from the voices of all ten subjects, we used the following statistical measurements. First, relative changes of CQ, denoted by r_{CQ}, and NAQ, denoted by r_{NAQ}, between the clean conditions and the noisy conditions were computed as follows:

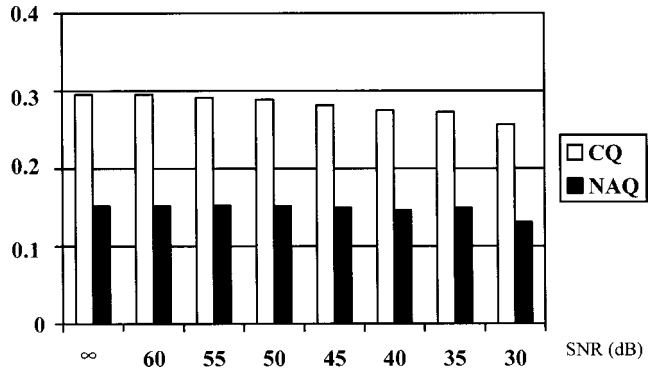


FIG. 4. Closing quotient (CQ, white bars) and normalized amplitude quotient (NAQ, black bars) as a function of signal-to-noise-ratio (SNR). The first value on the *x* axis (SNR$=\infty$) corresponds to the case in which no noise was added to the glottal flow. Female speaker, normal phonation.

$$r_{\text{CQ}} = 100\% \cdot \frac{|\text{CQ}_{\text{SNR}} - \text{CQ}_{\infty}|}{\text{CQ}_{\infty}}, \tag{5}$$

$$r_{\text{NAQ}} = 100\% \cdot \frac{|\text{NAQ}_{\text{SNR}} - \text{NAQ}_{\infty}|}{\text{NAQ}_{\infty}}, \tag{6}$$

where CQ_{SNR} and NAQ_{SNR} denote the closing quotient and the normalized amplitude quotient, respectively, computed with a finite SNR value, i.e., SNR$=$60,55,...,30 dB, and CQ_{∞} and NAQ_{∞} denote the closing quotient and the normalized amplitude quotient, respectively, computed in clean conditions, i.e., SNR$=\infty$. Second, in order to analyze how cycle-to-cycle variation of the quotients varies as a function of SNR, we computed the coefficient of variation (Wilks, 1962) for CQ, denoted by μ_{CQ}, and for NAQ, denoted by μ_{NAQ}, in noisy conditions as follows:

$$\mu_{\text{CQ}} = 100\% \cdot \frac{\text{s.d.}_{\text{CQ,SNR}}}{m_{\text{CQ,SNR}}}, \tag{7}$$

$$\mu_{\text{NAQ}} = 100\% \cdot \frac{\text{s.d.}_{\text{NAQ,SNR}}}{m_{\text{NAQ,SNR}}}, \tag{8}$$

where $\text{s.d.}_{\text{CQ,SNR}}$ and $m_{\text{CQ,SNR}}$ denote standard deviation and mean, respectively, of the closing quotient, and $\text{s.d.}_{\text{CQ,SNR}}$ and $m_{\text{CQ,SNR}}$ denote standard deviation and mean, respec-

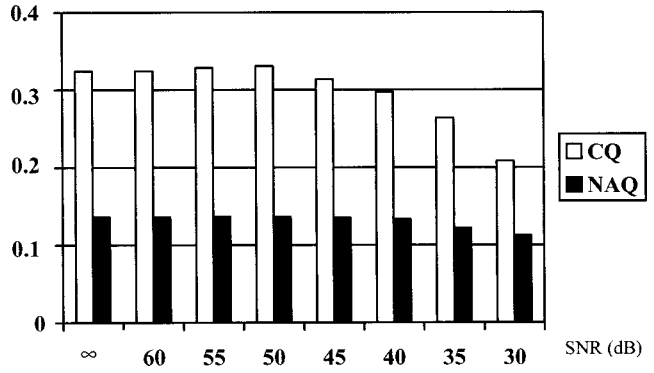


FIG. 5. Closing quotient (CQ, white bars) and normalized amplitude quotient (NAQ, black bars) as a function of signal-to-noise-ratio (SNR). The first value on the *x* axis (SNR$=\infty$) corresponds to the case in which no noise was added to the glottal flow. Male speaker, normal phonation.

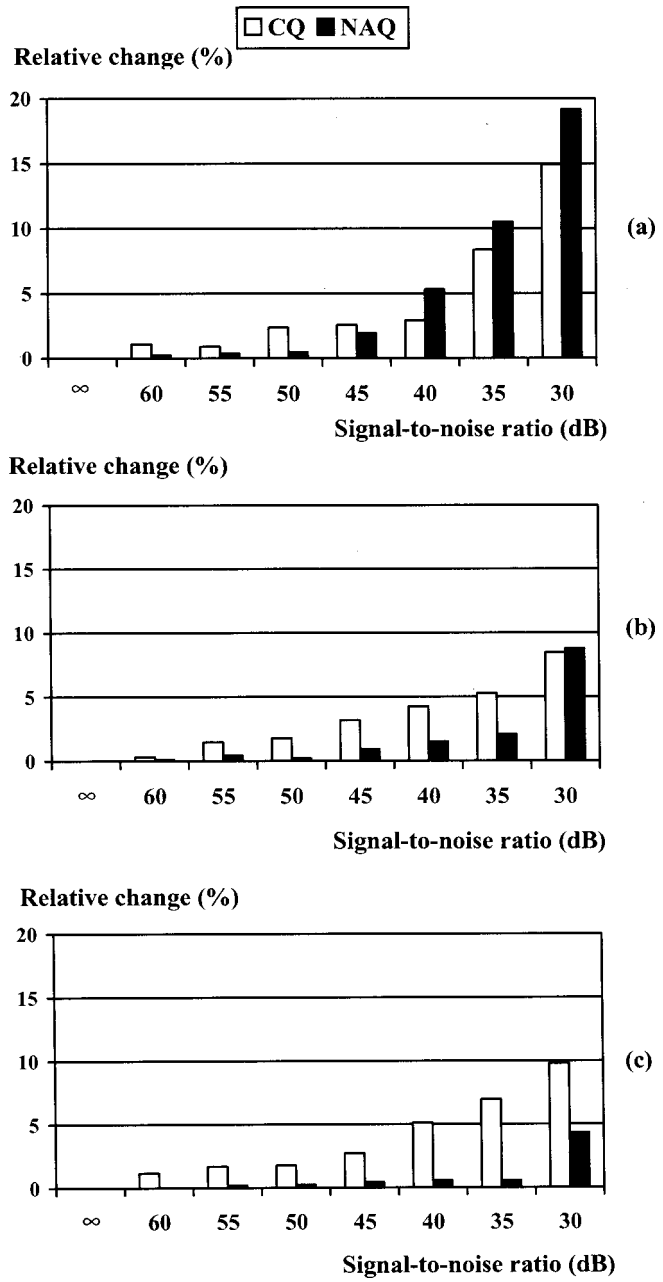


FIG. 6. Relative change of closing quotient (CQ, white bars) and normalized amplitude quotient (NAQ, black bars) as a function of signal-to-noise-ratio (SNR). The first value on the x axis (SNR$=\infty$) corresponds to the case in which no noise was added to the glottal flow. Relative change was determined for CQ and NAQ using Eq. (5) and Eq. (6), respectively. Female speakers ($n=5$), phonation type: breathy (a); normal (b); pressed (c).

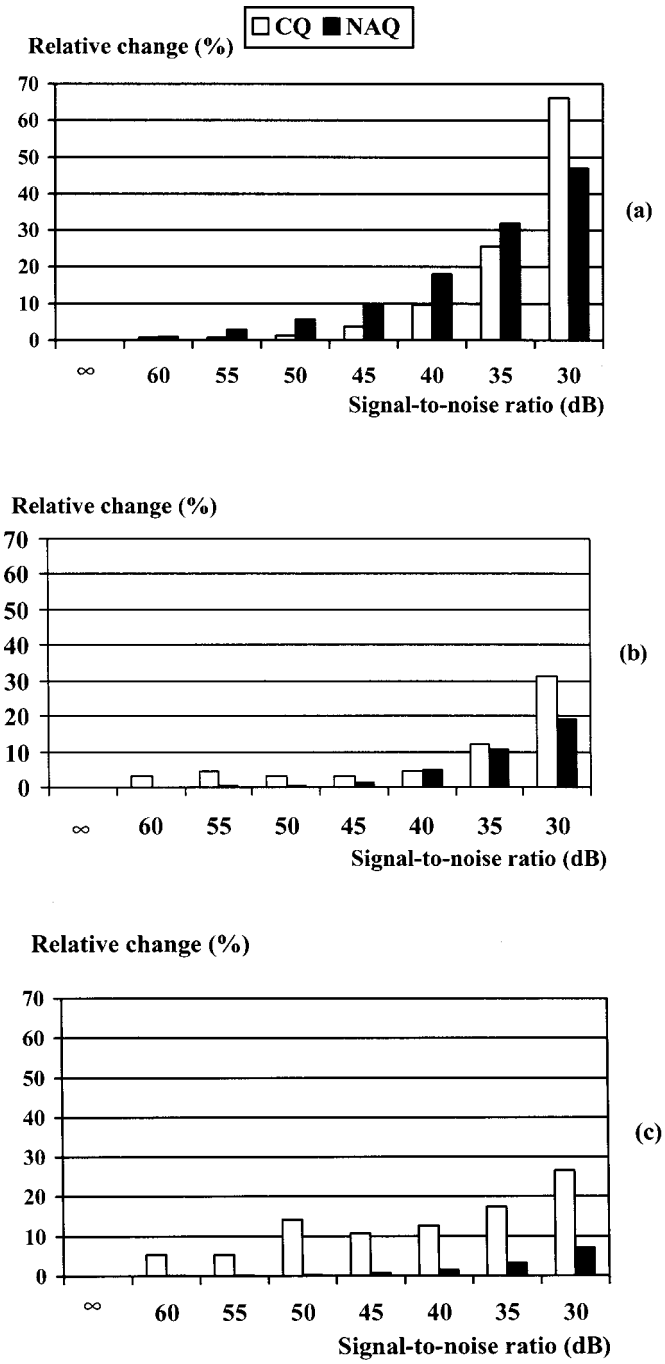


FIG. 7. Relative change of closing quotient (CQ, white bars) and normalized amplitude quotient (NAQ, black bars) as a function of signal-to-noise-ratio (SNR). The first value on the x axis (SNR$=\infty$) corresponds to the case in which no noise was added to the glottal flow. Relative change was determined for CQ and NAQ using Eq. (5) and Eq. (6), respectively. Male speakers ($n=5$), phonation type: breathy (a); normal (b); pressed (c).

tively, of the normalized amplitude quotient when parametrization was computed over ten glottal periods using a finite SNR value, i.e., SNR$=60,55,...,30$ dB.

The relative change of CQ and NAQ between the clean and noisy conditions is shown in Fig. 6 and Fig. 7 for female and male subjects, respectively. From these pictures it can be seen that r_{NAQ} is less than r_{CQ} in speech samples produced in normal and pressed phonation for all the values of SNR. (The sole exception is SNR value 30 dB in normal phonation of female subjects where r_{NAQ} was slightly larger than r_{CQ}.) This implies that NAQ changed less than CQ when extraction of the parameter has been distorted by adding noise to the glottal pulse form. In the case of breathy phonation, though, the result was different: the change of NAQ due to added noise was larger than that of CQ for female voices in three SNR values (40, 35, and 30 dB) and for male voices in six SNR values (60, 55, 50, 45, 40, and 35 dB).

Coefficient of variation is shown as a function of SNR in Fig. 8 and Fig. 9 for female and male subjects, respectively. Since the glottal waveforms to be parametrized were constructed by concatenating ten identical periods, it is natural that the cycle-to-cycle variation is zero for both CQ and

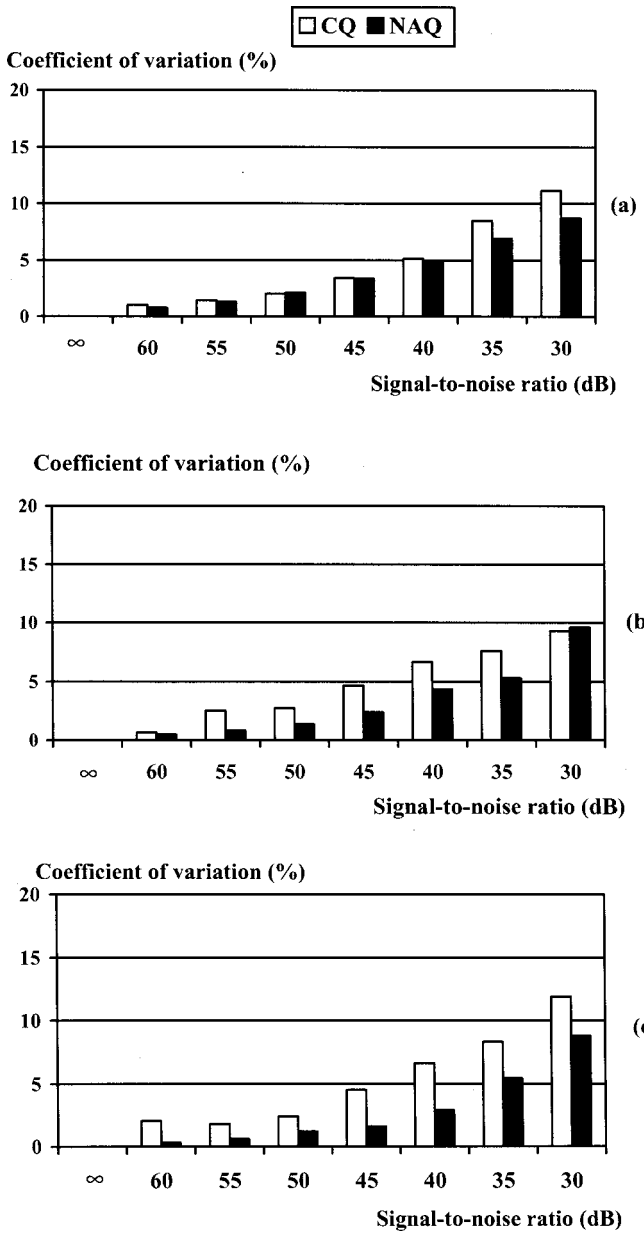


FIG. 8. Coefficient of variation for closing quotient (CQ, white bars) and normalized amplitude quotient (NAQ, black bars) as a function of signal-to-noise-ratio (SNR). The first value on the x axis (SNR$=\infty$) corresponds to the case in which no noise was added to the glottal flow. Coefficient of variation was determined for CQ and NAQ using Eq. (7) and Eq. (8), respectively. Female speakers ($n=5$), phonation type: breathy (a); normal (b); pressed (c).

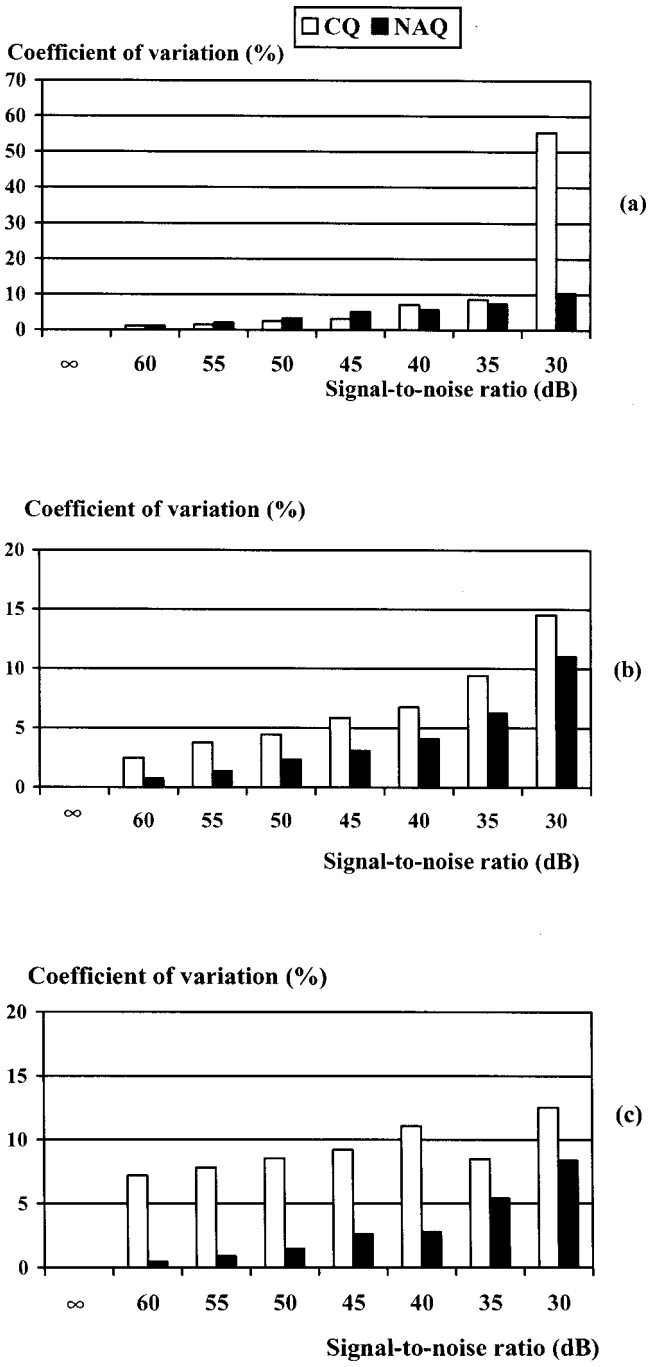


FIG. 9. Coefficient of variation for closing quotient (CQ, white bars) and normalized amplitude quotient (NAQ, black bars) as a function of signal-to-noise-ratio (SNR). The first value on the x axis (SNR$=\infty$) corresponds to the case in which no noise was added to the glottal flow. Coefficient of variation was determined for CQ and NAQ using Eq. (7) and Eq. (8), respectively. Male speakers ($n=5$), phonation type: breathy (a); normal (b); pressed (c).

NAQ when SNR equals infinity. However, when noise is added, both parameters show increased cycle-to-cycle variation. Coefficient of variation is below 10% for both parameters in almost all cases. The only major exception is CQ of breathy phonation for male voices [Fig. 9(a)] with SNR equal to 30 dB. In this case, the large value of μ_{CQ} was caused by occasional positive values of the flow derivative during the glottal closing phase due to the large amount of added noise. This, in turn, caused the extraction of the closing phase to yield values of large variation.

It can be seen that coefficient of variation is clearly larger for CQ than for NAQ. In the case of normal and pressed phonation, the value of μ_{CQ} was larger in all the SNR categories for both genders. (The sole exception is again SNR value 30 dB in normal phonation of females where μ_{NAQ} was slightly larger than μ_{CQ}.) Only in breathy phonation did μ_{NAQ} yield larger values than μ_{CQ}. This occurred in only one SNR category (SNR equal to 50 dB) for females and in three SNR categories (SNR equal to 55, 50, and 45 dB) for males.

Analysis of variance (ANOVA) of the mathematical package MATLAB (Mathworks Inc.) was also used to investi-

TABLE I. Mean, standard deviation and range of closing quotient (CQ) and normalized amplitude quotient (NAQ) in the three phonation types. Female speakers ($n=5$).

Parameter	Phonation type	Mean	Standard deviation	Range
CQ	Breathy	0.40	0.065	0.29–0.48
	Normal	0.29	0.037	0.26–0.36
	Pressed	0.26	0.036	0.22–0.32
NAQ	Breathy	0.22	0.039	0.15–0.27
	Normal	0.15	0.016	0.13–0.16
	Pressed	0.12	0.020	0.10–0.15

gate the effects of different factors on CQ and NAQ. First, one-way ANOVA with noise category as a factor was computed. Results showed a statistically significant influence of SNR on CQ [$F(7,232)=5.81$, $p<0.05$] but not on NAQ. Second, a three-way analysis of variance with the factors noise category, phonation type and gender was computed by taking into account noise categories with SNR$\geqslant$40, where effects of noise did not affect either of the two parameters too seriously. It was found that both CQ [$F(2,171)=300.8$, $p<0.05$] and NAQ [$F(2,171)=361.3$, $p<0.05$] were highly dependent statistically on the phonation type. However, there was no effect of gender either on CQ [$F(1,171)=0.88$, $p>0.05$] or NAQ [$F(1,171)=0.000\,17$, $p>0.05$]. In summary, these statistical analyses on noise-corrupted waveforms indicate that both of the parameters are able to categorize the type of phonation effectively, but NAQ is less vulnerable to distortion of the glottal pulse than CQ.

B. Behavior of NAQ in parametrization of glottal flows of different phonation types

The aim of the second part of our experiments was to analyze how NAQ behaves when it is used in parametrization of glottal flows representing three different phonation types (breathy, normal, and pressed). The data material of this part comprised all 30 glottal waveforms (10 speakers, 3 phonation types) obtained by inverse filtering. The glottal waveforms and their first derivatives were analyzed using the same extraction procedure described in Sec. III A. The values required for computation of CQ and NAQ were averaged over four consecutive glottal periods.

The results are shown in Table I and Table II for female and male voices, respectively. These data show that the mean value of NAQ decrease for both genders when phonation is changed along with the axis breathy–normal–pressed. It

TABLE II. Mean, standard deviation and range of closing quotient (CQ) and normalized amplitude quotient (NAQ) in the three phonation types. Male speakers ($n=5$).

Parameter	Phonation type	Mean	Standard deviation	Range
CQ	Breathy	0.45	0.046	0.38–0.51
	Normal	0.27	0.022	0.24–0.30
	Pressed	0.22	0.031	0.18–0.25
NAQ	Breathy	0.28	0.044	0.23–0.35
	Normal	0.13	0.022	0.11–0.17
	Pressed	0.09	0.011	0.08–0.11

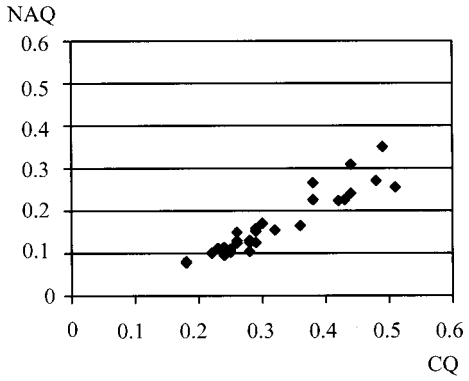


FIG. 10. Closing quotient (CQ) and normalized amplitude quotient (NAQ) expressed in the same plane for all the analyzed speech samples (10 speakers, 3 phonation types).

should be emphasized that this monotonic decrease of the parameter value occurred for all ten subjects analyzed. Statistical analyses (t-test, 95%) revealed that for male subjects the value of NAQ was significantly affected by the phonation type. The value of CQ, however, did not show a statistically significant difference between the normal and pressed male voices. For female subjects, both CQ and NAQ yielded a statistically significant difference between the breathy and normal voices, while neither of the two parameters showed a significant difference between samples produced using normal and pressed phonation.

In summary, all 30 analyzed voices are expressed in the same coordinate in Fig. 10. This figure indicates that correlation between CQ and NAQ is high (correlation coefficient equaled 0.94), even though voices from different phonation types and from both genders are pooled together. This result indicates that in the time-domain parametrization of the glottal closing phase the conventional CQ parameter can be replaced with NAQ, even though the analysis involves speech sounds of greatly different voice source characteristics. This, in turn, implies that the time-domain parametrization of the glottal closing phase can be improved by making it more straightforward and less vulnerable to distortion such as measurement noise.

IV. CONCLUSIONS

This study presents the normalized amplitude quotient (NAQ), which makes possible quantifying time-based features of the voice source from amplitude-domain measurements extracted from the glottal flow and its first derivative. The parameter is based on expressing the ratio between the amplitude of the ac flow and the negative peak amplitude of the flow derivative using the concept of equivalent rectangular pulse, a hypothetical signal with its center point located at the instant of the main excitation of the vocal tract. The ratio between the two amplitude values corresponds to the time length of the equivalent rectangular pulse when the area of the pulse is equal to the ac flow of the glottal pulse. The value of NAQ is determined by normalizing the time length of the equivalent rectangular pulse with respect to the duration of the fundamental period.

NAQ is a sequel to previous studies addressing parametrization of the glottal flow based on amplitude domain measurements (Fant and Lin, 1988; Fant *et al.*, 1994; Fant, 1995, 1997; Alku and Vilkman, 1996a, 1996b). The latter studies presents AQ as a straightforward method to parametrize glottal flows of different phonation types from two amplitude domain measures. Significantly, their studies did not discuss the fact that AQ, even though extracted from two amplitude values, yields a time-domain quantity. Moreover, there was no normalization of AQ with respect to the length of the fundamental period. Consequently, AQ values presented in Alku and Vilkman (1996a, 1996b) were different for female and male speakers. Obtaining a time-domain measure by computing the ratio between f_{ac} and d_{peak} was first presented in Fant *et al.* (1994). In their study, the ratio between the two amplitude values is called the effective declination time and it is geometrically interpreted as ''the projection on the time axis of a tangent to glottal flow at the point of excitation, limited by ordinate values of 0 and f_{ac}'' (Fant, 1997). In Fant (1995, 1997), the effective declination time is normalized by multiplying it by $F0/110$ (i.e., fundamental frequency, $F0$, of the voice divided by the approximated average fundamental frequency, 110 Hz, typical in male speech). This normalization yields a voice source parameter, denoted by R_d, that equals NAQ divided by 110.

Even though there is a close relationship between NAQ of the current study and R_d discussed in Fant (1995, 1997), the concept of the equivalent rectangular pulse is new in the parametrization of the glottal source. This concept constitutes an alternative to Fant's geometrical interpretation of the ratio between the amplitude of the ac flow and the negative peak amplitude of the flow derivative. Our new interpretation is motivated, first, because the equivalent rectangular pulse makes it easier to understand and visualize the role of the quotient between the two amplitude values rather than to deal with the ''projection on the time axis of a tangent to the glottal flow.'' Second, and more importantly, the computation of NAQ using the current approach allows for a straightforward comparison of two glottal flow pulses in terms of their time-domain features during the closing phase: the glottal pulses are transformed into the *same* simple functions (i.e., rectangular pulses) and the comparison is performed between these waveforms. Transforming waveforms of different shapes into rectangular functions has proven to be useful, for example, in telecommunications in comparing frequency characteristics of filters using noise-equivalent bandwidths (Carlson, 1986). Therefore, to make this idea known also in the time-domain parametrization of the glottal waveform is justified. Finally, we would like to point out that previous studies on R_d have not addressed widely the classification of the phonation type.

In computation of NAQ, the length of the equivalent rectangular pulse is determined using the maximum amplitude values of the flow and its derivative during one glottal cycle without requiring the extraction of the time instant of glottal closure. Therefore, computation of NAQ is straightforward, even though glottal flow waveforms contain distortion such as measurement noise that make the extraction of glottal closure complicated. Difficult conditions in data extraction were simulated in the present study by degrading glottal flows given by inverse filtering with additive Gaussian noise. The comparison between NAQ and its conventional counterpart, closing quotient (CQ), showed that the proposed new technique yielded a more robust parametrization method than CQ in normal, but especially in pressed phonation. However, in breathy phonation, in particular for male voices, the value of CQ changed less than NAQ when noise-distorted pulse forms were compared to the clean ones. (From the point of view of NAQ this is regrettable because time instants of breathy phonation are typically the most difficult to extract reliably and, hence, their extraction would require improved parametrization methods.) The reason why NAQ is more robust against noise than CQ in normal and pressed phonation types but not so much in breathy phonation is explained by the shape of glottal flow derivative at the instant (t_{peak}) of its negative peak. In normal and in pressed phonation, the derivative curve at t_{peak} is very sharp and of a large amplitude, which implies that the value of d_{peak} changes little, even though the flow waveform is somewhat distorted by noise. However, in breathy phonation the waveform of the flow derivative at d_{peak} is smooth and the value d_{peak} is low, which makes it more vulnerable to the effects of noise.

Since NAQ is computed using two amplitude values, both of which are extracted at a single time instant, it is possible for the accuracy of the quotient to deteriorate when the glottal flows are severely affected by noise. Distortion of the flow signal even at a single time instant during the glottal closing phase might, in the worst case, cause a large error in the value of the negative peak of the derivative. (This results typically in an increased value of d_{peak}, which in turn reduces the value of NAQ.) To alleviate the distortion caused by instantaneous noisy peaks of the flow derivative, it could be possible to use a procedure where a glottal flow model represented by predefined mathematical functions is first fit to the flow waveform over the entire length of the closing phase. The value of NAQ could then be computed by using the derivative of this mathematical function instead of using d_{peak} extracted at a single time instant of the original, distorted derivative.

By analyzing voices produced in three different phonation types, the study showed that NAQ values were, on average, approximately 50% smaller than the corresponding CQ values. It was also proven that there is a high correlation between NAQ and CQ. In addition, the value of NAQ showed a monotonic decrease for all ten analyzed subjects when phonation was changed from breathy to normal and then to pressed. Hence, NAQ reflects changes in the voice register. In this respect, the behavior of NAQ is similar to that of CQ. In measuring the time-domain features of the glottal closing phase, however, NAQ takes into account only the energetically decisive portion of the flow derivative in the vicinity of t_{peak}. This principle can be considered more justified than measuring the entire length of the glottal closing phase embedded in the computation of CQ if the derivative waveform is smooth (and consequently of minor importance energetically) in the beginning and in the end of the glottal closing phase.

In focusing on a subsection of the entire glottal closing phase, NAQ is similar to some of the time-domain parametrization schemes reported in previous studies (Fant, 1997; Sundberg *et al.*, 1993; Frölich *et al.*, 2001). In these studies, the parametrization of the glottal flow has been computed by defining the closing phase as a time span between the flow maximum and the negative peak of the flow derivative. Measuring this time span, which is denoted by T_{pp} in Sundberg *et al.* (1993), implies that the so-called return phase of the flow derivative has been ignored. Among these studies, the work by Sundberg *et al.* (1993) is closest to the current survey, because it also involves analyses on the mode of phonation. Unfortunately, the results obtained by Sundberg *et al.* (1993) cannot be directly compared to those of the current study, because breathy phonation was not analyzed in their experiments and they did not normalize T_{pp} with respect to the length of the fundamental period. However, Sundberg *et al.* (1993) reported that T_{pp} tended to be shorter in pressed voices than in normal voices of the same pitch. Hence, their results are in line with those of the current study, even though the parametrization methods, both of which focus on a subsection of the glottal closing phase, are different. Interestingly, it was reported by Sundberg *et al.* (1993) that in 24% of the cases T_{pp} increased when phonation was changed to pressed. Comparison of this finding to the result of the current study, which shows that there was a monotonic change of NAQ between the phonation types for all the utterances, suggests that NAQ is able to classify the phonation type more accurately than T_{pp}.

[1]It should be noticed in Eqs. (3) and (4) that the domain of AQ is time, because this quotient is defined as a ratio between a flow value and a value of the *time derivative* of the flow. If inverse filtering is based on digital signal processing, which is typical in voice source analysis today, the values of f_{ac} and d_{peak} are usually extracted from discrete-time waveforms that are expressed using integer numbers as the time variable. In this case, AQ in Eqs. (3) and (4) needs to be divided by the sampling frequency in order to express the parameter value in seconds.

[2]The resulting length of the rectangular pulse does not, of course, depend on the absolute time location of the pulse. However, to adjust the center point of the pulse to the instant of the negative peak of the flow derivative emphasizes that d_{peak} is the only amplitude value of the flow derivative needed in the computation of T_{erp}.

[3]Modeling a spectral resonance with a digital all-pole filter requires (at least) one complex conjugate pair of poles in the z domain. Since the signal bandwidth used in the current study was 4 kHz, and vowels have on average one formant per 1 kHz, the minimum order of the vocal tract filter is eight. However, using this small order of the vocal tract model sometimes results in formant ripple due to insufficient canceling of the vocal tract resonances. In general, the distortion caused by format ripple can be reduced by increasing the order of the all-pole filter. Then again, if too large a filter order is used, the glottal flow estimate can be distorted by smoothing caused by increased low-frequency amplification of the inverse filter. In order to minimize the effects of these two distortions in the current study, the order of the vocal tract all-pole filter was adjusted separately for each utterance by always starting with a filter order of eight and by increasing it to 10 or 12 when a reduction in the amount of formant ripple was required.

Alku, P., Strik, H., and Vilkman, E. (**1997**). ''Parabolic spectral parameter—A new method for quantification of the glottal flow,'' Speech Commun. **22**, 67–79.

Alku, P., and Vilkman, E. (**1994**). ''Estimation of the glottal pulseform based on discrete all-pole modeling,'' in Proceedings of the International Conference on Spoken Language Processing 1994 (Yokohama), 1619–1622.

Alku, P., and Vilkman, E. (**1996a**). ''A comparison of glottal voice source quantification parameters in breathy, normal, and pressed phonation of female and male speakers,'' Folia Phoniatr Logop **48**, 240–254.

Alku, P., and Vilkman, E. (**1996b**). ''Amplitude domain quotient for characterization of the glottal volume velocity waveform estimated by inverse filtering,'' Speech Commun. **18**, 131–138.

Carlson, B. (**1986**). *Communication Systems* (McGraw-Hill, Singapore), pp. 177–178.

Carlson, R., Fant, G., Gobl, C., Granström, B., Karlsson, I., and Lin, Q. (**1989**). ''Voice source rules for text-to-speech synthesis,'' in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 223–226.

Childers, D. G., and Lee, C. K. (**1991**). ''Vocal quality factors: Analysis, synthesis, and perception,'' J. Acoust. Soc. Am. **90**, 2394–2410.

Dromey, C., Stathopoulos, E. T., and Sapienza, C. M. (**1992**). ''Glottal airflow and electroglottographic measures of vocal function at multiple intensities,'' J. Voice **6**, 44–54.

El-Jaroudi, A., and Makhoul, J. (**1991**). ''Discrete all-pole modeling,'' IEEE Trans. Signal Process. **39**, 411–423.

Fant, G. (**1993**). ''Some problems in voice source analysis,'' Speech Commun. **13**, 7–22.

Fant, G. (**1995**). ''The LF-model revisited. Transformations and frequency domain analysis,'' Speech Transmission Laboratory, Quarterly Progress and Status Report, Royal Institute of Technology, Stockholm **2–3**, pp. 119–156.

Fant, G. (**1997**). ''The voice source in connected speech,'' Speech Commun. **22**, 125–139.

Fant, G., Kruckenberg, A., Liljencrants, J., and Båvegård, M. (**1994**). ''Voice source parameters in continuous speech. Transformation of LF-parameters,'' in Proceedings of the International Conference on Spoken Language Processing 1994 (Yokohama), pp. 1451–1454.

Fant, G., Liljencrants, J., and Lin, Q. (**1985**). ''A four-parameter model of glottal flow,'' Speech Transmission Laboratory, Quarterly Progress and Status Report, Royal Institute of Technology, Stockholm **4**, pp. 1–13.

Fant, G., and Lin, Q. (**1988**). ''Frequency domain interpretation and derivation of glottal flow parameters,'' Speech Transmission Laboratory, Quarterly Progress and Status Report, Royal Institute of Technology, Stockholm **2–3**, pp. 1–21.

Fröhlich, M., Michaelis, D., and Strube, H. (**2001**). ''SIM—Simultaneous inverse filtering and matching of a glottal flow model for acoustic speech signals,'' J. Acoust. Soc. Am. **110**, 479–488.

Hertegård, S., Gauffin, J., and Karlsson, I. (**1992**). ''Physiological correlates of the inverse filtered flow waveform,'' J. Voice **6**, 224–234.

Holmberg, E. B., Hillman, R. E., and Perkell, J. S. (**1988**). ''Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice,'' J. Acoust. Soc. Am. **84**, 511–529.

Howell, P., and Williams, M. (**1992**). ''Acoustic analysis and perception of vowels in children's and teenagers' stuttered speech,'' J. Acoust. Soc. Am. **91**, 1697–1706.

Monsen, R. B., and Engebretson, A. M. (**1977**). ''Study of variations in the male and female glottal wave,'' J. Acoust. Soc. Am. **62**, 981–993.

Moore, B. C. (**1982**). *An Introduction to the Psychology of Hearing* (Academic, London), p. 82.

Rothenberg, M. (**1973**). ''A new inverse-filtering technique for deriving the glottal air flow waveform during voicing,'' J. Acoust. Soc. Am. **53**, 1632–1645.

Strik, H., and Boves, L. (**1992**). ''On the relation between voice source parameters and prosodic features in connected speech,'' Speech Commun. **11**, 167–174.

Sulter, A. M., and Wit, H. P. (**1996**). ''Glottal volume velocity waveform characteristics in subjects with and without vocal training, related to gender, sound intensity, fundamental frequency, and age,'' J. Acoust. Soc. Am. **100**, 3360–3373.

Sundberg, J., Andersson, M., and Hultqvist, C. (**1999**). ''Effects of subglottal pressure variation on professional baritone singers' voice sources,'' J. Acoust. Soc. Am. **105**, 1965–1971.

Sundberg, J., Titze, I., and Scherer, R. (**1993**). ''Phonatory control in male singing: A study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source,'' J. Voice **7**, 15–29.

Wilks, S. S. (**1962**). *Mathematical Statistics* (Wiley, New York), p. 74.

Wong, D. Y., Markel, J. D., and Gray, Jr., A. H. (**1979**). ''Least-squares glottal inverse filtering from acoustic speech waveforms,'' IEEE Trans. Acoust., Speech, Signal Process. **27**, 350–355.

Learning to perceive speech: How fricative perception changes, and how it stays the same

Susan Nittrouer
Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 25 May 2001; accepted for publication 28 May 2002)

A part of becoming a mature perceiver involves learning what signal properties provide relevant information about objects and events in the environment. Regarding speech perception, evidence supports the position that allocation of attention to various signal properties changes as children gain experience with their native language, and so learn what information is relevant to recognizing phonetic structure in that language. However, one weakness in that work has been that data have largely come from experiments that all use similarly designed stimuli and show similar age-related differences in labeling. In this study, two perception experiments were conducted that used stimuli designed differently from past experiments, with different predictions. In experiment 1, adults and children (4, 6, and 8 years of age) labeled stimuli with natural /f/ and /θ/ noises and synthetic vocalic portions that had initial formant transitions varying in appropriateness for /f/ or /θ/. The prediction was that similar labeling patterns would be found for all listeners. In experiment 2, adults and children labeled stimuli with initial /s/-like and /ʃ/-like noises and synthetic vocalic portions that had initial formant transitions varying in appropriateness for /s/ or /ʃ/. The prediction was that, as found before, children would weight formant transitions more and fricative noises less than adults, but that this age-related difference would elicit different patterns of labeling from those found previously. Results largely matched predictions, and so further evidence was garnered for the position that children learn which properties of the speech signal provide relevant information about phonetic structure in their native language.
[DOI: 10.1121/1.1496082]

PACS numbers: 43.71.An, 43.71.Ft, 43.71.Es [CWT]

I. INTRODUCTION

Perception is the extraction of information about the affordances of things in the world, according to Eleanor Gibson (1991b). The information to be extracted is always there, Gibson tells us, ''like truth'' (1977/1991a, p. 474). This does not mean, however, that all perceivers extract the same information about events. What is extracted depends upon several factors, with Gibson listing three as most important: the species of the organism, developmental maturity, and learning. When it comes to understanding how children learn to perceive the speech signal, we are most interested in the role that learning plays in the acquisition of competency because learning is the only one of the three factors that we can influence. Research with speakers/listeners of different languages has demonstrated robustly that the information extracted from the speech signal is highly dependent on the native language of the individual (Strange, 1995, offers a substantive review of the literature on cross-linguistic speech perception). Clearly, then, children must learn what information should be extracted in their native language. But, what determines the information that the child needs to learn to extract? Gibson tells us that a guide to answering this question is to think about the goal of perception. For example, the goal of perception for wine and tea tasters is to recognize various flavors, and so they must learn to extract information about subtle variances in the substances creating those flavors. In the case of speech, the goal of perception is to apprehend linguistic structure. For the purposes of most speech perception studies, the structure of interest is phonetic structure, and so it is here. In matters of speech perception, children must learn to extract the information that will permit access to phonetic structure in their native language.

One final principle described by Gibson (1977/1991a) that guides our investigation of how children learn to perceive speech is the notion that there is increasing specificity of correspondence between the information in the world that is extracted and in what is perceived. According to this account, events can be perceived from birth, but without as much differentiation among them as there will be later. As the level of differentiation increases, as it must, the information extracted from the environment becomes more specifically related to the events being perceived: that is, the child learns what information is relevant.

The hypothesis that has emerged to capture these developmental changes as they relate to speech perception has been termed the ''developmental weighting shift'' (or DWS). This hypothesis suggests simply that the informational aspects of the signal that get weighted greatly (or, attended to greatly) change as children gain experience listening to and speaking their native language. Initially, children seem to attend largely to those acoustic properties that specify a changing vocal tract, which means the dynamic spectral changes known as formant transitions. Then, as the child becomes more skilled, attention shifts to acoustic properties that do not involve spectral change, properties such as silent gaps (specifying periods of vocal-tract closure), differences in voicing duration (specifying the voicing of syllable-final

stops), or periods of stable spectral information (specifying place of consonantal constrictions).

One series of experiments tracing this developmental shift has focused on /s/-vowel and /ʃ/-vowel syllables (Nittrouer, 1992; Nittrouer and Miller, 1997a, 1997b; Nittrouer and Studdert-Kennedy, 1987). For adult English speakers, the decision about which of these sibilants occurred syllable initially is largely based on the spectrum of the noise (e.g., Harris, 1958). As would be predicted from the above account, young children base their responses not so much on the noise spectrum (when compared to adults), but more on the formant transitions that arise from movement as the vocal tract goes from the consonantal constriction to the vocalic constriction. A developmental change has been observed in the relative amounts of attention (or weight) paid to these acoustic properties for children between $3\frac{1}{2}$ and $7\frac{1}{2}$ years of age. However, these experiments all used similarly designed stimuli that had fricative noises varying along acoustic continua and formant transitions largely appropriate for initial /s/ or /ʃ/. The purpose of the experiments reported here was to test predictions arising from the DWS using stimuli designed differently.

First, the DWS hypothesis predicts that adults and children should perform similarly for a labeling decision that adults make based primarily on formant transitions. That is, if perceptual strategies early in life put particular weight on formant transitions, we should be able to find phonetic contrasts that require no modification from those early strategies because paying attention to transitions is the most efficient strategy for apprehending phonetic structure. Harris (1958) provides evidence of just such a decision. In that experiment, noises from unvoiced /f/, /θ/, /s/, and /ʃ/, as well as their voiced counterparts, were cross spliced with vocalic portions that had formant transitions appropriate for one of these syllable-initial fricatives. Harris showed that adult listeners make decisions about whether the fricative was /f/ or /θ/ based largely on the formant transitions, while decisions about whether it was /s/ or /ʃ/ are based largely on the noise spectrum. Thus, the decision about whether a syllable-initial fricative is /f/ or /θ/ should be one for which competent English speakers rely on perceptual strategies that do not require modification from their earliest state, and so adults and children should show similar results on labeling tasks. To test this prediction, stimuli were used in which formant transitions varied along continua from those appropriate for a syllable-initial /θ/ to those appropriate for a syllable-initial /f/, and with either a /θ/ or /f/ noise. This arrangement of acoustic properties was, in a sense, opposite to that of the earlier /ʃ/ versus /s/ labeling experiments. In those earlier experiments, the fricative noises formed the continuum, and formant transitions were set appropriately for one or the other fricative condition. However, that arrangement of acoustic properties across stimuli is not possible in experiments designed to investigate /f/ versus /θ/ perception because there simply is not enough of a difference between the spectra of these noises to form a continuum.

The second prediction of the DWS tested in this study arose precisely from this difference in stimulus design. According to the DWS, even if we reverse the arrangement of acoustic properties in an /ʃ/ versus /s/ labeling experiment, children should continue to show evidence of weighting formant transitions more, and fricative noises less, than adults. This prediction was tested in a second experiment reported here, in which the arrangement of acoustic properties across stimuli was different from past experiments examining developmental changes in /ʃ/ versus /s/ perception (Nittrouer, 1992; Nittrouer and Miller, 1997a, 1997b; Nittrouer and Studdert-Kennedy, 1987). Here, acoustic properties were arranged as they were in the /f/ versus /θ/ experiment, with formant transitions varying along a continuum and two fricative noises set to favor one or the other fricative label.

In summary, two experiments were conducted to test rigorously the predictions of the DWS. In the first experiment examining /f/ versus /θ/ perception, listeners of all ages should show the same weighting strategies. In the second experiment examining /ʃ/ versus /s/ perception, a developmental decrease in the weight assigned to formant transitions should be found, with a developmental increase in the weight assigned to the fricative-noise spectra.

II. ANALYSIS OF PREVIOUS STUDY

Nittrouer and Miller (1997b) serves as an example of a study tracing the emergence of mature perceptual weighting strategies for /s/-vowel versus /ʃ/-vowel decisions. In the three experiments reported there, the fricative noise varied along a nine-step continuum from one appropriate for /ʃ/ to one appropriate for /s/. Four vocalic portions were used in each experiment: two /ɑ/ portions and two /u/ portions, with formant transitions appropriate for either a syllable-initial /ʃ/ or /s/. In experiment 1 these vocalic portions were natural. In experiments 2 and 3 they were synthetic, either with $F2$ alone varying as appropriate for syllable-initial /ʃ/ or /s/ (experiment 2), or with both $F2$ and $F3$ varying as appropriate for syllable-initial /ʃ/ or /s/ (experiment 3). Listeners labeled stimuli as starting with /ʃ/ or /s/, and responses were plotted as the proportion of ''s'' responses at each level of fricative noise, for each vocalic portion separately. Cumulative normal distributions from experiment 3 of that study are shown in Fig. 1 for adults, 7-, 5-, and $3\frac{1}{2}$-year-olds. Information about the relative weighting of the fricative noises and formant transitions can be gleaned from the shapes of the labeling functions and from the separation between those functions based on formant transitions: The steeper the functions, the more weight that was assigned to the fricative-noise spectrum; the greater the separation between functions for stimuli with /ʃ/ and /s/ transitions, the more weight that was assigned to the formant transitions.

Of course, estimating the relative weights assigned to the fricative noises and formant transitions in this way does not provide a precise metric. Another way to index these weights is by doing regression analyses, and that was done here. For data from each listener, the proportion of ''s'' responses given at each level of the two acoustic properties manipulated (fricative-noise spectrum and formant transitions) was correlated with these two acoustic properties, using the arcsine transforms of the proportions themselves. These analyses were done for the first and third experiments of Nittrouer and Miller (1997b) because these two experi-

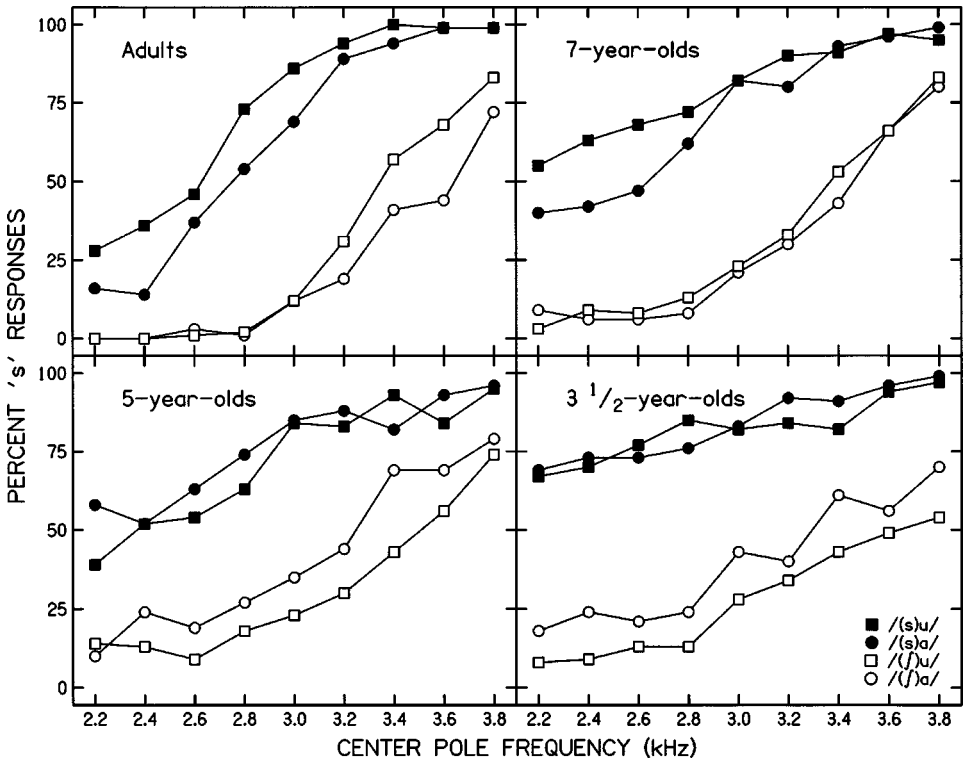


FIG. 1. Labeling functions for /ʃ/-vowel and /s/-vowel stimuli in which noise spectra varied along a nine-step continuum and formant transitions varied dichotomously. (Reprinted from Nittrouer and Miller, 1997b, experiment 3).

ments had both $F2$ and $F3$ varying as appropriate for syllable-initial /ʃ/ or /s/. Tables I and II show mean partial correlation coefficients derived for each age group in the two experiments, with standard deviations (s.d.'s) given in parentheses. Although the exact correlations for any age group differ across the studies, a general trend is easily discernible: With increasing age, correlation coefficients for noise increase while those for formant transitions decrease. Welch's equality of means tests were performed on correlation coefficients for each acoustic property to test the significance of these trends. This test is similar to analyses of variance (ANOVAs), but does not assume homogeneity of variance. For experiment 1 of Nittrouer and Miller, the effect of age was significant for correlations involving both noise, $W(2,35)=10.83$, $p<0.001$, and formant transitions, $W(2,38)=5.16$, $p=0.01$. For experiment 3, the age effect was also significant for both kinds of correlations: noise, $W(3,22)=5.67$, $p=0.005$; formant transitions, $W(3,18)=7.70$, $p=0.002$. In summary, these correlational analyses capture nicely the fact that experiments have reliably demonstrated a developmental decrease in the weight assigned to formant transitions and a developmental increase in the weight assigned to fricative-noise spectra in decisions of fricative identity for syllable-initial /ʃ/ and /s/.

TABLE I. Mean partial correlation coefficients for each group in experiment 1 of Nittrouer and Miller (1997b). Standard deviations (s.d.'s) are given in parentheses.

	$4\frac{1}{2}$-year-olds	7-year-olds	Adults
Experiment 1			
Fricative noise	0.683	0.770	0.826
	(0.126)	(0.101)	(0.061)
Formant transitions	0.459	0.366	0.324
	(0.135)	(0.139)	(0.133)

III. EXPERIMENT 1: /f/ VERSUS /θ/ PERCEPTION

The purpose of this experiment was to test the prediction that adults and children would show similar perceptual strategies in labeling /f/-vowel and /θ/-vowel syllables because listeners of all ages would be found to weight formant transitions heavily, but not the noise spectra. This similarity in weighting strategies should give rise to correlation coefficients that do not differ across age groups.

A. Method

1. Listeners

To participate, listeners were required to pass a hearing screening of the pure tones 0.5, 1.0, 2.0, 4.0, and 6.0 kHz, presented at 25 dB HL. American English was the native language of all adults participating, and all children lived in monolingual English home environments. Children needed to score at or above the 30th percentile on the Goldman–Fristoe Test of Articulation (Goldman and Fristoe, 1986), and they needed to have negative histories of otitis media, defined as less than six episodes during the first 2 years of life.

TABLE II. Mean partial correlation coefficients for each group in experiment 3 of Nittrouer and Miller (1997b). s.d.'s are given in parentheses.

	$3\frac{1}{2}$-year-olds	5-year-olds	7-year-olds	Adults
Experiment 3				
Fricative noise	0.367	0.581	0.599	0.664
	(0.182)	(0.135)	(0.200)	(0.147)
Formant transitions	0.774	0.589	0.619	0.554
	(0.139)	(0.162)	(0.167)	(0.148)

Adults were required to demonstrate at least an 11th grade reading level on the Wide-Range Achievement Test-Revised (Jastak and Wilkinson, 1984). Meeting these requirements were seventeen 4-year-olds, thirteen 6-year-olds, fourteen 8-year-olds, and 13 adults. However, four of these 4-year-olds were subsequently dismissed because they were unable to label reliably the best exemplars of the category labels. The mean ages (and ranges) of the children included in the study, given in years; months, were 4; 3 (3; 11 to 4; 6), 6; 1 (5; 11 to 6; 4), and 8; 2 (7; 11 to 8; 5). The mean age of adults was 30 years, with the range between 21 years and 40 years.

2. Equipment and materials

All testing took place in a sound-attenuated booth, with the computer that controlled the experiment in an adjacent room. The hearing screening was done with a Welch Allen TM262 audiometer and TDH-39 earphones. Stimuli were stored on a computer and presented by way of a Data Translation DT-2801A digital-to-analog converter, a Frequency Devices 901-F filter, a Crown D-75 amplifier, and AKG-K141 headphones. Responses were recorded with a keyboard connected to the computer. Four hand-drawn pictures (8 in.×8 in.) were used to represent each response label (''tha,'' ''thu,'' ''fa,'' or ''fu''). These pictures were of animate objects, given the names of the response labels. Recorded stories were presented via a Nakamichi MR-2 audiocassette player, with a Tascam PA-30B amplifier and AKG-K141 headphones. Hand-drawn story books accompanied the taped stories, and the object named with the response label figured prominently on every page. Gameboards with ten steps were also used with children: they moved a marker to the next number on the board after each block of test stimuli. Cartoon pictures were used as reinforcement and were presented on a high-graphics 14-in. color monitor after completion of each block of stimuli. A bell sounded while the color graphics were being presented and served as additional reinforcement.

3. Stimuli

Stimuli consisted of natural /f/ and /θ/ noises combined with synthetic /ɑ/ and /u/ vocalic portions. In order to create stimuli that matched natural productions of /fɑ/, /θɑ/, /fu/, and /θu/, five tokens of each of these syllables were recorded by four male speakers. None of these speakers had any history of speech or language problems, and all passed a hearing screening of the frequencies 0.5, 1.0, 2.0, and 4.0 kHz presented at 20 dB HL. All were native speakers of American English, with Midwestern dialects. Five measurements were obtained from these samples: relative amplitude of the noise portion of the syllable, compared to the vocalic portion (dB), duration of the noise (ms), the first spectral moment of the noise (kHz), $F2$ for the first pitch period (Hz), and $F3$ for the first pitch period (Hz). Means (and s.d.'s) of each of these measures for each syllable type are shown in Table III.

Four fricative noises were used for stimulus creation because /f/ and /θ/ noises separated from natural fricative-/ɑ/ syllables were combined with the synthetic /ɑ/ portions, and /f/ and /θ/ noises separated from natural fricative-/u/ syllables were combined with the synthetic /u/ portions. The four noises used were all from the same speaker and closely approximated mean amplitude and first moment measurements for their respective fricative-vowel combination from the acoustic analysis. Spectra of these four noises are shown in Fig. 2. Each was truncated to 190 ms for the purpose of stimulus creation.

TABLE III. Mean measurements across tokens produced by four adult, male speakers of the syllables /θɑ/, /fɑ/, /θu/, and /fu/ (5 tokens of each syllable per speaker). Five measurements were made: relative amplitude of the noise to the vocalic portion (dB), duration of the noise (ms), the first spectral moment (kHz), $F2$ for the first pitch period (Hz), and $F3$ for the first pitch period (Hz). s.d.'s are given in parentheses.

	/ɑ/		/u/	
	/θ/	/f/	/θ/	/f/
Relative amplitude	−27.3	−29.5	−25.9	−27.9
	(3.0)	(3.3)	(3.5)	(3.1)
Duration	176	191	212	198
	(20)	(34)	(39)	(27)
First moment	8.00	7.68	7.88	7.86
	(0.61)	(0.68)	(0.39)	(0.47)
$F2$	1365	1092	1496	1337
	(67)	(46)	(180)	(216)
$F3$	2517	2423	2391	2268
	(187)	(220)	(143)	(141)

Eighteen synthetic vocalic portions were made using the Sensyn Laboratory Speech Synthesizer: nine each of /ɑ/ and /u/. These vocalic portions were created at a 20-kHz sampling rate to match the sampling rate of the noises. Stimuli for both continua varied from one appropriate for /θ/-vowel to one appropriate for /f/-vowel. Parameter settings were based on the acoustic measures of the natural vocalic portions in the acoustic analysis. For all vocalic portions, duration was 300 ms, and $f0$ started at 120 Hz, falling throughout

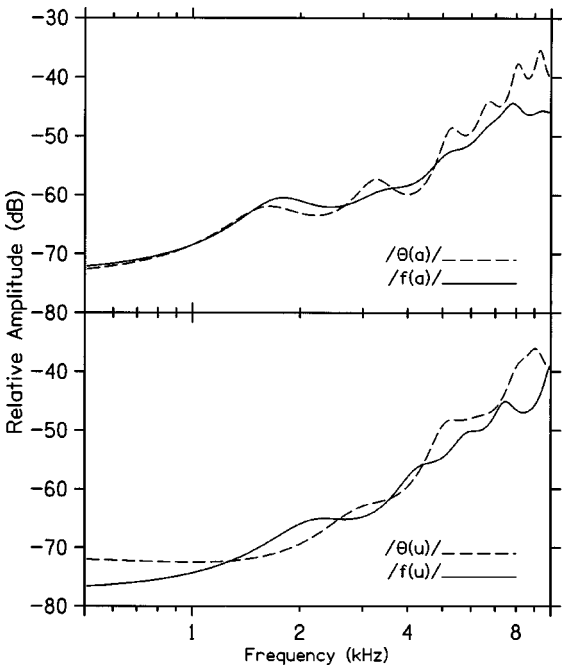


FIG. 2. Spectra of /f/ and /θ/ noises from /ɑ/ and /u/ contexts, spoken by a male adult. These noises were used in experiment 1.

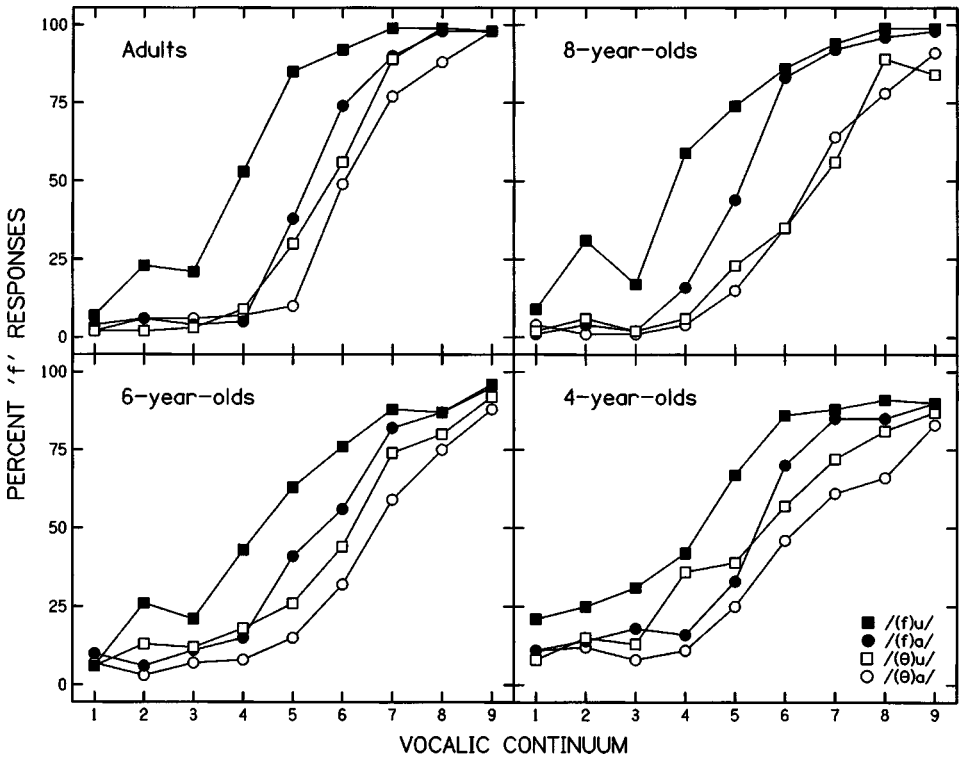


FIG. 3. Labeling functions for /f/-vowel and /θ/-vowel stimuli in which formant transitions varied along continua and noise spectra varied dichotomously (experiment 1).

the portion to an offset frequency of 90 Hz. The vocalic portion was always 22 dB greater in amplitude than the noise portion. For /ɑ/ portions, $F1$ began at 560 Hz and rose to a steady-state frequency of 750 Hz over the first 70 ms. $F2$ onset formed a continuum varying in 40-Hz steps from 1410 Hz (most /θ/-like) to 1090 Hz (most /f/-like). $F2$ changed over the first 70 ms from its onset frequency to a steady-state frequency of 1180 Hz. $F3$ onset varied in nine steps from a frequency of 2535 Hz (most /θ/-like) to 2415 Hz (most /f/-like) in 15-Hz steps. This formant changed over the first 70 ms from the onset frequency to a steady-state frequency of 2480 Hz. Each of these nine portions was combined with each of the /f/ and /θ/ noises taken from the /ɑ/ context, making 18 stimuli.

For /u/ portions, $F1$ was 350 Hz at onset and fell very gradually throughout the vocalic portion to an offset frequency of 330 Hz. $F2$ onset varied in 50-Hz steps from 1520 Hz (most /θ/-like) to 1120 Hz (most /f/-like). $F2$ fell throughout the vocalic portion to an offset frequency of 940 Hz. $F3$ onset varied in 20-Hz steps from 2400 Hz (most /θ/-like) to 2240 Hz (most /f/-like). $F3$ changed throughout the portion to an offset frequency of 2250 Hz. These vocalic portions were combined with the /f/ and /θ/ noises taken from the /u/ context.

4. Procedures

All screening tasks were completed first. For the 4- and 6-year-olds the next step was the presentation of tape-recorded stories about the labels they would be using in the labeling task (either ''fa'' and ''tha'' or ''fu'' and ''thu''). Each of these four stories was roughly 3 min long, and the label was the name of an animate object that served as the main character in the story. These stories were heard twice, once with natural speech and once with synthetic speech. The stories served two purposes. First, they allowed children greater opportunity to learn the stimulus labels. Second, they allowed children to acclimate to synthetic speech prior to testing.

Next, the labeling task was presented. Stimuli with /ɑ/ and /u/ portions were presented separately, making the listener's task a binary choice. All participants first heard ten practice items (five each of /f/-vowel and /θ/-vowel), which were the best exemplars of each category. For example, the best exemplar of ''fa'' was the stimulus with the natural /f/ noise and the synthetic vocalic portion in which $F2$ and $F3$ onsets were 1090 and 2415 Hz, respectively. To proceed to testing, subjects were required to label correctly at least nine of these ten practice stimuli.

During testing, the stimuli in each vowel set were presented ten times each in randomized blocks of 18. To have their data included in the analyses, participants needed to achieve at least 80% accurate responses for the best exemplars. This requirement insured that only participants who maintained attention during the task were included. Children kept track of their progress by moving the marker to the next space on the gameboard after they completed each block of stimuli. The order of presentation of the /ɑ/ and /u/ stimuli was randomized across listeners.

B. Results

Figure 3 shows the labeling functions for all four groups. Clearly, there is less of a developmental change in the labeling of these /f/-vowel and /θ/-vowel stimuli than in the labeling of the /s/-vowel and /ʃ/-vowel stimuli seen in Fig. 1, from experiment 3 of Nittrouer and Miller (1997b): The functions across age groups appear similarly steep, with similar separations among the functions.

TABLE IV. Mean partial correlation coefficients for each group in experiment 1 (/f/ versus /θ/). s.d.'s are given in parentheses.

	4-year-olds	6-year-olds	8-year-olds	Adults
Fricative noise	0.202	0.200	0.293	0.173
	(0.099)	(0.099)	(0.111)	(0.070)
Formant transitions	0.822	0.836	0.828	0.875
	(0.067)	(0.065)	(0.50)	(0.045)

Table IV shows the mean partial correlation coefficients for each age group in experiment 1, for fricative noise and formant transitions. The coefficients for formant transitions appear similar across age groups, and the statistical analysis supports that observation: there is no significant effect of age.[1] For fricative noise, however, the 8-year-olds show larger correlation coefficients than the other three groups. In keeping with that observation, a significant main effect of age was found, $W(3,27)=3.62$, $p=0.026$. Because of this finding, *post hoc t*-tests were done on these data. Results of these tests (using separate variances) are shown in Table V. Here, we see that 8-year-olds did indeed show a weighting of the fricative noise that was greater than that of all three other groups.

C. Discussion

The prediction for this experiment was that there would be no differences in the weights assigned to the fricative-noise spectra and formant transitions by listeners of different ages because all listeners would base responses largely on formant transitions, and not very greatly on fricative-noise spectra. To a great extent, this prediction was met. The one finding that did not meet the prediction was that 8-year-olds weighted the fricative noise more than any of the other three groups. This result may be aberrant in this study, and not representative of 8-year-olds' strategies in general. However, it might also represent an instance of overgeneralization. That is, the overarching hypothesis being examined by this work is that initial weighting strategies for speech perception focus on acoustic properties associated with movement: i.e., formant transitions. Then, as the child gains experience with a native language, these strategies are modified. One kind of modification, it is proposed, involves learning to look to acoustic detail at the syllable margins for clues to phonetic structure. In this instance it may be that children overgeneralize this strategy as it is developing, using it in contexts that would best be served by the strategy that they had used earlier. Certainly overgeneralization of newly learned linguistic forms is recognized in language development (e.g., de Villiers and de Villiers, 1978; Slobin, 1973).

TABLE V. Results of *post hoc t*-tests for experiment 1.

	Noise		
	df	*t*	*p*
4- vs 6-year-olds	23	0.06	ns
4- vs 8-year-olds	24	−2.24	0.034
4- vs adults	21	0.84	ns
6- vs 8-year-olds	24	−2.30	0.030
6-year-olds vs adults	21	0.78	ns
8-year-olds vs adults	22	3.30	0.003

Another interesting result from this experiment was that, although not statistically significant, adults had the highest correlation coefficients for formant transitions, out of the four age groups, and the lowest correlation coefficients for fricative-noise spectra. Although it was anticipated that adults would base judgments of fricative identity for /f/ versus /θ/ greatly on formant transitions, rather than on fricative noises, the prediction had been that the relative weights would be the same as those of young children. Instead, the weighting strategies of adults favored formant transitions slightly more than those of the children. Perhaps this finding reveals the extent to which highly honed, mature perceptual strategies match the weight assigned to each acoustic property to the amount of information it provides about phonetic structure in the language.

IV. EXPERIMENT 2: /s/ VERSUS /ʃ/ PERCEPTION

Experiment 1 showed that adults and children perform similarly when the phonetic judgment to be made is one for which we expect adults and children to use similar weighting strategies. This finding lends support for the conclusion that results of earlier studies using /s/-vowel and /ʃ/-vowel stimuli (showing that children weight formant transitions more and fricative-noise spectra less than adults) represent real perceptual differences across listener groups. However, it is difficult to make direct comparisons between experiment 1 (with /f/ and /θ/) and the earlier experiments (with /s/ and /ʃ/) because the arrangement of acoustic properties differed across the two kinds of studies. Specifically, the fricative noise was the property varied along a continuum in the earlier work, with formant transitions coming from natural syllables or from synthetic vocalic portions fashioned after natural syllables. As a result, formant transitions were always set dichotomously. In experiment 1, vocalic portions were synthesized with formant transitions varying along continua, and natural noises were used, and so were intrinsically dichotomous. The purpose of this experiment was to devise a similar experiment using /s/ and /ʃ/ so that results could be more readily compared.

Of course, realizing the utility in designing such an experiment is one thing, actually doing so is another. Extensive pilot testing showed that it was not possible to use stimuli with completely natural /s/ and /ʃ/ noises (and vocalic portions varying along a continuum) because to do so resulted in many listeners (particularly adults) responding according to the noise alone. Besides, it would not necessarily be desirable to use natural /s/ and /ʃ/ noises, with the large spectral difference that there is between them. Instead, a test more equivalent to the /f/ versus /ʃ/ experiment would be had if stimuli were designed with roughly the same acoustic difference between the /s/ and /ʃ/ noises as there was for the /f/ and /θ/ noises in experiment 1. That is, there were just a few dB difference between the major spectral peaks of those /f/ and /θ/ noises, and so stimuli in this experiment were designed with just a few dB difference between the major peaks of these /s/ and /ʃ/ noises. As in experiment 1, vocalic portions

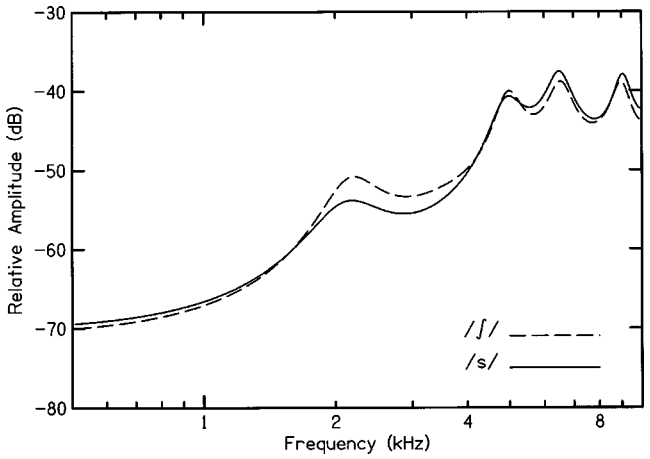


FIG. 4. Spectra of the noises used in experiment 2.

were synthesized so that the formant transitions at onset spanned a continuum: in this case, from those appropriate for a preceding /ʃ/ to those appropriate for a preceding /s/.

A. Method

1. Listeners

Listeners of the same age groups as those in experiment 1 participated in this experiment, and were required to meet the same criteria. Meeting these criteria were 12 4-year-olds, 13 6-year-olds, 19 8-year-olds, and 12 adults. However, four 4-year-olds and one 8-year-old were subsequently dismissed because they were unable to label reliably the best exemplars of each category. The mean ages (and ranges) of children were 4; 7 (4; 3 to 4; 9), 6; 3 (5; 9 to 6; 8), and 8; 0 (7; 10 to 8; 7). The mean age of adults was 31 years, with the range between 25 years and 39 years.

2. Equipment and materials

The same equipment was used in this experiment as in experiment 1.

3. Stimuli

Hybrid stimuli consisting of fricative noises, created from natural /s/ and /ʃ/, and synthetic vocalic portions were used. The noises used in this experiment were not completely appropriate for /ʃ/ and /s/. Instead, one of the noises used in this experiment had a spectrum more similar (than the other noise) to /ʃ/, while the other noise had a spectrum more similar to /s/. These two noises were selected from a continuum created from natural /ʃ/ and /s/ noises, and used in experiment 1 of Nittrouer and Miller (1997a). These noises were created by adjusting the amplitudes of natural /ʃ/ and /s/ noises to desired levels (depending on how /ʃ/-like or /s/-like the resulting noise should be), and then adding them together. In this case, the more /ʃ/-like noise was created by combining the original /s/ and /ʃ/ noises at equal amplitudes; the more /s/-like noise was created by combining the original /s/ and /ʃ/ noises at a 2/1 (respectively) amplitude ratio. The spectra of these two resulting noises are shown in Fig. 4. The first moment of the /ʃ/-like noise was 6.80 kHz, and the first moment of the /s/-like noise was 7.21. Both were 100 ms long.

Eighteen synthetic vocalic portions were created using Sensyn Laboratory Speech Synthesizer: nine each of /ɑ/ and /u/. These vocalic portions were created at a 20-kHz sampling rate, to match the sampling rate of the noises. For all vocalic portions, duration was 260 ms, and $f0$ started at 120 Hz, falling throughout the portion to an offset frequency of 90 Hz. The vocalic portion was always 11 dB greater in amplitude than the noise portion. This amplitude difference

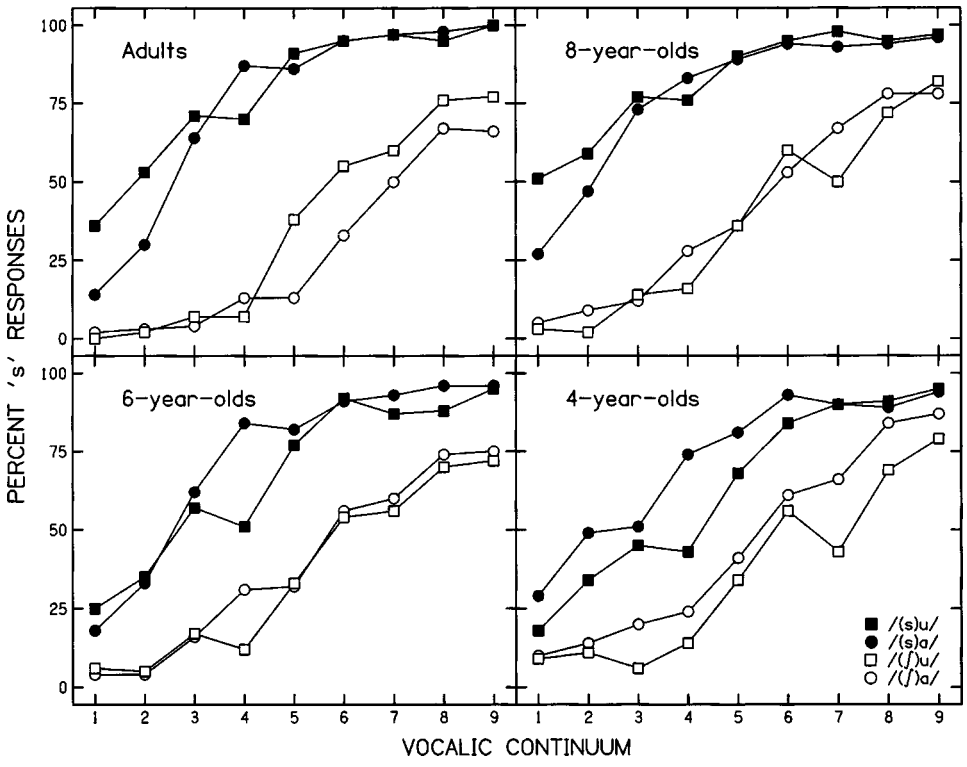


FIG. 5. Labeling functions for /s/-vowel and /ʃ/-vowel stimuli in which formant transitions varied along continua and noise spectra varied dichotomously (experiment 2).

TABLE VI. Mean partial correlation coefficients for each group in experiment 2 (/ʃ/ versus /s/). s.d.'s are given in parentheses.

	4-year-olds	6-year-olds	8-year-olds	Adults
Fricative noise	0.370	0.498	0.582	0.604
	(0.094)	(0.120)	(0.157)	(0.151)
Formant transitions	0.772	0.710	0.657	0.646
	(0.065)	(0.079)	(0.139)	(0.141)

was the mean difference found for speech samples collected from speakers as part of an earlier analysis (Nittrouer, 1995; Nittrouer, Studdert-Kennedy, and Neely, 1996). For all /ɑ/ portions, $F1$ began at 450 Hz and rose to a steady-state value of 650 Hz over the first 50 ms. $F2$ onset formed a nine-step continuum varying in 40-Hz steps from 1570 Hz (most /ʃ/-like) to 1250 Hz (most /s/-like). $F2$ changed over the first 100 ms from its onset value to a steady-state value of 1130 Hz. $F3$ onset varied in nine 58-Hz steps from a value of 2000 Hz (most /ʃ/-like) to 2464 Hz (most /s/-like). This formant changed over the first 100 ms from the onset value to a steady-state value of 2300 Hz.

For /u/ portions, $F1$ was 250 Hz throughout. $F2$ onset varied in 40-Hz steps from 1800 Hz (most /ʃ/-like) to 1480 Hz (most /s/-like). $F2$ fell throughout the vocalic portion to an offset value of 850 Hz. $F3$ onset varied in 40-Hz steps from 2200 Hz (most /ʃ/-like) to 2520 Hz (most /s/-like). $F3$ fell through the first 130 ms to an offset frequency of 2100 Hz.

4. Procedures

Procedures were the same for this experiment as for experiment 1.

B. Results

Figure 5 shows the labeling functions for all four groups, and generally gives the impression that the separation between functions, depending on fricative noise, increases with increasing age. Functions appear similar across age groups in terms of steepness. Table VI shows mean partial correlation coefficients for each age group, for both fricative noise and formant transitions. As with the partial correlation coefficients reported for experiments 1 and 3 of Nittrouer and Miller (1997b), we find a developmental increase in the weight assigned to the fricative noise, and a developmental decrease in the weight assigned to the formant transitions. The statistical analyses supported these observations: significant age effects were found both for the fricative noise, $W(3,22)=7.55$, $p=0.001$, and for the formant transitions, $W(3,23)=3.60$, $p=0.029$. To investigate these results further, *post hoc* *t*-tests were conducted, and these results are shown in Table VII. The only clearly significant differences among age groups are found for 4-year-olds versus every other group. The comparison of 6-year-olds versus adults for noise was marginally significant. Nonetheless, there are linear developmental trends across age groups for both fricative-noise spectrum and formant transitions.

C. Discussion

The predictions for this experiment were that a developmental increase in the weight assigned to the fricative-noise spectrum and a developmental decrease in the weight assigned to formant transitions would be found. Both predictions were met.

V. GENERAL DISCUSSION

These experiments were designed to test specific predictions arising from the hypothesis that children's perceptual weighting strategies change as they learn what information they should be extracting from the speech signal in their native language in order to apprehend phonetic structure. Specifically, the DWS proposes that young children initially focus their perceptual attention on general movement of the vocal tract, as conveyed by patterns of changing formant frequencies. This perceptual strategy would meet the needs of novice language users who are just learning the fundamentals of how to move their own vocal tracts for the purposes of communication: Early vocalizations do not show the detail, particularly at syllable margins, that mature speech demonstrates (e.g., Oller, 1986). The perceptual strategy suggested here is consistent also with the suggestion that children's word representations are not as precisely phonetic as those of adults, resulting in more globally organized lexicons and poorer retention of linguistic material in working memory (e.g., Charles-Luce and Luce, 1990, 1995; Nittrouer and Miller, 1999). As children gain experience with their native language, it is proposed, perceptual attention (i.e., weight) gradually shifts to take advantage of the properties of the acoustic signal that are most informative regarding phonetic structure. At least one aspect of this developmental shift, it is suggested, entails progressively greater weight being assigned to acoustic properties that convey information about details of production, such as precise shapes of consonantal constrictions. Consequently, this maturing perceptual

TABLE VII. Results of *post hoc* *t*-tests for experiment 2.

	Noise			Formant transitions		
	df	*t*	*p*	*df*	*t*	*p*
4- vs 6-year-olds	15	−2.63	0.019	14	1.88	0.081
4- vs 8-year-olds	18	−4.13	<0.001	21	2.82	0.010
4- vs adults	16	−4.17	<0.001	16	2.65	0.017
6- vs 8-year-olds	28	−1.68	0.104	27	1.36	ns
6-year-olds vs adults	21	−1.93	0.067	16	1.39	ns
8-year-olds vs adults	24	−0.39	ns	23	0.21	ns

trend not only helps the child access the phonetic structure of the message, it also helps the child refine her own productions.

In general, predictions for this study were met. For the /f/-vowel and /θ/-vowel stimuli, listeners of all ages showed similar weighting strategies for both the fricative-noise spectra and formant transitions. This result was predicted based on the finding of Harris (1958) showing that adults heavily weight formant transitions in making this decision. For the /s/-vowel and /ʃ/-vowel stimuli in which the formant transitions formed the acoustic continua and noise spectra varied dichotomously, weighting strategies resembled those of earlier studies in which the stimuli were designed differently: With increasing age came an increase in the weight assigned to the fricative-noise spectra and a decrease in the weight assigned to the formant transitions. Consequently, the conclusion seems warranted that language perceivers learn what information is usually available in the signal, and look to the relevant signal properties for that information.

In conclusion, the experiments described here provide general support for the hypothesis that children's perceptual strategies for speech are modified as they gain experience with a native language. Specifically, children learn what information in the signal must be extracted in order to apprehend phonetic structure in the language they are acquiring. Thus, learning to perceive speech in a mature manner can be viewed within a framework of general perceptual learning.

ACKNOWLEDGMENTS

This work was supported by Research Grant No. R01 DC00633 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health. The author gratefully acknowledges the assistance of Marnie E. Miller and Sandy Estee in the collection and analysis of data.

[1]Throughout this manuscript, precise statistical results will be given when $p < 0.10$. Otherwise, results will be reported simply as nonsignificant.

Charles-Luce, J., and Luce, P. A. (**1990**). ''Similarity neighbourhoods of words in young children's lexicons,'' J. Child Lang. **17**, 205–215.

Charles-Luce, J., and Luce, P. A. (**1995**). ''An examination of similarity neighbourhoods in young children's receptive vocabularies,'' J. Child Lang. **22**, 727–735.

deVilliers, J. G., and deVilliers, P. A. (**1978**). *Language Acquisition* (Harvard University, Cambridge, MA).

Gibson, E. J. (**1991a**). ''How perception really develops: A view from outside the network,'' in *An Odyssey in Learning and Perception* (MIT Press, Cambridge, MA), pp. 411–491. [Reprinted from *Basic Processes in Reading: Perception and Comprehension*, edited by D. Laberge and S. J. Samuels (Erlbaum, Hillsdale, NJ, 1977), pp. 155–173.]

Gibson, E. J. (**1991b**). *An Odyssey in Learning and Perception* (MIT Press, Cambridge, MA).

Goldman, R., and Fristoe, M. (**1986**). *Goldman Fristoe Test of Articulation* (American Guidance Service, Circle Pines, MN).

Harris, K. S. (**1958**). ''Cues for the discrimination of American English fricatives in spoken syllables,'' Lang. Speech **1**, 1–7.

Jastak, S., and Wilkinson, G. S. (**1984**). *The Wide Range Achievement Test-Revised* (Jastak Associates, Wilmington, DE).

Nittrouer, S. (**1992**). ''Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries,'' J. Phonetics **20**, 1–32.

Nittrouer, S. (**1995**). ''Children learn separate aspects of speech production at different rates: Evidence from spectral moments,'' J. Acoust. Soc. Am. **97**, 520–530.

Nittrouer, S., and Miller, M. E. (**1997a**). ''Developmental weighting shifts for noise components of fricative-vowel syllables,'' J. Acoust. Soc. Am. **102**, 572–580.

Nittrouer, S., and Miller, M. E. (**1997b**). ''Predicting developmental shifts in perceptual weighting schemes,'' J. Acoust. Soc. Am. **101**, 2253–2266.

Nittrouer, S., and Miller, M. E. (**1999**). ''The development of phonemic coding strategies for serial recall,'' Appl. Psycholing. **20**, 563–588.

Nittrouer, S., and Studdert-Kennedy, M. (**1987**). ''The role of coarticulatory effects in the perception of fricatives by children and adults,'' J. Speech Hear. Res. **30**, 319–329.

Nittrouer, S., Studdert-Kennedy, M., and Neely, S. T. (**1996**). ''How children learn to organize their speech gestures: Further evidence from fricative-vowel syllables,'' J. Speech Hear. Res. **39**, 379–389.

Oller, D. K. (**1986**). ''Metaphonology and infant vocalization,'' in *Precursors of Early Speech*, edited by B. Lindblom and R. Zettersröm (Stockton, New York), pp. 21–36.

Slobin, D. I. (**1973**). ''Cognitive prerequisites for the development of grammar,'' in *Studies of Child Language Development*, edited by C. A. Ferguson and D. I. Slobin (Holt, Rinehart, and Winston, New York), pp. 175–208.

Strange, W. (editor). (**1995**). *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (York, Baltimore).

Word recognition in competing babble and the effects of age, temporal processing, and absolute sensitivity[a)]

Karen B. Snell,[b)] Frances M. Mapes, Elizabeth D. Hickman, and D. Robert Frisina
International Center for Hearing and Speech Research, Rochester Institute of Technology, 52 Lomb Memorial Drive, Rochester, New York 14623-5604

(Received 15 August 2001; revised 9 November 2001; accepted 18 April 2002)

This study was designed to clarify whether speech understanding in a fluctuating background is related to temporal processing as measured by the detection of gaps in noise bursts. Fifty adults with normal hearing or mild high-frequency hearing loss served as subjects. Gap detection thresholds were obtained using a three-interval, forced-choice paradigm. A 150-ms noise burst was used as the gap carrier with the gap placed close to carrier onset. A high-frequency masker without a temporal gap was gated on and off with the noise bursts. A continuous white-noise floor was present in the background. Word scores for the subjects were obtained at a presentation level of 55 dB HL in competing babble levels of 50, 55, and 60 dB HL. A repeated measures analysis of covariance of the word scores examined the effects of age, absolute sensitivity, and temporal sensitivity. The results of the analysis indicated that word scores in competing babble decreased significantly with increases in babble level, age, and gap detection thresholds. The effects of absolute sensitivity on word scores in competing babble were not significant. These results suggest that age and temporal processing influence speech understanding in fluctuating backgrounds in adults with normal hearing or mild high-frequency hearing loss. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1487841]

PACS numbers: 43.71.Lz, 43.66.Mk, 43.66.Sr [DOS]

I. INTRODUCTION

This study was designed to clarify whether speech understanding in noise is related to temporal processing as measured by the detection of gaps in noise bursts. Several studies suggest that such a relationship exists. Tyler *et al.* (1982) measured speech recognition in noise with a phonetically balanced, monosyllabic word test. Gap detection thresholds obtained with noise bursts were correlated with word recognition in noise scores in younger normal-hearing and older hearing-impaired subjects. Similarly, Glasberg and Moore (1989) measured gap detection thresholds with noise bursts centered at 0.5, 1.0, and 2.0 kHz and determined speech reception thresholds in noise for subjects with hearing loss. Speech thresholds in noise were significantly related to the detection of gaps in noise bursts (though not to the detection of gaps in tones), whereas speech thresholds in quiet were primarily related to absolute thresholds. Recently, Snell and Frisina (2000) measured the detectability of gaps carried by noise bursts and speech reception thresholds in babble in younger and older adults with normal hearing or mild hearing loss. Gap detection thresholds were related to speech understanding in babble in the younger, but not the older, subjects. In contrast, a recent study by Strouse *et al.* (1998) indicated that speech perception in noise is unrelated to gap detection thresholds obtained with sinusoidal carriers. Gap detection thresholds were obtained at low sensation levels with tone bursts centered at 1 kHz in normal-hearing younger and older adults. Speech perception in noise was measured as the masking level difference for spondees. Correlations between the gap detection thresholds obtained with tone bursts and the masking level difference were not significant.

These disparate findings probably reflect, in part, differences in methodology in both the gap detection paradigm and the measures of speech understanding. However, since both temporal processing (as measured in a gap detection paradigm) and speech understanding in noise show age-related changes, the disparate findings probably also reflect differences in the age of the subjects in the various studies. For example, Barsz *et al.* (2002) recently reported age-related changes in behavioral and neural measures of auditory temporal processing in humans and mice. Age effects included increases in gap detection thresholds and increases in within-group variability as well as decreased saliency of suprathreshold gaps. Three earlier studies also demonstrated age-related deficits in gap detection thresholds in older humans with normal hearing or mild high-frequency hearing loss (Snell, 1997; Strouse *et al.*, 1998; He *et al.*, 1999). In these studies, the greatest age-related deficits in gap detection occurred in complex listening conditions. For example, He *et al.* reported that age had a significant effect on gap thresholds when the placement of the gap was close to carrier onset or offset rather than at the temporal midpoint of the carrier.

Similar age-related deficits in word recognition in noise have been reported. For example, Dubno *et al.* (1984) explored the effects of 12-talker babble on word recognition in 18 younger and 18 older subjects whose audiometric thresholds between 250 and 4 kHz were less than 15 dB HL. They reported that the older subjects required more advantageous speech-to-babble ratios than the young subjects required for equivalent performance. More recently, Stuart and Phillips (1996) studied word recognition in an interrupted broadband

[a)]Portions of this work were presented at the Midwinter Meeting of the Association for Research in Otolaryngology, St. Petersburg Beach, Florida, February, 2001.

[b)]Electronic mail: kbsncp@rit.edu

noise in 12 younger normal-hearing adults and 12 older adults with equivalent word recognition scores in quiet but elevated audiometric thresholds. The effects of noise level and age on word recognition in noise were significant, as was the interaction between noise level and age. That is, decreasing the speech-to-noise ratio had a more deleterious effect on word recognition in older adults than in younger adults. These authors, like Dubno *et al.* (but for different reasons), concluded that the poorer word scores could not be attributed to the small difference in high-frequency sensitivity. The authors suggested that fluctuating, rather than continuous, background noise may be more likely to reveal a temporal processing deficit.

The relationship between speech understanding in noise and temporal processing is further complicated by hearing loss. Several studies have demonstrated that hearing loss influences gap detection. For example, Glasberg *et al.* (1987) studied gap detection in eight older subjects with unilateral hearing loss. The gap detection thresholds for the ears with hearing loss were larger that those for the normal ears even when presented at a level producing asymptotic performance. Florentine and Buus (1984) studied gap detection in six normal hearing (age range=20–50 years) and seven hearing-impaired (age range=20–57 years) subjects. Hearing losses of 50 to 75 dB in the 4- to 8-kHz range often resulted in much poorer gap detection thresholds. However, no orderly decrease in temporal resolution with increasing hearing loss was found. Similarly, numerous studies have demonstrated the effects of hearing loss on speech understanding (e.g., Humes *et al.*, 1994; Frisina and Frisina, 1997).

In summary, the relationship between speech understanding in noise and temporal processing is unclear. It is possible that speech understanding in noise and temporal processing are unrelated. That is, both may decline with age but do so independently. While some studies suggest this may be the case, others suggest otherwise. The disparate results appear to reflect both differences in experimental paradigm and differences in the age and absolute sensitivity of the subjects. The purpose of this study was to clarify whether problems in word recognition in noise experienced by adults with normal hearing or mild hearing loss reflect their ability to detect gaps in noise bursts. To this end, adults spanning a wide age range with normal hearing or mild high-frequency hearing loss were selected as subjects. Conversational levels of open-set word lists (rather than ten selected spondees as in Snell and Frisina, 2000) and competing babble were used to measure word recognition across a range of speech-to-babble ratios. The bandwidth and spectrum level of the noise burst gap carrier approximated those of conversational speech more so than would the bandwidth and spectrum level of a sinusoidal carrier. Finally, the gap stimuli were placed closer to the onset of the carrier (rather than closer to the midpoint as in our previous (Snell and Frisina, 2000) study to simulate characteristics of gaps as they naturally occur at word and phoneme boundaries in speech. Thus, the results could more easily be related to everyday communication problems of adults.

II. METHOD

A. Subjects

Fifty adults between the ages of 18 and 88 years with normal hearing or mild high-frequency hearing loss served as subjects. To facilitate comparisons between the groups, the oldest age at which adults were classified as ''younger'' was selected so that the age ranges for the younger and older groups were comparable. Thus, 22 adults who ranged in age from 18 to 52 years (mean age=31.4 years, s.d. =12.1 years) were classified as ''younger'' subjects and 28 adults who ranged in age from 55 to 88 years (mean age =68.7 years, s.d.=6.5 years) as ''older'' subjects. Subjects were staff, faculty, retired faculty, students, spouses, and alumni recruited from local university communities. Subject were paid for participating. Most (26 of 28) of the older adults had previously participated in temporal processing and speech perception experiments. In contrast, only 12 of the 22 younger adults were experienced subjects.

Potential subjects were screened using the following criteria. A passing score on a screening test of mental competency was required. Pure-tone averages (PTAs) and audiometric thresholds between 0.25 and 2 kHz in the test ear were 25 dB HL or less. Audiometric thresholds at 4 and 8 kHz were no greater than 40 dB HL. Right- and left ear pure-tone averages (PTAs) differed by less than 15 dB and hearing losses were sensorineural as demonstrated by tympanometry and bone conduction thresholds. At 55- and 70-dB HL presentation levels, NU-6 word recognition in quiet scores were 96% or better in the test ear.

Table I contains the mean audiometric thresholds (ANSI, 1989) and standard deviations for the test and nontest ears. Note that the mean thresholds of the test ears of both age groups fall within the range of normal hearing. The mean pure-tone average (PTA) of the younger group was 8.6 dB HL; the mean PTA of the older group was 12.7 dB HL. Also shown in Table I are the high-frequency averages (HFAs) computed by averaging the audiometric thresholds at 4 and 8 kHz. None of the younger subjects (but nine of the older subjects) had HFAs greater than 20 dB HL. The audiometric thresholds and threshold averages for the test ear were analyzed in a series of two-group t-tests (SYSTAT, 1997) to estimate group differences in absolute sensitivity.[1] The results indicated that the audiometric thresholds at 2, 4, and 8 kHz and the high-frequency average (HFA) were significantly poorer in the older group than in the younger group ($p<0.05$).

In summary, the mean PTAs for both groups fell within the range of normal hearing, but significant group differences in absolute sensitivity were present between 2 and 8 kHz. These differences reflected the mild, high-frequency hearing losses in nine of the older listeners as well as the elevated (though still normal) hearing thresholds of the other 20 older subjects.

B. Stimuli and procedure

1. Gap detection task

The gaps were carried by digitally generated, gated noise bursts with low- and high-frequency cutoffs of 10 and

TABLE I. Mean audiometric thresholds (dB HL) and standard deviations (dB) for the subject groups. Pure-tone averages (PTAs) were computed as the average of the audiometric thresholds (dB HL) at 500, 1000, and 2000 Hz. High-frequency averages (HFAs) were computed as the average of the audiometric thresholds (dB HL) at 4000 and 8000 Hz.

		Audiometric threshold (dB HL) Frequency (Hz)							
	Age (y)	250	500	1000	2000	4000	8000	Pure-tone average (dB HL)	High-frequency average (dB HL)
Younger subjects ($N=22$)									
Test ear									
Mean	31.4	9.0	9.4	9.4	7.0	7.5	4.7	8.6	6.1
s.d.	12.1	5.4	4.8	6.3	6.8	6.5	9.3	5.2	6.9
Nontest ear									
Mean		8.6	8.8	5.9	5.3	5.5	3.8	6.6	4.6
s.d.		7.3	5.8	6.1	6.1	5.1	11.3	4.8	7.1
Older subjects ($N=28$)									
Test ear									
Mean	68.7	12.5	12.9	12.0	13.2	17.3	20.5	12.7	18.9
s.d.	6.5	6.6	7.5	6.1	6.3	7.6	11.6	5.5	7.6
Nontest ear									
Mean		14.8	16.4	13.2	15.5	23.4	26.6	15.1	25.0
s.d.		9.0	9.1	9.9	8.2	13.1	15.8	7.8	11.9

6010 Hz (TDT AP1). The noise bursts were transduced by a 16-bit D/A converter (TDT DA1), and attenuated to an overall level of 75 dB SPL (TDT PA4). To minimize the contribution of auditory fibers tuned to high frequencies where thresholds for our subject groups were poorly matched, a high-frequency masker without a temporal gap was gated on and off with the noise bursts. The spectrum level of the high-frequency masker was fixed at 20 dB below the spectrum level of the noise burst and its cutoff frequencies were 6 and 12 kHz. The noise bursts were mixed (TDT SM3) with a continuous white-noise (TDT WG1) floor that had been attenuated (TDT PA 4) to an overall level of 45 dB SPL ($N_O = 4$ dB) with a bandwidth limited by the frequency response of the earphone. Thus, the gaps were partially filled by the continuous white-noise floor as well as the gated high-frequency masker. The output of the mixer was led to a headphone buffer (TDT HB4) and then to a single earphone (Beyer DT48) mounted in a circumaural cushion. Overall stimulus duration (defined as the interval between first and last nonzero points on the carrier waveform) was fixed at 150 ms. The duration of the leading noise burst (defined as the interval between first and last nonzero points on the leading noise burst) was held constant throughout a run at 10 ms. Note the onset of the gap fell close to the first point of full amplitude on the waveform envelope. The rise portion of the leading noise burst and the fall portion of the trailing noise burst were shaped by 10-ms cosine-squared envelopes. Gap onsets and offsets were shaped by 1-ms cosine-squared envelopes to minimize stimulus artifacts. The reference noise bursts in the other intervals also consisted of a 10-ms leading noise burst and a trailing noise burst separated by a small (1 ms), constant, but nondetectable gap. Also shaped by a 1-ms cosine-squared envelope, these gaps in the reference stimuli minimized the possibility that detection could be based on stimulus artifacts unique to the signal. Subjects sat in a double-walled sound booth (Acoustic Systems). Gap thresholds were determined in a three-interval, forced-choice (3IFC) procedure using 75-trial runs. At the beginning of each run, the subjects listened to samples of the reference noise burst paired with a signal noise burst containing a clearly audible gap until familiar with the stimuli. During a run, the duration of the gap was adaptively varied using a three-down, one-up rule to estimate the 79.4-percent point on the psychometric function (Levitt, 1971). The step size was logarithmically varied. Feedback was provided after each response and the interstimulus intervals were 450 ms. Five replications were completed by each subject. The first replication was discarded as practice and the mean of the last four replications was used to estimate the gap detection threshold in ms.

2. Speech task

Commercially available CD recordings (Auditec NU-6 with Four Talker Noise) of the phonetically balanced NU-6 word lists (1A, 2A, 3A, and 4A) and four-talker babble were used as speech materials. The test items are recorded on the first channel of the CD by a male with a General American dialect. Each test word is preceded by the carrier phrase ''Say the word.'' Separate recordings by three female talkers and one male talker were mixed to produce the four-talker babble on a second channel. Peaks in the fine structure of the four-talker babble amplitude waveform were observed to occur randomly every ms or so, whereas modulation peaks in the envelope were observed at random intervals between 50 and 250 ms. Occasional words and phrases spoken by one female and one male talker were intelligible. The subjects were instructed to ignore the babble and to attend to only the test words. The two-channel output of a CD player (Sony CDP-591, Denon DN-C680) was led to two channels of an audiometer (Grason Stadler GSI 61), attenuated, and then presented through insert earphones (Etymotic ER3). The speech signal was presented to the subject's test ear at a fixed presentation level of 55 dB HL. Subjects listened to the

speech signal with their right ear unless the left ear had been used in previous experiments, but the competing four-talker babble was presented in phase to both ears. This listening condition somewhat resembles how a listener might orient a preferred ear towards a speaker in a fluctuating noise background. The babble level was varied among 50-, 55-, and 60-dB HL producing speech-to-babble ratios of +5, 0, and −5, respectively, in the test ear.

Subjects sat in a single-walled sound booth (Acoustic Systems). In each of the three babble conditions they listened to one 50-word, NU-6 list (1A, 2A, 3A, or 4A). The list order and the babble levels were randomized for each subject. Most subjects completed the three lists in a single session of 45 min or less. Word scores in quiet for the two presentation levels were obtained using randomized lists for each subject during the screening process. If a month or more elapsed before collection of the word scores in babble for any subject, the quiet conditions were repeated. In this case, memory effects were minimized by obtaining the word scores in quiet after the babble conditions using a different randomization of the list from the −5-dB S/B babble condition. Word scores in percent correct were transformed into rationalized arcsine units (rau) to make them suitable for parametric statistical analysis (Studebaker, 1985).

III. RESULTS

A. Gap detection in noise

A one-way analysis of variance of the gap detection thresholds was completed with age as the main factor. The results indicated that the main effect of group was significant [$F(1,48)=13.396$, $p<0.05$]. That is, the mean gap detection threshold of the younger age group (5.6 ms) was significantly smaller than the mean gap detection threshold of the older age group (9.6 ms). Correlations among age, gap detection threshold, and high-frequency average within each age group were insignificant except for the correlation between age and high-frequency average in the younger group ($r=0.72$, $p<0.05$).

B. Word recognition in four-talker babble

Mean word scores for the two age groups on the NU-6 word lists are shown in Table II. The two columns to the left contain the mean word scores in percent correct and standard deviations obtained in quiet. At a presentation level of 55 dB HL, the mean scores for the younger and older groups are 99.3 (s.d.=1.5) and 99.5 (s.d.=1.2) percent correct, respectively. At a presentation level of 70 dB HL, the mean scores for the younger and older groups are 99.8 (s.d.=0.9) and 99.0 (s.d.=1.6), respectively. Shown in the three columns on the right in Table II are the mean word scores obtained in the three babble conditions. The mean group word scores range from 24.1 to 79.0 percent correct. Note that even at the lowest (most advantageous) level of background babble, the mean word score of the older group appears to be poorer than that of the younger group. As the babble level increases, the mean word scores appear to decrease in both groups but more so for the older group.

TABLE II. Mean NU-6 word recognition scores (percent correct) and standard deviations for the two age groups. The second and third columns contain the mean scores obtained at speech signal levels of 55 and 70 dB HL in quiet. The three columns to the right contain scores obtained at the same speech signal level of 55 dB HL, but in the presence of three levels (50, 55, and 60 dB HL) of competing four-talker babble.

	Quiet background		Babble background		
	Speech signal presentation level (dB HL)		Four-talker babble level (dB HL)		
	55	70	50	55	60
Younger subjects ($N=22$)					
Mean	99.3	99.8	79.0	63.1	40.5
s.d.	1.5	0.9	6.7	9.4	9.0
Older subjects ($N=28$)					
Mean	99.5	99.0	71.6	51.1	24.1
s.d.	1.2	1.6	9.2	10.1	8.3

Figure 1 shows the relationship between age and the word scores in the three levels of babble. For each subject, age in years is shown on the abscissa and score in percent correct on the ordinate. As the level of the babble increases, the individual word scores appear to decrease regardless of subject age. The Hadi outlier procedure (Hadi, 1994) was used to identify and eliminate potential outliers but no outliers were detected. All three Pearson correlation coefficients between age and word scores were significant ($p<0.05$) and are shown in Fig. 1. Note that the strength of the relationship between word scores and age appears to increase as babble level increases.

The word scores (rau) obtained in background babble were analyzed in a repeated measures ANOVA (SYSTAT, 1997) with one between (age group) and one within (babble level) factor. The results were significant for the main effects of age group [$F(1,48)=43.578$, $p<0.05$] and babble level [$F(2,96)=381.351$, $p<0.05$], and for the interaction between age group and babble level [$F(2,96)=4.117$, $p<0.05$]. *Post hoc* contrasts between groups at each of the three babble levels revealed significant differences ($p<0.05$) between groups at each of the three babble levels.

C. The contributions of temporal and absolute sensitivity to word recognition in babble

A repeated measures analysis of covariance (ANCOVA) identifies the variability in the dependent variable attributable to independent variables or covariates other than the grouping factor (Winer, 1971). To determine the contributions of temporal and absolute sensitivity to word recognition in babble, the analysis was repeated, this time with pure-tone average (PTA), high-frequency average (HFA), and gap detection threshold introduced as covariates. The results of the ANCOVA are shown in Table III. The results were significant for the main effect of age group [$F(1,45)=15.606$, $p<0.05$], the main effect of babble level [$F(2,90)=33.335$, $p<0.05$], and the covariate effect of gap detection threshold [$F(1,45)=5.382$, $p<0.05$]. However, the covariate effect of high-frequency average [$F(1,45)$

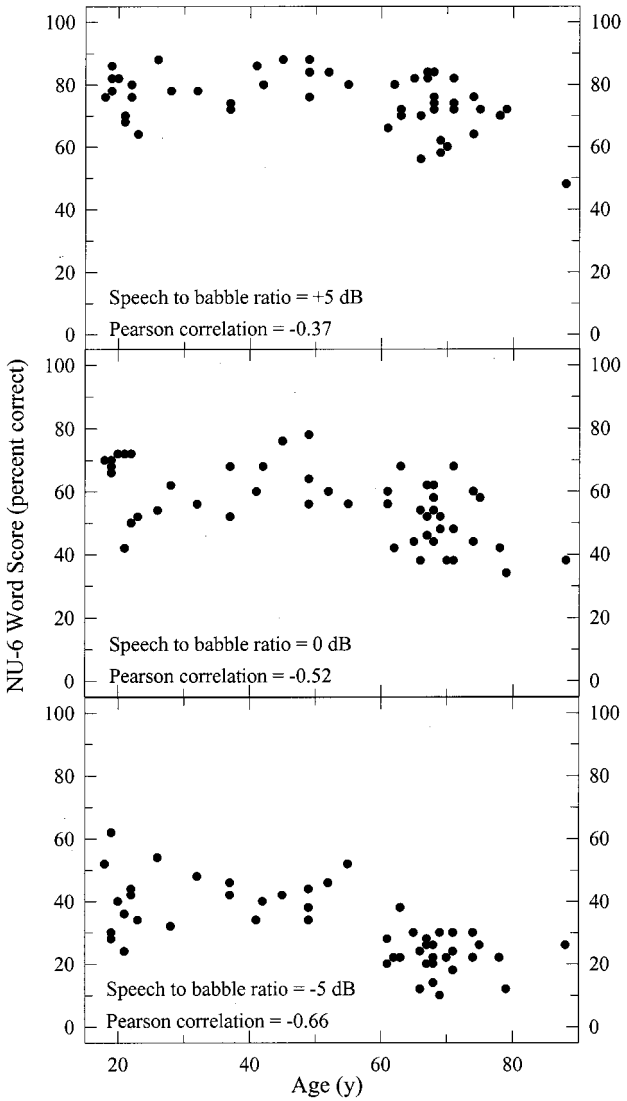


FIG. 1. Three scatter plots for the younger and older subjects showing age in years on the abscissa and NU-6 word scores in percent correct on the ordinate. The NU-6 word lists were presented at an overall level of 55 dB HL. The upper, middle, and lower plots show the effects of competing babble at overall levels of 50, 55, and 60 dB HL, respectively.

$=0.099$, $p>0.05$] and the covariate effect of pure-tone average [$F(1,45)=3.397$, $p>0.05$] were not significant. Also, none of the interactions was significant.

Figure 2 shows the relationship between word recognition for the three babble levels and gap detection thresholds. The younger and older subjects are indicated by open triangles and filled squares, respectively. The Hadi outlier procedure (Hadi, 1994) was used to identify and eliminate potential outliers, but no outliers were detected. The three Pearson correlation coefficients (-0.39, -0.56, and -0.36) were significant ($p<0.05$) when Bonferroni probabilities were used to adjust the probabilities for multiple comparisons.

To summarize, older subjects with normal absolute sensitivity or mild high-frequency hearing loss and excellent speech understanding in quiet demonstrate age-related deficits in word recognition across a range of competing babble levels. Temporal processing, as estimated by gap detection thresholds, contributes significantly to word recognition in babble in these adults with good hearing. Absolute sensitivity does not.

TABLE III. Results of an analysis of covariance (ANCOVA) for the NU-6 word scores (rau). The results for the between-subject factors indicate that the main effect of age group and the covariate effect of the gap detection threshold were significant. The effects of the other two covariates, pure-tone average and high-frequency average, were not significant. The results for the within-subject factors indicate that only the main effect of babble level was significant. None of the interactions was significant.

Source	df	MS	F-ratio	*P*
Between subjects				
Age group	1	1644.691	15.606	<0.05
Gap detection threshold	1	567.255	5.382	<0.05
Pure-tone average	1	358.032	3.397	>0.05
High-frequency average	1	10.413	0.099	>0.05
Error	45	105.389		
Within subjects				
Babble level	2	1960.737	33.335	<0.05
Babble level* Age group	2	150.297	2.555	>0.05
Babble level* Gap detection threshold	2	106.151	1.805	>0.05
Babble level* Pure-tone average	2	4.327	0.074	>0.05
Babble level* High-frequency average	2	24.863	0.423	>0.05
Error	90	58.818		

D. Effects within each age group

A repeated measures analysis was completed on the word scores for the 28 older subjects. Babble level [$F(2,48)=11.090$, $p<0.05$] and gap detection threshold [$F(1,24)=4.928$, $p<0.05$] had significant effects on word scores. Neither the main effects of PTA [$F(1,24)=3.629$, $p>0.05$] and HFA [$F(1,24)=0.657$, $p>0.05$] nor any of the interactions were significant. Correlation coefficients between the word scores at the three babble levels and the pure-tone average (PTA) and high-frequency average (HFA) ranged from -0.34 to -0.11 for the older group; none of the coefficients achieved significance [$p>0.05$].

A similar repeated measures analysis on the word scores for the 22 younger subjects was performed. Babble level had a significant effect [$F(2,36)=46.003$, $p<0.05$] on word scores. However, the effects of gap detection threshold [$F(1,18)=0.903$, $p>0.05$], PTA [$F(1,18)=0.858$, $p>0.05$], and HFA [$F(1,18)=1.323$, $p>0.05$] were not significant. Also, none of the interactions was significant. Correlation coefficients between the word scores at the three babble levels and the PTA and HFA ranged between -0.40 and 0.38 for the younger group and none of the coefficients was significant [$p>0.05$].

E. Summary of results

In this sample of adults with normal hearing or mild, high-frequency hearing loss, word recognition scores at moderate and high conversational levels in a quiet background were excellent. Individual word scores in quiet varied between 96- and 100-percent correct at conversational speech levels. Adding babble at three different levels resulted in poorer word recognition for younger and older age groups. Results of a two-way repeated measures analysis of variance

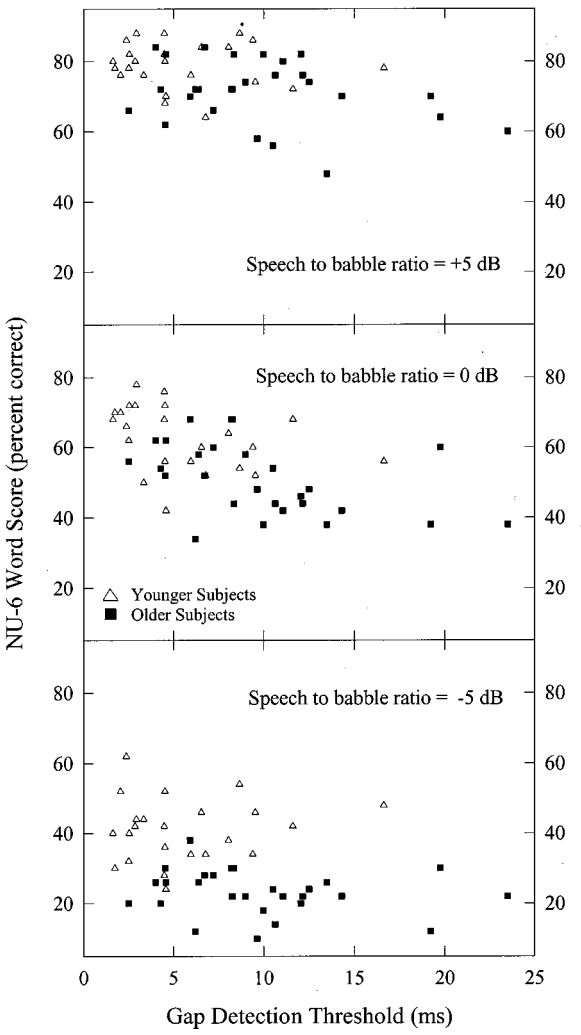


FIG. 2. Three scatter plots for the younger and older subjects showing gap detection thresholds in milliseconds on the abscissa and NU-6 word scores in percent correct on the ordinate. The open triangles represent younger subjects; the filled squares represent older subjects. The NU-6 word lists were presented at an overall level of 55 dB HL. As in Fig. 1, the upper, middle, and lower plots show the effects of competing babble at overall levels of 50, 55, and 60 dB HL, respectively.

indicated that the level of babble had a significant effect on word scores, as did age group and the interaction between age group and level. When the audiometric and gap detection thresholds were added to the model as covariates, the effect of gap detection was significant but the effects of the threshold averages were not. When the scores of the older subjects were analyzed separately, gap detection threshold was a significant factor in predicting speech understanding in babble, whereas absolute sensitivity was not. In contrast, when the scores of the younger subjects were analyzed separately, neither gap detection threshold nor absolute sensitivity were significant factors in predicting speech understanding in babble. These results suggest that in adults with normal speech understanding in quiet, difficulties in understanding speech in background babble can be attributed partially to deficits in temporal processing.

IV. DISCUSSION

While many of the presumed age-related deficits in speech understanding in quiet probably reflect losses in absolute sensitivity, other factors may contribute to poorer speech understanding in noise (e.g., Bacon *et al.*, 1994; Divenyi and Haupt, 1997). Several studies (Glasberg and Moore, 1989; van Rooij *et al.*, 1992) suggest that speech recognition in quiet and speech recognition in noise are only loosely related. If so, important factors underlying speech recognition in quiet and in noise may differ. In this study, gap detection thresholds obtained with noise-burst carriers were shown to point to a subject's ability to process transient, temporal aspects of speech in background babble. The strength of this relationship may reflect the use of a noise burst as the carrier in the gap task, the placement of the gap close to noise-burst onset, the use of a fluctuating background in the speech task, and the use of relatively large samples of younger and older subjects with good or excellent hearing.

A. The role of carrier bandwidth and other stimulus characteristics

As several investigators have observed, gaps in tonal carriers bear little resemblance acoustically to naturally occurring gaps in speech. In contrast, noise-like carriers often mark gaps in speech, and so an individual's ability to detect gaps in noise is more likely to point to his ability to detect gaps in speech. The results of this study are consistent with those of three previous studies using noise-burst carriers (Tyler *et al.*, 1982; Glasberg and Moore, 1989; Snell and Frisina, 2000). The stronger relationship between absolute sensitivity and speech perception in noise noted in the two earlier studies (but not in this study) reflects the much wider range in audiometric thresholds of those subjects.

While similar in many respects to our earlier study (Snell and Frisina, 2000), considerable methodological differences are present. For example, in the previous study, the gap was placed at the midpoint of the noise-burst carrier whereas here the gap was placed close to the carrier onset. Consistent with the results of Snell and Hu (1999), the mean gap thresholds of the older age subjects exceeded those of the younger subjects by 4 ms at this placement (compared to less than 1 ms in the Snell and Frisina, 2000 study). This additional deficit in temporal processing for older subjects seen close to the waveform onset may have contributed to the stronger relationship between temporal and speech processing demonstrated for the older subjects in this study. Also, in the previous study the speech items were a closed set of 11 spondees and the level of babble was adaptively varied. This resulted in highly disadvantageous speech-to-babble ratios (-12 to -4-S/B) that fall largely outside the range of presentation levels used in this study. Finally, note the different temporal characteristics of the babble. The four-talker babble used in this study generated more modulation in the amplitude envelope compared to the 12-talker babble in the previous study. The relatively higher word scores of the younger subjects may reflect their taking more advantage of the ''dips'' in the background babble.

B. The roles of hearing loss and age

When the effects of age are studied in younger and older subjects with similar audiograms, absolute sensitivity is

minimized as a confounding factor. Similarly, when the effects of hearing loss are studied in a within-subjects design with unilateral hearing losses, age is eliminated as a confounding factor. In contrast, when gap detection thresholds of older subjects with significant hearing loss are compared to those of younger subjects with normal hearing, the independent effects of age and hearing loss are hard to discern.

V. CONCLUSIONS

Twenty years ago, Dirks *et al.* (1982) noted that most conversations do not occur in quiet, but rather in a variety of complex backgrounds. When experimental speech tasks are performed in quiet or in a constant background noise, generalizing from research findings to problems in everyday communication may be difficult. While conversations in background noise fluctuate in overall level, average background noise levels in indoor public places (e.g., stores, hospitals, parties, and classrooms) are commonly between 45 and 65 dB(A) (Pearsons, Bennett, and Fidell, 1977; Webster and Snell, 1983) with the ratio between conversational speech and competing speech babble often maintained between −5 dB and +5 dB or better (Webster and Klumpp, 1962; Pearsons, Bennett, and Fidell, 1977). Evaluating performance at higher levels of competing babble or in much poorer signal-to-noise ratios may be less helpful in understanding everyday problems simply because in such disadvantageous conditions even adults with excellent hearing quickly give up trying to maintain a conversation. Also, when compressed or distorted speech utterances are used as speech materials, generalizing from research findings to problems in everyday communication can be difficult. In contrast, in this study the speech items and background babble were natural speech productions presented at typical conversational levels. It follows that these results indicating that deficits in speech understanding are related to temporal processing have implications for problems in everyday communication.

In this study we show that in adults with normal or near-normal hearing, speech processing in fluctuating backgrounds is related to the ability to detect gaps carried by a noise burst with an upper cutoff frequency of 6 kHz. Other studies, however, suggest that significant hearing loss may weaken the relationship between temporal processing and speech perception in noise. Unfortunately, the mechanism of this interaction remains unclear. A more complete understanding of speech processing in fluctuating backgrounds requires continued consideration of how hearing loss interacts with temporal processing in temporally complex backgrounds. Finally, much of the variability in word recognition scores in fluctuating backgrounds remains unexplained even after the influences of age, temporal processing, and absolute sensitivity are considered. Other factors such as contralateral suppression, spatial hearing, and memory and attention deserve continued research.

ACKNOWLEDGMENTS

This research was supported by the Rochester International Center for Hearing and Speech Research and a grant from NIA (No. P01-AG09524). We are grateful to the editor and the reviewers for their insightful comments. Portions of this work were presented at the Midwinter Meeting of the Association for Research in Otolaryngology, St. Petersburg Beach, Florida, February 2001.

[1]Throughout this study, separate variance t-tests or the Greenhouse–Geiser (G–G) statistic were used when necessary to adjust the probabilities for failure of the compound symmetry assumption. Similarly, Bonferroni probabilities were used to adjust the probabilities for multiple comparisons.

ANSI (**1989**). ANSI S3.6-1989, ''Specifications for audiometers'' (American National Standards Institute, New York).

Auditec NU-6 with Four Talker Noise, Cafeteria Noise, and Speech Spectrum Noise. (**1984**). Auditec of St. Louis, 2515 South Big Bend Blvd., St. Louis, MO 63143.

Bacon, S. P., Opie, J. M., and Montoya, D. Y. (**1994**). ''Speech recognition thresholds in temporally complex backgrounds: Effects of hearing loss and noise masking,'' J. Acoust. Soc. Am. **95**, 2993.

Barsz, K., Ison, J. R., Snell, K. B., and Walton, J. P. (**2002**). ''Behavioral and neural measures of auditory temporal acuity in aging humans and mice,'' Neurobiol. Aging (in press).

Dirks, D. D., Morgan, D. E., and Dubno, J. R. (**1982**). ''A procedure for quantifying the effects of noise on speech recognition,'' J. Speech Hear Disord. **47**, 114–123.

Divenyi, P. L., and Haupt, K. M. (**1997**). ''Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss. III. Factor representation,'' Ear Hear. **18**, 189–201.

Dubno, J. R., Dirks, D. D., and Morgan, D. E. (**1984**). ''Effects of age and mild hearing loss on speech recognition in noise,'' J. Acoust. Soc. Am. **76**, 87–96.

Florentine, M., and Buus, S. (**1984**). ''Temporal gap detection in sensorineural and simulated hearing impairments,'' J. Speech Hear. Res. **27**, 449–455.

Frisina, D. R., and Frisina, R. D. (**1997**). ''Speech recognition in noise and presbycusis: Relations to possible neural mechanisms,'' Hear. Res. **106**, 95–104.

Glasberg, B. R., and Moore, B. C. J. (**1989**). ''Psychoacoustic abilities of subjects with unilateral and bilateral cochlear impairments and their relationship to the ability to understand speech,'' Scand. Audiol. Suppl. **32**, 1–25.

Glasberg, B. R., Moore, B. C. J., and Bacon, S. P. (**1987**). ''Gap detection and masking in hearing-impaired and normal-hearing subjects,'' J. Acoust. Soc. Am. **81**, 1546–1556.

Hadi, A. S. (**1994**). ''A modification of a method for the detection of outliers in multivariate samples,'' J. R. Stat. Soc. Ser. B. Methodol. **56**, No. 2, 393–396.

He, N., Horwitz, A. R., Dubno, J. R., and Mills, J. H. (**1999**). ''Psychometric functions for gap detection in noise measured from young and aged subjects,'' J. Acoust. Soc. Am. **106**, 966–978.

Humes, L. E., Watson, B. U., Christensen, L. A., Cokely, C. G., Halling, D. C., 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.

Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Pearsons, K. S., Bennett, R. L., and Fidell, S. (**1977**). ''Speech levels in various noise environments,'' U.S. Environmental Protection Agency Report, EPA-600/1-77-025.

Snell, K. B. (**1997**). ''Age-related changes in temporal gap detection,'' J. Acoust. Soc. Am. **101**, 2214–2220.

Snell, K. B., and Hu, H.-L. (**1999**). ''The effect of temporal placement on gap detectability,'' J. Acoust. Soc. Am. **106**, 3571–3577.

Snell, K. B., and Frisina, D. R. (**2000**). ''Relationships among age-related differences in gap detection and word recognition,'' J. Acoust. Soc. Am. **107**, 1615–1626.

Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, D. W. (**1998**). ''Temporal processing in the aging auditory system,'' J. Acoust. Soc. Am. **104**, 2385–2399.

Stuart, A., and Phillips, D. P. (**1996**). ''Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-hearing, and presbyacusic listeners,'' Ear Hear. **17**, 478–489.

Studebaker, G. A. (**1985**). ''A rationalized arcsine transform,'' J. Speech Hear. Res. **28**, 455–462.

SYSTAT for Windows: Statistics, Version 7.0 edition, SPSS, Inc., Chicago, IL, 1997.

Tyler, R. S., Summerfield, Q., Wood, E. J., and Fernandes, M. A. (**1982**). ''Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **72**, 740–752.

van Rooij, J. C. G. M., and Plomp, R. (**1992**). ''Auditive and cognitive factors in speech perception by elderly listeners. III. Additional data and final discussion,'' J. Acoust. Soc. Am. **91**, 1028–1033.

Webster, J. C., and Klumpp, R. G. (**1962**). ''Effects of ambient noise and nearby talkers on a face-to-face communication task,'' J. Acoust. Soc. Am. **34**, 936–941.

Webster, J. C., and Snell, K. B. (**1983**). ''Noise levels and the speech intelligibility of teachers in classrooms,'' J. Acad. Rehab. Aud. **16**, 234–255.

Winer, B. J. (**1971**). *Statistical Principles in Experimental Design* (McGraw-Hill, New York).

Captive dolphins, *Tursiops truncatus*, develop signature whistles that match acoustic features of human-made model sounds

Jennifer L. Miksis
Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543 and Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island 02882

Peter L. Tyack
Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

John R. Buck
Department of Electrical and Computer Engineering & School for Marine Science and Technology, University of Massachusetts Dartmouth, 285 Old Westport Road, North Dartmouth, Massachusetts 02747-2300 and Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 30 May 2001; revised 7 April 2001; accepted 28 May 2002)

This paper presents a cross-sectional study testing whether dolphins that are born in aquarium pools where they hear trainers' whistles develop whistles that are less frequency modulated than those of wild dolphins. Ten pairs of captive and wild dolphins were matched for age and sex. Twenty whistles were sampled from each dolphin. Several traditional acoustic features (total duration, duration minus any silent periods, etc.) were measured for each whistle, in addition to newly defined flatness parameters: total flatness ratio (percentage of whistle scored as unmodulated), and contiguous flatness ratio (duration of longest flat segment divided by total duration). The durations of wild dolphin whistles were found to be significantly longer, and the captive dolphins had whistles that were less frequency modulated and more like the trainers' whistles. Using a standard t-test, the captive dolphin had a significantly higher total flatness ratio in 9/10 matched pairs, and in 8/10 pairs the captive dolphin had significantly higher contiguous flatness ratios. These results suggest that captive-born dolphins can incorporate features of artificial acoustic models made by humans into their signature whistles. [DOI: 10.1121/1.1496079]

PACS numbers: 43.80.Ka, 43.80.Ev, 43.60.Qv [WA]

I. INTRODUCTION

Bottlenose dolphins, *Tursiops truncatus*, produce frequency-modulated signals called whistles. Caldwell and Caldwell (1965) presented data suggesting that each dolphin within a captive group produced an individually distinctive whistle, which they called a ''signature whistle.'' Four independent research groups have studied signature whistles in a total of 132 captive (Caldwell *et al.*, 1990; Janik *et al.*, 1994; Janik and Slater, 1998) and 90 wild dolphins (Sayigh *et al.*, 1990, 1995, 1999; Smolker *et al.*, 1993) for a total of 222 individuals. The setting in which it has been easiest to identify whistles involves temporarily isolating one dolphin. In this setting, the isolated dolphin tends to produce a long series of sound, repeating a whistle over and over. Human judges rate spectrograms of these whistles as similar within an individual's repertoire and highly distinctive across individuals.

In the 1960s, when David and Melba Caldwell initially introduced the signature whistle hypothesis, they believed that each dolphin was only capable of producing one kind of stereotyped whistle, and they called any whistle other than an individual's signature whistle an ''aberrant'' whistle. However, in settings other than isolation, dolphins produce many other whistle types, such as simple upsweeps, downsweeps, and sinusoidal patterns of modulation, along with variable contours that may not be repeated precisely (Tyack, 1986; Janik and Slater, 1998). The tendency for signature whistles to be produced during isolation suggested to Janik and Slater (1998) that these whistles function as contact calls.

Sayigh *et al.* (1990) provide evidence that bottlenose dolphins produce signature whistles that develop within a few years of age, and that are stable for decades. At the same time, however, adult dolphins retain remarkable abilities to imitate manmade whistle-like sounds (Caldwell and Caldwell, 1972; Richards *et al.*, 1984). Bottlenose dolphins may spontaneously imitate sounds within a few seconds after the first exposure (Herman, 1980), or after only a few exposures (Reiss and McCowan, 1993). Dolphins can also be trained to imitate manmade whistle-like sounds (Evans, 1967; Richards *et al.*, 1984; Sigurdson, 1993). After experience with imitation training, some dolphins have learned to imitate a model sound immediately after it was first presented. Once a dolphin learns to imitate a sound, the imitation can be incorporated into its vocal repertoire, and the dolphin can produce the sound even when it does not hear the model. This means that dolphins have an open vocal repertoire that can change throughout the lifespan.

Little is known, however, about the role of imitation or vocal matching in the development of signature whistles.

Most terrestrial mammals appear to inherit acoustic features of their calls (Janik and Slater, 1997). There is usually a strong similarity between calls of parents and offspring, even if they are cross fostered (Owren *et al.*, 1993). By contrast, most dolphin calves develop signature whistles quite different from their parents. Sayigh *et al.* (1995) found that roughly half of bottlenose dolphin sons develop whistles similar to their mothers, but nearly all daughters and the other half of sons develop different whistles. The lack of similarity in signature whistles of dolphins and their mothers couples with evidence that young dolphins may develop whistles similar to those of a foster mother to suggest that the auditory environment may be more important than inheritance in determining the structure of a dolphin's signature whistle (Tyack and Sayigh, 1997).

Some of the most convincing data for vocal learning in whistle ontogeny stem from comparing the whistles of calves to sounds present in their acoustic environments. Tyack and Sayigh (1997) discuss evidence that two dolphins born in a captive community pool developed whistles very similar to a whistle used by trainers. Most dolphin trainers use a dog whistle to signal a dolphin that it can approach for food after performing a requested task. The sound from these dog whistles is a high-frequency tone with little frequency modulation.

This paper examines the hypothesis that whistle development in captive calves may be influenced by exposure to these flat unmodulated whistles used by trainers. We matched pairs of captive and wild dolphins by age and sex and compared the modulation of whistles produced by these two groups in order to test whether dolphins born in captivity are more likely than wild dolphins to have whistles with little frequency modulation like the unmodulated trainer's whistle. Dolphins were analyzed as pairs matched for age and sex because Caldwell *et al.* (1990) report a significant increase in frequency modulation with age in captive bottlenose dolphins, and because Sayigh *et al.* (1995) report strong differences in mother–offspring comparisons of signature whistles of sons vs daughters.

II. METHODS

A. Whistle recordings

Whistles from wild-born dolphins were recorded from a population of Atlantic bottlenose dolphins off the coast of Sarasota, Florida, during capture–release sessions. For a complete description of the study site and the Sarasota capture–release projects, see Scott *et al.* (1990) and Sayigh *et al.* (1990). Whistles used in this study were obtained from recordings that cover a 7-year period from 1986–1992. Whistles from the Sarasota population were recorded while the dolphin was placed on a raft for processing. During this time a suction cup hydrophone (designed by Tyack; Tyack, 1985) was affixed to the animal's head slightly behind the blowhole. Recordings made prior to 1989 were made with a Sony TC-D5 or a Marantz PMD-430 stereo tape recorder and used Maxell UDXLII cassette tapes. This system had a frequency response of 30–15 000 Hz. Recordings made after 1989 were made with a Panasonic AG-6400 stereo hi-fi VCR using standard VHS tapes, and the frequency response measured 20–20 000 Hz.

Whistles from captive-born dolphins were recorded at the Miami Seaquarium over a span of 4 years, 1988–1991. Eight individuals were recorded at the aquarium's ''Top Deck'' pool, and two individuals were recorded at the ''Flipper'' lagoon. Top Deck is a cylindrical outdoor pool approximately 6 meters in depth and 27 meters in diameter. The Flipper lagoon is a manmade seawater pool which extends into the Biscayne Bay. Over the course of the study the number of dolphins in Top Deck ranged from 10–14 individuals. Five dolphins were present in the Flipper lagoon. All animals were born in Top Deck. Adults in both pools performed daily shows for the public, and all the dolphins were habituated to interactions with trainers.

Recordings of the Miami Seaquarium dolphins were made using a Realistic hi-fi VHS recorder and Scotch T-120 cassettes or with a Sony TCD3M stereo cassette recorder and Maxell UDXLII tapes. One channel recorded underwater sounds from a modified U.S. Navy sonobuoy mounted in the pool. A microphone fed into the second channel for comments. The frequency response was limited by the hydrophone in both systems and was approximately 100–15 000 Hz. The whistles from the wild and captive dolphins were recorded with different equipment, but comparisons of whistle contours made with contact hydrophones in air vs far-field underwater show little difference in the contour.

Dolphins in Top Deck were viewed from either a floating platform or an underwater window during recording sessions. Animals in the Flipper pool were observed from a dock during recordings. Dolphins producing sound were identified through synchronized blowhole movement with the onset of whistle production, bubble streams from the blowhole during the whistle duration, or whistles audible at the surface of the water that could be localized in air. While signature whistles are typically recorded in an isolation setting, our recordings of whistles identified from captive dolphins swimming within large community pools in the Miami Seaquarium were consistent with signature whistles in that each animal was identified producing a distinctive contour.

B. Analysis techniques

Captive and wild-born dolphins were matched for age and sex and totaled ten pairs (Table I). Half of the pairs were male, half were female, and the ages ranged from 1–35 years. We used a sample of 20 signature whistles from each of these 20 dolphins. The sample size of dolphins was restricted by the number of dolphins that we had recorded and which were born in a facility where we knew that the dolphin had heard trainers' whistles as it developed its signature whistle. The number of whistles from each dolphin was limited by the difficulty of identifying whistles of captive dolphins using the methods employed in this paper. Weeks of observational effort often yielded only a few tens of whistles identified to a particular individual. The maximum number of whistles that could be identified from three of the captive animals was 20; hence, the sample size of 20 whistles was selected to maintain uniformity throughout the study.

TABLE I. Age, sex, and animal identifications of the ten pairs of wild and captive dolphins matched for age and sex.

Animal name	Year of birth	Captive dolphin code	Sex	Age (years)	Wild dolphin code	Year of birth	Animal ID number
Samantha	1988	C1	F	2	W1	1986	FB55
Noel	1988	C2	M	2	W2	1987	FB22
Dawn	1975	C3	F	15	W3	1971	FB90
Dancer	1980	C4	F	10	W4	1979	FB25
Ivan	1983	C5	M	6	W5	1980	FB32
Tori	1990	C6	F	1	W6	1986	FB1
Bebe	1956	C7	F	35	W7	1953	FB153
Sundance	1990	C8	M	1	W8	1988	FB50
PJ	1990	C9	M	1	W9	1988	FB2
Shadow	1985	C10	M	4	W10	1985	FB150

Captive–wild pairs were constructed by randomly choosing a wild counterpart for each captive subject from an established database of Sarasota recordings (Sayigh *et al.*, 1995). To prevent bias, pairs were selected before any spectrograms were made. The initial selection and pairing of individuals remained constant throughout the study. Spectrograms were created from whistles digitized on a Kay Elemetrics Corp. model 5500 digital signal processing system at 81 920 Hz with an upper frequency limit of 32 kHz and a dynamic range setting of 42 dB. Spectrograms of whistles from the Sarasota recordings were made from the tapes at a rate of approximately one whistle every 60 s. No such sampling protocol was established for whistles from the captive dolphins because only a limited number of whistles were identified from captive individuals.

Five frequency parameters were measured by hand from the spectrograms (Table II). Minimum, maximum, start, and end frequencies were measured directly from the spectrogram for each wild and captive whistle. Frequency range was calculated from the difference of maximum and minimum frequencies. Three different trainer's whistles were also measured for the same parameters. Each frequency and duration parameter for wild and captive pairs presented in Table II reflects the mean of 20 whistles. Values for the trainers' whistles were based on four recordings of the first kind of whistle used for training and two recordings each for the second and third kinds of trainers' whistles. The total $n=8$ for the trainers' whistles reflects an exemplar from each of the eight trainers using whichever of the three kind of whistles they habitually used.

Contours of the fundamental frequency were obtained using a previously reported contour extractor for *Tursiops truncatus* whistles (Buck and Tyack, 1993). This contour extractor computes the short-time Fourier transform (STFT) (Oppenheim *et al.*, 1999) for each time interval, or block, of the whistle. The contour extractor chooses the fundamental frequency for each block independently of the other blocks. The fundamental frequency is generally chosen to be the frequency with the strongest energy, except for a few heuristic rules to handle brief segments when the second harmonic contained more energy than the fundamental. These segments most frequently occurred during the start or end of a whistle.

For all the whistles in this study, the STFT blocks were 1024 samples long, corresponding to 12.5 ms, and there was no overlap between adjacent blocks. The FFT size was also 1024 points, yielding an FFT bin size of 80 Hz. Figure 1 shows the time-domain signal, spectrogram, and the extracted contour for three signals—one wild animal whistle, one captive animal whistle, and one trainer's whistle. Since the whistles from wild dolphins were recorded using a contact hydrophone, they tended to have a high signal-to-noise ratio (SNR). Whistles from the captive dolphins tended to

TABLE II. Frequency parameters of whistles measured by hand from the whistle contours. The values for each dolphin are based on the mean of $n=20$. The values for the trainers whistles represent means for each of the three different trainer whistles, as well as a group mean for all of the trainers' whistle recordings ($n=8$).

	Minimum frequency (Hz)			Maximum frequency (Hz)			Start frequency (Hz)			End frequency (Hz)			Frequency range (Hz)		
Pair #	Trainer	Captive	Wild	Trainer	Captive	Wild	Trainer	Captive	Wild	Trainer	Captive	Wild	Trainer	Captive	Wild
1	4720	3920	4400	6 880	12 800	14 560	4880	6 000	4 640	5280	4 640	8 880	2160	8 880	10 160
2	6560	3200	8080	9 600	13 040	22 880	6720	4 720	8440	6720	4 240	22 160	3040	9 840	14 800
3	7680	2720	4240	10 560	16 960	17 760	7840	4 880	5520	7840	3 120	13 360	2880	14 240	13 520
4		5360	7040		17 120	22 640		8 240	11 520		6 320	18 880		11 760	15 600
5		6000	3920		17 520	17 360		9 360	7040		14 320	21 440		11 520	13 440
6		4080	5840		8 880	17 920		4 880	6640		4 560	17 680		4 800	12 080
7		3920	3520		17 280	20 800		4 720	3920		8 480	9 040		13 360	17 280
8		4320	4080		8 720	14 240		4 720	4320		5 520	8 720		4 400	10 160
9		4000	4880		10 640	16 640		4 800	7680		5 040	6 560		6 640	11 760
10		5760	4240		17 620	17 360		10 800	6880		15 760	8 960		11 760	13 120
Average	**5920**	**4328**	**5024**	**8480**	**14 048**	**18 216**	**6080**	**6 312**	**6660**	**6280**	**7 200**	**13 568**	**2560**	**9 720**	**13 192**

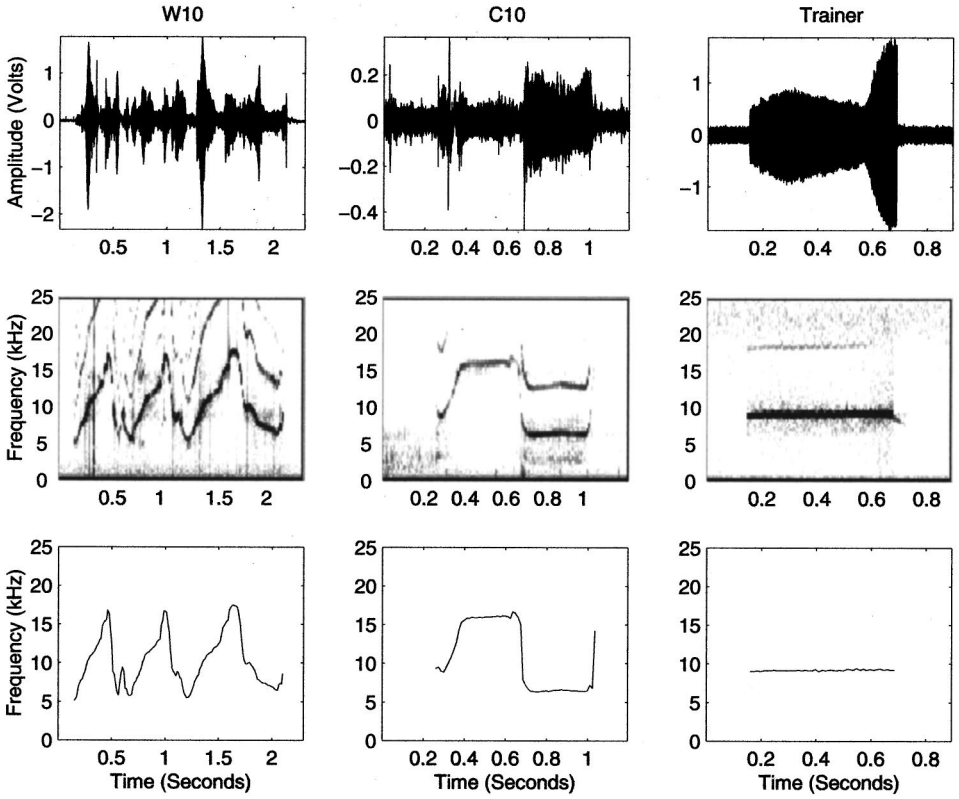


FIG. 1. Wave form (top row), spectrogram (middle row), and contour (bottom row) for a randomly selected whistle from animal W10 (left column), animal C10 (middle column), and the trainer (right column). Note the similarities in the spectrogram and contour between C10 and the trainer.

have a lower SNR. For all three signals, the extracted contour follows the fundamental contour of the whistle.

C. Estimating modulation rate of whistle contours

This study focused on flatness, or the lack of frequency modulation, in signature whistles. A lack of modulation corresponds to a small or zero slope in the extracted contour of frequency vs time. We defined a time interval as containing an unmodulated or flat whistle segment if the slope of the contour over this interval was below a chosen modulation threshold μ. The method for choosing μ is described in more detail below. Practically, the measure of flatness implemented for comparison to μ is the difference of frequencies in adjacent time bins divided by the time interval between the bins. The resolution of this estimate is limited by the FFT bin size (80 Hz) divided by the STFT block length (12.5 ms). The minimum change in frequency that this approach could detect is 80 Hz in 12.5 ms or 80 Hz/0.0125 s=6400 Hz/s.

This difference of frequencies approximates the slope or time derivative of the frequency contour. This approximation is necessary, because the derivative of a discrete-time signal is, strictly speaking, not well defined. Theoretically, the best estimate of the derivative would be to sample the derivative of the bandlimited interpolation of the original contour, also called bandlimited differentiation. For this study, the first backward difference (Oppenheim *et al.*, 1999) was used as an approximation to this bandlimited derivative. Specifically, if $x[n]$ is the sampled frequency contour for $0 \leqslant n \leqslant N-1$, the derivative signal $y[n]$ is defined to be

$$y[n] = \frac{x[n]-x[n-1]}{\Delta T}, \tag{1}$$

for $1 \leqslant n \leqslant N-1$, where $\Delta T = 1024/81\,920 = 12.5$ ms is the time interval between the samples of the contour. The backward difference is not in general a good approximation to a bandlimited differentiator. However, if the signal $x[n]$ was highly oversampled such that most of the energy in its discrete-time Fourier transform $X(e^{j\omega})$ is concentrated at low frequencies, the backward difference is a reasonable approximation to bandlimited differentiation. The contours studied in this paper are sufficiently oversampled for this approximation to be valid. The errors introduced at moderate to high ω, discrete-time frequency, will not be significant. Once the slope is above our modulation threshold, the segment is considered to be modulated, and not flat. The specific value of the slope is not of interest.

The durations of whistle contours were also measured in order to calculate the ratio of flat segment durations to the duration of the whole whistle. Two different whistle durations were measured. The total duration was measured from the beginning of the whistle to the last element of the contour. Many whistles such as those from C4 and W6 in Fig. 2 had silent gaps within the whistle. A ''gap-free'' duration was calculated by subtracting the duration of these gaps from the total duration.

Two quantitative measures of the amount of the whistle which is unmodulated are computed from the backward difference $y[n]$. The first, referred to as total flatness, is the ratio of flat segment durations to total gap-free duration; total flatness can also be defined as the fraction of the values of $|y[n]| \leqslant \mu$. Intuitively, this total flatness measure quantifies the percentage of the duration of the whistle which consists of unmodulated signal. In the extreme case of a wholly unmodulated signal, this number should be 1. The second mea-

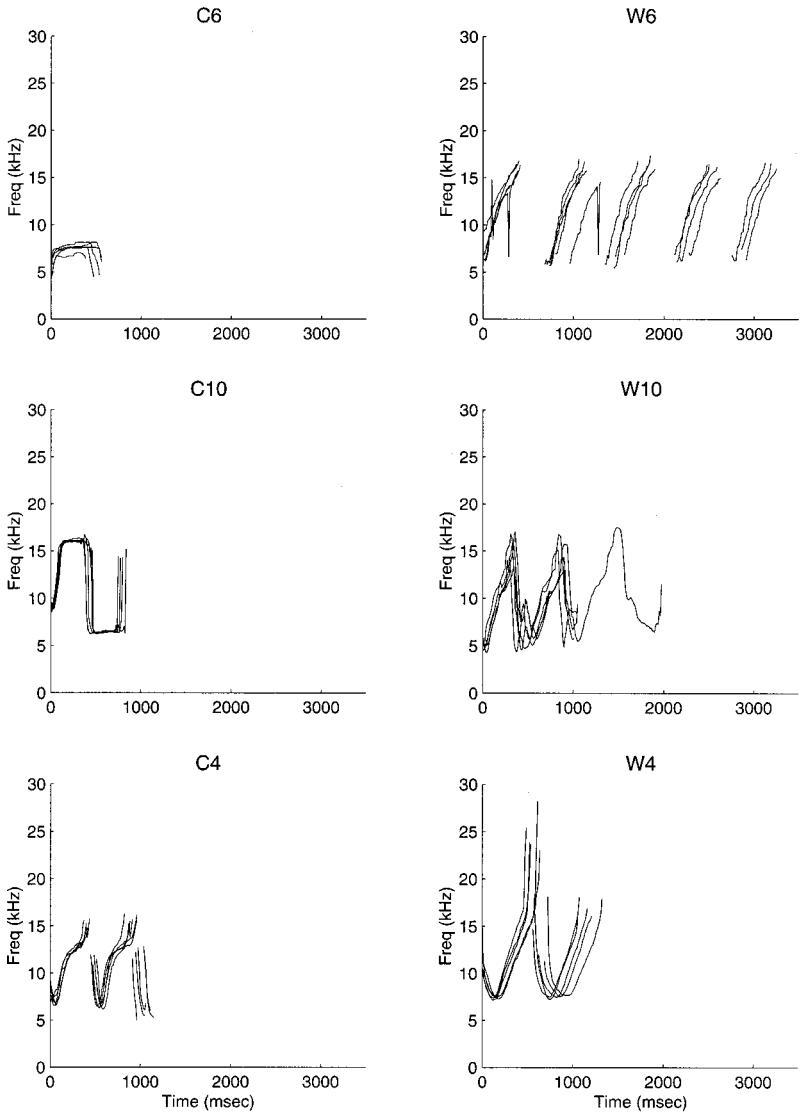


FIG. 2. Contour comparisons for three captive–wild matched pairs. Each graph contains five randomly selected contours from the specified animal. Out of the ten pairs studied, C4–W4 is the only pair in which the wild animal has a longer flat segment in its signature whistle.

sure, referred to as contiguous flatness, is the ratio of the duration of the longest contiguous segment with $|y[n]| \leqslant \mu$ to the total gap-free duration of the $y[n]$. This measure quantifies the tendency to make whistles with one long unmodulated segment, opposed to several short unmodulated segments. A pathological whistle which gives $y[n]=2\mu$ for n odd and $y[n]=0$ for n even would give roughly 0.50 for the total flatness, but only $1/N$ for the contiguous flatness, since all the flat segments are only one sample long. For the signals analyzed in this study, these measures are generally correlated, with signals scoring high in one metric also scoring high in the other.

D. Selection of flatness threshold

The average of both the total flatness and contiguous flatness ratios over 20 whistles were computed as a function of μ for six animals (three pairwise comparisons) to determine an appropriate choice of threshold. The results of these comparisons are shown in Fig. 3. As expected, all of these figures show the proportion of whistle duration monotonically increasing with increasing threshold μ. The proportion of whistle duration judged as flat asymptotically approaches 1 as μ gets very large, indicating that with a very broad threshold, all of almost every signal is considered flat. The trainers' whistles are more flat than most of the animals at almost any nonzero threshold.

For both the C6–W6 and C10–W10 comparisons in Fig. 3, the captive animal ranks higher than the wild animal on both flatness metrics for any choice of the threshold μ. Consequently, the results of this study are robust to different choices of threshold for a reasonable range of μ. The C10–W10 comparison data match the visual impressions obtained in Fig. 1. Specifically, the trainer's whistle is very flat, the captive whistle mostly flat, and the wild whistle highly modulated. Note that for $\mu > 3\times10^4$, even a highly modulated whistle like W10 in Fig. 1 scores over 0.5 on the ratio of flat duration test. This indicates that choices of μ in this range are insufficiently discriminatory of features which appear unmodulated in the spectrograms and contours shown in Fig. 1. The C4–W4 comparison indicates the signals are roughly comparable in flatness over most meaningful values of μ, with a slight edge in favor of the captive animal in the total flatness. However, it is clear from the bottom row of

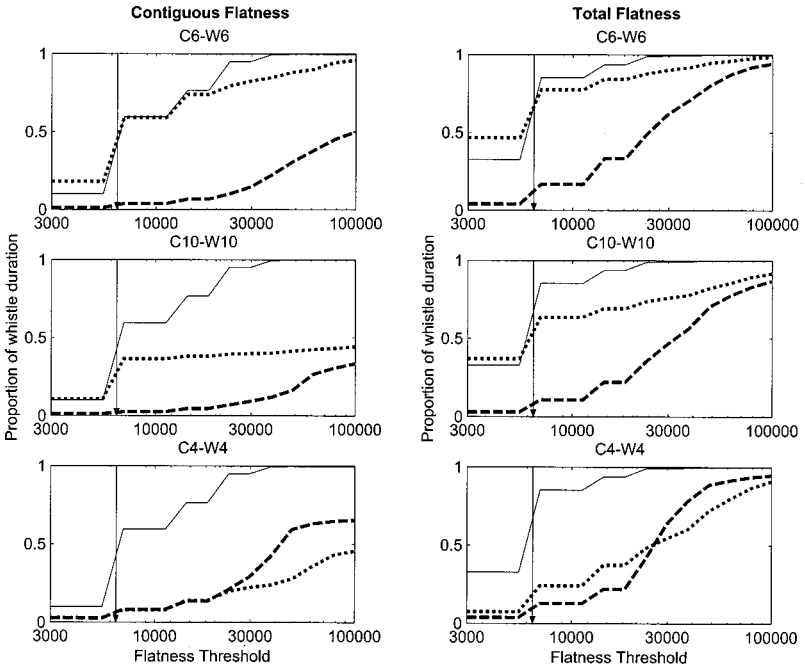


FIG. 3. Comparison of proportion of whistle duration judged to be flat as a function of varying the flatness threshold. This comparison was used to select one flatness threshold for the rest of the study. These flatness comparisons were made for three matched pairs of whistles from wild and captive dolphins, and the results are compared to the trainer's whistle. Solid lines indicate the trainer's whistle. Lines with short dashes represent captive born animals, and lines with long dashes represent wild born animals. Vertical arrows mark the 6400-Hz/s flatness threshold. The left column contains comparisons of contiguous flatness (duration of longest flat segment divided by total duration). The right column contains comparisons of total flatness (percentage of whistle scored as unmodulated).

Fig. 3 that neither of these signals is as unmodulated as the trainer's whistle.

Our working definition of a flat signal throughout the Results section is one whose slope is less than or equal to one frequency bin (80 Hz) per time block (12.5 ms), equivalent to a modulation rate of 80 Hz/12.5 ms=6400 Hz/s. This is the minimum nonzero value possible for μ, as it is the resolution limit for the STFT used in the experiment. Choosing $\mu=0$ is undesirable for the following reason. If the fundamental frequency of an unmodulated signal fell exactly between two FFT bins, e.g., 8040 Hz, the resulting contour would jitter between these bins due to instrument noise, causing apparent changes of frequency at this modulation rate, causing the perfectly unmodulated whistle with 0 slope to be scored as if the slope were 6400 Hz/s. Hence, to avoid this problem the minimum reasonable threshold for an unmodulated signal is ± 1 FFT bin/block=6400 Hz/s. This value is indicated by a vertical arrow in Fig. 3.

E. Statistical analysis

Statistical testing of whether captive whistles differed from wild whistles within each pairing was performed using two techniques. The acoustic features were compared using a standard *t*-test (Zar, 1996). For the duration measurements (Tables III and IV), the observed data were transformed using a log transformation to ensure normality. The flatness measures (Tables V and VI) used an arcsine transformation to ensure normality.

The results were also analyzed using a nonparametric pooled sign test. This test compares the values of the acoustic feature for all possible captive and wild whistle pairs for each matched pair of dolphins. Since 20 whistles were sampled from each dolphin, and each wild whistle can be compared to all 20 of the matched captive whistles, there was a total of 400 pairwise comparisons of wild vs captive whistles for each pair. The total number of times the captive whistle had a greater value of the acoustic feature than the wild whistle was tallied. The binomial *p*-value indicates the probability of obtaining a result at least as extreme as the observed tally if the captive and wild populations had the same distribution for the feature in question. This probability was computed using the normal approximation to the binomial distribution for large numbers of trials (Zar, 1996). This test can be considered to be a Fisher permutation (Efron and Tibshirani, 1994) or exact (Conover, 1999) test on the result of a sign test on the pooled data. The advantage of using the number of sign test comparisons as a test statistic was that it allowed exact computation of the observed result's significance. This eliminated the need for evaluating large numbers of permutations of the data required by Fisher's test for most test statistics.

The computation of *p*-values for each result varied depending on the feature and table in question. For the durations (Tables III and IV), the null hypothesis was that the populations have the same durations. Consequently, the binomial *p*-values were computed for a two-tailed distribution. For the flatness measures (Tables V and IV), the null hypothesis was that the wild dolphin whistles were at least as flat as the captive dolphin whistles. The alternative hypothesis was that the captive dolphins have flatter whistles than the wild dolphins. Because of the polarity of these hypotheses, a one-tailed distribution was used to compute the *p*-values for Tables V and VI.

III. RESULTS

Table II shows that the whistles of wild dolphins had much higher values of maximum frequency, end frequency, and frequency range than either whistles of captive dolphins or trainers' whistles. Visual inspection of the whistles in Fig. 1 showed the trainer's whistle was the least modulated in frequency compared to whistles of captive or wild dolphins,

TABLE III. Total durations of whistles including interloop intervals. The means and variances in columns 3 and 4 reflect raw data values. The *t*-statistics and *p*-values in columns 5 and 6 were obtained using a log transformation [log(*x*)] to ensure normality. The counts of pairs in which total duration was greater for the wild member of the pair are listed in column 7. Column 8 lists the probability from the binomial distribution of obtaining a result at least as anamalous as the observed tally if the captive and wild dolphins are drawn from the same distribution.

Captive–wild match	Sample size	Mean (s)	Variance	*t* stat	*t* stat *p*-value	Duration tally (wild>captive)	Binomial *p*-value
Trainer	8	0.4156	0.0062				
C1	20	1.1781	0.1141	1.09	0.28	137/400	$<10^{-6}$
W1	20	1.0725	0.1451				
C2	20	1.2412	0.1151	0.04	0.96	221/400	0.036
W2	20	1.2518	0.1552				
C3	20	0.8675	0.37	−1.13	0.27	261/400	$<10^{-6}$
W3	20	0.9194	0.0777				
C4	20	1.1531	0.1423	1.81	0.079	160/400	$<10^{-4}$
W4	20	0.945	1.1002				
C5	20	0.9431	0.0118	−1.48	0.15	250/400	$<10^{-6}$
W5	20	1.06	0.0682				
C6	20	0.54	0.0134	−3.74	$<10^{-6}$	284/400	$<10^{-6}$
W6	20	1.555	1.0701				
C7	20	0.9818	0.128	−2.63	0.012	278/400	$<10^{-6}$
W7	20	1.355	0.2464				
C8	20	0.7662	0.036	0.93	0.36	156/400	$<10^{-4}$
W8	20	0.71	0.0575				
C9	20	0.4064	0.0036	−12.5	$<10^{-6}$	400/400	$<10^{-6}$
W9	20	0.905	0.0496				
C10	20	0.7731	0.0015	−6.17	$<10^{-6}$	380/400	$<10^{-6}$
W10	20	1.3038	0.1541				

and that the captive whistle was less modulated than the wild whistle. The measurements of frequency range of the trainer and dolphin whistles followed a similar pattern. The trainers' whistles had the narrowest frequency range, and the wild whistles had the widest frequency range. Our initial hypothesis suggested the broader frequency range corresponded to a higher degree of frequency modulation, indicating that wild dolphin whistles had more frequency modulation than whistles of dolphins born in a captive environment where an unmodulated trainer's whistle was heard. To verify this sug-

TABLE IV. Gap-free whistle durations excluding interloop intervals. The means and variances in columns 3 and 4 reflect raw data values. The *t*-statistics and *p*-values in columns 5 and 6 were obtained using a log transformation [log(*x*)] to ensure normality. The counts of pairs in which gap-free duration was greater for the wild member of the pair are listed in column 7. Column 8 lists the probability from the binomial distribution of obtaining a result at least as anomalous as the observed tally if the captive and wild dolphins are drawn from the same distribution.

Captive–wild match	Sample size	Mean (s)	Variance	*t* stat	*t* stat *p*-value	Duration tally (wild>captive)	Binomial *p*-value
Trainer	8	0.4156	0.0062				
C1	20	1.1651	0.1154	3.36	0.0018	80/400	$<10^{-6}$
W1	20	0.8424	0.0525				
C2	20	1.2356	0.116	2.48	0.017	102/400	$<10^{-6}$
W2	20	0.9987	0.121				
C3	20	0.8569	0.5596	0.57	0.57	220/400	0.046
W3	20	0.6799	0.0572				
C4	20	0.9837	0.1097	0.82	0.42	205/400	0.62
W4	20	0.8988	0.0828				
C5	20	0.9294	0.016	−1.65	0.11	251/400	$<10^{-6}$
W5	20	1.06	0.0682				
C6	20	0.54	0.0134	−2.61	0.013	278/400	$<10^{-6}$
W6	20	0.9612	0.3145				
C7	20	0.6274	0.0645	−5.75	$<10^{-5}$	363/400	$<10^{-6}$
W7	20	1.3106	0.2394				
C8	20	0.7662	0.036	0.98	0.33	154/400	$<10^{-5}$
W8	20	0.71	0.0575				
C9	20	0.4064	0.0036	−11.41	$<10^{-6}$	400/400	$<10^{-6}$
W9	20	0.8041	0.0339				
C10	20	0.7731	0.0015	−5.83	$<10^{-6}$	380/400	$<10^{-6}$
W10	20	1.2662	0.1037				

TABLE V. Total flatness measurements (percentage of whistle flatness) for signature whistles of wild and captive-born animals. The *t*-statistics and *p*-values in columns 5 and 6 were obtained using an arcsine transformation [arcsin(x)] to ensure normality. The counts of pairs in which total flatness was greater for the captive member of the pair are listed in column 7. Column 8 lists the probability from the binomial distribution of obtaining a result at least as anomalous as the observed tally if the captive and wild dolphins are drawn from the same distribution.

Captive–wild match	Sample size	Mean	Variance	*t* stat	*t*-stat *p*-value	Flat tally (captive>wild)	Binomial *p*-value
Trainer	8	0.853	0.0225				
C1	20	0.294	$4.77E-03$	2.44	0.0095	275/400	$<10^{-6}$
W1	20	0.246	$2.75E-03$				
C2	20	0.235	$1.62E-03$	14.53	$<10^{-6}$	400/400	$<10^{-6}$
W2	20	0.0525	$6.26E-04$				
C3	20	0.197	$4.14E-03$	1.2	0.12	219/400	0.029
W3	20	0.173	$5.24E-03$				
C4	20	0.24	$1.33E-03$	13.04	$<10^{-6}$	398/400	$<10^{-6}$
W4	20	0.128	$2.90E-04$				
C5	20	0.613	$4.37E-03$	16.93	$<10^{-6}$	400/400	$<10^{-6}$
W5	20	0.299	$2.16E-03$				
C6	20	0.777	$2.80E-03$	31.79	$<10^{-6}$	400/400	$<10^{-6}$
W6	20	0.116	$2.40E-03$				
C7	20	0.115	$4.08E-03$	3.57	$<10^{-3}$	313/400	$<10^{-6}$
W7	20	0.0601	$7.36E-04$				
C8	20	0.689	$5.29E-03$	5.98	$<10^{-6}$	377/400	$<10^{-6}$
W8	20	0.525	$9.45E-03$				
C9	20	0.355	$1.10E-02$	10.82	$<10^{-6}$	400/400	$<10^{-6}$
W9	20	0.119	$9.14E-04$				
C10	20	0.636	$6.58E-04$	40.71	$<10^{-6}$	400/400	$<10^{-6}$
W10	20	0.106	$1.30E-03$				

gestion, we conducted a quantitative analysis of frequency modulation using the signal processing techniques described in the Methods section.

Figure 2 shows five contours selected randomly from the 20 whistles analyzed from three of the ten captive–wild pairs used for the comparisons in this paper. These contours were simply aligned by starting each whistle at time 0. The C6–W6 comparison was selected to illustrate a particularly unmodulated captive whistle, C10–W10 was more typical, and C4–W4 was the one example of a wild animal that had

TABLE VI. Contiguous flatness measurements (ratio comparing the duration of the longest contiguous flat segment to the total duration) for signature whistles of wild and captive-born animals. The *t*-statistics and *p*-values in columns 5 and 6 were obtained using an arcsine transformation [arcsin(x)] to ensure normality. The counts of pairs in which contiguous flatness was greater for the captive member of the pair are listed in column 7. Column 8 lists the probability from the binomial distribution of obtaining a result at least as anomalous as the observed tally if the captive and wild dolphins are drawn from the same distribution.

Captive–wild match	Sample size	Mean	Variance	*t* stat	*t*-stat *p*-value	Flat tally (captive>wild)	Binomial *p*-value
Trainer	8	0.595	0.118				
C1	20	0.0763	$6.60E-04$	1.6	0.059	272/400	$<10^{-6}$
W1	20	0.0642	$6.71E-04$				
C2	20	0.0821	$5.77E-04$	13.34	$<10^{-6}$	400/400	$<10^{-6}$
W2	20	0.0163	$3.04E-05$				
C3	20	0.0974	$3.22E-03$	2.67	0.0055	275/400	$<10^{-6}$
W3	20	0.0617	$6.61E-04$				
C4	20	0.0777	$1.14E-03$	−0.69	0.25	186/400	0.92
W4	20	0.0825	$6.70E-04$				
C5	20	0.363	0.237	9.67	$<10^{-6}$	400/400	$<10^{-6}$
W5	20	0.0737	$5.76E-04$				
C6	20	0.591	0.02	19.87	$<10^{-6}$	400/400	$<10^{-6}$
W6	20	0.0376	$3.53E-04$				
C7	20	0.0547	$1.13E-03$	4.26	$<10^{-4}$	333/400	$<10^{-6}$
W7	20	0.021	$3.12E-04$				
C8	20	0.277	$9.59E-03$	2.85	0.0035	288/400	$<10^{-6}$
W8	20	0.2	$5.81E-03$				
C9	20	0.172	$5.28E-03$	8.02	$<10^{-6}$	395/400	$<10^{-6}$
W9	20	0.0541	$4.39E-04$				
C10	20	0.364	$6.89E-04$	59.28	$<10^{-6}$	400/400	$<10^{-6}$
W10	20	0.0248	$5.56E-05$				

a longer flat segment than its captive pair. We discuss the C6 and C10 whistles in more detail to illustrate differences between our total flatness and contiguous flatness measures. The longest unmodulated segments of the five C6 whistles had durations of 181.7, 342.4, 232.3, 329.8, and 304.5 ms. The total flatness scores for these C6 whistles were 46.9%, 41.3%, 64.1%, 60.9%, and 61.4%. The C10 whistles, on the other hand, had two relatively flat sections at different frequencies. Even though the C10 whistles scored 58.7%–64.7% in total flatness, their longest unmodulated segments were 221.5–295.6 ms, producing contiguous flatness scores of 27.7%–39.3%. Both total flatness and contiguous flatness measures were retained for later analysis since they measure different features associated with unmodulated whistles.

Some signature whistles had a structure with repeated segments, while others did not contain repetitions of a contour. For example, the five whistles of C6 and C10 illustrated in Fig. 2 did not have segments that were repeated a variable number of times. On the other hand, the whistles of W6 and W10 did repeat a variable number of times. The whistles from W6 had between three and five repetitions of a basic upsweep, while four of the five versions of W10's contour had two repetitions. The fifth W10 contour had three repetitions, as can be seen by the single contour line between 1000 and 2000 ms on the middle right cell of Fig. 2. This distribution of repeated vs nonrepeated whistles was similar in the wild and captivity. Of the 81 Sarasota dolphins analyzed in Sayigh (1992) 22 (27%) have nonrepetitive whistles (Tyack and Sayigh, 1997). Of the 126 captive dolphins in the Caldwell *et al.* (1990) sample, 25 (20%) were not recorded repeating segments. In general, nonrepetitive signature whistles tended to be shorter in duration than those with repeated elements, but since the proportions of these whistle types were similar in captive and wild settings, there was no *a priori* reason to expect a systematic difference in duration across our matched pairs.

The first statistical tests we performed compared the durations of whistles from wild and captive dolphins. We had no *a priori* predictions about differences in whistle duration, and the comparisons of whistle durations yielded mixed results depending on the statistic measured and test used. For total whistle duration, including gaps (Table III), the captive dolphin had shorter whistles in seven out of the ten pairs. The *t*-test only showed significant ($p \leqslant 0.05$) differences for four of the ten pairs, and in all four of the significant pairs, the captive dolphin whistles were shorter. The binomial test (column 7 of Table III) also indicated that the captive dolphin had shorter whistles for seven of the ten pairs, but the binomial test showed a significant difference ($p \leqslant 0.05$) in whistle duration for all ten pairs (column 8 of Table III). Similar results occurred for the durations excluding any silent periods in the whistle (Table IV). Comparison of mean durations (column 3 of Table IV) indicated that five pairs showed the wild dolphin having longer whistles, and five pairs showed the captive dolphin having longer whistles. Six of the pairs generated significant *t*-test values at $p < 0.05$. Four of the six significant pairs had longer whistles for the wild dolphin. The binomial test indicated that eight of the ten pairs had a highly significant ($p < 0.001$) difference in length. Of these eight highly significant pairs, five had longer whistles for the wild dolphin. Pooling all binomial tests showed a significant ($p < 10^{-6}$) difference in the durations for both total duration and gap-free duration, with wild dolphins tending to have longer whistles.

The primary hypothesis of this study was that dolphins born in a captive setting where they hear trainers' whistles would have flatter whistles than matched animals that never heard such whistles. The whistles of captive-born animals were significantly flatter than whistles of wild-born animals, especially for the measure of total flatness. Comparison of the mean total flatness between wild and captive pairs (column 3 in Table V) showed that the captive animal in all ten pairs had the higher mean total flatness. In 9/10 of these pairs, the difference in total flatness tested by a *t*-test was significant to the $p < 0.05$ level; the exception was the C3–W3 comparison. Comparison of the mean contiguous flatness between wild and captive pairs (column 3 in Table VI) showed that the contiguous flatness was greater for the captive-born animals than the wild-born animals in 9/10 of the pairs. The exception was the C4–W4 pair; the differences were not statistically significant for this C4–W4 pair and the C1–W1 pair judging by the *t*-test ($p = 0.05$). All other pairs, in which contiguous flatness was greater for the captive member of the pair, were significant to the $p < 0.05$ level for the *t*-test.

The binomial statistics also supports the hypothesis that wild dolphins produce more highly modulated whistles than their captive counterparts. Examining total flatness, all ten pairings produced tallies (column 7 of Table V) where more than half of the comparisons indicated a less modulated whistle for the captive animal. For contiguous flatness, 9/10 of the pairings had a majority of the comparisons with a flatter whistle for the captive animal (column 7 of Table VI). In fact, the captive animals had less modulated whistles by perfect tallies of 400/400 for 5/10 of the total flatness pairings and 4/10 of the contiguous flatness pairings. These perfect scores indicate that for those pairs of animals, every captive dolphin whistle was less modulated than every wild dolphin whistle. The *p*-values for the comparison tallies (column 8 of Tables V and VI) also supported our primary hypothesis. Nine of the ten pairs had highly significant ($p < 10^{-6}$) tallies for both total flatness and contiguous flatness, and the only insignificant difference for contiguous flatness involved the only pair with a higher flatness score for wild animals. Pooling all binomial tests showed a significant ($p < 10^{-6}$) difference in both total flatness and continuous flatness, with captive dolphins tending to have less modulated whistles.

The most conservative statistical analysis would treat each matched pair of animals as a unit of analysis for statistical testing ($n = 10$). For the total flatness measure, all ten of the captive members of the pair had higher mean flatness values and higher flat tallies for the binomial test. The significance of this 10/10 result as a binomial test was $p < 0.001$. For contiguous flatness, the mean flatness score and flat tallies were higher for captives in 9/10 of the pairs, giving an overall significance of $p = 0.0107$. Consequently, we saw that all four of our overall statistical tests using the pair

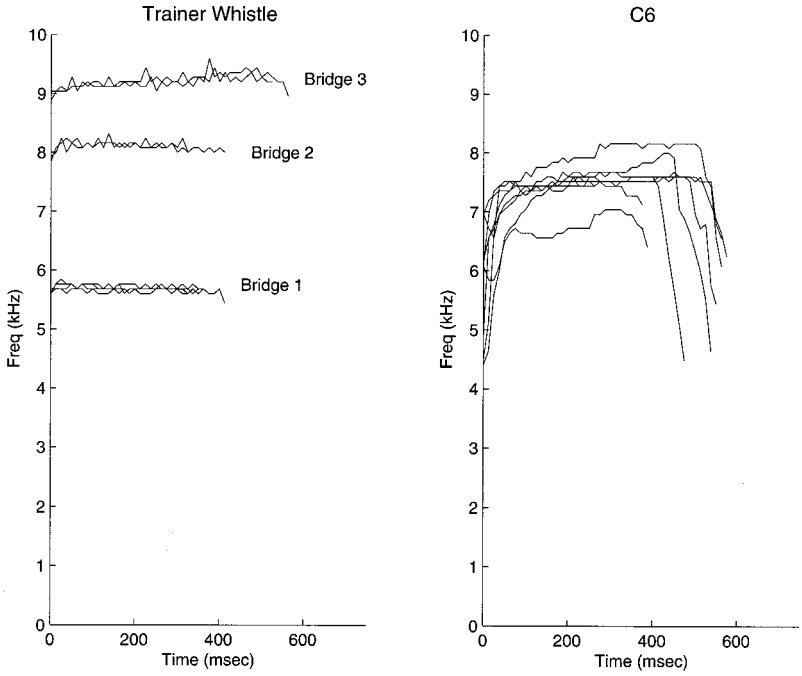


FIG. 4. Contour comparisons of eight examples of the whistles of three trainers and eight whistles from the dolphin, C6, whose whistle was most similar to the trainer's whistles, C6. The trainers' whistles and whistles from C6 were aligned by starting all whistles at time 0.

as the unit of analysis (mean flatness and flat tallies for both flatness measures) yielded a significant result for the combined data set. The combined results support the hypothesis that the captive dolphins overall have less modulated, flatter whistles than wild dolphins.

We have shown that the whistles of captive dolphins show a much better match than do wild dolphins to the lack of frequency modulation seen in trainers' whistles. How similar were the whistles of captive dolphins to the timing and frequency of the trainer's whistle? Figure 4 shows contours of three different trainers' whistles along with the whistles of C6, the captive dolphin with a whistle closest to the trainer's whistle in terms of the modulation measures. This figure shows that while all trainers' whistles were unmodulated, the three different whistles used by the eight trainers tended to have a different absolute frequency. The whistle of C6 was closest to the intermediate trainer's whistle. Comparing the contours of C6's whistles (all of which contain one loop) to the trainer's whistle, the whistles appeared similar in overall flatness and duration. The means of the contiguous flatness within C6's and the trainer's whistles were not significantly different (C6 mean$=0.591$ $n=20$; trainer mean$=0.595$ $n=8$; $t=-0.55$; $p=0.58$). However, the trainers' whistles tended to have a higher mean total flatness over the entire call than the whistles of C6 (C6 mean $=0.775$; trainer mean$=0.853$; $t=-2.73$; $p=0.011$). Because C6 had a one-loop signature whistle, the durations were measured from the start to end of the whistle; since there were no gaps, the gap-free duration equaled the total duration. The mean duration of trainer's whistles ($n=8$) was 0.415 s, while the mean duration of C6's whistles ($n=20$) was 0.540 s. A p-value of 0.013 showed that the trainer's whistle was significantly shorter than C6's whistles.

IV. DISCUSSION

The results of this comparison support the hypothesis that young dolphins born hearing a synthetic whistle in captivity incorporate acoustic features of this synthetic whistle into their own signature whistles. However, there are several minor imperfections in the study design that suggest benefits for future work to reach a more definitive conclusion. Recordings from wild and captive dolphins were not made in exactly the same contexts, and Janik *et al.* (1994) demonstrated slight changes in frequency parameters of signature whistles in different contexts in captivity. On the other hand, both wild (Sayigh *et al.*, 1990) and captive (Caldwell *et al.*, 1990) dolphins produce signature whistles with similar overall contours when stranded in air vs swimming freely, and Sayigh *et al.* (1990) reported that the contour of whistles emitted by Sarasota dolphins recorded in rafts was similar to those emitted from the same animal while freely swimming in a net corral and immediately after release. We know of no results suggesting that the context of restraint would reduce the probability of producing unmodulated whistles, but it would be a better study design to control for behavioral context.

Another problem with this study was that all of the captive dolphins came from one facility. This complicates interpretation of the results where some calves were born soon after other calves. For example, as Table I indicates, Samantha and Noel were born in the same year, and Tori, Sundance, and PJ were born in the same year. These dolphins were all born in the same pool and were all in the pool together for some months after the births of Tori, Sundance, and PJ. This means that, for example, if Samantha had imitated the trainer's whistle and was producing an unmodulated whistle, we cannot discriminate whether Noel, Tori, etc. developed unmodulated whistles through imitating the trainers or imitating Samantha. However, even if this were happening, the data clearly indicate that young dolphin calves are incorporating features of acoustic models prominent in their natal environment.

A third potential problem with interpreting these results stems from a limitation in the synthetic models provided to

young calves: the only synthetic whistle tested has been unmodulated whistles used by trainers. Captive dolphins are raised in a more confined and reverberant environment than wild dolphins. It is plausible that this reverberant environment played a role in the captive animals developing shorter or less modulated whistles than their wild counterparts. However, the impact of reverberation is difficult to assess without a clear understanding of the function of the whistles and which acoustic features are perceptually relevant. A highly modulated whistle would be better for detecting and localizing the source in a reverberant environment. For communication and discrimination, the impact of the environment is less clear. One possible hypothesis is that the reverberant tank environment would encourage young dolphins to develop unmodulated whistles at a nonresonant frequency to minimize the impact of the resonances of the tank on the amplitude of the signal. Consideration of the physical acoustics and the modes of the tank argues against this hypothesis. For the 5–10-kHz frequency range, the resonances of the pool are less than 30 Hz apart in frequency. The observed captive dolphin whistles do not demonstrate less than 30 Hz of variability, so this hypothesis is unlikely to affect modulation of whistles in resonant pools. It is difficult to say conclusively how the reverberation would impact the dolphins' ability to discriminate different fundamental contours. One way to control for these effects of reverberation would be to compare captive dolphins from similar pools, one group of which has been exposed to trainers' whistles and another group of which has never heard them.

All of the above problems would be solved by an expanded study of similar design in which several different facilities were tested, with each new calf in each new facility being exposed to a novel synthetic whistle that is not present in the pre-exposure whistle repertoire of dolphins in the natal pool. In this kind of multifacility setting, it would also be possible to test for the social vs the acoustic salience of the model sound. When trainers blow their whistle, they are signaling to a dolphin that it has performed the correct behavior and can return to station for a reward. Dolphins respond strongly and immediately to this signal. In the Top Deck pool where the captive dolphins in this study were recorded, the trainers only blow their whistle during limited training sessions or presentations to the public. The trainer's whistle is much less frequent than many of the dolphin whistles heard in the pool. This suggests that the intense reinforcing quality of the signal, and the strong predictable responses it evokes, may be as important as any purely acoustic qualities it may have. This hypothesis could be tested using a yoked control design in which the trainers used their whistles for training in one pool, while the exact same pattern of acoustic presentation of whistles, with no associated training or reinforcement, was presented in another pool.

An expanded study with a larger sample size would also be able to address alternative hypotheses of how dolphin age and length of time spent in captivity may affect the acoustic structure of whistles. It is possible that older dolphins exposed to the trainer's whistle for several years would have incorporated more features of the trainer's whistle into their own whistles, but our results do not support this hypothesis. The oldest captive dolphins, C3 (15 years) and C7 (35 years), had the lowest total flatness scores in the captive category. This agrees with the results of Caldwell *et al.* (1990), who report a significant increase in frequency modulation with age in captive bottlenose dolphins. Studies of dolphins born in the wild and brought into captivity at varying ages could address the question of whether there is a sensitive period after which dolphins are less likely to incorporate acoustic features of the trainers' whistles. Tyack (1986) presented evidence that signature whistles of two wild-born dolphins brought into captivity at about 5 years of age maintained their relatively modulated whistle contours for up to 7 years in captivity, even though they occasionally imitated the trainer's whistle.

This study stemmed from a serendipitous ability to use the trainer's whistle as a synthetic stimulus presented during whistle development. The evidence that calves match the trainers' whistles not only provides strong evidence for vocal learning in whistle development, but also suggests that more controlled and systematic presentation of synthetic stimuli is a promising experimental paradigm for further studies on the role of learning in vocal development of marine mammals.

ACKNOWLEDGMENTS

J.R.B. gratefully acknowledges the support of the National Science Foundation Ocean Sciences CAREER Award 9733391. P.L.T. acknowledges the support of the Office of Naval Research Grant N00014-87-K-0236 and NIH Grant 5 R29 NS25290 for supporting the collection of the data used in this study and NIH Grant R01 DC04191 for support in preparation of the manuscript. The authors would also like to thank the two anonymous reviewers for their helpful comments on an earlier manuscript draft. This is Contribution Number 10040 from the Woods Hole Oceanographic Institution and Contribution Number 990902 from the University of Massachusetts Dartmouth School for Marine Science and Technology. All research performed complied with all current U.S. laws and permits.

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.

Caldwell, M. C., and Caldwell, D. K. (**1965**). ''Individualized whistle contours in bottlenosed dolphins (*Tursiops truncatus*),'' Nature (London) **207**, 434–435.

Caldwell, M. C., and Caldwell, D. K. (**1972**). ''Vocal mimicry in the whistle mode by an Atlantic bottlenosed dolphin,'' Cetology **9**, 1–8.

Caldwell, M. C., Caldwell, D. K., and Tyack, P. L. (**1990**). ''A review of the signature whistle hypothesis for the Atlantic bottlenose dolphin, *Tursiops truncatus*,'' in *The Bottlenose Dolphin: Recent Progress in Research*, edited by S. Leatherwood and R. Reeves (Academic, San Diego), pp. 199–234.

Conover, W. J. (**1999**). *Practical Nonparametric Statistics*, 3rd ed. (Wiley, New York).

Efron, B., and Tibshirani, R. J. (**1994**). *An Introduction to the Bootstrap* (CRC Press, Boca Raton).

Evans, W. E. (**1967**). ''Vocalization among marine mammals,'' in *Marine Bioacoustics*, edited by W. N. Tavolga (Pergamon, Oxford), Vol. 2, pp. 159–186.

Herman, L. M. (**1980**). ''Cognitive characteristics of dolphins,'' in *Cetacean Behavior: Mechanisms and Functions*, edited by L. M. Herman (Wiley-Interscience, New York), pp. 363–429.

Janik, V. M., Dehnhardt, G., and Todt, D. (**1994**). ''Signature whistle variations in a bottlenosed dolphin, *Tursiops truncatus*,'' Behav. Ecol. Sociobiol. **35**, 243–248.
Janik, V. M., and Slater, P. J. B. (**1997**). ''Vocal learning in mammals,'' Adv. Study Behav. **26**, 59–99.
Janik, V. M., and Slater, P. J. B. (**1998**). ''Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls,'' Anim. Behav. **56**, 829–838.
Oppenheim, A. V., Schafer, R. W., with Buck, J. R. (**1999**). *Discrete-time Signal Processing*, 2nd ed. (Prentice-Hall, Englewood Cliffs, NJ).
Owren, M. J., Dieter, J. A., Seyfarth, R. M., and Cheney, D. L. (**1993**). ''Vocalizations of rhesus (*Macaca mulatta*) and Japanese (*Macaca fuscata*) macaques cross-fostered between species show evidence of only limited modification,'' Dev. Psychobiol. **26**(7), 389–406.
Reiss, D., and McCowan, B. (**1993**). ''Spontaneous vocal mimicry and production by bottlenose dolphins (*Tursiops truncatus*): Evidence for vocal learning,'' J. Comp. Psych. **107**, 301–312.
Richards, D. G., Wolz, J. P., and Herman, L. M. (**1984**). ''Vocal mimicry of computer-generated sounds and vocal labeling of objects by a bottlenosed dolphin, *Tursiops truncatus*,'' J. Comp. Psych. **8**, 10–28.
Sayigh, L. S. (**1992**). ''Development and functions of signature whistles of free-ranging bottlenose dolphins, *Tursiops truncatus*,'' Ph.D. dissertation, MIT/WHOI Joint Ph.D. Program, Woods Hole, MA.
Sayigh, L. S., Tyack, P. L., Wells, R. S., and Scott, M. D. (**1990**). ''Signature whistles of free-ranging bottlenose dolphins, *Tursiops truncatus*: Stability and mother-offspring comparisons,'' Behav. Ecol. Sociobiol. **26**, 247–260.
Sayigh, L. S., Tyack, P. L., Wells, R. S., Scott, M. D., and Irvine, A. B. (**1995**). ''Sex difference in whistle production in free-ranging bottlenose dolphins, *Tursiops truncatus*,'' Behav. Ecol. Sociobiol. **36**, 171–177.
Sayigh, L. S., Tyack, P. L., Wells, R. S., Solow, A., Scott, M. D., and Irvine, A. B. (**1999**). ''Individual recognition in wild bottlenose dolphins: A field test using playback experiments,'' Anim. Behav. **57**, 41–50.
Scott, M. D., Wells, R. S., and Irvine, A. B. (**1990**), ''A long-term study of bottlenose dolphins on the west coast of Florida,'' in *The Bottlenose Dolphin Recent Progress in Research*, edited by S. Leatherwood and R. Reeves (Academic, San Diego), pp. 235–244.
Sigurdson, J. (**1993**). ''Whistles as a communication medium,'' in *Language and Communication: Comparative Perspectives*, edited by H. L. Roitblat, L. M. Herman, and P. Nachtigall (Erlbaum, Hillsdale, NJ), pp. 153–173.
Smolker, R. A., Mann, J., and Smuts, B. B. (**1993**). ''Use of signature whistles during separations and reunions by wild bottlenose dolphin mothers and infants,'' Behav. Ecol. Sociobiol. **33**, 393–402.
Tyack, P. (**1985**). ''An optical telemetry device to identify which dolphin produces a sound,'' J. Acoust. Soc. Am. **78**, 1892–1895.
Tyack, P. (**1986**). ''Whistle repertoires of two bottlenosed dolphins, *Tursiops truncatus*: Mimicry of signature whistles?'' Behav. Ecol. Sociobiol. **18**, 251–257.
Tyack, P. L., and Sayigh, L. S. (**1997**). ''Vocal learning in cetaceans,'' in *Social Influences on Vocal Development*, edited by C. T. Snowdon and M. Hausberger (Cambridge University Press, Cambridge), pp. 208–233.
Zar, J. H. (**1996**). *Biostatistical Analysis*, 3rd ed. (Prentice-Hall, Upper Saddle River, NJ).

The relative detectability for mice of gaps having different ramp durations at their onset and offset boundaries

James R. Ison[a)]
Department of Brain and Cognitive Sciences and Division of Otolaryngology, Department of Surgery, University of Rochester, Rochester, New York 14627

Jason Castro, Paul Allen, and Tracy M. Virag
Department of Brain and Cognitive Sciences, University of Rochester, Rochester, New York 14627

Joseph P. Walton
Division of Otolaryngology, Department of Surgery, University of Rochester, Rochester, New York 14627

(Received 17 September 2001; revised 28 April 2002; accepted 3 May 2002)

The effect on gap detectability of varying noise fall time (FT) and rise time (RT) of the gap boundary ramps was examined in mice using reflex modification audiometry, measuring inhibition of acoustic startle reflexes by variously shaped gaps just preceding reflex expression. In experiment 1 ($n=12$) inhibition increased up to near-asymptotic values with longer FT (0, 1, 2, 3, 5, or 10 ms) and QT (quiet time, 0 to 13 ms), with a 2:1 trade-off between FT and QT. In experiment 2 ($n=24$) inhibition increased for any RT above 0 ms (2, 3, 5, or 7 ms) if QT=1 ms, but diminished with increased RT when QT=3 or 8 ms. Enhanced detectability for subthreshold gaps by longer ramps results from their extending the apparent gap duration. The negative effect of increased RT for threshold gaps suggests the importance for gap detection of the stronger neural responses to sharp edges at the end of the gap shown previously in the mouse inferior colliculus. These effects are specific to gaps: inhibition for fixed (70-dB SPL) or varied level pulses (30 to 60 dB) was unaffected by varying the ramped edges (experiments 3 and 4, $n=9$). © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1490352]

PACS numbers: 43.80.Lb, 43.66.Gf, 43.66.Mk [WA]

I. INTRODUCTION

The ability to detect brief, temporal decrements in a background noise provides a simple, convenient, and useful experimental paradigm for studying in both humans and in laboratory animals the acoustic variables and the characteristics of the listener that may be important in the temporal processing of more complex acoustic signals (see, for example, Michelsen, 1985). Many investigators have analyzed the effects on gap detection of individual differences between subjects, such as age and degree of hearing impairment (Florentine and Buus, 1984; Irwin and McAuley, 1987; McCroskey and Kidder, 1980; Salvi and Arehole, 1985; Snell, 1997; Trehub *et al.*, 1995; Walton *et al.*, 1997; and others), and the acoustical characteristics of the gap carrier and the gap itself, including the spectral composition, the bandwidth, and the level of the carrier, and the depth of a partially filled gap (for example, Buus and Florentine, 1985; Eddins *et al.*, 1992; Ison *et al.*, 1998; Forrest and Green, 1987; Penner, 1975; Snell *et al.*, 1994). With rare exception, all of these experiments on gap detection in both laboratory animals and human listeners have used gaps with near-instantaneous onsets and offsets. Such abrupt transitions must characterize very few natural sounds, and at least for this reason alone, besides their intrinsic interest, further study of the effects of variation in onset and offset ramps is warranted. In the present experiments we varied the rise and fall time of brief gaps embedded in noise, using the method of reflex modification audiometry (Young and Fechter, 1983) to measure the behavioral effects of gaps in noise presented at subthreshold and suprathreshold gap durations.

There are but three published experiments describing the effects on gap detection of varying the temporal details of the envelope of a gap in a noise carrier, these appearing in very different experimental paradigms. Leitner *et al.* (1993) used reflex modification audiometry to study gap detection in rats, in an experiment in which gaps having different durations of quiet time (QT: 0 to 10 ms) and different rise–fall times (linear RT/FT: 0.25 to 2.5 ms) were used as the probe stimuli, to show that adding ramps to small gaps increased their inhibitory effect on startle behavior. Allen *et al.* (2002) provided psychophysical evidence in humans for the beneficial effect of adding different rise or fall times (linear RT or FT: 2 to 8 ms) to brief gaps (QT: 1 to 4 ms), which enhanced detection probability and reaction speeds. Barsz *et al.* (1998) measured neural activity in single units of the central nucleus of the inferior colliculus (IC) of mice in response to gaps having different onset and offset ramps (cosine-squared RT and FT: 0.5 to 16 ms) to find that adding a brief ramp to a brief gap increased its effect, while longer ramps of 8 and 16 ms led to reduced responses at the end of the gap and increased neural gap thresholds. These diverse sets of data are in general agreement for the positive effects of brief gaps and brief ramps, but the psychophysical data obtained in humans by Allen *et al.* (2002) provided no indication that longer values of FT or RT had any detrimental effect on gap detection,

a)Electronic mail: ison@bcs.rochester.edu

as had been found by Barsz *et al.* (1998) for neural thresholds in mice. To the contrary, both stimulus detection and response speeds in humans were enhanced by adding long as well as brief ramps to the terminal boundary of periods of quiet within the gap.

Thus, at present the overall generality and theoretical significance of these data depicting the effects of temporal variation at the boundary edges of gaps in noise on gap detection across the several research domains are uncertain. The present experiments followed the general procedures of Leitner *et al.* (1993), save that here mice rather than rats were used as subjects. Following Allen *et al.* (2002), we varied RT and FT independently over a range that included the short ramp durations that had positive effects in all of the research reported to date as well as the longer durations that had negative consequences in the work of Barsz *et al.* (1998). Reflex modification audiometry is an especially appropriate method to examine the potential correspondence of behavioral psychophysical measures in animals and IC electrophysiology. The phenomenon on which it is based, namely, the inhibition of the ASR by immediately prior acoustic events, depends on the integrity of the IC (Leitner and Cohen, 1985), and electrical stimulation of the IC inhibits the acoustic startle reflex (ASR) with a time course consistent with the idea that the inhibitory pathway has an obligatory passage through the IC (Li *et al.*, 1998). Thus, if the effect of varied gap envelopes on neuronal activity in the IC of the mouse has a direct correspondence in mouse behavior, then it would be most likely to appear in experiments concerned with the inhibitory effect of gaps on the acoustic startle reflex. Two pulse experiments were also conducted in order to determine whether the behavioral effect of ramps in gap envelopes would be the same or greater than that of ramped noise pulses: Barsz *et al.* (1998) had reported that ramps in the gap envelope had a greater neural effect than ramps in the envelope of a noise pulse.

II. METHODS

A. Subjects

The subjects for the two gap experiments were 24 mice (14 male, 10 female), the F1 hybrid offspring of CBA/CaJ males and C57BL/6J females, approximately 10 weeks of age. Twelve mice were tested in the first experiment, while all 24 were tested in the second. The subjects for the two subsequent pulse experiments were 14 mice (11 male, 3 female), all of the CBA/CaJ strain, approximately 4 months of age. Four mice were tested in both experiments, the others used in only one experiment, for a count of 9 in each. All were group housed in the University of Rochester vivarium in a constant climate and a 12/12 h normal L/D cycle, and were tested in the daytime. Food and water were available *ad libitum* save during the test session. All procedures were approved by the University of Rochester Committee on Animal Resources.

B. Apparatus

A mouse was confined for testing in an acrylic plastic cage, 5 cm wide, 7 cm long, and 4 cm high, with slotted sides and roof for free sound penetration. The cage was mounted on a suspended acrylic platform to which an accelerometer was attached, and placed in an anechoic chamber (Eckel Corp., Cambridge, MA, model number 555-250-3) housed within a sound-attenuating room (IAC, Bronx, NY), one mouse being tested at a time. The startle stimulus was a 110-dB (peak-to-peak, SPL linear scale), 20-ms duration noise burst provided by a white-noise generator, gated through an electronic switch with <0.2 ms rise and fall times, then amplified and delivered through a high-frequency tweeter with maximum output at 16 kHz and a 5-dB/octave roll-off. The gap carrier (rms 70-dB SPL, linear scale) in the first two experiments and the prepulse in the last two was provided by the same white-noise generator, gated through a second electronic switch with variable rise and fall times, and then amplified and delivered through a Panasonic high-frequency leaf tweeter that varied by no more than ± 6 dB over a range of 2 to 100 kHz. The ambient noise level in the chamber was less than 25 dB SPL for all frequencies above 125 Hz. Sound levels were measured with a $\frac{1}{4}$-in. Bruel and Kjaer microphone, model 4135, connecting to a measuring amplifier, Bruel and Kjaer, model 2610.

The acoustic startle reflex in rodents consists of a rapid sequence of rostral to caudal movements beginning with eye and pinna responses and a brief extension of the limbs, followed by a generalized flexion response that brings the body together in a compact hunched posture (Davis, 1984; Horlington, 1968). The response is usually quantified in rodents by the force it exerts on the floor of the housing cage, which is detected by the accelerometer attached to its substrate. The onset latency of the response measured in the EMG can be as brief as 5 to 7 ms (Ison *et al.*, 1973) in the rat, while in the present apparatus the onset of the recorded response is typically 10 to 14 ms after the onset of the eliciting noise burst. Figure 1 shows schematics of two stimulus conditions at the top, a control ''no-gap'' condition on the left, and a gap condition on the right. The lower functions show the traces of the resulting two responses. (See the legend for details.) Stimulus presentation and response measurements were under computer control.

C. Procedure

All experimental tests began with the mouse being placed in the chamber for a 5-min period, in background noise (experiments 1 and 2), or in quiet (experiments 3 and 4). For the gap experiments the mice then received trials in which the startle stimulus was presented in the presence of an uninterrupted noise to provide a control baseline response, and other trials in which the startle stimulus followed a brief gap in the noise. Twelve subjects received 8 test days (those mice tested in experiments 1 and 2), and 12 received 6 test days. At least one rest day intervened between each test. For the pulse experiments the startle stimuli were presented in quiet, sometimes alone, sometimes following a brief noise burst. Five mice received just 1 test day, for the first pulse experiment; 5 mice received 2 test days, for the second pulse experiment; and 4 mice received 3 test days total, for both experiments. The various stimulus conditions were always given in 11 randomized blocks of trials, in which each con-

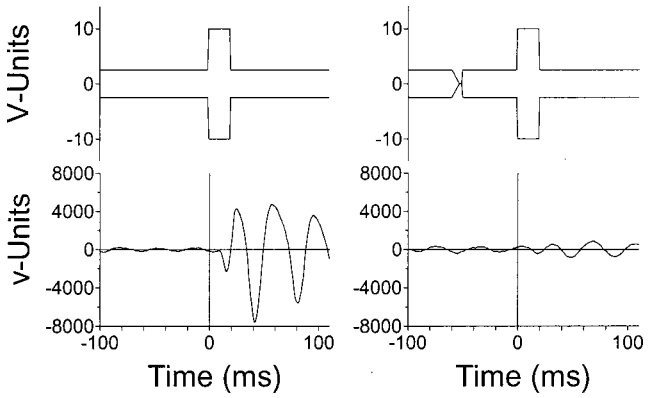


FIG. 1. At the top the graphic describes the stimulus contours for two trials as they would be seen on an oscilloscope measuring the input to the speakers (the relative voltages of the background noise at 70-dB rms SPL and the startle stimulus at 110-dB PtoP SPL are not drawn to scale). At the bottom the corresponding traces of two amplified responses of the accelerometer are shown, as recorded by the computer. The left pair shows a baseline control trial in which the startle stimulus followed 50 ms after a ''gap'' of 0-ms duration, this yielding a relatively large control response (3400 voltage units). The right pair shows a prepulse trial in which the startle stimulus followed a gap of 10-ms total duration, with a 7-ms ramped-noise offset followed by a 3-ms period of quiet ending in an abrupt noise onset. The startle reflex provides an impulse to the stabilimeter which then rings at its resonant frequency, and the measure of response amplitude is the rms voltage in the response over a 100-ms period. The startle stimulus presented 50 ms later provided a relatively small response (483 v-units).

dition was presented before any condition was repeated, and the intertrial intervals averaged 20 s (15 to 27 s).

The major graphic and statistical analyses of variance (ANOVA) of the data used mean relative levels of inhibition, in which each subject's mean response for any particular stimulus condition (ASR_s) was expressed as the complement of its amplitude relative to that of its mean control response (ASR_c), that is $(1-ASRn_s/ASR_c$; given as ''1-%C'' in the figures). The first block of trials was not used in the analyses but was presented to allow adaptation to the experimental conditions. The probability levels associated with the *F* values were calculated using the Huynh–Feldt adjustments for degrees of freedom in repeated measures designs.

D. Experimental designs

1. Pretraining for gap detection

The inhibitory effect of weak prestimuli given at lead times beyond about 30 ms increases in rodents with experience with the testing conditions (Crofton *et al.*, 1990; Ison and Bowen, 2000), and for this reason the main gap experiments in this report were preceded by two series of preliminary tests, following the procedures previously used in rats to provide experience with gap detection. All mice first received 3 test days in which a startle stimulus was given alone, or preceded by a gap in a 70-dB noise of 10-ms duration and RFT of 0 ms (actual ~0.2 ms) having various lead times, 10, 15, 20, 30, 40, 60, 110, 160, and 210 ms, gap onset to startle onset. They then received a fourth pretraining day in which the startle stimulus was given alone, or preceded by a gap with varying durations, 1, 2, 3, 4, 6, 8, 10, and 15 ms, RT/FT=0 ms, the gap ending 50 ms prior to the onset of the startle stimulus.

2. Effect of variation in noise FT at gap onset on reflex inhibition

This experiment examined the effect of variation in noise FT at the start of the gap with the noise RT at the end of the gap fixed at 0 ms (see Fig. 1). The startle stimulus was presented 50 ms after the end of the gap. The subjects ($n=12$) were tested over 2 days. There were 13 different stimulus conditions: one in which the startle stimulus was presented alone (i.e., gap duration=0 ms, this condition given twice as often as any other to provide a more reliable control baseline); four conditions in which the gaps had rise and fall times both set at a nominal 0 ms, and their QT=1, 2, 3, or 13 ms; three conditions in which the total gap duration was 3 ms, made up of QT=0, 1, or 2 ms, paired with FT=3, 2, and 1 ms; and five conditions in which QT=3 ms, and FT=1, 2, 3, 5, or 7 ms (and thus the total gap duration was 4, 5, 6, 8, or 10 ms). An additional ''no-stimulus'' control condition was given in each block, in which the accelerometer output was integrated over a 100-ms period without a preceding stimulus to provide an activity baseline.

3. Effect of variation in noise RT at gap offset on inhibition

This experiment examined the effects of varied noise RT at the end of the gap, and all gap conditions were run with a 0-ms noise FT at the beginning of the gap. A schematic of a gap with a particular nonzero rise time would look like the right-hand upper graph in Fig. 1, save that the gap would begin with an abrupt noise offset and end with a ramped noise onset. The startle stimulus was presented 50 ms after the beginning of the ramped noise onset at the end of the gap. The subjects ($n=24$, 12 of which had served in the preceding experiment) were tested on 2 days. This experiment included 15 stimulus conditions: one in which the startle stimulus was presented alone (again twice as frequently as the other conditions); four conditions in which the gaps had RT=0 ms and QT=1, 3, 8, or 15 ms; four conditions in which QT=1 ms, and RT=2, 3, 5, or 7 ms (and the total gap was thus 3, 4, 6, or 8 ms); two conditions in which QT=3 ms, and RT=5 or 7 ms (total gap times of 8 or 10 ms); and four conditions in which QT=8 ms, and RT=2, 3, 5, or 7 ms (total gap times of 10, 11, 13, or 15 ms). Spontaneous activity was integrated over a 100-ms period without a preceding startle stimulus once in each block to provide the activity baseline.

4. Effect of variation in noise pulse RT/FT on inhibition

The preliminary stimulus was a 70-dB noise pulse, with onset and offset ramps of 0, 5, or 10 ms. The total duration of the prepulse with the 0-ms RT/FT was 20 ms, but the total duration of the prepulse with the 5-ms RT/FT was 25 ms and the prepulse with the 10-ms RT/FT had a duration of 30 ms. Variation in total duration was intended to equate acoustic energy in each pulse. The interval between the onset of the prepulse and the onset of the startle stimulus was maintained at 50 ms. The procedures were the same as used above for

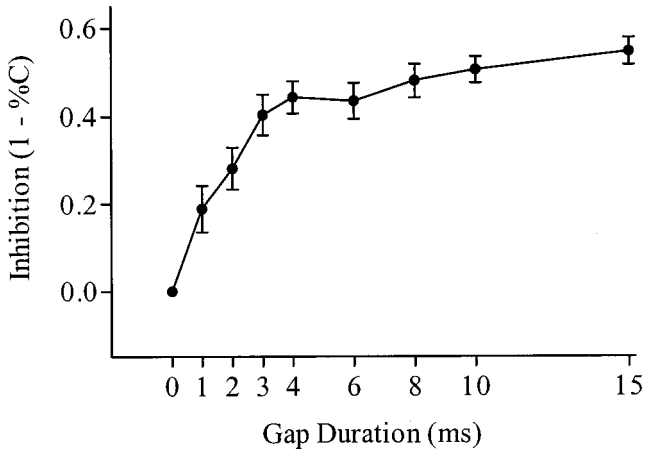


FIG. 2. Means (±SEM) for relative inhibition for gaps from 0 to 15 ms in duration in the 24 mice used in experiments 1 and 2. Note the rapid increase in the level of inhibition from 0 to 4 ms and the slow increase thereafter. The mean threshold for inhibition defined for each subject as the duration which provided at least 50% of the maximum level of inhibition was 2.46 ms.

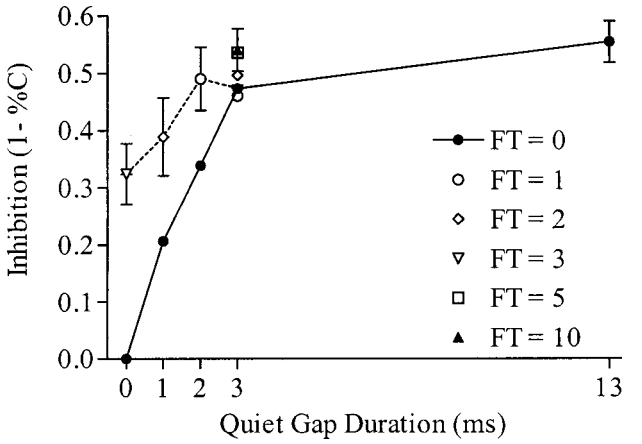


FIG. 3. Means for relative inhibition in experiment 1, for gaps having different periods of absolute quiet of 0 to 13 ms, and additional periods of relative quiet during their linear fall time of noise offset at the start of the gap, from 0 to 10 ms. FT always increased inhibition, especially for short quiet periods. The SEM bars at 13 ms are the average across all QT=0 conditions, and at the FT conditions describe the SEM for the within-S differences with the QT condition.

the gap experiments save that no pretraining was thought to be necessary in experiment 3 for the 70-dB prepulse.

5. Effect of variation in noise pulse RT/FT and noise pulse level on inhibition

The preliminary stimulus was a noise pulse. Its RT/FT had two values, 0 and 10 ms, each presented at four levels, namely, 30, 40, 50, and 60 dB. Because weak prepulses were used, 2 test days were run for this experiment. All other procedures were the same as those described above.

III. RESULTS

A. Preliminary test for gap detection

Figure 2 shows the mean relative response and the standard error of the mean (M±SEM) to the startle stimulus when preceded by gaps of varied duration. These data are similar to those obtained in the rat (i.e., Ison and Bowen, 2000) and in human listeners run under similar conditions (Ison and Pinckney, 1983), especially in their reaching a near-asymptotic level of inhibition of about 50% in relative inhibition units at about 4 ms. The mean gap threshold in these data was 2.46 (±0.27) ms, this value defined for each mouse as the minimal gap duration at which inhibition was at least 50% of its maximum. This value is just slightly greater than the mean threshold of 2.03 ms found in phasic on-neurons in the mouse IC by Walton *et al.* (1997).

B. Effect of variation in noise FT at gap onset on inhibition

Figure 3 shows the mean values for the inhibitory effects of gaps having QT of 0 to 13 ms and FT varying overall from 0 to 10 ms. The main gap function obtained with the FT set at 0 ms approximated that shown in Fig. 2, with the strength of inhibition rising rapidly as QT increased from 0 to 3 ms, and then showing a small additional increase with a QT of 13 ms: overall $F(3/33)=15.96$, $p<0.01$. Increasing the FT at the beginning of gaps having any particular QT strengthened their inhibitory effect, especially for relatively small gap durations: thus, compared to gaps with FT of 0 ms and QT of 0, 1, or 2 ms, comparable gaps with the same QT but FT of 3, 2, or 1 ms, respectively, in each case had a significantly stronger effect, $p<0.05$, for Bonferroni multiple comparison *t*-tests following a significant overall ANOVA. These data agree with those of Allen *et al.* (2002) in human listeners in their showing that adding a nonzero FT to a subthreshold or near-threshold QT increased its salience. They show also that substituting an FT for some portion of QT in these relatively brief gaps had the expected opposite effect of diminishing inhibition: thus, for a series of gaps with a total gap duration (QT+FT) of 3 ms, the effect of a gap with the QT of 3 ms was greater than that a gap with a QT of 1 ms and FT of 2 ms, which in turn was more inhibitory than a gap with a QT of 0 ms and FT of 3 ms (all $p<0.05$). The effect of adding FT to gaps with a QT of 3 ms was generally positive, though modest in size: the ANOVA of the apparent stronger inhibitory effect of FT for gaps with QT of 3 ms provided a significant linear increasing effect on inhibition, $F(1/11)=8.06$, $p<0.05$.

In summary, this experiment found that adding noise-offset ramps to the beginning of gaps provided a graded increase in their inhibitory effect according to the duration of the ramp, and never weakened gap inhibition even for suprathreshold gap durations. The overall pattern in these data is similar to that reported for human gap detection by Allen *et al.* (2002).

C. Effect of variation in noise RT at gap offset on inhibition

Figure 4 shows the mean values for the inhibitory effects of gaps having QT of 0 to 15 ms, and RT of 0 to 7 ms. The gap duration function for gaps with RT of 0 ms again approximates those shown in Figs. 2 and 3 in its rapid increase in inhibition from 0 to 3 ms and a slow increase thereafter: $F(4/92)=91.21$, $p<0.01$. For gaps with QT of 1 ms, increasing RT from 0 ms to values ranging from 2 to 7 ms

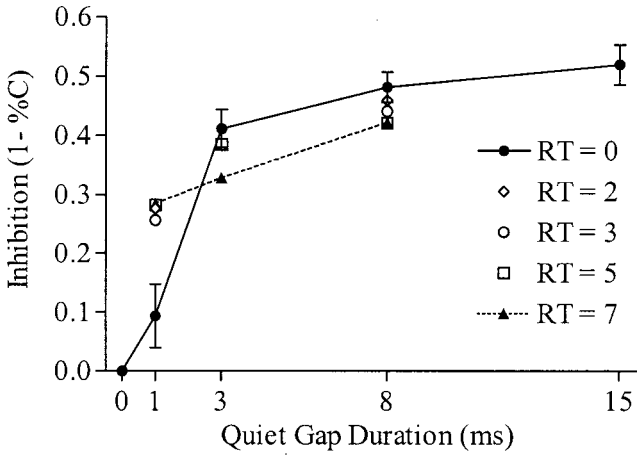


FIG. 4. Means for relative inhibition in experiment 2, for gaps having periods of absolute quiet of 0 to 15 ms, and additional periods of relative quiet during their linear rise time of noise onset at the end of the gap, from 0 to 7 ms. RT increased inhibition for gaps with a 1-ms period of quiet but decreased inhibition for gaps with 3- or 8-ms periods of quiet. The SEM bars at 15 ms are the average across all QT=0 conditions, and at the various RT conditions describe the SEM for the within-S differences with the QT condition.

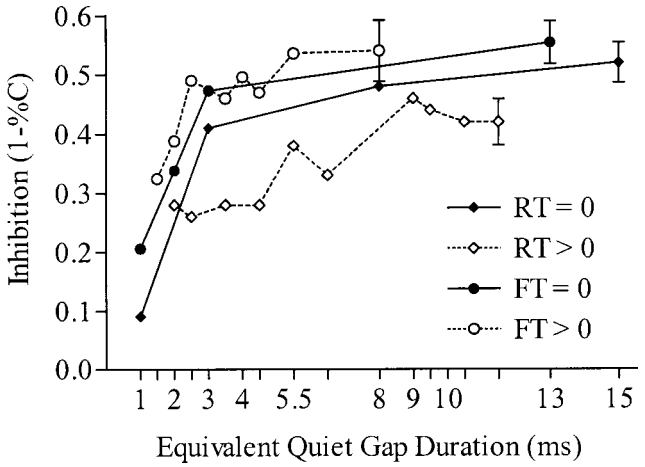


FIG. 5. Mean inhibition for both FT and RT conditions in experiments 1 and 2 plotted against "Equivalent quiet gap duration" calculated by adding half the FT or RT period to the period of absolute quiet. Note that the FT data are well described by the equivalent quiet duration, while adding RT to the gap reduces the level of inhibition below that expected from this calculation. The SEM bars are the average for all of the conditions on that particular function.

increased their inhibitory effect, $F(4/92)=7.46$, $p<0.01$. Subsequent *t*-tests using Dunnett's method for comparisons having a common control value showed that all gaps having a nonzero RT differed from the gap with RT of 0 ms, $p<0.01$, but, as is evident in Fig. 4, none of these gaps with a nonzero RT differed from each other, $F<1$. In contrast to this positive outcome of increasing RT when QT was equal to 1 ms, inspection of the data obtained with gaps having QT of 3 or 8 ms reveals that increasing the RT above 0 ms in these conditions had the opposite effect of reducing the strength of inhibition. The ANOVA of the effect of RT on gaps having QT=3 ms provided a significant decreasing linear trend for RT, $F(1/23)=5.29$, $p<0.05$, as well as a significant effect overall, $F(2/46)=3.79$, $p<0.05$. The gap with an RT of 7 ms, and, thus, a total gap duration of 10 ms, provided significantly less inhibition than the gap with an RT of 0 ms and a total gap duration of just 3 ms, $p<0.05$. An ANOVA of all gaps with a QT of 8 ms also provided a significant linear decreasing trend in inhibition with increased RT, $F(1/23)=7.55$, $p<0.05$, and there were significant differences between gaps with RT of 5 and 7 ms contrasted with the gap with an RT of 0 ms, $p=0.05$.

In summary, this experiment found that when a ramped noise onset was added to the end of a subthreshold quiet period then gap inhibition was increased, but by a fixed amount regardless of the RT; in contrast, when the same RT was added to the end of a longer threshold or suprathreshold gap, then inhibition was reduced. The overall pattern in these data is similar to that provided by Barsz *et al.* (1998) in their study of neural activity in the mouse IC.

D. A comparison of the inhibitory effects of adding RT and FT to a fixed quiet period

Allen *et al.* (2002) found that for human listeners the increment in either response speed or the probability of detection produced by increasing the duration of an onset or offset ramp at the beginning or end of the gap was precisely matched by increasing the duration of absolute quiet by one-half that value, and that the functions for RT and FT overlapped throughout their entire range. Figure 5 shows the data of both of the present gap experiments aligned along a common dimension of "equivalent quiet gap time," which was calculated by adding one-half of the RT or FT to the duration of quiet within the gap. The solid lines describe the mean levels of inhibition for the two sets of control gaps with 0-ms ramped onsets and offsets across gap duration, which generally lie within 1 SEM of each other. The dotted lines describe the mean levels of inhibition for gaps with either ramped noise offset at their beginning (FT>0 ms, open circles) or ramped noise onset at their end (RT>0 ms, open diamonds). The 2:1 trade-off function between ramp time and quiet time to determine an equivalent quiet time provides a good fit between the FT functions (solid versus open circles), but not for the RT functions (solid versus open diamonds). Ramps placed at the beginning of the gap and those placed at its end have very different effects, with the disadvantage in relative inhibitory strength for the noise onset ramp placed at the end of the gap beginning at the 2-ms equivalent quiet time.

E. Effect of varied RT/FT on inhibition by noise pulses

Figure 6 (top panel) shows the inhibiting effects of 70-dB noise pulses presented 50 ms before a startle stimulus, the pulses having RT/FT of 0, 5, or 10 ms. These prestimuli inhibited the response on average by about 75%, their SEM overlapped, and the ANOVA of the effect of RT/FT on inhibition provided an $F<1$. Noise pulses provided more inhibition of startle than did gaps for a carrier of the same intensity. This finding suggested that the examination of RT/FT in pulses should be extended to lower pulse intensities that would have inhibitory effects approximating those of gaps, and perhaps also a similar degree of sensitivity to variation in RT/FT. The lower portion of Fig. 6 shows the effect of varying the RT/FT of the prepulse, at 0 and 10 ms, across

different pulse intensities, from 30 to 60 dB SPL. Inhibition increased with level as anticipated, $F(3/24)=59.87$, $p<0.01$. There was a small difference in inhibition in six of the nine mice favoring the 0-ms RT/FT for the moderate 40- and 50-dB prepulse levels that provided approximately the same inhibition values as the asymptotic gaps, but for the group as a whole this difference was not significant, and the overall effect of RT/FT provided $F<1$.

The purpose of these two experiments was to determine if the effects on reflex inhibition of varying RT and FT at the boundaries of a gap in noise background would be duplicated if the inhibiting stimulus was a noise pulse in a quiet background. Barsz *et al.* (1998) reported that responsivity of onset cells in the IC of the mouse was sensitive to noise-burst rise time, but that increasing the RT of the noise marking the end of the gap had a more profound effect than variation in the RT of a noise burst presented in quiet. Varying the RT/FT of a noise pulse does not affect reflex inhibition in rats (Ison, 1978), but the present experiments were necessary to extend these findings to the mouse. In summary, neither of these experiments found that varying the RT/FT of noise pulses had a significant effect on the strength of reflex inhibition in mice.

IV. DISCUSSION

The present set of data confirms the general conclusion apparent in the small number of prior publications on this problem, that variation in the temporal structure of gap envelopes provided by changing the durations of the noise offset and onset ramps at its boundaries affects both threshold detection and the salience of suprathreshold gaps (Allen *et al.*, 2002; Barsz *et al.*, 1998; Leitner *et al.*, 1993). Lengthening the duration of the noise offset ramp at the start of gaps consistently strengthened its inhibitory effect on startle behavior in the mouse for a range of near-threshold quiet periods, entirely in accord with the increased probability of detection of gaps and increased response speed in the similar condition reported for human listeners by Allen *et al.* (2002). In the human psychophysical data, the terms of the 2:1 tradeoff between quiet gap duration and ramp time indicated that the underlying variable determining the effect of FT on gap salience was the magnitude of the dip in the acoustic energy integrated across the duration of the gap, and the same effect for FT was seen here for reflex inhibition in the mouse. These FT data are consistent with the reports of Forrest and Green (1987) in human listeners and Ison *et al.* (1998) in mice, both showing that directly manipulating the amplitude of this dip by varying the noise floor within the gap systematically varied its detectability, and are consistent also with a loudness-detector model of gap detection developed by Florentine *et al.* (1999). Further, in the present experiment the expected complement to this effect was also demonstrated, in the data showing that a gap is a less effective stimulus if the noise offset ramp is not added to the gap, but instead extends into its quiet period.

For human listeners, increasing FT and RT provided precisely the same beneficial effect on the detectability of gaps having a set QT, but in the present study FT and RT had different effects: an increase in FT was always beneficial,

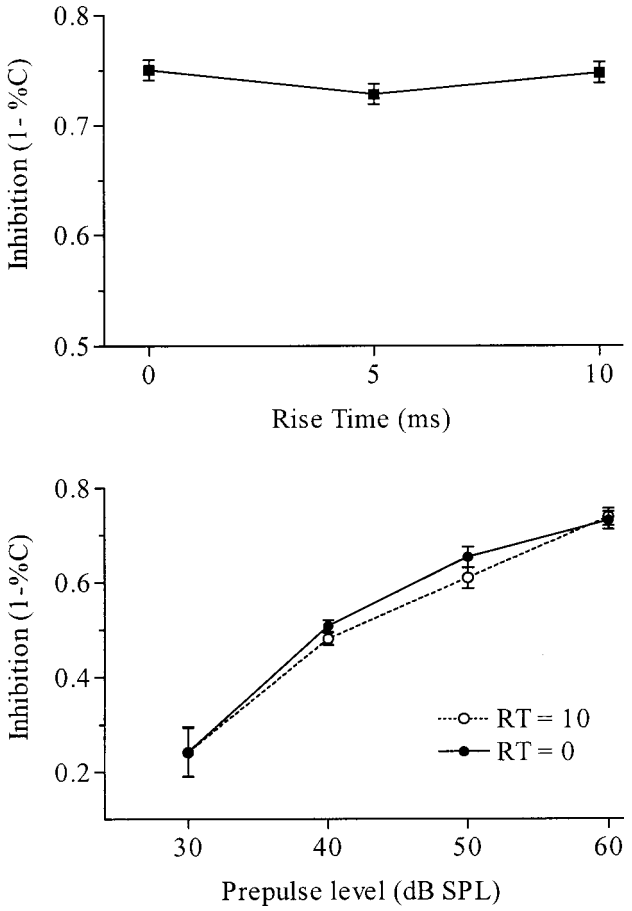


FIG. 6. Means (±SEM) for relative inhibition plotted against rise/fall time of a noise prepulse (experiment 3) or against level of a noise prepulse with rise/fall time as a parameter (experiment 4). Note that RFT has little effect on inhibition by noise pulses.

while an increase in RT increased inhibition for a gap with a subthreshold QT, but diminished the inhibitory effect of gaps with threshold or suprathreshold QT. While the observed effect of increasing RT depended on the duration of the quiet period, the pattern obtained in these data indicates that variation in RT always had two opposing effects, the difference in their expression reflecting the diminished importance of an increased quiet duration for longer gaps, given the compressive nonlinearity of the function relating gap detection to gap duration. For gaps having a brief quiet period, any gradual noise-onset ramp was more effective than the abrupt noise onset, but the strength of inhibition did not then increase with the longer RT values, as it did for longer FT values. This difference suggests that the positive and negative consequences of increasing RT beyond the abrupt offset condition were balanced in strength even when the beneficial effect of increasing the duration of the relative quiet period was near its peak. These data for RT are consistent with the physiological data collected by Barsz *et al.* (1998), showing that neural gap thresholds increased with longer RT/FT. They are contrary to the psychophysical data of Allen *et al.* (2002) collected in human listeners who demonstrated increased salience for gaps ending with a longer noise RT, and, overall, equivalent effects of variation in the duration of onset and offset ramp durations.

V. CONCLUSIONS

The effect obtained by Barsz *et al.* (1998) in the physiological experiment was attributed to the reduction in the strength of the onset response in phasic cells of the inferior colliculus with increased duration of the ramped noise onset at the end of the gap. The fact that increasing RT diminishes gap detectability agrees with the theoretical description of gap detection provided by Plomp (1964), as being determined by the vigor of the onset response at the end of the gap. He applied this model to the results of psychophysical experiments in human listeners in which gap duration and the level of the noise marker at the end of the gap were manipulated, but given the neural onset data of Barsz *et al.* (1998) obtained in the mouse IC, plus the understanding that the inhibition of the startle reflex by preceding acoustic events depends on IC activity (Leitner and Cohen, 1985; Li *et al.*, 1998), his model also provides one possible account of the present finding of diminished behavioral effects for gaps having longer duration onset ramps at their end. These data obtained in mice are consistent also with a suggestion by Florentine *et al.* (1999) that gap detection is primarily based on detecting a loudness perturbation at the output of an auditory filter for brief gaps, but benefits as well from the neural response to noise onset at the end of the gap, especially for suprathreshold gaps and for noise onsets with a short rise time.

The question then remains why the behavioral data we obtain here with mice do not agree with those provided by human listeners working with similar stimuli, as reported by Allen *et al.* (2002). The answer may lie in differences in the methodological details of these experiments or in species differences in the way that rise time is encoded at the neural level, but it is important to point out that there were striking similarities between the data sets, despite these differences in species and methods. First, adding a ramped noise offset to the beginning of a quiet period increased the detectability of gaps and their suprathreshold salience, as indicated by a greater strength of reflex inhibition in mice, and by an increased probability of detection and increased response speed in humans. Second, the general shape of the functions relating performance to the ''effective quiet period'' were similar across the experiments and the trade-off for combining the noise-offset ramp duration with actual quiet time was identical. Third, lengthening the duration of the noise-onset ramp at the end of the gap did have a positive effect in both studies: in fact, the single difference in outcome was that this last manipulation had only a positive effect for human listeners, but an additional negative effect on gap salience in the behavior of the mouse that had no counterpart for human gap detection.

It is possible that the critical difference between the two sets of data is that reflex inhibition appears to depend heavily on the responsivity of cells in the inferior colliculus, and that the neural effects of gaps at this level of the IC (but not pulses presented in quiet, which affect many types of cells) are particularly dependent on there being a rapid rise in the carrier level at the end of the gap. In contrast, one may speculate that the detection responses of human listeners must depend on the results of additional processing at higher neuronal levels that might assign a greater weight to other features of the gap, for example, its effective quiet time. It would be interesting then to see if reflex modification audiometry applied to the human listener (as in Ison and Pinckney, 1983) might reveal the same decrement in the inhibitory effect of a gap that ended with a slowly rising noise onset as seen here in the mouse.

ACKNOWLEDGMENTS

This research was supported by USPHS Research Grant No. AG09524, and by the Rochester International Center for Hearing and Speech Research. A preliminary report of these findings was given before the Association for Research in Otolaryngology, St. Petersburg Beach, FL, February 2000.

Allen, P. D., Virag, T. M., and Ison, J. R. (**2002**). J. Acoust. Soc. Am. (submitted).

Barsz, K., Benson, P. K., and Walton, J. P. (**1998**). ''Gap encoding by inferior collicular neurons is altered by minimal changes in signal envelope,'' Hear. Res. **115**, 13–26.

Buus, S., and Florentine, M. (**1985**). ''Gap detection in normal and impaired listeners: The effect of level and frequency,'' in *Time Resolution in Auditory Systems*, edited by A. Michelsen (Springer, London), pp. 159–179.

Crofton, K. M., Dean, K. F., Sheets, L. P., and Peele, D. (**1990**). ''Evidence for the involvement of associative conditioning in reflex modification of the acoustic startle response with gaps in a background noise,'' Psychobiol. **18**, 467–474.

Davis, M. (**1984**). ''The mammalian startle response,'' *Neural Mechanisms of Startle Behavior*, edited by R. C. Eaton (Plenum, New York), Chap. 10, pp. 287–351.

Eddins, D. A., Hall, J. W., and Grose, G. H. (**1992**). ''The detection of temporal gaps as a function of frequency region and absolute noise bandwidth,'' J. Acoust. Soc. Am. **91**, 1069–1077.

Florentine, M., and Buus, S. (**1984**). ''Temporal gap detection in sensorineural and simulated hearing impairment,'' J. Speech Hear. Res. **27**, 449–455.

Florentine, M., Buus, S., and Geng, W. (**1999**). ''Psychometric functions for gap detection in a yes–no procedure,'' J. Acoust. Soc. Am. **106**, 3512–3520.

Forrest, T. G., and Green, D. M. (**1987**). ''Detection of partially filled gaps in noise and the temporal modulation transfer function,'' J. Acoust. Soc. Am. **82**, 1933–1943.

Horlington, M. (**1968**). ''A method for measuring acoustic startle response latency and magnitude in rats: Detection of a single stimulus effect using latency measurements,'' Physiol. Behav. **3**, 839–844.

Irwin, R. J., and McAuley, S. F. (**1987**). ''Relations among temporal acuity, hearing loss, and the perception of speech distorted by noise and reverberation,'' J. Acoust. Soc. Am. **81**, 1557–1565.

Ison, J. R. (**1978**). ''Reflex inhibition and reflex elicitation by acoustic stimuli differing in abruptness of onset and peak intensity,'' Anim. Learn Behav. **6**, 106–110.

Ison, J. R., Agrawal, P., Pak, J., and Vaughn, W. J. (**1998**). ''Changes in temporal acuity with age and with hearing impairment in the mouse: A study of the acoustic startle reflex and its inhibition by brief decrements in noise level,'' J. Acoust. Soc. Am. **104**, 1696–704.

Ison, J. R., and Bowen, G. P. (**2000**). ''Scopolamine reduces sensitivity to auditory gaps in the rat, suggesting a cholinergic contribution to temporal acuity,'' Hear. Res. **145**, 169–176.

Ison, J. R., McAdam, D. W., and Hammond, G. R. (**1973**). ''Latency and amplitude changes in the acoustic startle reflex of the rat produced by variation in auditory prestimulation,'' Physiol. Behav. **10**, 1035–1039.

Ison, J. R., and Pinckney, L. A. (**1983**). ''Reflex inhibition in humans: Sensitivity to brief silent periods in white noise,'' Percept. Psychophys. **34**, 84–88.

Leitner, D. S., and Cohen, M. E. (**1985**). ''Role of the inferior colliculus in the inhibition of acoustic startle in the rat,'' Physiol. Behav. **34**, 65–70.

Leitner, D. S., Hammond, G. R., Springer, C. P., Ingham, K. M., Mekilo, A. M., Bodison, P. R., Aranda, M. T., and Shawaryn, M. A. (**1993**). ''Parameters affecting gap detection in the rat,'' Percept. Psychophys. **54**, 395–405.

Li, L., Priebe, R. P., and Yeomans, J. S. (**1998**). ''Prepulse inhibition of acoustic or trigeminal startle of rats by unilateral electrical stimulation of the inferior colliculus,'' Behav. Neurosci. **112**, 1187–1198.
McCroskey, R. L., and Kidder, H. C. (**1980**). ''Auditory fusion among learning disabled, reading disabled, and normal children,'' J. Learn Disabil **13**, 69–76.
Michelsen, A. (**1985**). *Time Resolution in Auditory Systems* (Springer, London).
Penner, M. J. (**1975**). ''Persistence and integration: Two consequences of a sliding integrator,'' Percept. Psychophys. **18**, 114–120.
Plomp, R. (**1964**). ''Rate of decay of auditory sensation,'' J. Acoust. Soc. Am. **36**, 277–282.
Salvi, R. J., and Arehole, S. (**1985**). ''Gap detection in chinchillas with temporary high-frequency hearing loss,'' J. Acoust. Soc. Am. **77**, 1173–1177.
Snell, K. B. (**1997**). ''Age-related changes in temporal gap detection,'' J. Acoust. Soc. Am. **101**, 2214–2220.
Snell, K. B., Ison, J. R., and Frisina, D. R. (**1994**). ''The effects of signal frequency and absolute bandwidth on gap detection in noise,'' J. Acoust. Soc. Am. **96**, 1458–1464.
Trehub, S. E., Schneider, B. A., and Henderson, J. L. (**1995**). ''Gap detection in infants, children, and adults,'' J. Acoust. Soc. Am. **98**, 2532–2541.
Walton, J. P., Frisina, R. D., Ison, J. R., and O'Neill, W. E. (**1997**). ''Neural correlates of behavioral gap detection in the inferior colliculus of the young CBA mouse,'' J. Comp. Physiol. [A] **181**, 161–176.
Young, L. S., and Fechter, L. D. (**1983**). ''Reflex inhibition procedures for animal audiometry: A technique for assessing oto-toxicity,'' J. Acoust. Soc. Am. **73**, 1686–1693.

Auditory temporal resolution in birds: Discrimination of harmonic complexes

Robert J. Dooling[a)]
Department of Psychology, University of Maryland, College Park, Maryland 20742

Marjorie R. Leek
Army Audiology and Speech Center, Walter Reed Army Medical Center, Washington, DC 20307

Otto Gleich
ENT Department, University of Regensburg, 93042 Regensburg, Germany

Micheal L. Dent
Department of Psychology, University of Maryland, College Park, Maryland 20742

(Received 13 November 2001; revised 15 May 2002; accepted 23 May 2002)

The ability of three species of birds to discriminate among selected harmonic complexes with fundamental frequencies varying from 50 to 1000 Hz was examined in behavioral experiments. The stimuli were synthetic harmonic complexes with waveform shapes altered by component phase selection, holding spectral and intensive information constant. Birds were able to discriminate between waveforms with randomly selected component phases and those with all components in cosine phase, as well as between positive and negative Schroeder-phase waveforms with harmonic periods as short as 1–2 ms. By contrast, human listeners are unable to make these discriminations at periods less than about 3–4 ms. Electrophysiological measures, including cochlear microphonic and compound action potential measurements to the same stimuli used in behavioral tests, showed differences between birds and gerbils paralleling, but not completely accounting for, the psychophysical differences observed between birds and humans. It appears from these data that birds can hear the fine temporal structure in complex waveforms over very short periods. These data show birds are capable of more precise temporal resolution for complex sounds than is observed in humans and perhaps other mammals. Physiological data further show that at least part of the mechanisms underlying this high temporal resolving power resides at the peripheral level of the avian auditory system. [DOI: 10.1121/1.1494447]

PACS numbers: 43.80.Lb, 43.66.Gf, 43.64.Tk [WA]

I. INTRODUCTION

Bird vocalizations are known to be extremely complex acoustic signals. This observation, along with the well-known differences between birds and mammals in the anatomy and physiology of the peripheral and central auditory systems, has led some to suggest that birds must have extremely fine temporal processing abilities (Carr and Friedman, 1999; Greenewalt, 1968; Konishi, 1969; Pumphrey, 1961; Schwartzkopff, 1968). Notwithstanding such indications, however, direct psychophysical measures of temporal processing such as detection of gaps in noise, temporal integration, duration discrimination, and temporal modulation transfer functions have all shown that birds are, overall, not more sensitive to the temporal features of acoustic signals than are other vertebrates (Dooling, 1982; Dooling *et al.*, 2000; Dooling and Haskell, 1978; Dooling and Searcy, 1981, 1985; Fay, 1988; Klump and Maier, 1989).

For the most part, the psychophysical tests that have failed to reveal differences between mammals and birds in temporal resolution have used simple stimuli and involved slow, overall changes in amplitude (the waveform envelope) rather than rapid pressure variations that carry acoustic information (the temporal fine structure). It may be that such stimuli do not provide an adequate test of the limits of temporal resolution in the avian ear. Many birds produce and learn complex tonal or harmonic vocalizations that involve rapid modulations in frequency and amplitude (for reviews, see Kroodsma and Miller, 1982, 1996), some of which are inaudible to humans. It is of interest to know whether the kinds of changes that occur in these complex sounds are within the perceptual capabilities of the species of birds that learn and use them as communication signals. As one example, the harmonic vocalizations of the zebra finch (*Taeniopygia guttata*) have very short fundamental periods of about 1.5 ms, shorter than most estimates of temporal resolution in the human auditory system (Viemeister and Plack, 1993). Any acoustic information produced by waveform fine structure within these periods is undoubtedly lost to humans, but may be available to zebra finches and other birds.

Taking human speech as an example, acoustic variability in intensity, frequency, and time, both in steady-state utterances like vowels and in rapidly changing sounds such as consonants and diphthongs, provides information to the listener regarding the speaker's individual identification, emotional state, and intended message. While we are much more familiar with human speech than with bird vocalizations, even the most cursory analysis of vocalizations of various

a)Electronic mail: dooling@psyc.umd.edu

 0001-4966/2002/112(2)/748/12/$19.00

bird species shows that acoustic differences among calls include changes in spectrum, waveform envelope, and temporal fine structure both within and across vocalizations of individuals (as, for example, the fine structure within the periods of harmonic zebra finch vocalizations). To the extent that these changes in acoustic characteristics are distinguishable to the intended communication target, they have the potential of being communicatively relevant. Historically, analyses of bird vocalizations have focused extensively on spectral features, much less on global temporal or envelope modulations occurring over the duration of vocalizations, and not at all on variations in temporal fine structure (Kroodsma and Miller, 1982, 1996). Perception of temporal fine structure may be more relevant to the problems of acoustic communication than previously thought, given the recent findings on the degree of fine motor control in avian vocal production (Brainard and Doupe, 2001; Fee *et al.*, 1998; Tchernichovski *et al.*, 2001; Yu and Margoliash, 1996). In the following series of experiments, the abilities of three species of birds to discriminate temporal envelope and temporal fine structure in complex sounds have been assessed. These sounds have similar spectra but differ systematically in envelope and fine structure. Both behavioral and electrophysiological experiments have been undertaken in a converging approach to defining those temporal limits and to providing preliminary evidence concerning the physiological bases of the perceived differences among stimuli. In experiment 1, birds were required to discriminate harmonic complexes constructed with frequency components all starting in cosine phase or all in randomly selected phases, with fundamental periods varying between 10 and 1 ms. Complexes within pairs of stimuli to be discriminated thus have equivalent long-term spectra, but different temporal waveform structures. Moreover, discrimination must be based on increasingly short segments of the stimuli as the period shrinks. A second experiment retested the birds' ability to use intraperiod structure in the waveforms of harmonic complexes, but envelope information was essentially removed as a cue from the stimuli, and only fine structure remained as a basis for making discriminations. Within a stimulus pair, the members to be discriminated are the time reverse of one another, with harmonic component phases selected according to an algorithm that systematically increases or decreases phase (and instantaneous frequency within the periods). In all, birds of three different species were trained by operant conditioning with food reward to discriminate within pairs of harmonic complexes over several different fundamental frequencies. The results for birds were compared to the results from humans tested on identical stimuli. Finally, cochlear microphonic (CM) and compound action potential (CAP) recordings from the three bird species to the same harmonic complexes used in behavioral tests revealed correlates to the species differences in sensitivity to temporal fine structure obtained behaviorally.

II. EXPERIMENT 1—DISCRIMINATION OF COSINE AND RANDOM PHASE HARMONIC COMPLEXES

In the first behavioral experiment, discrimination of harmonic stimuli with two kinds of phase selection was measured to determine whether, when frequency and amplitude information are held constant, the shape of the temporal waveform alone can provide different perceptions. By using harmonic complexes with a range of fundamental frequencies, we can also determine the limits of the duration of the harmonic periods that will support this discrimination. The discrimination task in each case was between a harmonic complex with each of the components starting in cosine phase (resulting in a highly peaked waveform) with seven different harmonic complexes with all components starting in randomly selected phases.

A. Materials and methods

1. Subjects

Three budgerigars (*Melopsittacus undulatus*) were tested on the discrimination of cosine phase harmonic complexes from random phase harmonic complexes. These birds were either bought commercially or hatched at the University of Maryland and housed in individual cages in a vivarium at the University of Maryland. The birds were kept on a normal day/night cycle correlated with the season at approximately 90% of their free-feeding weights. Animal housing and care met all standards of the University of Maryland Animal Care and Use Committee. All birds had hearing within normal limits for their species, as shown by their audiograms (Dooling *et al.*, 2000). Three humans were also tested. They were experimenters working in the laboratory at the time of the experiment, reported no history of hearing disorders, and had absolute thresholds at audiometric test frequencies better than 20 dB HL (*re* ANSI, 1989).

2. Stimuli and procedures

The stimuli were harmonic complexes consisting of a set of equal-amplitude harmonic components of a given fundamental frequency, with frequencies ranging from 200 (or the fundamental frequency) to 5000 Hz [see Fig. 1(a)]. The actual number of harmonic components varied with the fundamental frequency, which was either 200, 400, 800, or 1000 Hz (fundamental periods of 5, 2.5, 1.25, and 1 ms). For each fundamental frequency, one complex was generated with all components in cosine starting phase, and seven different complexes were produced with each component phase selected randomly from a rectangular distribution ranging from 0 to 2π radians. A different set of random phases was selected for each complex at each of the four fundamental frequencies. Within a given fundamental frequency, each stimulus had identical long-term frequency spectra, but differently shaped temporal waveforms. The waveforms were 260 ms in duration including 20-ms cosine^2 onset and offset ramps. Figure 1(b) shows examples of cosine-phase and random-phase stimuli for two of the fundamental frequencies used in this experiment.

The waveforms were created digitally, at a sampling rate of 40 kHz, using software provided by Tucker Davis Technologies (TDT, Gainesville, FL) to combine frequencies in the correct phases and amplitudes, followed by an inverse

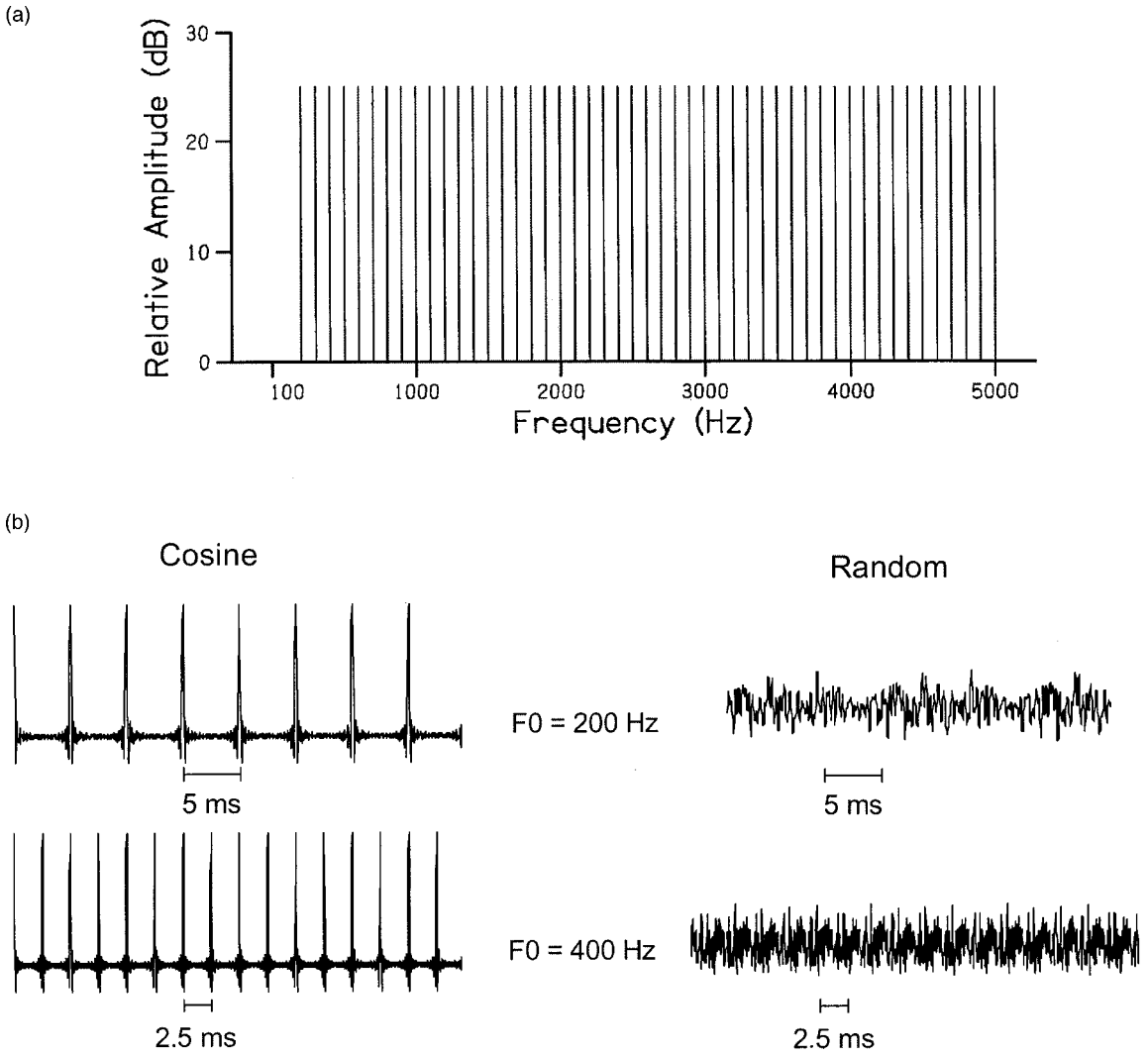


FIG. 1. (a) Schematic of the spectrum of one of the harmonic complexes used in this study. The frequencies in the stimuli ranged from 200 to 5000 Hz. (b) Temporal waveforms of harmonic complexes with the phase of each component starting in either cosine phase (left) or a randomly-selected phase (right).

fast Fourier transform (FFT) to create the waveforms. They were created off-line and stored as files for playback during the experiments.

3. Behavioral testing apparatus

The birds were tested in a wire cage ($23\times25\times16$ cm^3) mounted in a sound-isolation chamber (Industrial Acoustics Company, IAC-3). A response panel consisting of two microswitches with light-emitting diodes (LEDs) was mounted on the wall of the test cage just above a food hopper. The bird could trip the microswitch by pecking at the LED. The left microswitch and LED served as the observation key while the right microswitch and LED served as the report key. The behavior of the animals during test sessions was monitored by a video camera system (Sony HVM-322).

Test sessions were controlled by a Pentium PC computer. The digital stimuli were output to an overhead loudspeaker (KEF Electronics, Holliston, MA, model 80C), located 25 cm above the bird's head. Stimuli were presented through Tucker-Davis modules at a sampling rate of 40 kHz and presented at 80 dB SPL. Stimulus calibration was performed using a Larson Davis (Provo, UT) sound level meter (model 824). Stimulus intensities were measured with a $\frac{1}{2}$-in. microphone attached to the sound level meter via a 3-m extension cable. The microphone was placed in front of the response keys in the approximate position occupied by the bird's head during testing. Stimulus intensities were measured several times during these experiments to ensure that stimulus levels remained constant and that the entire audiometric system remained calibrated.

4. Training and testing procedures

Birds were trained by standard operant auto-shaping procedures to peck at the left LED (observation key) during a repeating background of sound (i.e., the presentation of a cosine phase complex at a rate of 2/s, or an interstimulus interval of 240 ms) until a new stimulus (i.e., the presentation of one of the random phase harmonic complexes with the same fundamental frequency) was presented alternately with the background sound, and then to peck the right LED (report key) when the change was detected. If the bird pecked the report key within 2 s of this alternating pattern, the food hopper was activated for 2 s, allowing the bird to obtain food reinforcement. During each experimental ses-

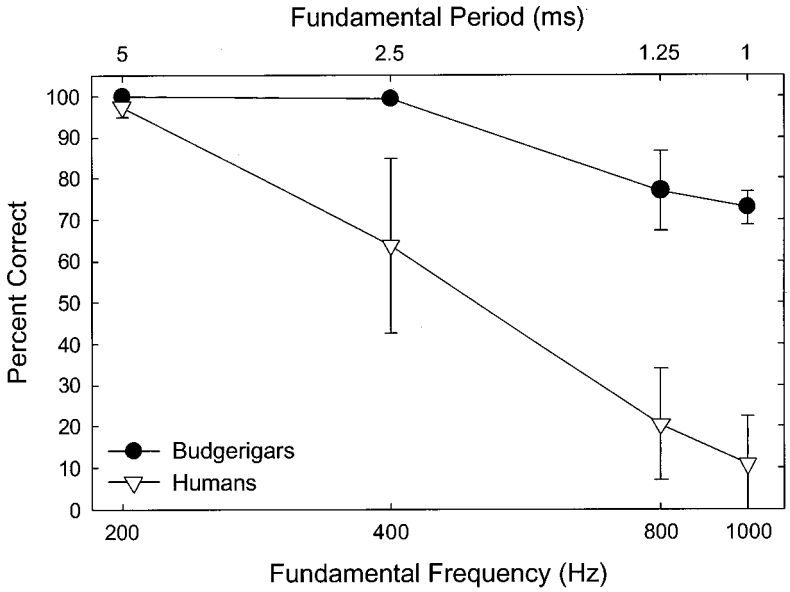


FIG. 2. Results from three budgerigars and three humans tested on the cosine phase versus random phase waveform discrimination at different fundamental frequencies. Performance is shown as percent correct discriminations and error bars are the between-subject standard errors.

sion, subjects listened to only one of the sets of fundamental frequency complexes and were required to discriminate between the cosine phase and the random phase complexes within the set. The harmonic complexes of all four fundamental frequencies (i.e., 200, 400, 800, and 1000 Hz) were tested in a random order and a different random order was used for each bird.

The bird initiated a trial by pecking repeatedly on the observation key. The time between pecking the observation key (i.e., the start of a trial) and the beginning of an alternating sound pattern was randomized between 2 and 7 s. Hit rate was defined as the proportion of target trials (when the background alternated with one of the random phase targets) on which the bird pecked the report key within 2 s of the alternating sound pattern. A failure to peck the report key within 2 s of sound alternation was recorded as a miss for a target trial and a correct rejection in the case of a sham trial. Following a miss, no reinforcement was given and a new trial sequence was initiated. Thirty percent of all trials were sham trials in which the target sound was the same as the repeating background sound. A peck to the report key during a 2-s sham trial was recorded as a false alarm, and the lights in the test chamber were extinguished while the repeating background continued. The length of this time-out period was normally 5 s, but varied according to an individual bird's behavior, with longer time-out periods imposed if birds began developing higher false alarm rates. Sessions with a total false alarm rate of 16% or higher were discarded. Fewer than 20% of all sessions across birds were discarded for this reason. The mean false alarm rate across birds was 5%. The birds were typically tested in one to two daily sessions consisting of about 100 trials each, until percent correct values stabilized, and then testing continued for another 200 trials. Final percent correct discrimination values were taken as the mean percent correct over the last 200 trials.

The human listeners were laboratory staff members, tested with the same stimuli as the birds, using a standard/two-alternative forced-choice procedure. Stimuli were presented over earphones, at a level of 80 dB SPL. For each fundamental frequency, comparisons were made between the cosine-phase stimulus and each of the seven random-phase stimuli in blocks of 40 trials each. On each trial, the cosine-phase complex was presented, followed by the same stimulus and the comparison stimulus, in random order, separated by 300 ms of silence. Subjects were asked to indicate which of the second or third presentation on a trial was different from the standard presentation. After the subject touched a marked area on a touch screen terminal to indicate a response, correct answer feedback was provided, and the next trial was initiated after 500 ms. Percent correct responses were averaged across the seven random-phase stimuli for each fundamental frequency. These values, ranging from chance performance at 50% correct to perfect performance (100%) were subsequently scaled to the range of 0%–100% for comparison with the bird data.

There is always some concern when comparing data for human subjects taken under earphones with animal behavioral data measured in a sound field (see Leek *et al.*, 2000, for a discussion of this problem). To be assured that these differences in measurement were unlikely to materially affect the data reported in this and the following experiments, humans were also tested informally on selected stimuli in the chamber used to test the birds. For these trials, the test cage was removed from the chamber and the human subject stood with their head in the sound chamber in the approximate location of the test cage, with one ear pointed toward the speaker at a distance roughly corresponding to the distance between the speaker and the bird during testing. The same software, procedures, and stimuli were used that were employed in testing the birds. Generally similar results were obtained in these tests when compared with results under earphones. These procedures provided added assurance of the validity of comparing birds tested in free field with humans tested under earphones.

B. Results

Figure 2 shows that budgerigars can discriminate cosine phase harmonic complexes from random phase complexes at

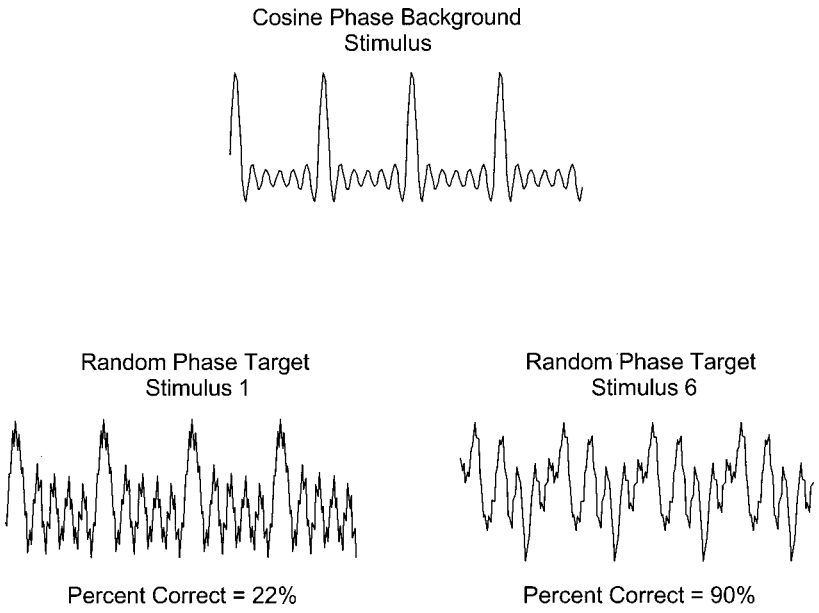


FIG. 3. Several periods of the cosine phase harmonic complex and two of the random phase harmonic complexes, all with a fundamental frequency of 800 Hz. The random phase harmonic complex on the left had a waveform shape more similar to the cosine phase harmonic complex, and overall discrimination was poorer than that for the random phase harmonic complex on the right, which had a shape less similar to the cosine phase harmonic complex.

much higher fundamental frequencies (shorter periods) than can humans. The budgerigars performed at 100% up to 400 Hz, while humans had dropped to about 65% at that fundamental frequency. Even at the highest fundamental frequency of 1000 Hz, the budgerigars' performance was superior to that of humans at 400 Hz. A two-way repeated measures (species×fundamental frequency) ANOVA showed a significant effect of species [$F(1,4)=19.32$, $p<0.05$] and fundamental frequency [$F(3,12)=13.69$, $p<0.001$], but there was no significant interaction between the variables [$F(3,12)=3.41$, $p>0.05$]. *Post-hoc* tests using a Bonferroni *t*-test showed that budgerigars were better than humans at fundamental frequencies ranging from 400 to 1000 Hz ($p<0.05$). The results with humans are consistent with those of Patterson (1987), who found that human subjects could discriminate a cosine-phase waveform from a random-phase waveform as long as the fundamental frequency of the waveforms was below about 400–500 Hz.

By chance, some of the random phase selections produced waveforms with envelopes similar to cosine phase waves. Examples of a cosine phase harmonic complex and two random phase harmonic complexes with fundamental frequencies of 800 Hz are shown in Fig. 3. Performance for the birds was worse on the random phase complexes that were similar to the envelope shape of the cosine phase complex (e.g., random phase target 1 PC=22%) than on random phase complexes that were dissimilar (e.g., random phase target 6 PC=90%). As a measure of similarity between waveforms, a cross-correlation between the cosine-phase waveform and each of the random phase stimuli at fundamental frequencies of 800 and 1000 Hz was calculated (these were the only fundamental frequencies that had sufficient error rates to produce meaningful correlations). In general, as the cosine- and random-phase waveforms increased in similarity, discrimination performance decreased. There was a significant negative correlation between the similarity of the cosine- and random-phase complexes, and the birds' discrimination accuracy ($r=-0.52$, $p<0.05$). The relationship between waveform similarity and discrimination argues that the birds were relying to some extent on characteristics of the temporal waveform in discriminating cosine phase versus random phase complexes.

III. EXPERIMENT 2—DISCRIMINATION BETWEEN POSITIVE AND NEGATIVE SCHROEDER WAVEFORMS

The discrimination of cosine from random phase harmonic complexes was driven, at least in part, by characteristics of the temporal waveform which included both the envelope and the within-period fine structure. Perceptual differences among these stimuli may also include differences in loudness. There are two additional phase selections that can be used to disambiguate the influence of these waveform characteristics. These phase selections essentially hold envelope information constant across stimuli to be discriminated, but reverse the temporal fine structure between stimuli. These Schroeder-phase waveforms have recently been used in a series of behavioral experiments on auditory masking and reveal significant differences in hearing between birds and humans (Dooling *et al.*, 2001; Leek *et al.*, 2000). In humans, but not in birds, Schroeder-phase harmonic complexes constructed with monotonically increasing (positive Schroeder) or decreasing (negative Schroeder) component phases are differentially effective as maskers, even though they have essentially identical temporal envelopes and long-term spectra. One explanation for the similarity in masking effectiveness of these harmonic complexes in birds is that the different masker waveforms were indiscriminable. The present experiment tests this explanation. Further, the temporal limitations on birds' abilities to discriminate harmonic complexes using fine structure alone were tested using pairs of Schroeder-phase waveforms created with different fundamental frequencies.

A. Materials and methods

1. Subjects

Three zebra finches, three budgerigars, and three canaries (*Serinus canaria*) were used as subjects in this experi-

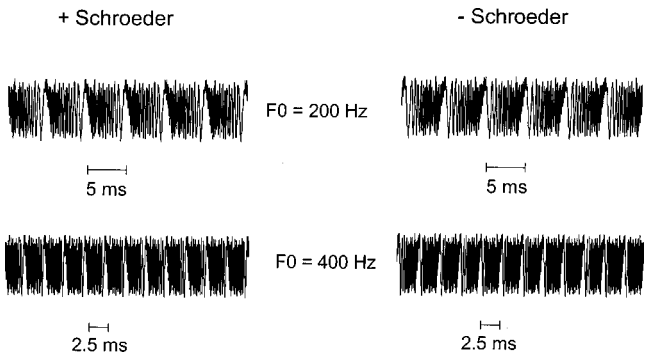


FIG. 4. Temporal waveforms of positive- and negative-Schroeder-phase harmonic complexes with a fundamental frequency of 200 and 400 Hz.

ment. Three humans, researchers in the laboratory, were also tested.

2. Stimuli

Waveforms were constructed in a manner similar to those used in experiment 1, but with component starting phases selected according to an algorithm developed by Schroeder (1970). The component amplitudes were equal, and the frequency range of the stimuli was from 200 (or the fundamental frequency) to 5000 Hz, as was used in experiment 1. Seven pairs of these harmonic complexes were produced, with fundamental periods ranging from 6.6 ms (fundamental frequency of 150 Hz) to 1 ms (fundamental frequency of 1000 Hz) in duration. Figure 4 shows examples of negative and positive Schroeder-phase waveforms for two of the fundamental frequencies used here. The phases of the components were monotonically increasing (positive Schroeder complex) or decreasing (negative Schroeder complex) with harmonic number, resulting in instantaneous frequencies that fell or rose monotonically across each period. The acoustic differences between members of a pair of these complexes are limited to temporal fine structure: all waveforms have a flat envelope and, within a pair defined by the fundamental frequency, have identical long-term spectra. The waveforms were 260 ms in duration including 20-ms cosine^2 onset and offset ramps.

The birds were tested in similar procedures as those described in experiment 1, and used the same psychoacoustic paradigm and experimental chambers. In this experiment, either the positive- or negative-phase waveform was selected as the repeating background, and the other Schroeder-phase wave of the same fundamental frequency was the target sound. Both Schroeder phases were tested as background and target, and the values were averaged for each fundamental frequency. Testing continued for harmonic complexes of all seven fundamental frequencies in random order from 150 to 1000 Hz. The stimuli were generated and output in a similar manner as that described in experiment 1.

Humans were tested under the same conditions (as near as could be) and using the same procedures as for the birds. Humans stood leaning into the small test chamber so that their head was directly under the speaker in the approximate location of the bird's head during testing. A small hand-held button box was used to signal observation and report, analogous to the bird's key pecking. The same stimulus presentation and psychophysical procedure was used as for the birds.

B. Results

Figures 5(a)–(c) show the performance of individual birds of each species tested on the positive/negative Schroeder waveform discriminations at different fundamental frequencies. All birds were able to discriminate between positive and negative Schroeder harmonic complexes at fundamental frequencies up to at least 600 Hz. Budgerigars and canaries showed some difficulty discriminating at the highest fundamental frequencies (800 and 1000 Hz), while the zebra finches discriminated easily between the positive and negative Schroeder waveforms even at the highest fundamental frequency.

Figure 5(d) shows the average of all birds from each species, and the average of three humans. Large differences are evident among the species: human listeners begin having difficulty making these discriminations when the fundamental period becomes shorter than about 3 ms, budgerigars and canaries do much better, and zebra finches have little difficulty even over periods as short as 1 ms. A two-way repeated measures (species×fundamental frequency) ANOVA was conducted. Results showed a significant effect of species $[F(3,8)=38.82,\ p<0.001]$, fundamental frequency $[F(4,32)=53.63,\ p<0.001]$, and a significant interaction between the variables $[F(12,32)=14.12,\ p<0.001]$. *Post-hoc* tests using a Bonferroni *t*-test showed that all three species of birds were better than the humans ($p<0.05$).

Clearly, birds have better resolution of temporal fine structure than humans, notwithstanding some earlier reports of similar performance on other temporal processing tasks. As far as we know, these are the only comparative data available which directly address the question of sensitivity to temporal fine structure in complex sounds and they raise questions about both the perceptual differences among birds and between birds and humans. For this reason, it is of interest to explore the possible mechanisms underlying these species differences by using physiological techniques that may localize responses within the peripheral auditory system.

IV. EXPERIMENT 3—COCHLEAR MICROPHONIC AND COMPOUND ACTION POTENTIALS

The locus of the mechanisms underlying the species' differences in behavioral discrimination thresholds for temporal fine structure are unclear, but there are dramatic differences between birds and mammals at all levels of the auditory system (Carr and Code, 2000; Gleich and Manley, 2000). In fact, much of the past evidence for enhanced temporal resolution in birds has been indirect, with explanations that appeal to the fine structure of vocalizations or invoke either anatomical and physiological aspects of cochlear processing (Greenewalt, 1968; Konishi, 1969; Pumphrey, 1961; Schwartzkopff, 1968). As a first step in exploring the physiological bases of species differences observed psychophysi-

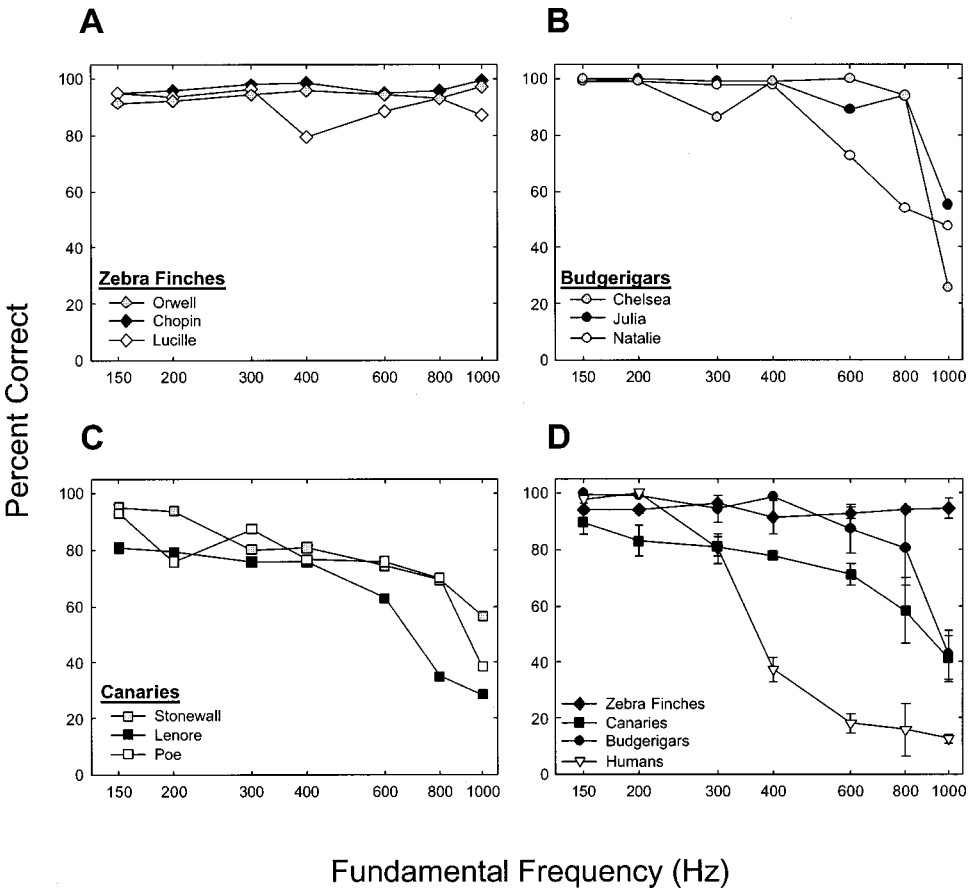


FIG. 5. Results from individual zebra finches (a), budgerigars (b), and canaries (c), tested on the positive/negative Schroeder waveform discrimination at different fundamental frequencies. Average results from the three species of birds and three humans are shown in (d). Error bars are the between-subject standard errors.

cally in experiments 1 and 2, cochlear microphonic potentials (CMs) and compound action potentials (CAPs) from the VIIIth nerve were recorded from the round window in the birds' ears in response to some of the same stimuli used in the behavioral discrimination studies. For a mammalian comparison, similar measures were also taken in the Mongolian gerbil (*Meriones unguiculatus*). If the negative and positive Schroeder complexes that were discriminable at high fundamental frequencies in the behavioral studies with birds do, in fact, generate differential cochlear microphonics or compound action potentials in birds, but not in gerbils, it would suggest that the search for mechanisms accounting for the differences observed behaviorally between birds and humans might begin with consideration of the auditory periphery.

A. Materials and methods

1. Subjects

Two budgerigars, two zebra finches, two canaries, and two Mongolian gerbils were used in these experiments.

2. Procedures

Birds were anesthetized with injections of 20 mg/kg xylazine and 40 mg/kg ketamine into the breast muscle, and gerbils with 15 mg/kg xylazine and 60 mg/kg ketamine injected intraperitoneally. Additional doses of anesthetics (50% of the initial dose) were supplemented as needed, determined either by occasional foot pinch or increased muscle noise in the electrical recordings (generally every 30 to 60 min).

Surgical procedures to gain access to the cochlea have been previously described in detail for birds (Manley *et al.*, 1985) and gerbils (McGuirt *et al.*, 1995). Feathers and hair were removed from the head and around the external ear opening. An incision in the skin along the midline of the skull exposed the bone and it was cleared of connective tissue and dried. A small screw was cemented on the surface of the skull with dental cement to allow precise and stable placement of the head in a holding device and reproducible positioning of the ear canal opening in relation to the speaker.

In birds, the base of the cochlea was exposed by a dorsolateral approach. The core of a thin teflon coated silver wire (0.005 in., WPI) was exposed at the end and inserted through a tiny hole in scala tympani to give direct electrical access to the perilymph. The teflon insulation sealed the hole and prevented leak of perilymph. A subdermal needle inserted into a neighboring neck muscle served as reference and grounding was performed by a needle inserted into the skin of the neck further caudal (standard platinum alloy, subdermal needle electrodes, Grass; E-2B). The bird was placed in a body-shaped styrofoam block with a temperature sensor in contact with the breast muscle. A dc heating pad was used to cover the bird and keep the body temperature close to 40 °C.

In gerbils, an incision behind the external ear and careful dissection of the muscle layer exposed the bulla and a small piece of bone was removed from the bulla to gain access to the niche of the round window. The exposed end of a teflon

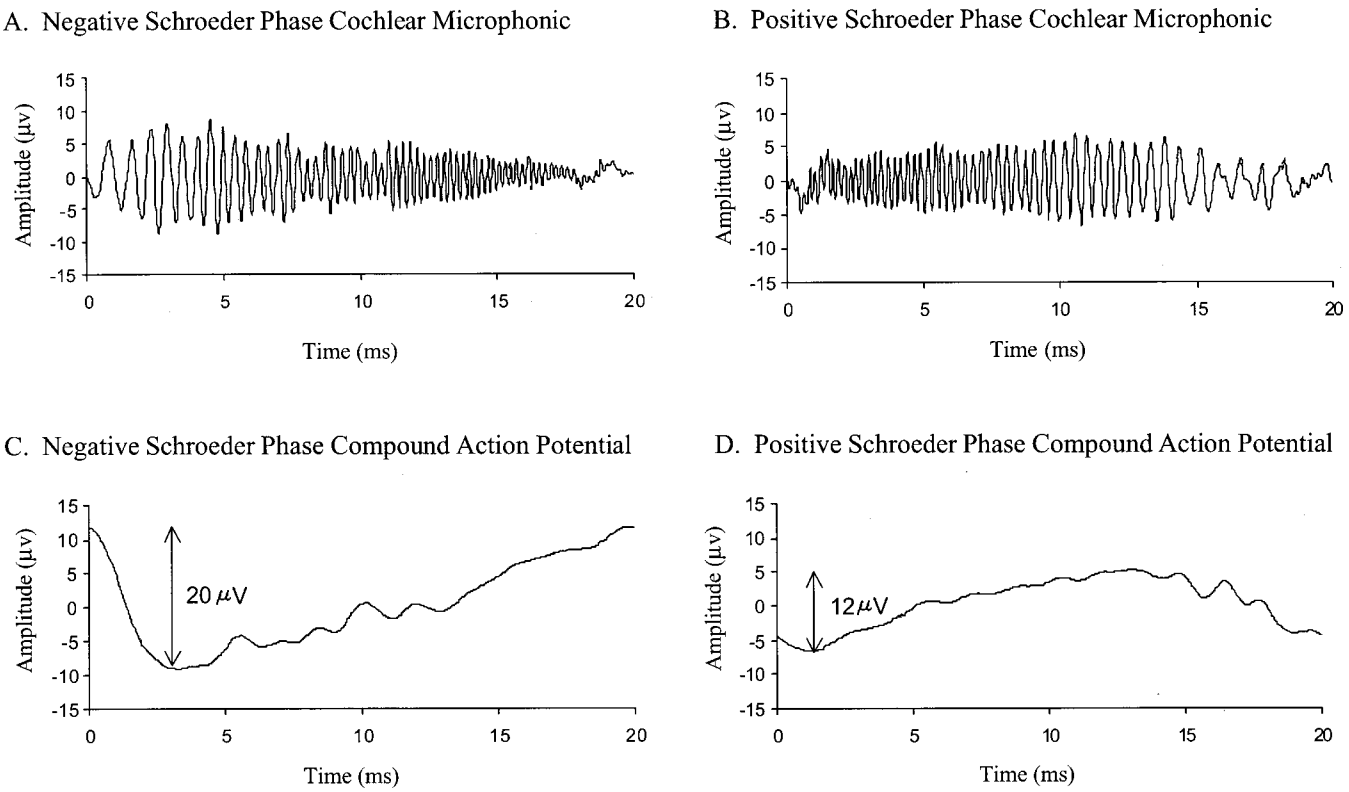


FIG. 6. Examples of CM (a,b) and CAP (c,d) responses to positive and negative Schroeder harmonic complexes, collapsed across periods, for one budgerigar.

insulated silver wire was hooked to the bony ridge of the round window niche and fixed with super glue to the exposed external surface of the bulla. Reference and grounding electrodes were placed as in birds. As with the birds, gerbils were also placed onto a styrofoam block in direct contact with the temperature sensor and covered by the dc heating blanket. The temperature was kept between 37 °C and 38 °C.

Cochlear responses were recorded with a low impedance digital amplifier (TDT, HS4/DB4) using the 60-Hz noise rejection, but no additional filtering. All subjects were placed in the sound field such that their head was at a precise location in the free field 40 cm from the speaker. The stimulus waveforms were fed through a DA1 digital-analog converter, a PA4 programmable attenuator, and a HB6 transducer, which directly drove the speaker (KEF SP 3235, Model 60S, KEF Electronics of America Inc., Holliston, MA). The electrodes were connected to the HS4 Headstage that amplified and digitized the signal before sending it over fiber optic cables to the DB4 Digital Biological Amplifier. A TG6 timing generator was used to synchronize A/D and D/A conversion at a sample rate of 40 kHz.

3. Stimuli

The stimuli were the same as those used in the Schroeder-phase behavioral experiments. The fundamental frequencies of the stimuli ranged from 50 to 1000 Hz. In addition to this set of normal stimuli, a set of inverted stimuli was generated to isolate cochlear microphonic (CM) and compound action potential (CAP) responses, as described below. They were played at 80 dB SPL, and calibrated at the location of the animal's head during the experiment as described for the behavioral experiments. Each stimulus was also recorded using a small microphone placed at that location, and the outputs of the speakers in response to each stimulus waveform were verified off-line.

4. Data collection and reduction

The response to each stimulus was averaged over 124 presentations repeated at a rate of 2/s. After each normal signal presentation, the response to the inverted version of the stimulus was recorded. The CAP response component was derived by adding the response traces obtained for the normal and the inverted stimulus (canceling the CM component) and scaling the resulting response amplitude by half. This derived neural response component was then subtracted from the response trace to the normal stimulus to derive the CM response.

The potentials in this experiment were then further prepared as follows. Taking a response segment, which included the plateau of the stimulus but omitted the rise/fall time (i.e., 220 ms in the middle of the response from 20 to 240 ms within the total 260-ms stimulus), an average CM and CAP period response was obtained by averaging across fundamental periods of the 100-ms segment of the response waveform. From this averaged period response, both the root mean square (rms) of the cochlear microphonic and the peak-to-peak amplitude of the CAP were examined. Examples of a CM and CAP from a budgerigar in response to a negative-Schroeder harmonic complex and a positive-Schroeder harmonic complex are shown in Figs. 6(a)–(d).

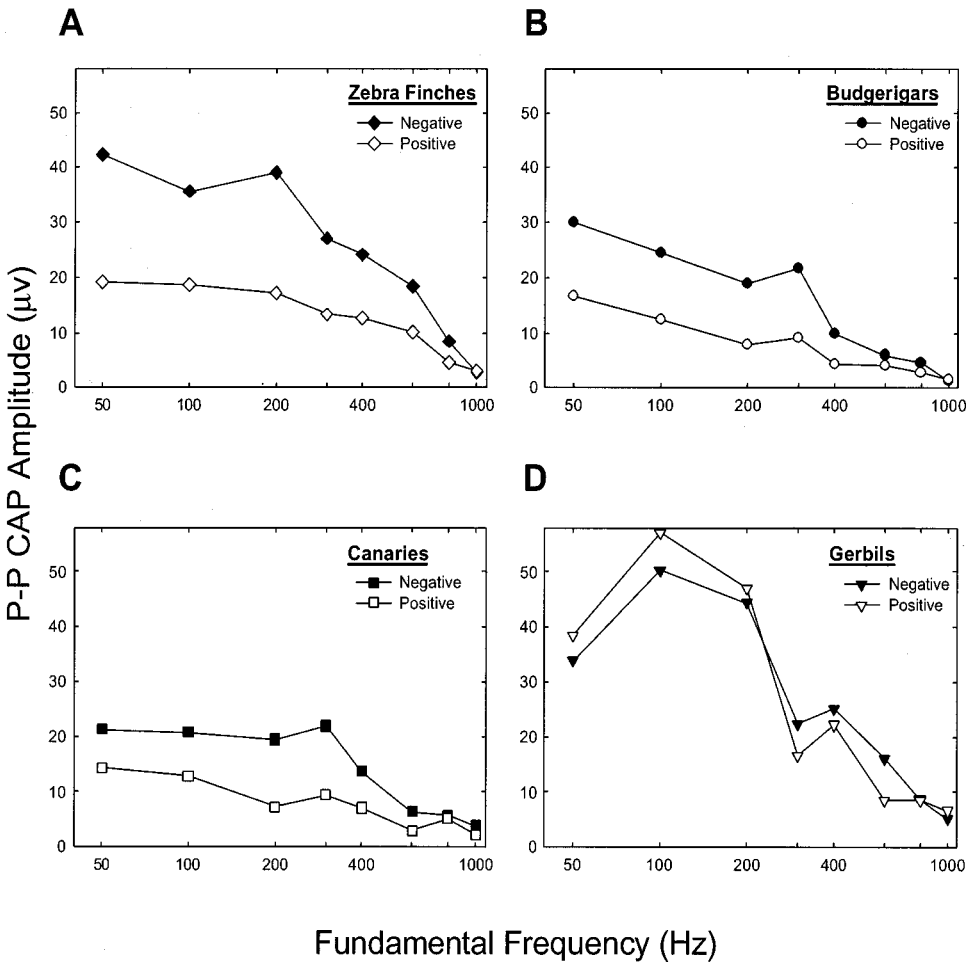


FIG. 7. Average CAP amplitude responses to positive and negative Schroeder complexes as a function of fundamental frequency for zebra finches (a), budgerigars (b), canaries (c), and gerbils (d).

B. Results

In all four species, the CM shape approximately followed the acoustic waveform shape, and the CM amplitude was independent of the fundamental frequency [see Figs. 6(a) and (b)]. In general, gerbils showed much larger CM amplitudes than birds. There was no difference in CM amplitude between negative- and positive-phase waveforms within each fundamental frequency pair for either birds or gerbils. A three-way ANOVA on the rms amplitude values of the cochlear microphonic showed that there was a significant effect of species [$F(3,64)=266.57$, $p<0.001$] but no effect of fundamental frequency [$F(7,64)=0.79$, $p>0.05$] or phase [$F(1,64)=0.97$, $p>0.05$]. No interactions were significant ($p>0.05$). *Post-hoc* tests using a Bonferroni *t*-test showed that rms amplitude of the CM in gerbils was significantly different from that in all species of birds ($p<0.05$), but was not significantly different across bird species ($p>0.05$).

The amplitude of the CAP, however, did vary in several interesting ways. There was a significant decrease for all species in the amplitude of the CAP with increasing fundamental frequency [Figs. 7(a)–(d)]. Gerbils showed no systematic difference in the amplitude of the CAP to positive versus negative Schroeder complexes over the entire range of fundamental frequencies, while all three species of birds showed a much larger CAP to the negative Schroeder complexes than the positive complexes at low fundamental frequencies. A three-way ANOVA showed a significant effect of species [$F(3,64)=21.92$, $p<0.001$], phase [$F(1,64)=35.24$, $p<0.001$], and fundamental frequency [$F(7,64)=27.69$, $p<0.001$]. The interaction between species and fundamental frequency was also significant [$F(3,64)=3.16$, $p<0.001$], but none of the other interactions was significant ($p>0.05$).

The results from the CAP measurements differ from the behavioral measures in that at the highest fundamental frequency tested (1000 Hz), there were no differences between responses to positive and negative Schroeder stimuli for any of the three species. However, paralleling the behavioral responses, as fundamental frequency increased, the differential response to positive and negative Schroeder waveforms persisted longer in finches (800–1000 Hz) than canaries (600–800 Hz), or budgerigars (400–600 Hz). This suggests that CAP responses may be partially related to the discrimination precision observed behaviorally, but that there are contributions from other peripheral or higher auditory processes.

V. DISCUSSION

We have shown that birds can discriminate between synthetic harmonic complexes that differ only in temporal fine structure over extremely short fundamental periods, and that they demonstrate differences in the VIIIth nerve compound action potentials that support this detail of auditory analysis in the synchronization of neural firing. Overall, the three species of birds were able to discriminate between several

types of harmonic complexes with higher fundamental frequencies than humans. These experiments show that birds have an ability to discriminate the temporal fine structure of complex sounds that is two to three times better than the limits shown for humans.

These results are surprising in view of the comparative data showing that birds are similar to mammals in temporal resolving power (see Dooling *et al.*, 2000). Only in a few of these measures is there a hint of consistent differences between birds and mammals. A comparison of temporal modulation transfer functions (TMTFs) of several mammals and several species of birds reveals that humans are more sensitive to modulation-based changes in intensity than birds at low modulation frequencies. Although birds and humans show a similar cutoff of performance at high modulation frequencies, the difference in intensity resolving power at low modulation rates results in a shorter time constant for birds compared to humans (Dooling *et al.*, 2000; Dooling and Searcy, 1981). Gap detection thresholds at very low intensity levels are also generally better in birds than mammals (Dooling *et al.*, 2000), probably reflecting the tendency of mammalian tuning curves to narrow with decreasing sound pressure levels. Taken together, results point to a need for future investigation into the possibility that enhanced sensitivity to temporal fine structure of complex sounds may be a distinguishing feature of the avian auditory system. A more complete evaluation of this hypothesis would require testing, along with more birds, many more mammals, especially those with auditory systems that appear to be specialized for processing auditory temporal information as, for instance, many species of bats (see review in Moss and Simmons, 1996).

Gerbils are becoming popular models of mammalian hearing for both simple and complex sounds (Heffner and Heffner, 1988; Ryan, 1976; Sinnott and Mosteller, 2001), though it is still the case that much less is known about their auditory capabilities than the birds in these experiments. Thus, it is not clear why the physiological data taken from the gerbils do not match the expectations based on behavioral data from the other mammal used in these experiments (humans). Behavioral data from humans suggests easy discrimination between the positive- and negative-Schroeder-phase waveforms at low fundamental frequencies, but the compound action potentials presented here for the gerbil show little difference between the two types of waveforms. Clearly, these results stand in stark contrast to findings for all three bird species that show a difference in CAP amplitude up to nearly 800 Hz paralleling the behavioral performance. The difference in CAP between birds and gerbils clearly demonstrates a difference in cochlear processing. Behavioral experiments are needed to determine whether gerbils truly cannot discriminate sounds differing in only temporal fine structure at the low fundamental frequencies used in our physiological tests. If they can make such discriminations behaviorally, then the coding of temporal fine structure must involve something other than the synchronous firing of the VIIIth nerve in gerbil. Alternatively, the lack of CAP asymmetry in gerbil might have no relation to the gerbil's ability to discriminate temporal fine structure in behavioral experiments.

A. Mechanisms underlying temporal resolution

Cochlear frequency analysis in mammals is often modeled as classical linear filtering with the broader filter bandwidths found in higher frequency regions supporting better temporal analysis and narrower filters in lower frequency regions resulting in poorer temporal analysis (Viemeister and Plack, 1993). Individual frequency components of a harmonic complex that are more widely spaced than the frequency analyzing channels in the auditory system significantly reduce the ability to perceive the temporal characteristics of the complex. When the fundamental frequency, and therefore component spacing, is such that several components fall within an auditory channel, the temporal properties of the sound can be used for discrimination. Discrimination of the random from the cosine phase waves and between the Schroeder-phase complexes demonstrated by the human listeners in this study reflected this limitation, as has been described repeatedly in the literature using tasks such as phase discrimination in three-component complexes (e.g., Goldstein, 1967), perception of pitches generated by spectral edges in harmonic complexes (Kohlrausch and Houtsma, 1992), and other studies of random versus cosine-phase discrimination in harmonic complexes (Patterson, 1987). These reports in the literature, as well as the data presented here, suggest that the ability to discriminate temporal cues in these harmonic stimuli is severely compromised when the fundamental frequency is greater than about 400–500 Hz.

Interestingly, the above relationship between component spacing, filter bandwidths, and temporal resolving power seems not to hold for the birds. By most accounts, birds have cochlear frequency resolution slightly worse than observed in humans (Dooling *et al.*, 2000; Sachs *et al.*, 1978) but tuning curves of some bird auditory nerve fibers are actually more sharply tuned than some mammals like cats and guinea pigs (see review in Gleich and Manley, 2000). Human frequency resolution estimated from suppression of distortion product otoacoustic emissions indicates resolution as good as or slightly worse than avian frequency resolution (see review in Gleich and Manley, 2000). Thus, a frequency domain analysis in the avian cochlea at reported levels conflicts with the ability of birds to make these waveform discriminations at fundamental frequencies exceeding 600 Hz. At such high fundamental frequencies, birds are making discriminations between two sounds that differ in fine structure over time intervals as small as 1 ms—much faster than any estimate of monaural temporal resolution capacity of humans. This basic difference in perception between humans and birds calls into question conventional views of the relationship between frequency resolution and temporal acuity within harmonic complexes, and points to the need for further comparative investigations of pitch, timbre, and frequency perception across species.

B. The relationship between perception and production

One reason that the differences among the species of birds in the ability to discriminate between the waveform shapes of harmonic complexes are intriguing is because of differences among the species in the characteristics of their species-specific vocalizations. Zebra finches are the only one of these three bird species whose calls and songs are strongly and almost exclusively harmonic, with fundamental frequencies approaching 600–700 Hz (see, for example, Zann, 1984). Interestingly, much is also known about the acoustic characteristics of the vocalizations of canaries and budgerigars (see, for example, Farabaugh and Dooling, 1996; Guttinger *et al.*, 1978; Lavenex, 1999). From these studies, as well as from our own casual observation of songs and calls of these species, it is clear that zebra finch calls have the most broadband harmonic structure, and that budgerigar vocalizations in general contain a greater proportion of such acoustic structure than do canary vocalizations. This order parallels the species differences in the ability to discriminate temporal fine structure over exceedingly small time windows, with zebra finches showing the largest perceptual and CAP differences between waveform shapes among the birds.

These apparent parallels between perception and production shown here suggest further comparative and developmental studies may be useful. It might be that these parallels arise through some form of co-evolution as has been suggested for human speech (Liberman *et al.*, 1967). One can imagine various ontogenetic strategies. One possibility is that the avian ear and auditory system in general excel at discrimination of temporal fine structure. Species differences may emerge as developing vocalizations fail to exploit initial sensitivities during critical periods, and these sensitivities are then lost in adulthood. A similar argument has been made regarding the development of human languages in infants by Kuhl and her colleagues (e.g., Kuhl, 2000). Another possibility is that the avian ear and auditory system remain highly plastic throughout life and that some bird species may learn that relevant information is coded in the fine temporal structure of various signals, leading to an enhanced sensitivity to temporal variation for these species.

With the exception of sound localization in birds, these data are the first to demonstrate behaviorally the use of fine temporal processing abilities for perception of acoustic stimuli with some similarities to communication signals. These findings of enhanced temporal resolution may have particular relevance for understanding acoustic communication, song learning, and individual recognition in birds. For zebra finches, in particular, confirmation of precise temporal coding in the auditory system (Janata and Margoliash, 1999; Sen *et al.*, 2001; Theunissen and Doupe, 1998) parallels recent findings of precise coding of temporal information in the motor circuits and syringeal machinery controlling song in this species (Brainard and Doupe, 2001; Fee *et al.*, 1998; Tchernichovski *et al.*, 2001; Vicario, 1991; Yu and Margoliash, 1996).

In summary, these data show temporal resolution in the processing of acoustic communication signals in birds well beyond the limits typically reported for humans and other mammals. Synchronized activity in the VIIIth nerve also reveals a sensitivity to waveform shape of such stimuli that may be unique to birds, suggesting a radical and unexpected difference in the coding of these stimuli in the peripheral auditory system of birds and mammals. At the level of acoustic communication, the analysis of bird vocalizations is usually done in the spectral domain and rarely extends to the level of temporal fine structure. The size of a species' communication repertoire and the degree of vocal learning is judged by these traditional types of analyses. If it turns out that birds typically perceive and make use of fine temporal detail in their complex vocalizations that is beyond the range of human capabilities, they may have much larger vocabularies than previously thought.

ACKNOWLEDGMENTS

This work was supported by NIH Grant No. DC-00198 to RJD and No. DC-00626 to MRL. MLD was supported by NSRAs from NIH (DC-00046 and MH-12698). We are very grateful to J. Cucina for his assistance in analyzing the physiological data. Special thanks also to B. Brittan-Powell, B. Lohr, and M. Mavilia for their help on various aspects of these studies. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

ANSI (**1989**). ANSI S3.6-1989, "Specifications for audiometers" (American National Standards Institute, New York).

Brainard, M. S., and Doupe, A. J. (**2001**). "Postlearning consolidation of birdsong: stabilizing effects of age and anterior forebrain lesions," J. Neurosci. **21**, 2501–2517.

Carr, C. E., and Code, R. A. (**2000**). "The central auditory system of reptiles and birds," in *Comparative Hearing: Birds and Reptiles*, edited by R. J. Dooling, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 197–248.

Carr, C. E., and Friedman, M. A. (**1999**). "Evolution of time coding systems," Neural Comput. **11**, 1–20.

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. 95–130.

Dooling, R. J., and Haskell, R. J. (**1978**). "Auditory duration discrimination in the parakeet (*Melopsittacus undulatus*)," J. Acoust. Soc. Am. **63**, 1640–1643.

Dooling, R. J., and Searcy, M. H. (**1981**). "Amplitude modulation thresholds for the parakeet (*Melopsittacus undulatus*)," J. Comp. Physiol. **143**, 383–388.

Dooling, R. J., and Searcy, M. H. (**1985**). "Temporal integration of acoustic signals by the budgerigar (*Melopsittacus undulatus*)," J. Acoust. Soc. Am. **77**, 1917–1920.

Dooling, R. J., Lohr, B., and Dent, M. L. (**2000**). "Hearing in birds and reptiles," in *Comparative Hearing: Birds and Reptiles*, edited by R. J. Dooling, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 308–359.

Dooling, R. J., Dent, M. L., Leek, M. R., and Gleich, O. (**2001**). "Masking by harmonic complexes in three species of birds: Psychophysical thresholds and cochlear responses," Hear. Res. **152**, 159–172.

Farabaugh, S. M., and Dooling, R. J. (**1996**). "Acoustic Communication in Parrots: Laboratory and Field Studies of Budgerigars, *Melopsittacus undulatus*," in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Cornell U.P., Ithaca, NY), pp. 97–117.

Fay, R. R. (**1988**). *Hearing in Vertebrates: A Psychophysics Databook* (Hill-Fay, Winnetka, IL).

Fee, M. S., Shraiman, B., Pesaran, B., and Mitra, P. P. (**1998**). "The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird," Nature (London) **395**, 67–71.

Gleich, O., and Manley, G. A. (**2000**). ''The hearing organ of birds and crocodilia,'' in *Comparative Hearing: Birds and Reptiles*, edited by R. J. Dooling, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 70–138.
Goldstein, J. L. (**1967**). ''Auditory spectral filtering and monaural phase perception,'' J. Acoust. Soc. Am. **41**, 458–479.
Greenewalt, C. H. (**1968**). *Bird Song: Acoustics and Physiology* (Smithsonian Institute, Washington, DC).
Guttinger, H. R., Wolffgramm, J., and Thimm, F. (**1978**). ''The relationship between species specific song programs and individual learning in songbirds: A study of individual variation in songs of canaries, greenfinches, and hybrids between the two species,'' Behaviour **65**, 241–262.
Heffner, R., and Heffner, H. (**1988**). ''Sound localization and the use of binaural cues in the gerbil,'' Behav. Neurosci. **102**, 422–428.
Janata, P., and Margoliash, D. (**1999**). ''Gradual emergence of song selectivity in sensorimotor structures of the male zebra finch song system,'' J. Neurosci. **19**, 5108–5118.
Klump, G. M., and Maier, E. H. (**1989**). ''Gap detection in the starling (*Sturnus vulgaris*), I: Psychophysical thresholds,'' J. Comp. Physiol. **164**, 531–539.
Kohlrausch, A., and Houtsma, A. J. M. (**1992**). ''Pitch related to spectral edges of broadband signals,'' Philos. Trans. R. Soc. London, Ser. B **336**, 375–381.
Konishi, M. (**1969**). ''Time resolution by single auditory neurons in birds,'' Nature (London) **222**, 566–567.
Kroodsma, D. E., and Miller, E. H. (**1982**). *Acoustic Communication in Birds, Vol. 2* (Academic, New York).
Kroodsma, D. E., and Miller, E. H. (**1996**). *Ecology and Evolution of Acoustic Communication in Birds* (Cornell U.P., Ithaca, NY).
Kuhl, P. K. (**2000**). ''A new view of language acquisition,'' Proc. Natl. Acad. Sci. U.S.A. **97**, 11850–11857.
Lavenex, P. B. (**1999**). ''Vocal production mechanisms in the budgerigar (*Melopsittacus undulatus*): The presence and implications of amplitude modulation,'' J. Acoust. Soc. Am. **106**, 491–505.
Leek, M. R., Dent, M. L., and Dooling, R. J. (**2000**). ''Masking by harmonic complexes in budgerigars (*Melopsittacus undulatus*),'' J. Acoust. Soc. Am. **107**, 1737–1744.
Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (**1967**). ''Perception of the speech code,'' Psychol. Rev. **74**, 431–461.
Manley, G. A., Gleich, O., Oeckinghaus, H., and Leppelsack, H.-J. (**1985**). ''Activity patterns of cochlear ganglion neurones in the starling,'' J. Comp. Physiol., A **157**, 161–181.
McGuirt, J. P., Schmiedt, R. A., and Schulte, B. A. (**1995**). ''Development of cochlear potentials in the neonatal gerbil,'' Hear. Res. **84**, 52–60.
Moss, C. F., and Simmons, J. A. (**1996**). ''Perception along the axis of target range in the echolocating bat,'' in *Neuroethological Studies of Cognitive and Perceptual Processes*, edited by C. F. Moss and S. J. Shettleworth (Westview, Boulder, CO), pp. 253–279.
Patterson, R. D. (**1987**). ''A pulse ribbon model of monaural phase perception,'' J. Acoust. Soc. Am. **82**, 1560–1586.
Pumphrey, R. J. (**1961**). ''Sensory Organs: Hearing,'' in *Biology and Comparative Anatomy of Birds*, edited by A. J. Marshall (Academic, New York), pp. 69–86.
Ryan, A. (**1976**). ''Hearing sensitivity of the Mongolian gerbil,'' J. Acoust. Soc. Am. **59**, 1222–1226.
Sachs, M. B., Sinnott, J. M., and Hienz, R. D. (**1978**). ''Behavioral and physiological studies of hearing in birds,'' Fed. Proc. **37**, 2329–2335.
Schroeder, M. R. (**1970**). ''Synthesis of low-peak-factor signals and binary sequences with low autocorrelation,'' IEEE Trans. Inf. Theory **16**, 85–89.
Schwartzkopff, J. (**1968**). ''Structure and function of the ear and the auditory brain areas in birds,'' in *Hearing Mechanisms in Vertebrates*, edited by A. V. S. De Reuck and J. Knight (Little, Brown, Boston), pp. 41–59.
Sen, K., Theunissen, F. E., and Doupe, A. J. (**2001**). ''Feature analysis of natural sounds in the songbird auditory forebrain,'' J. Neurophysiol. **86**, 1445–1458.
Sinnott, J. M., and Mosteller, K. W. (**2001**). ''A comparative assessment of speech sound discrimination in the Mongolian gerbil,'' J. Acoust. Soc. Am. **110**, 1729–1732.
Tchernichovski, O., Mitra, P. P., Lints, T., and Nottebohm, F. (**2001**). ''Dynamics of the vocal imitation process: How a zebra finch learns its song,'' Science **291**, 2564–2569.
Theunissen, F. E., and Doupe, A. J. (**1998**). ''Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVc of male zebra finches,'' J. Neurosci. **18**, 3786–3802.
Vicario, D. S. (**1991**). ''Contributions of syringeal muscles to respiration and vocalization in the zebra finch,'' J. Neurobiol. **22**, 63–73.
Viemeister, N. F., and Plack, C. J. (**1993**). ''Time analysis,'' in *Human Psychophysics*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 116–154.
Yu, A. C., and Margoliash, D. (**1996**). ''Temporal hierarchical control of singing in birds,'' Science **273**, 1801–2.
Zann, R. (**1984**). ''Structural variation in the zebra finch distance call,'' Z. Tierpsychol. **66**, 328–345.

Erratum: "Geoacoustic inversion for fine-grained sediments" [J. Acoust. Soc. Am. 111, 1560–1564 (2002)]

Charles W. Holland[a)]
SACLANT Undersea Research Centre, Vitale San Bartolomeo 400, 19138 La Spezia, Italy

(Received 10 May 2002; accepted for publication 10 May 2002)

[DOI: 10.1121/1.1490131]

PACS numbers: 43.20.El, 43.30.Pc, 43.30.Ma, 43.10.Vx

Figure 7 was incorrect as printed. A corrected version appears here.

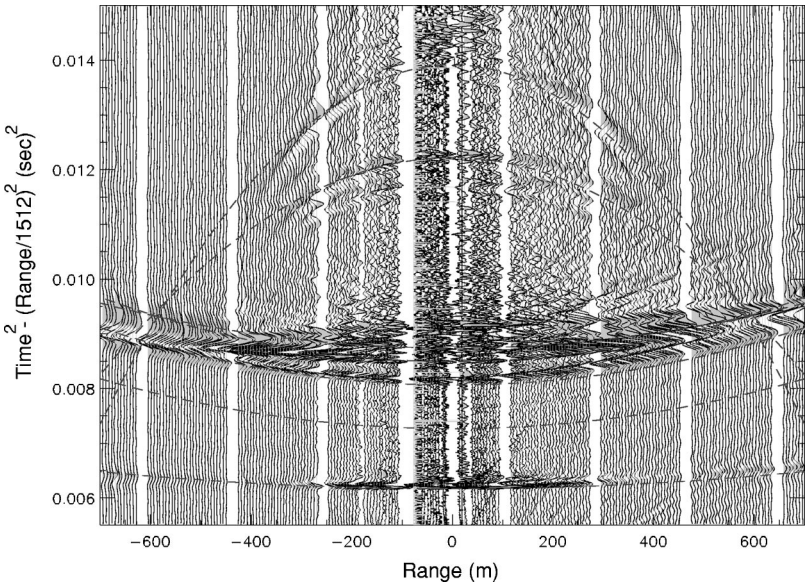


FIG. 7. Time domain data with horizon picks (dashed hyperbolae) on the Malta Plateau. The reflection from the silty-clay interface is the first arrival (at about 0.006 s^2). Note the decrease in amplitude and change of phase due to the angle of intromission at about 440-m range, corresponding to 15 degrees. The reflection data of Figs. 2 and 6 were calculated using a time window of $5e$-4 s^2 around the silty-clay interface.

[a)]Present address: Pennsylvania State University, Applied Research Lab, P.O. Box 30, State College, PA 16804.

Erratum: "Kramers-Kronig relations applied to finite bandwidth data from suspensions of encapsulated microbubbles" [J. Acoust. Soc. Am. 108, 2091–2106 (2000)]

Joel Mobley and Kendall R. Waters
Laboratory for Ultrasonics, Department of Physics, Washington University, St. Louis, Missouri 63130-4899

Michael S. Hughes
Mallinckrodt, Inc., 675 McDonnell Boulevard, Hazelwood, Missouri 63042

Christopher S. Hall and Jon N. Marsh
Laboratory for Ultrasonics, Department of Physics, Washington University, St. Louis, Missouri 63130-4899

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 5 May 2002; accepted for publication 18 May 2002)

[DOI: 10.1121/1.1494808]

PACS numbers: 43.35.Bf, 43.10.Vx

Equations (3), (13), (14), (17), (A6), and (A7) in Ref. 1 are corrected below. The results given in the paper[1] were calculated using the correct forms given below and are thus unaffected by these errors.

$$\alpha(\omega)=\lim_{|\Omega|\to\infty}\left(\left[-\frac{1}{\pi}\int_0^{\Omega}\frac{\frac{\omega'}{c(\omega')}-\frac{\omega}{c(\omega)}}{\omega'-\omega}d\omega'-\frac{1}{\pi}\int_0^{\Omega}\frac{\frac{\omega'}{c(\omega')}+\frac{\omega}{c(\omega)}}{\omega'+\omega}d\omega'\right]\right.$$

$$+\left[\alpha(\omega_0)+\frac{1}{\pi}\int_0^{\Omega}\frac{\frac{\omega'}{c(\omega')}-\frac{\omega_0}{c(\omega_0)}}{\omega'-\omega_0}d\omega'+\frac{1}{\pi}\int_0^{\Omega}\frac{\frac{\omega'}{c(\omega')}+\frac{\omega_0}{c(\omega_0)}}{\omega'+\omega_0}d\omega'\right]$$

$$\left.+(\omega-\omega_0)\left[\frac{d}{d\omega}\alpha(\omega)\Big|_{\omega=\omega_0}+\frac{1}{\pi}\int_0^{\Omega}\mathrm{Im}\,\Lambda_2(\omega',\omega_0)d\omega'+\frac{1}{\pi}\int_0^{\Omega}\mathrm{Im}\,\Lambda_2(-\omega',\omega_0)d\omega'\right]\right), \tag{3}$$

$$\alpha(\omega)=\left(-\frac{1}{\pi}\int_{\omega_a}^{\omega_b}\frac{\frac{\omega'}{c(\omega')}-\frac{\omega}{c(\omega)}}{\omega'-\omega}d\omega'-\frac{1}{\pi}\int_{\omega_a}^{\omega_b}\frac{\frac{\omega'}{c(\omega')}+\frac{\omega}{c(\omega)}}{\omega'+\omega}d\omega'\right)$$

$$+\left(\alpha(\omega_0)+\frac{1}{\pi}\int_{\omega_a}^{\omega_b}\frac{\frac{\omega'}{c(\omega')}-\frac{\omega_0}{c(\omega_0)}}{\omega'-\omega_0}d\omega'+\frac{1}{\pi}\int_{\omega_a}^{\omega_b}\frac{\frac{\omega'}{c(\omega')}+\frac{\omega_0}{c(\omega_0)}}{\omega'+\omega_0}d\omega'\right)$$

$$+(\omega-\omega_0)\left(\frac{d}{d\omega}\alpha(\omega)\Big|_{\omega=\omega_0}+\frac{1}{\pi}\int_{\omega_a}^{\omega_b}\mathrm{Im}\,\Lambda_2(\omega',\omega_0)d\omega'+\frac{1}{\pi}\int_{\omega_a}^{\omega_b}\mathrm{Im}\,\Lambda_2(-\omega',\omega_0)d\omega'\right), \tag{13}$$

$$\alpha(\omega)=\left(-\frac{1}{\pi}\int_{\omega_a}^{\omega_b}\frac{\frac{\omega'}{c(\omega')}-\frac{\omega}{c(\omega)}}{\omega'-\omega}d\omega'-\frac{1}{\pi}\int_{\omega_a}^{\omega_b}\frac{\frac{\omega'}{c(\omega')}+\frac{\omega}{c(\omega)}}{\omega'+\omega}d\omega'\right)$$

$$+\left(\alpha(\omega_0)+\frac{1}{\pi}\int_{\omega_a}^{\omega_b}\frac{\frac{\omega'}{c(\omega')}-\frac{\omega_0}{c(\omega_0)}}{\omega'-\omega_0}d\omega'+\frac{1}{\pi}\int_{\omega_a}^{\omega_b}\frac{\frac{\omega'}{c(\omega')}+\frac{\omega_0}{c(\omega_0)}}{\omega'+\omega_0}d\omega'\right), \tag{14}$$

$$\frac{d}{d\omega}\alpha(\omega)\Big|_{\omega=\omega_0}+\frac{1}{\pi}\int_{\omega_a}^{\omega_b}\mathrm{Im}\,\Lambda_2(\omega',\omega_0)d\omega'+\frac{1}{\pi}\int_{\omega_a}^{\omega_b}\mathrm{Im}\,\Lambda_2(-\omega',\omega_0)d\omega', \tag{17}$$

$$\alpha(\omega)=\alpha(\omega_0)+(\omega-\omega_0)\alpha'(\omega_0)-\frac{(\omega-\omega_0)^2}{\pi}\int_0^{\infty}\frac{\mathrm{Im}\,\Lambda_2(\omega',\omega_0)-\mathrm{Im}\,\Lambda_2(\omega,\omega_0)}{\omega'-\omega}d\omega'$$

$$-\frac{(\omega-\omega_0)^2}{\pi}\int_0^{\infty}\left(\frac{\frac{\omega'}{c(\omega')}+\frac{\omega_0}{c(\omega_0)}-(\omega'+\omega_0)\frac{d}{d\omega}\frac{\omega}{c(\omega)}\Big|_{\omega=\omega_0}}{(\omega'+\omega_0)^2}+\mathrm{Im}\,\Lambda_2(\omega,\omega_0)\right)\frac{d\omega'}{\omega'+\omega}, \tag{A6}$$

and

$$\frac{\omega}{c(\omega)}=\frac{\omega_0}{c(\omega_0)}+(\omega-\omega_0)\frac{d}{d\omega}\frac{\omega}{c(\omega)}\Big|_{\omega=\omega_0}-\frac{(\omega-\omega_0)^2}{\pi}\int_0^{\infty}\frac{\mathrm{Re}\,\Lambda_2(\omega',\omega_0)-\mathrm{Re}\,\Lambda_2(\omega,\omega_0)}{\omega'-\omega}d\omega'$$

$$-\frac{(\omega-\omega_0)^2}{\pi}\int_0^{\infty}\left(\frac{\alpha(\omega')-\alpha(\omega_0)+(\omega'+\omega_0)\alpha'(\omega_0)}{(\omega'+\omega_0)^2}+\mathrm{Re}\,\Lambda_2(\omega,\omega_0)\right)\frac{d\omega'}{\omega'+\omega}. \tag{A7}$$

[1] J. Mobley, K. R. Waters, M. S. Hughes, C. S. Hall, J. N. Marsh, G. H. Brandenburger, and J. G. Miller, "Kramers-Kronig relations applied to finite bandwidth data from suspensions of encapsulated microbubbles," J. Acoust. Soc. Am. **108**, 2091–2106 (2000).