CODEN: JASMAN

ISSN: 0001-4966

The Journal of the Acoustical Society of America

Vol. 114, No. 2 August 2003

LETTERS TO THE EDITOR

GENERAL LINEAR ACOUSTICS [20]

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 114, NO. 2, AUGUST 2003

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 114, NO. 2, AUGUST 2003

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 114, NO. 2, AUGUST 2003

CONTENTS—*Continued from preceding page*

BIOACOUSTICS [80]

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.*

Preliminary Notice: 146th Meeting of the Acoustical Society of America

The 146th Meeting of the Acoustical Society of America will be held Monday through Friday, 10–14 November 2003, at the Renaissance Austin Hotel, Austin, Texas, USA. A block of rooms has been reserved at the Renaissance Hotel.

For information about the technical program refer to the Meetings Section on the ASA Home Page at ⟨http://asa.aip.org⟩.

Technical Program

The technical program will consist of lecture and poster sessions. Technical sessions will be scheduled Monday through Friday, 10–14 November. The special sessions described below will be organized by the ASA Technical Committees.

Special Sessions

Acoustical Oceanography (AO)

Aubrey L. Anderson memorial session on acoustics of gas-bearing sediments (Joint with Underwater Acoustics)
Honoring Aubrey Anderson's contributions to the understanding of bubbly and hydrated sediment acoustics

Developments in multibeam sonar for water-column measurements
Multibeam sonar, which is widely used for bathymetric mapping and imaging, is seeing new, quantitative applications in water-column measurements. This session will add the exchange of information on new developments including applications to fish schools and air bubbles, among other scatterers and calibration

Animal Bioacoustics (AB)

Medical imaging techniques to understand auditory processing
(Joint with Signal Processing in Acoustics)
Imaging of auditory reception, sound production and related CNS structures using conventional and developmental imaging techniques, including CT, MRI, ultrasound, fiber optics, SPECT and fMRI to understand sound transduction mechanisms

Neurobiology of communications
Behavioral and neural mechanisms underlying communication by animals

Architectural Acoustics (AA)

Acoustic design of government buildings (Joint with Noise)
Architectural acoustics and building noise issues that frequently surface in various types of federal, state and local government buildings

Designing for speech—Both intelligibility and privacy
(Joint with Speech Communication)
If more than one person occupies a space, then intelligibility is necessary, and if more than two, then privacy is necessary

Electro-acoustic enhancement system in rooms for music
(Joint with Musical Acoustics)
New techniques, user requirements and examples of such systems

Forensic acoustics (Joint with Noise and Speech Communication)
Special analysis techniques and case histories in forensic acoustics

Small rooms—Big challenges
Making measurements, metrics and models work in small rooms

Sound quality—When sound is the essential quality
(Joint with Noise, Engineering Acoustics and Musical Acoustics)
Methods of evaluating the sound quality of this type of product

Honoring the contributions of Elmer Hixson
(Joint with Noise and Engineering Acoustics)
A review of the professional contributions of Elmer Hixson represented by technical papers from past students and professional colleagues

Biomedical Ultrasound/Bioresponse to Vibration (BB)

Multidisciplinary and nonconventional approaches in biomedical ultrasound
Forum to discuss novel, forgotten and nontraditional methods and techniques in biomedical and clinical applications of ultrasound

Tissue harmonic imaging
(Joint with Physical Acoustics)
Use of nonlinearly generated harmonic signals which result in higher image quality in diagnostic ultrasound

Ultrasound contrast agents (Joint with Physical Acoustics)
Use of micron sized bubbles injected into the blood stream to enhance ultrasound imaging

Topical meeting on shock waves in medicine
(Joint with Physical Acoustics)
A one-day colloquium and discussion on the topic ''Shock Waves in Medicine'' will be held. The session will focus on nonthermal biological effects of short-duration shock pulses. Four subtopic sessions will focus on an overview of shock waves in medicine, lithotripsy, orthopedic applications, and cell permeability. Each subtopic session will contain invited presentations followed by a moderated panel discussion.

Education in Acoustics (ED)

''Neat'' acoustics web sites
Presentation of web sites with descriptions and comments on relevance to acoustics

Take ''5's''
Bring your favorite teaching tips, demos, or teaching aids (5 min. max; no abstracts)

Engineering Acoustics (EA)

Hearing aid sound quality: Theory, myths, experience
(Joint with ASA Committee on Standards)
Discussion of goals and methods to achieve quantitative and perceived sound quality improvement in hearing aids

Musical Acoustics (MU)

Piano and bowed string instrument acoustics: A session honoring Gabriel Weinreich
Reports of current research in the acoustics of the piano and bowed strings, in recognition of the contributions of Gabriel Weinreich to these areas

Software for teaching musical acoustics (Joint with Education in Acoustics)
Descriptions and demonstrations of software and web sites useful in teaching musical acoustics

Where are they now? Current research by past student paper award winners
Past winners of the student paper award in musical acoustics report on their current or recent research activities

 0001-4966/2003/114(2)/551/5/$19.00

Wind instruments and nonlinearity (Joint with Physical Acoustics)
Acoustics of musical instruments, especially wind instruments, with emphasis on the role of nonlinearity

The University of Texas Wind Ensemble
Performance session

Noise (NS)

Acoustical design issues for health care industry
(Joint with Architectural Acoustics)
Acoustics and noise control for hospitals, nursing homes, clinics, emergency rooms, medical test equipment and medical manufacturing

Everything old: Adaptive reuse, acoustic *au jus*
(Joint with Architectural Acoustics)
Acoustic concerns inherent in adapting stale facilities to serve new functions while retaining their historic flavor

Building code noise compliance
(Joint with Architectural Acoustics and ASA Committee on Standards)
Building noise codes, means of measurements of noise regulation, measures to effect compliance

New technologies in active control
(Joint with Engineering Acoustics and Signal Processing in Acoustics)
Review the achievements in active control and present new approaches and achievements

Sound quality: For products over which one has no control
(Joint with Engineering Acoustics)
Sound quality of traffic noise, construction and my air-conditioner versus your air-conditioner

Physical Acoustics (PA)

Nonlinear acoustics in honor of David Blackstock
(Joint with Biomedical Ultrasound/Bioresponse to Vibration)
Topics in nonlinear acoustics associated with the work of David Blackstock

Signal Processing in Acoustics (SP)

Signal processing for aircraft noise (Joint with Noise)
Particular signal processing problems and techniques for evaluating noise produced by aircraft and their operations

Detection and classification in acoustics (Joint with Underwater Acoustics)
Recent advances in signal processing for detection of signals in various environments and frequency bands will be reported

Speech Communication (SC)

Issues in similarity and distinctiveness
Perceptual processing of similarity and distinctiveness is key for perception, cognition and spoken language processing

Speech intelligibility and AAC devices
(Joint with ASA Committee on Standards)
Factors that affect speech intelligibility with an ear towards improving the speech of AAC devices used by communicatively disabled persons

Statistical patterns in speech
How are models of speech perception constrained by distributional characteristics of the input, and can these characteristics explain evolution of language processes of perception?

Structural Acoustics and Vibration (SA)

Control of sound transmission through partitions
(Joint with Architectural Acoustics and Noise)
Recent developments in both passive and active control of sound transmission through walls, fuselages and other structures

Measurements of particle velocities
(Joint with Engineering Acoustics and Physical Acoustics)
Forum for sharing knowledge and state-of-the-art technologies to measure fluid particle velocities in an insonified fluid medium

Underwater Acoustics (UW)

Gradient array acoustics
(Joint with Engineering Acoustics and Signal Processing Acoustics)
Physics and signal processing issues of gradient array acoustics (e.g., superdirective processing, intensity arrays) from a range of technical areas and applications

The acoustics of bubbles in the marine boundary layer
(Joint with Acoustical Oceanography and Physical Acoustics)
Scattering and absorption from wave-induced bubbles in the marine boundary layer, and the generation of ambient noise by wave-induced bubbles

Other Technical Events

Distinguished Lecture

The Technical Committees on Architectural Acoustics and Noise will sponsor a distinguished lecture titled ''Dosage-effect analysis of community response to transportation noise a quarter century later'' presented by Sanford Fidell of BBN Technologies.

Hot Topics Session

A ''Hot Topics'' session sponsored by the Tutorials Committee is scheduled covering the fields of Acoustical Oceanography, Musical Acoustics and Noise.

Meeting Program

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

Exhibit

The instrument and equipment exposition, which will be conveniently located near the registration desk and meeting rooms, will open at the Renaissance Austin Hotel with a reception on Monday evening, 10 November, and will close on Wednesday afternoon, 12 November. The exposition will include computer-based instrumentation, 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. A special session on new and innovative techniques by exhibitors is also being planned. For further information, please contact the Exhibit Manager, Robert Finnegan, Advertising and Exhibits Div., AIP, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, Tel: 516-576-2433; Fax: 516-576-2481; E-mail: rfinneg@aip.org.

Paper Copying Service

Authors are requested to provide one paper copy of their projection material and/or paper(s) to the Paper Copies Desk upon arrival. The copy should contain material on one side only on $8\frac{1}{2}\times 11$ inch or A4 paper suitable for photocopy reproduction. Copies of available papers will be made for a nominal charge.

Tutorial Lecture

A tutorial presentation on Classroom Acoustics will be given by Peggy Nelson on Monday, 10 November, at 7:00 p.m. Lecture notes will be available at the meeting in limited supply. Those who register by 20 October are guaranteed receipt of a set of notes.

To partially defray the cost of the lecture a registration fee is charged. The fee is $15 for registration received by 30 September and $25 thereafter including on-site registration at the meeting. The fee for students with current ID cards is $7.00 for registration received by 30 September and $12.00 thereafter, including on-site registration at the meeting.

Short Course on Time Reversal

A short course on Time Reversal will be held on Sunday and Monday, 9 and 10 November, at the Renaissance Austin Hotel.

An acoustic Time Reversal Mirror (TRM) refocuses an incident acoustic field to the position of the original source regardless of the complexity of the medium between this ''probe'' source and the TRM. TRMs have now been implemented in a variety of physical scenarios from mega Hertz ultrasonics with order centimeter aperture size to hundreds/thousands of Hz in ocean acoustics with order hundred-meter aperture size. Common to this broad range of scales is a remarkable robustness as evidenced by observations at all scales that the more complex the medium between the probe source and the TRM, the sharper the focus. The potential for applications in many areas of acoustics is quite high.

The objective of this course is to provide the acoustical physics overview and description of the experimental implementation of time reversal and phase conjugate processes as related to ultrasonics and imaging, nondestructive testing, medical ultrasonics, propagation in random media, room acoustics, waveguides, and ocean acoustics.

The instructors, who have worked extensively in the area of time reversal acoustics, are Mathias Fink of the University of Paris and William Kuperman of the University of California, San Diego.

The registration fee is $250.00 and covers attendance, instructional materials and coffee breaks. The number of attendees will be limited so please register early to avoid disappointment. Only those who have registered by 20 October will be guaranteed receipt of instructional materials. There will be a $50 discount for registration made prior to 30 September. Full refunds will be made for cancellations prior to 30 September. Any cancellation after 30 September will be charged a $50 processing fee.

Forensic Acoustics Seminar How to Succeed as an Expert Witness

This seminar will be held all day on Friday, 10 November, at the Renaissance Austin Hotel.

Courtroom presentations throughout the nation are involving more and more expert testimony in acoustics. The legal system generally seeks to find qualified experts from the ranks of consultants, engineers, professors and physicians. The Acoustical Society of America and the National Council of Acoustical Consultants proudly co-sponsor this seminar which will present information about becoming an expert witness or improving your services as an expert witness. The seminar is offered as a follow-on to the special session on Forensic Acoustics held on Wednesday afternoon.

This seminar on succeeding as an expert witness on Forensic Acoustics is intended for practitioners in virtually all fields of acoustics. The legal system needs experts to provide clear, concise, correct information on subjects ranging from community noise impact to employee noise exposure, from assessment of hearing damage risk to the mechanisms leading to loss of hearing, from product sound levels to the sound levels of warning devices, from forensic phonetics to speaker recognition and identification, from transportation noise generation to the production of industrial noise levels and more. It is the objective of this seminar to acquaint the participants in the proper method of presentation, both in the courtroom and before prospective clients.

A special group of trial attorneys and specialists are assembled to share their knowledge in support of expert witness activity. These presenters will offer insight into how expert witnesses should approach their practice and what to expect from the attorneys who employ them. Presenters will be: Michael M. Barron, Esq., ''Choosing an Expert Witness from a Lawyer's Point of View;'' Jeff Civins, Esq., ''Investigations, Analysis and Presentations by Expert Witnesses;'' Kyle Rowley, Esq., ''Federal Rules of Evidence—Surviving a Daubert Hearing;'' Rosalie Hamilton, ''Marketing Your Expert Witness Services;'' and David M. Benjamin, Ph.D., ''Developing Active Listening Skills to Enhance Courtroom Communications.''

The book *Succeeding as an Expert Witness* by Harold A. Feder, Esq., forms a basis for the seminar and each seminar participant will receive a copy. Each participant will also receive an instructional video with tips, advice and examples for expert testimony given under deposition. Some of the topics covered in the seminar are: Usefulness of the expert witness; How an attorney chooses an expert witness; Participating in the trial strategy; Proper methods in testifying; What to expect under cross examination; Delivering testimony under deposition; How an expert is qualified by the court; Marketing your services as an expert.

The registration fee is $325.00 and covers attendance, textbook, instructional materials, coffee breaks and lunch. The number of attendees will be limited to 60 so please register early. Only those who have registered by 20 October will be guaranteed receipt of instructional materials. There will be a $50 discount for registration made prior to 30 September. Full refunds will be made for cancellations prior to 30 September. Any cancellation after 30 September will be charged a $50 processing fee.

Special Meeting Features

Student Transportation Subsidies

A student transportation subsidies fund has been established to provide limited funds to students to partially defray transportation expenses to meetings. Students presenting papers who propose to travel in groups using economical ground transportation will be given first priority to receive subsidies, although these conditions are not mandatory. No reimbursement is intended for the cost of food or housing. The amount granted each student depends on the number of requests received. To apply for a subsidy, submit a proposal (e-mail preferred) to be received by 30 September to: Jolene Ehl, ASA, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, Tel: 516-576-2359, Fax: 516-576-2377, E-mail: jehl@aip.org. The proposal should include your status as a student; whether you have submitted an abstract; whether you are a member of ASA; method of travel; if traveling by auto; whether you will travel alone or with other students; names of those traveling with you; and approximate cost of transportation.

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 Austin meeting. This award is designed for young professionals who have completed the doctorate in the past five years (not currently enrolled as a student), who plan to present a paper at the Austin meeting. Each award will be of the order of $200. 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: Lily M. Wang, University of Nebraska, 200B Peter Kiewit Institute, 1110 South 67th Street, Omaha, NE 68182-0681; Tel.: 402-554-2065; E-mail: lwang4@unl.edu. Deadline for receipt of applications is 30 September.

Students Meet Members for Lunch

The Education Committee has established a program for students to meet one on one with members of the Acoustical Society over lunch. The purpose is to make it easier for students to meet and interact with members at meetings. Each lunch pairing is arranged separately. Students who wish to participate may sign up by contacting David Blackstock, University of Texas at Austin, Mechanical Engineering Dept., 1 University Station C2200, Austin TX 78712-0292; Tel.: 512-471-3145; Fax: 512-471-1045; E-mail: dtb@mail.utexas.edu. The information needed in order to arrange each pairing is the student's name, university, department, status (graduate student or undergraduate), research field, interests in acoustics, and days that are free for lunch. The sign-up deadline is one week before the start of the meeting, but an earlier sign-up is strongly encouraged. The cost of the meal is the responsibility of each participant.

Plenary Session, Awards Ceremony, Fellows Suite, Fellows' Lunch and Social Events

Buffet socials with cash bar will be held on Tuesday and Thursday evenings at the Renaissance hotel.

The Plenary session will be held on Wednesday afternoon, 12 November, at the Renaissance hotel where Society awards will be presented and recognition of newly-elected Fellows will be announced.

A Fellows' Hospitality Suite will be open on Tuesday afternoon. Refreshments will be provided. A Fellows' Luncheon will be held on Thursday, 13 November, at 12:00 noon. Professor Steven Weinberg of the University of Texas at Austin Physics Department will be the speaker. Professor Weinberg was awarded the Nobel Prize for Physics in 1979, and is also well known for having written several books for general readers, including the

prize-winning *The First Three Minutes* (now translated into 22 languages), *The Discovery of Subatomic Particles*, and most recently, *Dreams of a Final Theory*. Each Fellow may bring one guest to the luncheon. Fellows should register online or download a registration form to purchase tickets for themselves and their guests.

Women in Acoustics Luncheon

The Women in Acoustics luncheon will be held on Wednesday, 12 November. Those who wish to attend this luncheon should Register online or download a registration form to purchase tickets for this luncheon. The fee is $15 (students $5) for preregistration by 30 September and $20 (students $5) thereafter including on-site registration at the meeting.

Transportation and Hotel Accommodations

Austin is served by Austin-Bergstrom International Airport (Identifier Code: AUS). A number of airlines fly in and out of Austin, offering nonstop service to 33 destinations as well as numerous connecting flights to other airport hubs. The passenger airlines serving Austin are America West, American, Continental, Delta, Frontier, Great Plains, Mexicana, Northwest, Southwest, and United. For flight information, local weather, and other information of interest, visit www.ci.austin.tx.us/austinairport.

Ground Transportation

Transportation from Austin-Bergstrom International Airport to the Renaissance Austin Hotel:

Major car rental companies. Rental car counters are located adjacent to the airport's baggage claim area.

SuperShuttle shared-ride, door-to-door service. The SuperShuttle ticket counter is located in the baggage claim area near the car rental counters. No reservation is necessary to ride from the airport to your accommodations. Vans are assigned on a first-come, first-served basis, and service is available 24 hours a day. The fare to the Renaissance is $15 per person, or $26 round-trip. Round-trip tickets must be purchased at the airport ticket counter. Please make your reservation for the return trip 24 hours in advance by checking with the concierge or front desk of your hotel, or by calling (512) 258-3826.

Taxicabs and limousines. These services may also be procured just outside the baggage claim area at the loading zone curb. Taxi fares are metered and should range from $35 to $45, depending on traffic. Limo fares should be verified with the operators before boarding.

Hotel Accommodations

The meeting and all functions will be held at the Renaissance Austin Hotel, located in the Arboretum on the northwest side of Austin. Please make your reservations directly with the hotel and ask for one of the rooms being held for the Acoustical Society of America (ASA). The reservation cut-off date for the special discounted ASA rates is 20 October 2003; after this date, the conference rate will no longer be available. See the section on Hotel Reservation information for information about the hotel and special ASA rates.

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, smoker or nonsmoker preference, by 10 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

Austin's climate is temperate, having 300 days of sunshine annually. In early November, average temperatures will range from 55 °F to 75 °F. The relative humidity will be about 65%, with rainfall possible—the city receives an average annual rainfall of 32.49 inches.

Hotel Reservation Information

A block of guest rooms at discounted rates has been reserved for meeting participants at the Renaissance Austin Hotel. Early reservations are strongly recommended. Note that the special ASA meeting rates are not guaranteed after 20 October 2003. You must mention the Acoustical Society of America when making your reservations to obtain the special ASA meeting rates.

The Renaissance Austin Hotel is situated in the scenic hills of northwest Austin. It is the cornerstone of the Arboretum, a 95-acre park-like development containing numerous stores, specialty shops, restaurants, movie theaters, and walking and jogging paths. The hotel offers both indoor and outdoor swimming pools, as well as a complete health club featuring the latest equipment, whirlpool, and sauna. Guest services include complimentary in-room coffee and newspaper, complimentary overnight shoe shine, 24-hour room service, laundry and valet services, fax service, and a gift shop.

For attendees who will be driving, complimentary self-parking is available; valet parking is also available for a daily fee. To drive to the Renaissance Austin Hotel from Austin Bergstrom International Airport, take Highway 71 West to U.S. Highway 183 North exit. Take 183 North to Capital of Texas Highway (Loop 360) exit (~17 miles). Turn left on Great Hills Trail, which is the second traffic light. Turn left on Jollyville Road (first traffic light beyond freeway).

Please make your reservation directly with the Renaissance Austin Hotel. When making your reservation, you must mention the Acoustical Society of America to obtain the special ASA meeting rates.

Renaissance Austin Hotel
9721 Arboretum Blvd.
Austin, TX 78759
Tel: 512-343-2626
Fax: 512-346-7945

Rates
Single: $143.00
Double: $153.00
Govt. Rate Single/Double (limited quantity): $85.00
Additional person: $130.00
Reservation cut-off date: 20 October 2003

General Information

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

Spouses and other visitors are welcome at the Austin meeting. The registration fee for accompanying persons is $35/$45. A hospitality room for accompanying persons will be open at the Renaissance Hotel from 8:00 a.m. to 11:00 a.m. each morning throughout the meeting where information about activities in and around Austin will be provided.

Entering Austin City Limits. Known for its casual and high-spirited nature, Austin is situated in the center of the Lone Star State. It stands as the gateway to the Texas Hill Country and the Highland Lakes. As the state capital and home to The University of Texas, the city supports a politically charged, culturally rich environment that's hip, trendy, and high tech.

Natural Appeal. Blessed with a temperate year-round climate and 300 days of sunshine a year, Austinites love their nature trails, parks, and wilderness preserves. In the very heart of downtown Austin, Town Lake offers an oasis bordered by 10 miles of hike-and-bike trails.

Historical View. Originally a buffalo hunting ground favored by Tonkawa Indians, Austin was permanently settled in 1838 as a trading post. Now, more than a million people live in the city named for Stephen F. Austin, who colonized Texas. In central Austin, larger-than-life history comes together at the Bob Bullock Texas State History Museum, which features intriguing artifacts, interactive exhibits, multimedia shows, and an

IMAX theatre. Also in downtown Austin, the Texas State Capitol is an imposing granite structure that's actually 14 feet taller than the nation's capitol; it's the largest domed statehouse in the country.

Musical Talent. In the city where the music never stops, live music plays at more than 100 venues on any given evening: a little blues, a little country, the beat of rock and roll, and even a few jazz licks. The largest concentration of music venues is found downtown in the Warehouse District and along Sixth Street, a six-block stretch of bars and restaurants. And Austin City Limits continues to be the best show in town! The long-running PBS series showcases American music and Texas music in particular. For ticket information, call (512) 471-4811 well in advance of the meeting.

Cultural Scene. Austin is one of only a few U.S. cities with professional ballet, symphony, opera, and theater companies. The city also offers 20 museums and dozens of galleries. Many of Austin's cultural resources are found on the campus of The University of Texas, such as the Blanton Museum of Art, the Harry Ransom Center, and the Lyndon B. Johnson Library and Museum, which remains the most visited of the nation's presidential libraries.

Registration Information

The registration desk at the meeting will open on Monday, 10 November, at the Renaissance Hotel. To register use the form in the printed call for papers or register online at ⟨http://asa.aip.org⟩.

If your registration is not received at the ASA headquarters by 20 October you must register on-site.

Registration fees are as follows:

Category	Preregistration by 30 September	Registration after 30 September
Acoustical Society Members	$270	$320
Acoustical Society Members One-Day	$135	$160
Nonmembers	$320	$370
Nonmembers One-Day	$160	$185
Nonmember Invited Speakers	Fee waived	Fee waived
Students (with current ID cards)	Fee waived	Fee waived
Emeritus members of ASA (Emeritus status preapproved by ASA)	$35	$45
Accompanying Persons (Spouses and other registrants who will not participate in the technical sessions)	$35	$45

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 (2004) 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 may register without charge.

NOTE: A $25 PROCESSING FEE WILL BE CHARGED TO THOSE WHO WISH TO CANCEL THEIR REGISTRATION AFTER 30 SEPTEMBER.

Members of the Local Committee for the Meeting

General Chair—Clark S. Penrod; Technical Program Chair—Evan K. Westwood; Hotel/Facilities—Lorri Polvado; Food Service/Social Events—B. J. Gatlin; Audio-Visual—Leonard Hebert; Registration—Jan Chambers; Accompanying Persons Program—Carlie Tilly; Signs—Leonard Hebert and Nanette Lemma; Treasurer—Joy Whitney; Deputy to the General Chair—James E. Stockton

USA Meetings Calendar

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

2003

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

2004

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

ACOUSTICAL NEWS—INTERNATIONAL

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

International Meetings Calendar

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

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.internoise2003)

25–29 **XIII Session of the Russian Acoustical Society**, Moscow, Russia. (Fax: +7 095 126 0100; Web: www.akin.ru)

September 2003

1–2 **First Congress of the Alps-Adria Acoustics Association (AAAA)**, Portoro, Slovenia. (Fax: +386 1 251 8567; Web: www.fs.uni-lj.si/sda)

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)

16–19 **Autumn Meeting of the Acoustical Society of Japan**, Nagoya, Japan. (Fax: +81 3 5256 1022; Web: wwwsoc.nii.ac.jp/asj/index-e.html)

18–19 **Surface Acoustics 2003**, Salford University, Manchester, UK. (Web: www.ioa.org.uk/salford2003)

23–25 **2nd International Symposium on Fan Noise**, Senlis, France. (Fax: +33 4 72 44 49 99; Web: www.fannoise2003.org)

October 2003

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

15–17 **Acoustics Week in Canada**, Edmonton, AB, Canada. (Fax: +1 780 414 6376; Web: caa-aca.ca/edmonton2003.html)

30–31 **Autumn Meeting of the Swiss Acoustical Society**, Basel, Switzerland. (Fax: +41 419 62 13; Web: www.sga-ssa.ch)

November 2003

5–6 **Institute of Acoustics (UK) Autumn Conference**, Oxford, UK. (Fax: +44 1727 850553; Web: ww.ioa.org.uk)

7–9 **Reproduced Sound**, Oxford, UK. (Fax: +44 1727 850553; Web: www.ioa.org.uk)

12–14 ***Tenth Asia-Pacific Vibration Conference (APVC 2003)**, Gold Coast, Queensland, Australia. (Web: www.apvc.net)

26–28 ***Tenth Mexican International Congress on Acoustics**, Puebla, Pua, Mexico. (Instituto Mexicano de Acustica, P.O. Box 12-1022, Col. Narvarte 03001 Mexico, D.F., Mexico; Fax: +52 555523 4742; e-mail: sberista@hotmail.com)

December 2003

10–12 **3rd International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications**, Firenze, Italy. (Fax: +39 55 479 6767; Web: www.maveba.org)

March 2004

17–19 **Spring Meeting of the Acoustical Society of Japan**, Atsugi, Japan. (Fax: +81 3 5256 1022; Web: wwwsoc.nii.ac.jp/asj/index-e.html)

22–25 **Joint Congress of the French and German Acoustical Societies (SFA-DEGA)**, Strasbourg, France. (Fax: +33 1 48 88 90 60; Web: www.sfa.asso.fr/cfadaga2004)

31–3 **International Symposium on Musical Acoustics (ISMA2004)**, Nara, Japan. (Fax: +81 77 495 2647; Web: www2.crl.go.jp/jt/a132/isma2004)

April 2004

5–9 **18th International Congress on Acoustics (ICA2004)**, Kyoto, Japan. (Fax: +81 66 879 8025; Web: www.ica2004.or.jp)

11–13 **International Symposium on Room Acoustics (ICA2004 Satellite Meeting)**, Hyogo, Japan. (Fax: +81 78 803 6043; Web: rad04.iis.u-tokyo.ac.jp)

May 2004

8–10 ***116th AES Convention**, Berlin, Germany. (Web: aes.org/events/116)

17–21 **International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004)**, Montréal, Canada. (Web: www.icassp2004.com)

June 2004

8–10 **Joint Baltic-Nordic Acoustical Meeting**, Mariehamn, Åland, Finland. (Fax: +358 09 460 224; e-mail: asf@acoustics.hut.fi)

July 2004

5–8 **7th European Conference on Underwater Acoustics (ECUA 2004)**, Delft, The Netherlands. (Fax: +31 70 322 9901; Web: www.ecua2004.tno.nl)

11–16 **12th International Symposium on Acoustic Remote Sensing (ISARS)**, Cambridge, UK. (Fax: +44 161 295 3815; Web: www.isars.org.uk)

August 2004

23–27 **2004 IEEE International Ultrasonics, Ferroelectrics, and Frequency Control 50th Anniversary Conference**, Montréal, Canada. (Fax: +1 978 927 4099; Web: www.ieee-uffc.org/index2-asp)

22–25 **Inter-Noise 2004**, Prague, Czech Republic. (Web: www.internoise2004.cz)

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; Web: www.spacustica.pt/novidades.htm)

November 2004

4–5 **Autumn Meeting of the Swiss Acoustical Society**, Rapperswil, Switzerland. (Fax: +41 419 62 13; Web: www.sga-ssa.ch)

August 2005

7–10 **Inter-Noise**, Rio de Janeiro, Brazil. (Details to be announced later)

 0001-4966/2003/114(2)/557/2/$19.00

28–2 *__Forum Acusticum Budapest 2005__, Budapest, Hungary. (Web: www.fa2005.org)

Preliminary Announcement

June 2008

23–27 **Joint Meeting of European Acoustical Association (EAA), Acoustical Society of America (ASA), and Acoustical Society of France (SFA)**, Paris, France. (Details to be announced later)

Institute of Acoustics awards Rayleigh Medal to Hugo Fastl

The Institute of Acoustics (U.K.) has awarded this year's Rayleigh Medal to Hugo Fastl of the Department of Human Machine Communication of the Technical University Munich. He received the Medal for ''outstanding contributions to acoustics.''

Professor Fastl, who started his earlier work with Professor Zwicker, has chosen to give the traditional Rayleigh Lecture on the theme ''From psychoacoustics to sound quality engineering.''

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).]*

Acoustic Communication in Insects and Anurans: Common Problems and Diverse Solutions

H. Carl Gerhardt and Franz Huber

University of Chicago Press, Chicago, 2002.
544 pp. Price: $100 (hardcover) ($35 paperback) ISBN: 0226288323 (hardcover) (0226288331 paperback).

In this book, two leaders in the area of animal bioacoustics summarize and synthesize the literature on acoustic communication in chorusing insects (crickets, katydids, grasshoppers, and cicadas) and anuran amphibians (frogs and toads). The conceptual framework of the book (cogently summarized in Table 1.1) is based on the idea that these groups of animals all communicate acoustically in mixed-species choruses, and face similar obstacles (small size, limited breeding periods, ectothermy) in using acoustic cues to guide their behavior. Direct comparisons of how these animals deal with ''common problems'' of signal recognition, signal/noise discrimination, and localization thus can lead to future breakthroughs in our understanding of the mechanisms and function of communication by sound.

Gerhardt and Huber emphasize a multidisciplinary neuroethological approach to acoustic communication, ranging from cellular mechanisms to evolutionary theory, and this is one of the strengths of their analyses. The authors discuss the information content of acoustic signals (Chap. 2), how they are produced, recognized, and localized (Chaps. 3, 5, 6, 7), their function as shown in preference and playback experiments (Chaps. 4, 8, 9, 10), and their evolutionary history (Chap. 11). Of course, with such a wide-ranging focus, some literature is omitted or only cursorily discussed, but references are generally complete and current to 2001. The arguments presented are tightly focused on the biological relevance of acoustic signaling for mating and territorial behaviors, as opposed to what the study of the auditory system of these animals can teach us about generalized (non-species-specific) aspects of auditory processing. For example, there is little attempt to compare or contrast neural coding of complex signals in the brains of chorusing insects and frogs with that in the brains of birds, fish, or mammals. Although Gerhardt and Huber address both function and mechanism, their conceptual emphasis are on behavioral and evolutionary issues, and how analysis of behavior can guide physiological questions.

Aside from the introductory first chapter, each of the ten subsequent chapters ends with a short summary and suggestions for further research. This is another strength of the book, as it provides clear identification of lacunae (both experimental and theoretical) in the literature. For example, the data on neural control of sound production (Chap. 3) is presented within a framework of command systems activating pattern-generating networks. This framework clearly applies to the insect literature, but there are insufficient data to determine if sound production in anurans can be reasonably described in this manner. This then is an open area for future research.

The critical coverage varies. Methodological issues are sometimes raised, but sometimes not. Additional information on particular topics not covered in the text is presented in ''boxes.'' An extremely useful feature of the book is the many tables and figures which directly provide comparative information (for example, Table 4.2, on behavioral preferences for temporal features of sounds), or which summarize the considerable literature in one group of animals (for example, Tables 9.1 and 9.2 on the evoked vocalization and satellite behaviors of male anurans). On the other hand, some of the figures are not well documented. For example, Fig. 2.3 summarizes within-male variation in acoustic properties in the advertisement calls of 24 species of anurans. Instead of providing documentation for the data set in the text or in an appendix (as was done for the comparisons in Fig. 10.1), readers are asked to contact Gerhardt directly for information on what species were included. This limits the archival relevance of the comparisons. A similar problem occurs with Fig. 5.12, which again lacks documentation, although in this instance readers are referred to a published paper to obtain the relevant information.

The book ends with three brief appendices on information theory, analysis of signals, and environmental acoustics, as well as a fourth appendix documenting the comparisons in Fig. 10.1. The first three appendices do not substitute for more complete sources. The index is thorough and easy to use.

I found several features of the coverage to be particularly insightful and well-argued. From their discussion of the literature on behavioral preferences for various features of acoustic signals, the authors conclude that animals probably make decisions based on multidimensional aspects of these signals, rather than just one. Although this may seem obvious, this distinction has been often ignored in the literature. Gerhardt and Huber emphasize the need for more behavioral studies, particularly field studies with more elaborate experimental protocols, to fully understand the communication system, and they argue that behavioral studies must take precedence over neurophysiological studies, in order to provide testable hypotheses to guide interpretation of mechanistic data. I would add that well-controlled psychophysical experiments, of which there are few, particularly in anurans, also are needed to provide a solid basis for interpreting both field and physiological data.

Similarly welcome is their critical evaluation of the influential concept of ''matched filtering'' that has guided much of the neurophysiological work on auditory processing in anurans. Gerhardt and Huber review the correlation of spectral tuning in the auditory nerve and frequency composition of advertisement calls in both insects and anurans, and conclude that these data are generally consistent with the idea of matched filtering. Still, some mismatches have been observed, and the authors point out some methodological factors that could affect interpretation of physiological data. For instance, because animals vocalize at levels well above threshold where the auditory system is nonlinear, emphasizing physiological responses at threshold levels (tuning curves) constrains our understanding of neural coding. Gerhardt and Huber are more critical of the literature relating behavioral preferences to neural responses in the central auditory system, and they argue that the evidence for matched filtering or selectivity in the temporal domain is weak. They conclude that single neurons that respond exclusively to conspecific signals probably do not exist, and they call for use of ensemble recording techniques and analysis of network characteristics to supplement the data on single cell recordings.

Restricting their analysis to chorusing insects and anurans, as the authors do, makes for a focused and coherent argument, but an obvious question one may ask is why songbirds, which also form choruses, are not included. The authors argue that chorusing insects and anurans are similar to each other in developing species-specific vocalizations in the absence of learning. Still, many species of mammals do not learn their vocalizations either, and there are clear innate features of bird song. Gerhardt and Huber acknowledge that experimental (rather than anecdotal) evidence that directly addresses experiential effects on vocalization development in anurans is lacking. In fact, some anuran tadpoles hear, and pass through lengthy larval stages in acoustically noisy environments, thus providing some substrate for possible effects of experience. Moreover, habituation is a form of learning, and it is difficult to argue for individual recognition in choruses (Chap. 9) without positing a role for experience. These caveats, of course, only emphasize the contributions Gerhardt and Huber have made in providing a conceptual foundation for guiding future research.

In the short time this book has been available, it has become a well-thumbed reference for my students, and several chapters have become the focus of our laboratory meetings. I highly recommend it for anyone interested in communication by sound.

ANDREA MEGELA SIMMONS
Departments of Psychology and Neuroscience
Brown University
Providence, Rhode Island 02912

 0001-4966/2003/114(2)/559/1/$19.00

ADVANCED-DEGREE DISSERTATIONS IN ACOUSTICS

Editor's Note: *Abstracts of Doctoral and Master's theses will be welcomed at all times. Please note that they must be limited to 200 words, must include the appropriate PACS classification numbers, and formatted as shown below. If sent by postal mail, note that they must be double spaced. The address for obtaining a copy of the thesis is helpful. Submit abstracts to: Acoustical Society of America, Thesis Abstracts, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, e-mail: asa@aip.org*

Evaluation of sound fields in a concert hall involving scattered reflections applying the subjective preference theory [43.55.Fw, 43.55.Br, 43.55.Ka]—Yukio Suzumura, *Graduate School of Science and Technology, Kobe University, Kobe, Japan, January 2003 (Ph.D.).* Convex tilted rear walls in a stage enclosure, an array of circular columns installed in front of walls, and triangular reflectors above the stage were newly adopted as scattering obstacles in an acoustic design of Tsuyama Music Cultural Hall, called ''Bell Folêt Tsuyama.'' The fundamental shape of the hall was designed using the theory of subjective preference. To calculate the effects of scattered reflections on a sound field in a real concert hall is extremely laborious. For this reason, the evaluation of effects of scattered reflections on the sound field in the hall was made experimentally by use of a $\frac{1}{10}$ acoustical scale. After construction of the hall, therefore, sound fields of the hall, which involves scattered reflections caused by the tilted convex rear, by the array of circular columns, and by the triangular reflectors, were measured using four orthogonal physical factors (LL, Δt_1, T_{sub}, IACC) described in the theory and the acoustical character of these scattering obstacles was clarified. Results clearly showed that these new attempts on scattered reflections substantially improved the quality of the sound field in the hall.

Thesis advisor: Yoichi Ando

Copies of this thesis written in English can be obtained from Yukio Suzumura. E-mail address: ysuzu11@lapis.plala.or.jp

On the temporal window of auditory-brain system in connection with subjective responses [43.60.Gk, 43.64.Bt]—Kiminori Mouri, *Graduate School of Science and Technology, Kobe University, Kobe, Japan, March 2003 (Dr.).* The human auditory-brain system processes information extracted from autocorrelation function (ACF) of the source signal and interaural cross correlation function (IACF) of binaural sound signals which are associated with the left and right cerebral hemispheres, respectively. The purpose of this dissertation is to determine the desirable temporal window (2T: integration interval) for ACF and IACF mechanisms. For the ACF mechanism, the visual change of $\Phi(0)$, i.e., the power of ACF, was associated with the change of loudness, and it is shown that the recommended temporal window is given as about $30(\tau_e)_{min}$ [s]. The value of $(\tau_e)_{min}$ is the minimum value of effective duration of the running ACF of the source signal. It is worth noticing from the experiment of EEG that the most preferred delay time of the first reflection sound is determined by the piece indicating $(\tau_e)_{min}$ in the source signal. For the IACF mechanism, the temporal window is determined as below: The measured range of τ_{IACC} corresponding to subjective angle for the moving image sound depends on the temporal window. Here, the moving image was simulated by the use of two loudspeakers located at $\pm 20°$ in the horizontal plane, reproducing amplitude modulated band-limited noise alternatively. It is found that the temporal window has a wide range of values from 0.03 to 1 [s] for the modulation frequency below 0.2 Hz.

Thesis advisor: Yoichi Ando

Copies of this thesis written in English can be obtained from Kiminori Mouri, 5-3-3-1110 Harayama-dai, Sakai city, Osaka 590-0132, Japan. E-mail address: km529756@aol.com

Subjective preference evaluation of sound fields by performing singers [43.55.Hy, 43.66.Mk, 43.75.Rs]—Dennis Noson, *Graduate School of Science and Technology, Kobe University, Kobe, Japan, March 2003 (Ph.D.).* A model of the auditory process is proposed for performing singers, which incorporates the added signal from bone conduction, as well as the psychological distance for subjective preference of the performer from the acoustic sound field of the stage. The explanatory power of previous scientific studies of vocal stage acoustics has been limited by a lack of an underlying theory of performer preference. Ando's theory, using the autocorrelation function (ACF) for parametrizing temporal factors, was applied to interpretation of singer sound field preference determined by the pair comparison method. Melisma style singing (no lyrics) was shown to increase the preferred delay time of reflections from a mean of 14 ms with lyrics to 23 ms without ($p<0.05$). The extent of the shift in preferred time delay was shown to be directly related to minima of the effective duration of the running ACF, $(\tau_e)_{min}$, calculated from each singer's voice. Voice matching experiments for singers demonstrated a strong overestimate of the voice outside the head compared with the singer's own voice (22.4 dB overestimate, $p<0.01$). Individual singer melisma singing delay preferences were compared for ''ah'' versus ''hum'' syllables, and the increased delay preference (41 ms) was shown to be correlated with $(\tau_e)_{min}$ ($r^2<0.68$, $p<0.01$). When the proposed bone conduction model was applied, using the measured overestimate of sound level of the singer's own voice *for each singer* (9.9 dB mean overestimate difference between ''ah'' and ''hum,'' $p<0.01$), the relationship of singer preference to $(\tau_e)_{min}$ was improved ($r^2=0.97$, $p<0.01$).

Thesis advisor: Yoichi Ando

Copies of this thesis are available from the author by inquiry at BRC Acoustics, 1741 First Avenue South, Seattle, WA 98134 USA. E-mail address: dnoson@brcacoustics.com

 0001-4966/2003/114(2)/560/1/$19.00

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*
HASSAN NAMARVAR, *Department of BioMed Engineering, University of Southern California, Los Angeles, California 90089*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prarie, Minnesota 55344*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
CARL J. ROSENBERG, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
WILLIAM THOMPSON, JR., *Pennsylvania State University, University Park, Pennsylvania 16802*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, Univ. of Rochester, Rochester, New York 14627*

6,503,580

43.35.Zc ACOUSTICALLY ENHANCED PAINT APPLICATION

Anthony A. Ruffa, assignor to The United States of America as represented by the Secretary of the Navy
7 January 2003 (Class 427/600); filed 30 July 2001

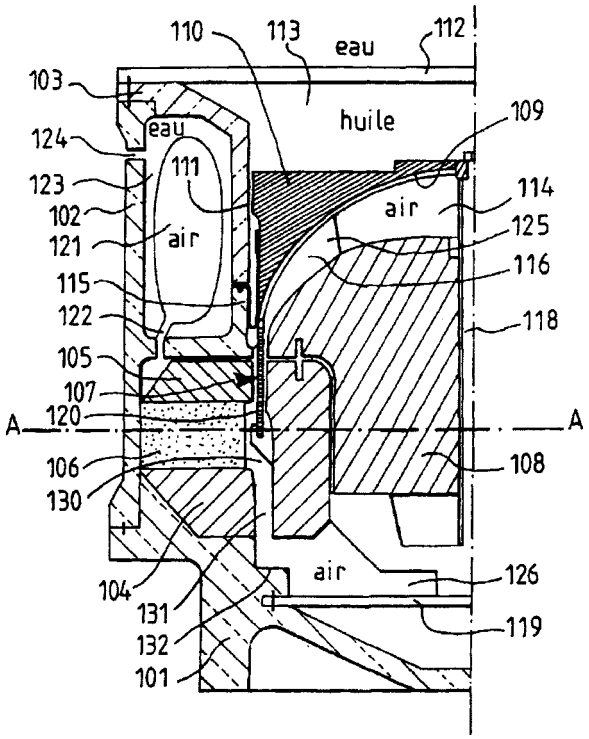
After a layer of liquid polymeric coating is applied to a substrate, an ultrasonic beam with acoustic pressures in excess of 190 dB in its main lobe is directed onto the coating. This is claimed to smooth any uneven surface features in the coating.—EEU

6,515,940

43.38.Dv ELECTRODYNAMIC TRANSDUCER FOR UNDERWATER ACOUSTICS

Vito Suppa *et al.*, assignors to Thales
4 February 2003 (Class 367/172); filed in France 26 May 2000

The cross section of an electrodynamic transducer, axisymmetric about a vertical center line, shown at the right-hand edge of the figure, reveals a horn-shaped radiating body **109** covered with some light rigid material **110**, such as syntactic foam, which makes the actual top radiating surface essentially flat. This radiating surface communicates with the surrounding water medium through a layer of oil **113** and a water-tight membrane **112**. The lower lip of the radiating body **109** supports a coil of wire **120** positioned in the thin air gap of a dc magnetic circuit **104**, **105**, and **106**. Hydrostatic pressure compensation is provided via an air-filled bladder **121** that communicates with the surrounding water via a series of holes **124** in the wall of the housing. The oil is separated from the internal air spaces by a thin membrane **115**. This design represents improvements over a previous but similar design in two ways that are intended to help dissipate heat build-up and thus increase efficiency. First, a series of vertical holes **131** is drilled through the pole pieces **104** and **105** to allow air to convect from a reservoir **126** near the bottom of the assembly up past the coil to the air space **114** near the top, where, upon cooling, it returns to the bottom reservoir **126** via some central



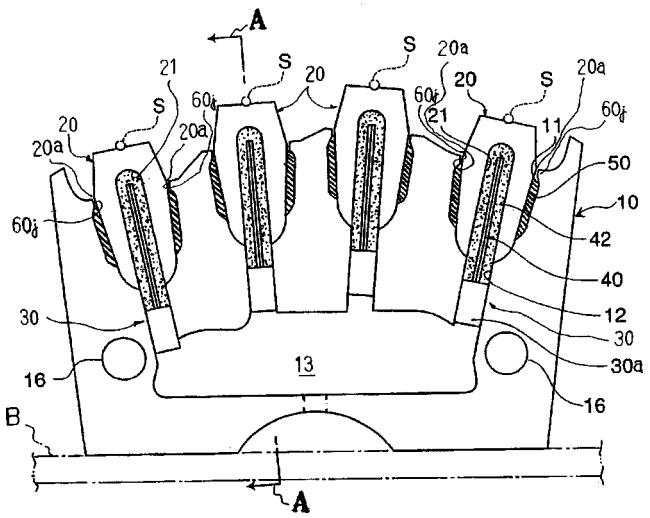
passageway. Second, metal masses (not shown in this view) are inserted between the several magnets **106** around the circumference of the assembly in order to conduct heat out to the surrounding water.—WT

6,515,214

43.38.Fx PICKUP UNIT INCORPORATED IN STRINGED INSTRUMENT FOR CONVERTING VIBRATIONS OF STRING TO ELECTRIC SIGNAL IN GOOD FIDELITY

Yojiro Takabayashi, assignor to Yamaha Corporation
4 February 2003 (Class 84/731); filed in Japan 27 April 2001

 0001-4966/2003/114(2)/561/15/$19.00



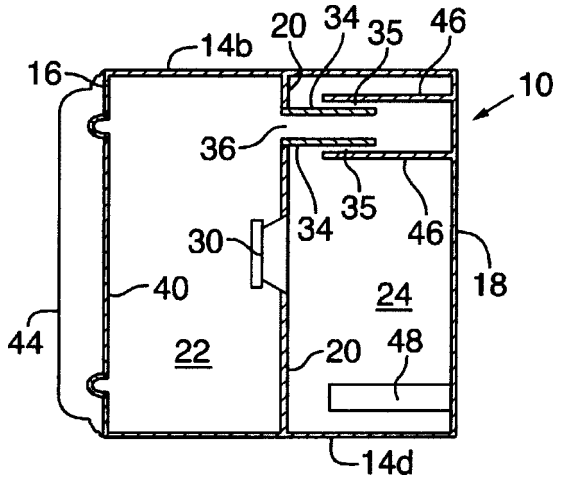
Electronic pickups for amplification can either be electromagnetic (if the strings are metallic) or microphones. Another proposed method is to use a piezoelectric transducer on each string and placed inside the bridge, as originally proposed by Barbera in 1989 (United States Patent 4,867,027). However, the author of this patent claims that the output of the piezoelectric sensors are too attenuated from the true signal. Accordingly, the transducers are modified so that the bimorph elements **40** are separated from the bridge **10** via the ''mediator'' **20** and the plastic coupling **42**.—MK

6,504,938

43.38.Ja DUAL-CHAMBER LOUDSPEAKER

Jeffrey S. Anderson *et al.*, assignors to Logitech Europe S.A.
7 January 2003 (Class 381/335); filed 6 October 2000

The text of this patent tends toward magical mysticism but the invention itself is not hard to understand. It is basically a conventional two-chamber, series-vented bandpass enclosure. Woofer **30** is mounted on partition **20** to drive front chamber **22** and rear chamber **24**. The two chambers



communicate through vent **36**. The resulting sound pressure in chamber **22** drives drone cone **40** to produce low-frequency sound. The only unusual feature appears to be concentric vent tubes **34–35**. This arrangement is said to allow easy adjustment for optimum tuning.—GLA

6,512,834

43.38.Kb ACOUSTIC PROTECTIVE COVER ASSEMBLY

Chad Anthony Banter and Brian G. Chapman, assignors to Gore Enterprise Holdings, Incorporated
28 January 2003 (Class 381/386); filed 7 July 1999

Microphones and loudspeakers are often used in exposed outdoor locations. If protective coverings are truly waterproof, they tend to be soundproof as well. The patent includes a good survey of the state of the art and then describes a weatherproof assembly ''that protects electronic devices from long-term exposure to liquid intrusion while providing equivalent or better sound attenuation than pre-existing acoustic covers.''—GLA

6,515,944

43.38.Md DISK PLAYBACK DEVICE AND METHOD OF CONTINUOUSLY PLAYING BACK DISKS WITHOUT A BREAK IN SOUND REPRODUCTION

Masanao Yoshida *et al.*, assignors to Sanyo Electric Company, Limited; Sanyo Tecnosound Company, Limited
4 February 2003 (Class 369/30.23); filed in Japan 2 March 1998

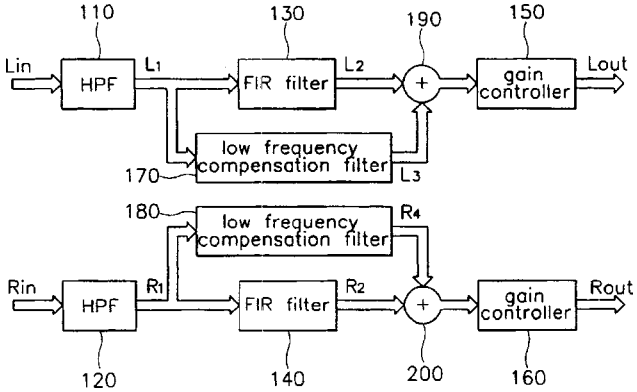
Readers of a certain age will recall turntables with spindles that dropped the next disk on top of the current disk for long playing recordings (like Beethoven's Ninth Symphony). Here, the inventors extend the idea, using a read-ahead memory to provide a window during which a new audio CD can be inserted in a single disk playback system and the playback would continue uninterrupted.—MK

6,504,933

43.38.Vk THREE-DIMENSIONAL SOUND SYSTEM AND METHOD USING HEAD RELATED TRANSFER FUNCTION

Dong-Ook Chung, assignor to Samsung Electronics Company, Limited
7 January 2003 (Class 381/1); filed in the Republic of Korea 21 November 1997

Finite impulse response filters with controllable gain are employed in a relatively simple arrangement intended to enhance two-channel stereo play-



back without adding loudspeakers. The two channels are processed separately so that existing directional cues can be maintained.—GLA

6,501,399

43.38.Vk SYSTEM FOR CREATING AND AMPLIFYING THREE DIMENSIONAL SOUND EMPLOYING PHASE DISTRIBUTION AND DUTY CYCLE MODULATION OF A HIGH FREQUENCY DIGITAL SIGNAL

Eldon Byrd, English, Indiana
31 December 2002 (Class 341/111); filed 16 May 2000

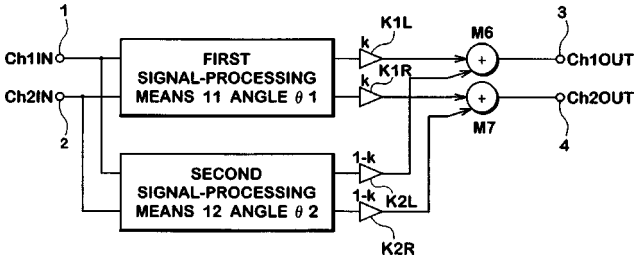
This patent describes an audio signal processing circuit that includes cross coupling, phase shifting, and harmonic enhancement. The processed signal is then converted into "duty cycle modulation of a high-frequency digital pulse (>43 kHz) in order to encode the information in a way that the ear and loudspeaker can precisely create the original information."—GLA

6,504,934

43.38.Vk APPARATUS AND METHOD FOR LOCALIZING SOUND IMAGE

Joji Kasai *et al.*, assignors to Onkyo Corporation
7 January 2003 (Class 381/17); filed in Japan 23 January 1998

Since the earliest days of commercial stereo reproduction inventors have searched for ways to create a wide stereo image from two closely spaced loudspeakers. For example, Harry F. Olson patented an arrangement in which sound from outwardly firing loudspeakers was redirected through acoustic prisms. In today's digital world, head related transfer functions are used to create phantom images electronically. This patent teaches that it is



relatively easy to create a convincing moving image if the movement is lateral, more difficult to create movement toward the listener, and almost impossible to create unambiguous movement away from the listener. An improved method is described which creates phantom images at predetermined angles and then makes use of a special coefficient to control appropriate cross coupling.—GLA

6,507,658

43.38.Vk SURROUND SOUND PANNER

Jonathan S. Abel and William Putnam, assignors to Kind of Loud Technologies, LLC
14 January 2003 (Class 381/17); filed 27 January 2000

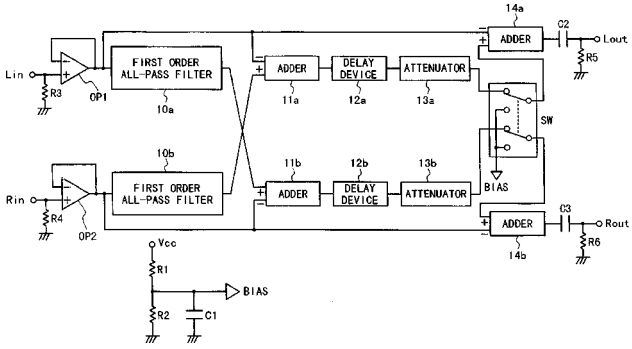
When stereophonic sound was added to motion pictures, three- and five-channel panning controls were developed to move phantom sources across the width of the movie screen. Later, to handle the requirements of quadraphonic recording, four-channel joystick panners were introduced. Today, the joystick panner or its software equivalent handles five channels, but it still attempts to locate a phantom source in terms of X-Y coordinates. The patent argues that better results can be achieved by controlling azimuth and source width instead.—GLA

6,507,657

43.38.Vk STEREOPHONIC SOUND IMAGE ENHANCEMENT APPARATUS AND STEREOPHONIC SOUND IMAGE ENHANCEMENT METHOD

Kenji Kamada *et al.*, assignors to Kabushiki Kaisha Kawai Gakki Seisakusho
14 January 2003 (Class 381/17); filed in Japan 20 May 1997

The circuit shown combines known stereo expansion techniques with interaural delay cancellation **12a**, **12b** and a few additional refinements. For



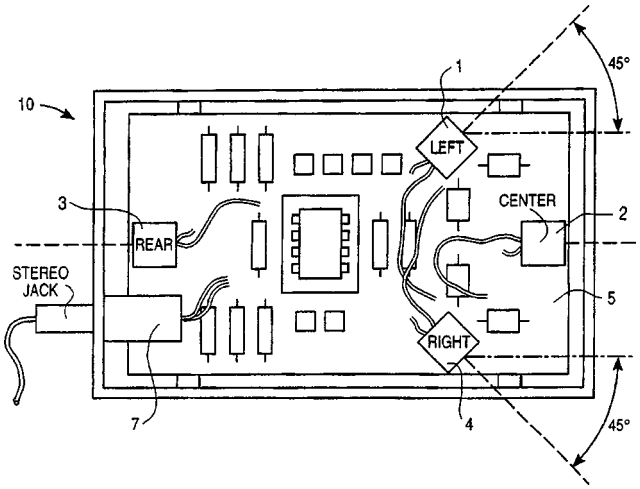
the most part, the various processing blocks use analog operational amplifiers although the delay devices can be analog or digital.—GLA

6,507,659

43.38.Vk MICROPHONE APPARATUS FOR PRODUCING SIGNALS FOR SURROUND REPRODUCTION

John J. Iredale and Roger S. Keller, assignors to Cascade Audio, Incorporated
14 January 2003 (Class 381/26); filed 22 November 2000

Wouldn't you like your portable video camera to automatically record five-channel surround sound? This patent teaches that the outputs of three or



more properly positioned microphones can be combined to produce a surround sound image compatible with current standard decoders.—GLA

6,515,939

43.38.Zp PULSE SAMPLED OPTICAL FIBER HYDROPHONE ARRAY (U)

Eugene L. Green *et al.*, assignors to The United States of America as represented by the Secretary of the Navy
4 February 2003 (Class 367/149); filed 4 September 1985

A fiber optic hydrophone array and its associated electronic circuitry are described for which the signal contaminating effects of light source intensity variation noise, lead induced phase noise, and light source phase noise can either be cancelled or just are not present.—WT

6,500,080

43.40.Tm GAME RACQUET WITH SEPARATE HEAD AND HANDLE PORTIONS FOR REDUCING VIBRATION

William D. Severa *et al.*, assignors to Wilson Sporting Goods Company
31 December 2002 (Class 473/521); filed 4 May 2001

Racquets devised according to this patent have separate head and handle portions, which are separated by a shock and/or vibration absorbing material, to which both are bonded.—EEU

6,500,279

43.40.Tm MATERIAL HAVING THE CAPACITY OF ABSORBING VIBRATION

Archer C. C. Chen, Taichung, Taiwan, Province of China
31 December 2002 (Class 148/325); filed 9 March 2001

The high-damping steel alloys described in this patent have double crystal structures of ferrite and pearlite. Loss factors of the order of 0.06 are cited, but no indications are given of how these were measured. The cited loss factors are about 60 times as great as the loss factor of a typical structural steel.—EEU

6,501,203

43.40.Tm VIBRATION CONTROL APPARATUS

Bjarni V. Tryggvason, assignor to Canadian Space Agency
31 December 2002 (Class 310/90.5); filed 1 June 2001

This apparatus, designed for use in micro-gravity environments in spacecraft, is essentially a two-stage isolation system employing magnetic levitation ''springs'' and actively controlled magnetic force actuators that maintain alignment in all six rigid-body degrees of freedom.—EEU

6,505,718

43.40.Tm VIBRATION DAMPING APPARATUS USING MAGNETIC CIRCUIT

Etsunori Fujita *et al.*, assignors to Delta Tooling Company, Limited
14 January 2003 (Class 188/267); filed in Japan 11 July 2000

An isolator with near zero stiffness over a double-amplitude of the order of 10 mm is comprised of a spring and magnetic circuit arrangement that act on a ferromagnetic block along a common axis. When the block is displaced downward from equilibrium, for example, the spring is compressed and the upward force it exerts on the block is increased. At the same time the upward (attractive) magnetic force that acts on the block is reduced, resulting in a near-constant net force on the block.—EEU

6,508,343

43.40.Tm VIBRATION DAMPER

Kazuhito Misaji *et al.*, assignors to Honda Giken Kogyo Kabushiki Kaisha; Tokai Rubber Industries, Limited
21 January 2003 (Class 188/379); filed in Japan 18 January 2000

The impact-type damper of this patent consists of a mass that is encapsulated in an elastic material and located in a housing, with small gaps between the outer surface of the mass and the inner surfaces of the housing. As the item to which the damper is attached vibrates, the mass slides along surfaces parallel to the direction of the vibration and impacts on surfaces perpendicular to it, thus removing energy from the primary vibration.—EEU

6,502,464

43.40.Yq ENERGY REDISTRIBUTION SYSTEM FOR A VIBRATING SYSTEM

Gregg K. Hobbs, Westminster, Colorado
7 January 2003 (Class 73/663); filed 29 September 2000

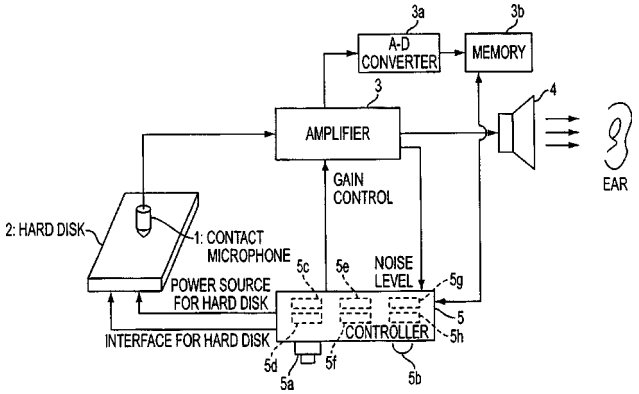
In certain shock tests, impacts are applied to a platform to which the test item is fastened, resulting in vibrations that occur preferentially in one direction and in a limited band of frequencies. This patent describes a system that consists essentially of a mass in an enclosure, which is attached to the test platform. Different shapes of the mass and different resilient coatings that may be applied to it produce different redistributions of the vibrational energy that occurs as the mass bounces back and forth against the sides of the enclosure.—EEU

6,523,413

43.40.Yq ACOUSTIC INSPECTION METHOD FOR HARD DISKS, ACOUSTIC INSPECTION SYSTEM FOR HARD DISKS, AND PROGRAM MEMORY MEDIUM

Takenori Hoshino *et al.*, assignors to Oki Electric Industry Company, Limited
25 February 2003 (Class 73/660); filed in Japan 28 December 1998

A device is described for detecting a blemish on the surface of a hard disk (particularly a stroke mark in the contact-start-stop area) without directly accessing the disk. An acoustic inspection is accomplished by detecting a vibration propagated through the hard disk housing which occurs when the head collides with the disk surface blemish. The inspection may be conducted during a period of time when the drive power is turned off but the disk continues to rotate by inertia.—DRR



6,510,738

43.40.Yq DEVICE AND METHOD FOR MEASURING VIBRATION

Chih-Kung Lee *et al.*, assignors to National Science Council
28 January 2003 (Class 73/579); filed in Taiwan, Province of China 9 August 1999

A vibration sensor according to this patent consists of a flexible element, such as a cantilever plate strip, to which there is attached a piezoelectric sensing layer. This layer is shaped so that its output is different for different modes of vibration of the flexible element, with the intent of obtaining a sensor with desired frequency bandwidth characteristics.—EEU

6,520,751

43.50.Gf VARIABLE DISPLACEMENT COMPRESSOR HAVING A NOISE REDUCING VALVE ASSEMBLY

Masaaki Fujita *et al.*, assignors to Sanden Corporation
18 February 2003 (Class 417/295); filed in Japan 4 April 2000

The aim of this device is to provide a variable displacement piston compressor capable of reducing the noise which results from repetition of fine movement of a valve body when opening the control valve at a very low flow rate. One embodiment of the device provides a suction port communicating with a suction chamber via a valve-controlled variable main channel and a bypass channel formed outside of the fluid damper. A fluid damper coupled to the main valve body damps vibration of the valve body.—DRR

6,520,280

43.55.Dt SYSTEM AND METHOD FOR WORKSPACE SOUND REGULATION

Paul Andrew Moskowitz, assignor to International Business Machines Corporation
18 February 2003 (Class 181/30); filed 31 January 2001

It is known that an ellipse will concentrate sound from one focal point to the other. By this invention, a work space is designed with reflecting walls in an elliptical arrangement in order to enhance communication between a worker at one focal point and a visitor or co-worker at the other. Because of the focusing effect, the two will communicate at a lower level, and reduce noise overall in a workspace environment.—CJR

6,520,288

43.55.Ti ACOUSTIC DOOR ASSEMBLY

Grant S. Quam *et al.*, assignors to Wenger Corporation
18 February 2003 (Class 181/287); filed 7 April 2000

This acoustic door assembly comprises an insulated acoustic door (with interior insulating layers) and a continuous cam-lift hinge along the entire length of the door, with a dual-magnetic seal at the perimeter of the door. The patent suggests that the continuous cam-lift hinge assures better support as well as better closure than several individual cam-lift hinges.—CJR

6,453,253

43.58.Gn IMPULSE RESPONSE MEASURING METHOD

Tsugio Ito, assignor to Yamaha Corporation
17 September 2002 (Class 702/77); filed in Japan 20 August 1998

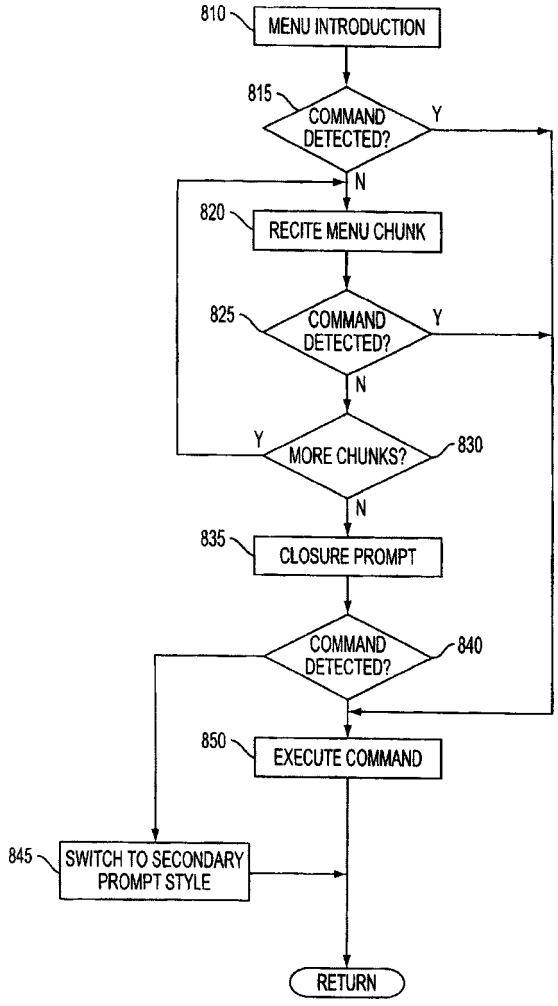
A method is described by which an impulse response, such as for an item of audio equipment, as measured in a typical small room may be averaged so as to obtain an impulse response essentially equivalent to what would be recorded in an anechoic chamber. The actual response signal is transformed to a complex spectrum and then portions of the spectrum are averaged. The averaged portions may be selected using either a constant frequency difference scale or a constant frequency ratio scale. The averaged complex spectrum is then inverse transformed to obtain the ''anechoic equivalent'' impulse response.—DLR

6,526,382

43.60.Qv LANGUAGE-ORIENTED USER INTERFACES FOR VOICE ACTIVATED SERVICES

Matthew John Yuschik, assignor to Comverse, Incorporated
25 February 2003 (Class 704/275); filed 7 December 1999

This patent describes a voice activated user interface (VAUI) based on an adaptive approach. The patent provides a significant systematic method to develop an adaptive VAUI which captures the semantics and syntax of



speech for each user. The system response is adaptively varied by means of an error prompt level selected on the basis of the accumulated number of user errors.—HHN

6,519,342

43.60.Qv METHOD AND APPARATUS FOR FILTERING AN AUDIO SIGNAL

Martin Opitz, assignor to AKG Akustische u. Kino-Gerate Gesellschaft m.b.H.
11 February 2003 (Class 381/61); filed in Germany 7 December 1995

At times, filtering of an audio signal is necessary to filter out room effects in recordings, to achieve a certain tone color, or to raise or lower the amplitude within certain frequency ranges. In the apparatus described here, an audio signal is digitized so that the duration of the sampling intervals is half or less of the period duration of the highest frequency to be expected in the audio signal. A digitized impulse response is made available in accordance with the desired filtering effect and a convolution sum is developed from the samples of the impulse response and of the audio signal. This has the result that (a) several adjacent samples of the impulse response define a shorter interval within the impulse response, (b) within the shorter interval, the samples of the impulse response are equated to a function of one or more of those samples, and (c) the steps (a) and (b) are repeated as necessary, with the requirement that the intervals do not overlap.—DRR

6,524,619

43.64.Gz DOSAGE FORMS USEFUL FOR MODIFYING CONDITIONS AND FUNCTIONS ASSOCIATED WITH HEARING LOSS AND/OR TINNITUS

Don C. Pearson and Kenneth T. Richardson, assignors to Chronorx, Incorporated
25 February 2003 (Class 424/472); filed 19 January 2001

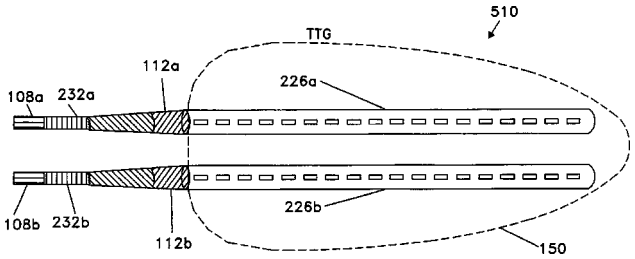
In this pharmaceutical patent, a number of active agents and formulations are recommended as being complementary in their physiological functions, especially as they relate to the quenching of free radicals, to the support of endothelial physiology, the reduction of hyperinsulinemia, and improvements in vascular health. The active components of the formulations were selected for inclusion in precise combinations specifically because they are claimed to improve physiological functions, thereby reducing a variety of risks associated with hearing loss and tinnitus, through improvement in local VIIIth nerve vascular health, modulation of conditions surrounding blood fluid dynamics, the consequences of hyperinsulinemia, and improvements in free radical defenses. A reduction of the potential for cochlear hair death, VIIIth nerve atrophy, and hearing loss is claimed.—DRR

6,456,886

43.64.Sj HUMAN CERABAL CORTEX NEURAL PROSTHETIC FOR TINNITUS

Matthew A. Howard III *et al.*, assignors to University of Iowa Research Foundation
24 September 2002 (Class 607/55); filed 5 June 1997

The detail section of this patent begins by stating the assumption that tinnitus consists of abnormally firing neurons. In view of this assumption, the device disclosed here would alter or suppress the abnormal firing patterns by stimulation of the tonotopically appropriate portions of the cerebral cortex or of the thalamus. The device consists of a multi-electrode probe



intended to be deployed along a tonotopic axis in a region of the auditory cortex or thalamus. Interestingly, the same therapy might be expected to be effective in the case of one of the alternative theories of tinnitus, namely that it represents an auto-initiated firing pattern created by the cortex in the absence of external stimulation.—DLR

6,517,480

43.64.Yp NEUROLOGICAL TESTING APPARATUS

Alvin Krass, Little Silver, New Jersey
11 February 2003 (Class 600/300); filed 4 October 2000

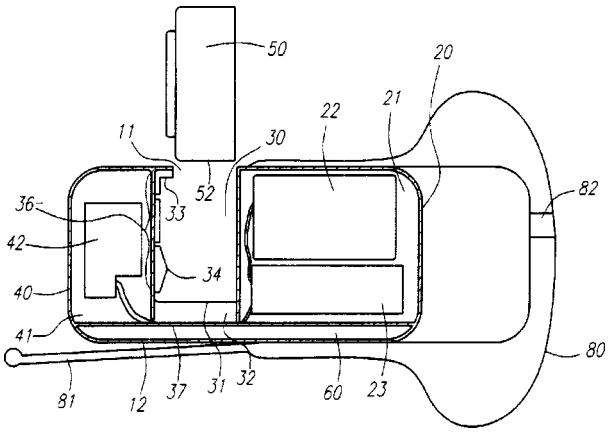
This apparatus is supposed to be a ''do-all'' device for performing neuropsychological tests. In addition to applying visual stimuli, the device contains a means of generating auditory sensory signals and a reaction detector such as a dynamometer, a touch screen, or a knob. Microprocessors are included to generate the sensory signals and automatically record neuropsychological responses.—DRR

6,516,074

43.66.Ts HEARING DEVICE WITH INTEGRATED BATTERY COMPARTMENT AND SWITCH

Owen D. Brimhall and Gregory N. Koskowich, assignors to Sonic Innovations, Incorporated
4 February 2003 (Class 381/322); filed 19 October 2000

An electronic hearing device is described in which the battery, when placed in the battery compartment, forms a portion of the outer wall of the device's housing. The battery compartment in the housing also functions as an integrated switch, which has a contact that engages the battery when the



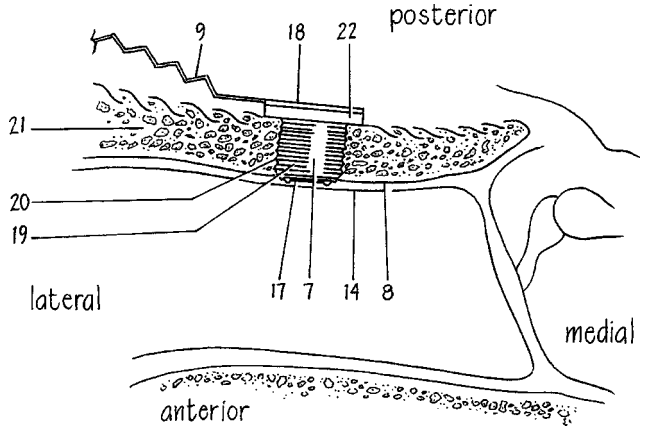
hearing device is deformed. The switch selects from several digital programs within the device. A conformal tip, used in combination with hearing aid retention and extraction means, surrounds the battery compartment and battery. A pathway in the housing provides a vent for air pressure equalization.—DAP

6,516,228

43.66.Ts IMPLANTABLE MICROPHONE FOR USE WITH A HEARING AID OR COCHLEAR PROSTHESIS

Peter G. Berrang *et al.*, assignors to Epic Biosonics Incorporated
4 February 2003 (Class 607/57); filed 7 February 2000

A totally implantable subminiature electret microphone, which is designed to be placed in the wall of the external auditory canal, has a bio-inert metallic membrane sound inlet **17** that is covered by the thin auditory canal skin **14**. The hermetically sealed microphone further consists of a cylindrical bio-inert housing **7** that has threaded and flange features to assist it in os-



siointegrating into the auditory canal bone **21** and to prevent migration into the auditory canal. The microphone output is facilitated, without compromising hermetic sealing, via lithographically formed laminated wires that are connected to electrical feedthroughs on the flange side of the housing.—DAP

6,522,988

43.66.Yw METHOD AND SYSTEM FOR ON-LINE HEARING EXAMINATION USING CALIBRATED LOCAL MACHINE

Zezhang Hou, assignor to Audia Technology, Incorporated
18 February 2003 (Class 702/122); filed 31 March 2000

This is a method for on-line audiological testing and hearing aid fitting. It is claimed that the on-line hearing loss testing can be self-performed without any specialized equipment (save for the audio components). In one embodiment, calibration parameters are determined by a calibration process on a local machine and participant parameters are produced by an on-line hearing loss test performed for the participating subject. Hearing loss for the participating subject is determined from the two parameter sets. Another approach entails on-line sound customization which can simulate hearing compensation, effectively constituting a hearing aid fitting process.—DRR

6,453,284

43.72.Ar MULTIPLE VOICE TRACKING SYSTEM AND METHOD

D. Dwayne Paschall, assignor to Texas Tech University Health Sciences Center
17 September 2002 (Class 704/208); filed 26 July 1999

The patent describes a neural network based processor for isolating the voice signal of a single talker from a composite signal containing multiple speakers' voices. An absolute magnitude difference function (AMDF) first isolates the pitch of each speaker's voice signal. The various pitch values are then tracked by a recursive neural network predictor. Although we are told that the AMDF pitch detector will work with multiple voices intermixed, no further detail is provided to substantiate that claim or of problems which may arise. The neural net organization and training is described in rather more detail.—DLR

6,453,043

43.72.Ew FLASH-CUT OF SPEECH PROCESSING FEATURES IN A TELEPHONE CALL

Lee Begeja *et al.*, assignors to AT&T Corporation
17 September 2002 (Class 379/406.03); filed 30 December 1998

Various speech enhancement processes are typically applied to the voice signal during normal telephone system operation. Some of these processes, such as filtering, are ''full on'' immediately, while others, such as echo cancellation and gain control, are adaptive and require a ''ramp-up'' time following activation. During this ramp-up interval, some of the adaptive processes may interfere with each other, degrading the speech output. According to the method described here, the ramp-up interval occurs using an off-line copy of the speech signal. Once all interacting processes have stabilized, a switchover occurs, inserting the processed signal onto the line at one instant, bypassing the ramp-up operation.—DLR

6,453,020

43.72.Kb VOICE PROCESSING SYSTEM

Jeremy Peter James Hughes *et al.*, assignors to International Business Machines Corporation
17 September 2002 (Class 379/88.04); filed in the United Kingdom 6 May 1997

This is an arrangement of voice processing equipment in a local area network (LAN) to handle speech recognition for a multi-line telephone response system. Of special interest are the control system for handling barge-in, where a user may interrupt a system prompt, and an echo cancellation (EC) circuit in the phone line input. When the user's voice input is detected, the EC circuit determines that it is not a line echo and any outgoing prompt is immediately terminated. The input signal is then routed via the LAN to the recognition system.—DLR

6,453,290

43.72.Kb METHOD AND SYSTEM FOR NETWORK-BASED SPEECH RECOGNITION

Christopher S. Jochumson, assignor to Globalenglish Corporation
17 September 2002 (Class 704/231); filed 4 October 1999

This patent deals with methods for rapid delivery and processing of voice signals in the Internet environment. All of the discussion pertains to the standard TCP/IP Internet protocols, with no mention of the newer ''real-time'' protocols, such as Voice over IP. Most of the patent text deals with issues of buffering the speech data signals. It is not at all clear that any solution is presented to the real issue of network transmission delays.—DLR

6,453,292

43.72.Ne COMMAND BOUNDARY IDENTIFIER FOR CONVERSATIONAL NATURAL LANGUAGE

Ganesh N. Ramaswamy and Jan Kleindienst, assignors to International Business Machines Corporation
17 September 2002 (Class 704/235); filed 28 October 1998

The task addressed by this patent is the problem of locating a command to be executed within a stream of content words, i.e., text to be acted upon by the commands. The method assumes that a high-quality recognizer operates on the speech stream as a first step. The sequence of recognized words is then examined for specific words and grammatical structures which can occur within a command or are likely to represent a command. A rec-

ognized command may be extracted from the word string or may be marked in some specific way as a flag for subsequent processing.—DLR

6,456,976

43.72.Ne MOBILE TERMINAL PROVIDED WITH SPEECH RECOGNITION FUNCTION FOR DIAL LOCKING

Takehiko Kuita, assignor to NEC Corporation
24 September 2002 (Class 704/273); filed in Japan 26 November 1998

This patent deals with some of the security issues involving cellular telephones. If the instrument is lost or stolen, calls may be fraudulently placed or valuable stored information may be accessed. Locking mechanisms are sometimes provided, but may be difficult to use and are thus frequently not activated. The patent discusses several strategies for using speech recognition capabilities to make the locking function easier to use.—DLR

6,456,977

43.72.Ne VOICE CONTROL MODULE FOR CONTROLLING A GAME CONTROLLER

Jong-Ding Wang, assignor to Primax Electronics Limited
24 September 2002 (Class 704/275); filed in Taiwan, Province of China 15 October 1998

This patent offers the prospect of a determined video game player not only hammering furiously on a firing button, but also yelling a desperate series of verbal commands at the game. Multiple ''levels'' of commands are considered, consistent with the demands of the game. For example, ''start'' and ''quit'' are presented as ''execute'' commands, while ''change weapon'' followed by ''rocket'' are considered as ''switch'' commands. There is no discussion of training, error handling, or any other of the realities of speech recognition.—DLR

6,515,209

43.75.Gh ADD-ON HEADSTOCK MASS DEVICE FOR A STRINGED MUSICAL INSTRUMENT

Richard Aspen Pittman, Sylmar, California
4 February 2003 (Class 84/453); filed 4 May 2001

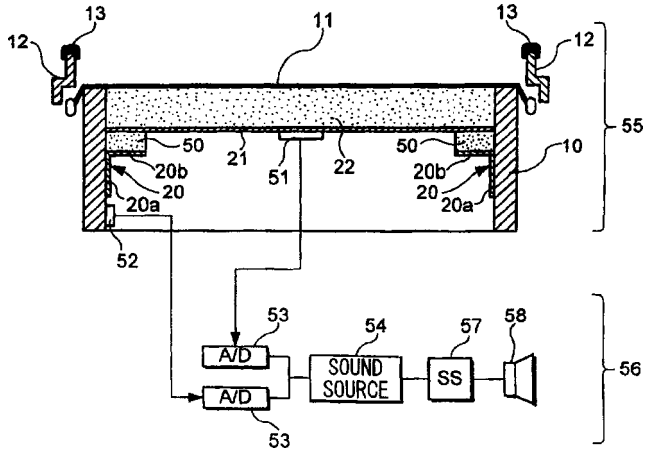
The inventor claims ''It has been recognized that the quality of sound is influenced by the effective mass in the instrument in the vicinity of the string support...at either end of the string.'' For the past 25 years or more, folk theory says that adding a C clamp on the headstock will increase sustain in plucked stringed instruments. Finally, someone has decided to patent a version of it. But since the weight of the proposed clamp is only 3.8 oz, wouldn't it be better to use depleted uranium?—MK

6,525,249

43.75.Hi DRUMHEAD AND MUTING STRUCTURE FOR ACOUSTIC AND ELECTRONIC PERCUSSION INSTRUMENTS

Yuichiro Suenaga, assignor to Yamaha Corporation
25 February 2003 (Class 84/411 R); filed in Japan 15 November 1999

There are two threads in this patent: First, drum heads should be made so they can be used fairly silently but without differences of ''feeling.'' Second, such a drum could also be used as an electronic drum transducer. In



the first usage, the head is porous, thereby changing the rebound; in the second, shown in the figure, a sensor **51** is connected to a support plate **21** that forms a sandwich with the head membrane **11**. The absorption layer **22** serves to dampen the hit.—MK

6,525,256

43.75.Wx METHOD OF COMPRESSING A MIDI FILE

Daniel Boudet *et al.*, assignors to Alcatel
25 February 2003 (Class 84/645); filed in France 28 April 2000

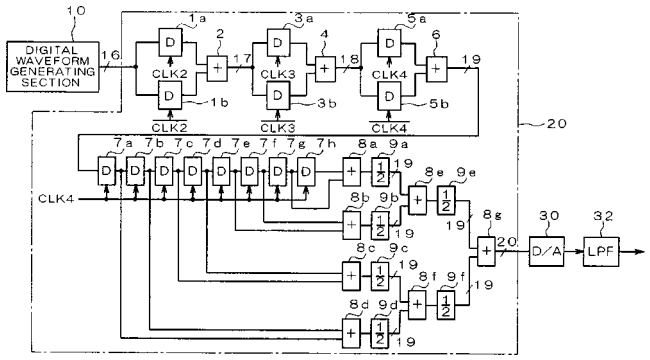
MIDI is a popular form of electronic score consisting principally of ''note on'' and ''note off'' commands. The patent proposes to compress a MIDI score for transmission. You might wonder why, given that MIDI is already fairly compressed. The answer lies in the target application: cellular telephony. Also included is a discussion of what to do when the limits of polyphony are reached at the terminal end.—MK

6,515,608

43.75.Yy DIGITAL-ANALOG CONVERTER AND METHOD, AND DATA INTERPOLATION DEVICE AND METHOD

Yukio Koyanagi, assignor to Yasue Sakai
4 February 2003 (Class 341/144); filed in Japan 18 June 1999

The issue is the output of a digital-to-analog (D/A) converter. It uses an interpolation method at the final output as shown in the figure. The output from the waveform section **10** is oversampled and then given to an eight-



sample sliding window **7**, which is then used to form the output sample and given to the final D/A converter **30**. Effectively, this is equivalent to a moving average filter applied to the digital input.—MK

6,519,547

43.80.Nd METHOD FOR DETERMINING ACOUSTIC IMPACT OF UNDERWATER ACOUSTIC SOURCES ON MARINE ANIMALS

Colin J. Lazauski *et al.*, assignors to The United States of America as represented by the Secretary of the Navy
11 February 2003 (Class 702/139); filed 23 January 2001

A procedure is presented to estimate the acoustic impact upon marine animals within a certain area. The procedure begins with information on the types and number of locations of acoustic sources to be used, the marine animals in that area, environmental characteristics of the area, and the existing environmental regulations relevant to the area. An acoustic model appropriate to the area is then chosen. Using the above information, the model generates a ''source footprint,'' e.g., contours of constant SPL or some other energy based criterion. This result is combined with time-weighted information about the marine animal distributions to compute the total number that might be impacted. Hence an acoustic test procedure, type of sources or their locations, or other criteria relevant to the number of marine animals impacted can be modified to reduce the acoustic impacts.—WT

6,513,621

43.80.Qf METHOD OF PRODUCING AND MAKING USE OF EAR TIPS HAVING A FILLED AIRTIGHT CHAMBER

Richard J. Deslauriers *et al.*, assignors to Doctors Research Group
4 February 2003 (Class 181/130); filed 15 August 2000

The preferred embodiment of this device consists of a liquid gelatin-filled ear tip with a unit for coupling to the patient's auditory canal through a center auditory passage. The ear tip is fabricated from a single piece of elastomeric material in the form of a sealed chamber filled with liquid gelatin. The resulting compliance permits the unit to configure to different outer ear canals. In addition, means are provided for coupling the ear tip to a stethoscope or another listening device.—DRR

6,514,201

43.80.Qf VOICE-ENHANCED DIAGNOSTIC MEDICAL ULTRASOUND SYSTEM AND REVIEW STATION

Jeffrey M. Greenberg, assignor to Acuson Corporation
4 February 2003 (Class 600/437); filed 29 January 1999

Apparently designed to accommodate die hard cyberphobes, this ultrasound diagnostic system contains voice-related interfaces so that a user can interact with an imaging system or review station by issuing verbal commands rather than using a mouse, keyboard, or other user interface that requires physical manipulation on the part of the user. The voice feedback system can also be used to allow the imaging system or review station to communicate with a patient.—DRR

6,514,204

43.80.Qf METHODS FOR ESTIMATING TISSUE STRAIN

Sheikh Kaisar Alam *et al.*, assignors to Riverside Research Institute
4 February 2003 (Class 600/442); filed 20 July 2001

A method for estimating tissue strain is proposed that entails transmitting ultrasonic signals into tissue and detecting the reflected signals. The tissue is then compressed to induce tissue strain. Ultrasonic signals are sent into the compressed tissue and another set of reflected signals is detected. The two sets of reflected signals are transformed into the spectral domain and the transform of the second set is scaled by a frequency scaling factor. The variance of the ratio of the scaled and the nonscaled spectral signals is then computed. The scaling factor is varied and this process is repeated to minimize the variance. Local tissue strain is then estimated from the frequency scaling factor at which the variance of the ratio is at a minimum.—DRR

6,514,211

43.80.Qf METHOD AND APPARATUS FOR THE NONINVASIVE DETERMINATION OF ARTERIAL BLOOD PRESSURE

Gail D. Baura, assignor to Tensys Medical, Incorporated
4 February 2003 (Class 600/490); filed 21 January 2000

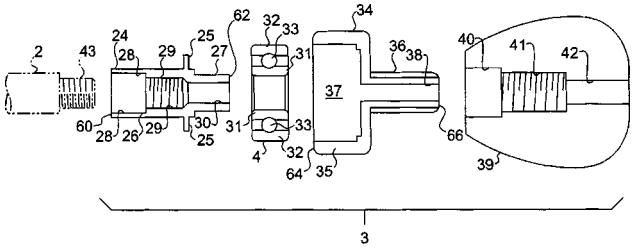
For the purpose of determining the mean arterial blood pressure, pressure and ultrasound transducers are placed over the radial artery of a patient's wrist. The ultrasound transducers transmit and receive acoustic energy so as to measure the blood velocity during periods of variable compression of the artery. During compression, the ultrasound velocity waveforms are recorded and processed through time-frequency analysis. The time at which the mean time-frequency distribution is maximal corresponds to the time at which the transmural pressure equals zero and thus the mean pressure read by the transducer equals the mean pressure within the artery. In another embodiment of the device, a wrist brace is disclosed that is said to be useful for measuring blood pressure. A method of continuously estimating systolic and diastolic pressure is also described.—DRR

6,514,213

43.80.Qf EARPIECE ASSEMBLY FOR A STETHOSCOPE

Takashi Moteki, assignor to Moteki Industries, Incorporated
4 February 2003 (Class 600/528); filed 3 October 2000

The purpose of this earpiece assembly is to provide a stethoscope which produces less undesired noise and to prevent injury to a user's ears.



The earpiece assembly includes a pipe **2** and a rotating bushing **32** which connects the pipe to an earpiece **37**. A sound path traverses from the earpiece through the bushing into the pipe.—DRR

6,514,220

43.80.Qf NON FOCUSSED METHOD OF EXCITING AND CONTROLLING ACOUSTIC FIELDS IN ANIMAL BODY PARTS

Hewlett E. Melton, Jr. *et al.*, assignors to Walnut Technologies
4 February 2003 (Class 601/2); filed 25 January 2001

This method takes advantage of the fact that body structures can act as resonators. For example, the brain vault is bounded by layers of differing acoustic impedance, thereby causing reflection of acoustic waves at these boundaries. At frequencies in the 0–500-kHz range, little attenuation of longitudinal waves occurs in the brain or skull. If reflection at a boundary layer is near total, then acoustic waves pass back and forth through tissue many times, creating a trapped mode resonator. The effect of ultrasound irradiance of a human or other animal body portion is claimed to be en-

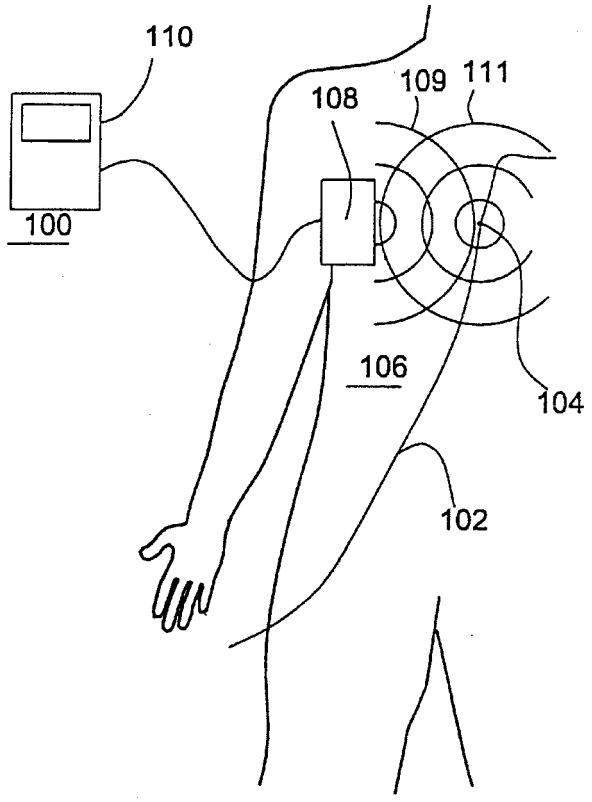
hanced by operating the body portion as a trapped mode resonator. The intensity and location of resonances within the body portion is controlled through such variables as the amplitude, frequency, and/or phase of the ultrasonic irradiation. This is said to minimize the overall energy required to be applied to the body portion in order to achieve a desired localized intensity field.—DRR

6,517,481

43.80.Qf METHOD AND SENSOR FOR WIRELESS MEASUREMENT OF PHYSIOLOGICAL VARIABLES

Bertil Hoek *et al.*, assignors to Radi Medical Systems AB
11 February 2003 (Class 600/300); filed 12 September 2001

A sensor and guide wire assembly incorporating a sensor element **104** is inserted into a patient's body. Either acoustic or electromagnetic waves are then emitted from an external source toward the sensor element. Upon receiving the incident waves, the sensor element enters into a resonating



state. The characteristics of the resonance correspond to the variable or variables to be measured. A transducer, located externally with respect to the patient's body, receives the scattered waves and transmits a resultant signal to a signal analyzer.—DRR

6,517,488

43.80.Qf MEDICAL DIAGNOSTIC ULTRASOUND SYSTEM AND METHOD FOR IDENTIFYING CONSTRICTIONS

John A. Hossack, assignor to Acuson Corporation
11 February 2003 (Class 600/454); filed 29 June 2000

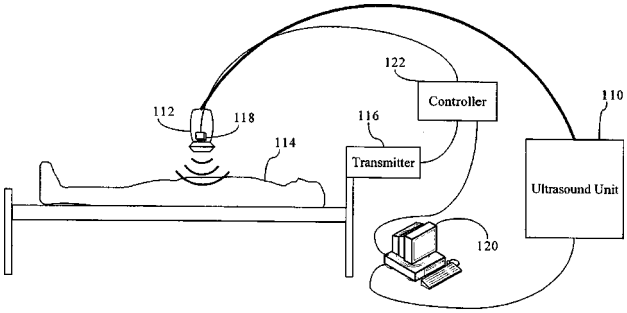
This patent refers to a medical diagnostic ultrasound system and a method for identifying constrictions in blood vessels. Total blood volume flow is measured in various locations along a vessel. According to a basic principle of fluid mechanics, the total volume flow must be conserved. Regions that appear to violate conservation of flow are identified as suspicious and are likely to be associated with a dropout artifact. Actual constrictions are established from the images in the regions where the geometry is not suspicious.—DRR

6,517,491

43.80.Qf TRANSDUCER WITH SPATIAL SENSOR

Karl E. Thiele and Showna Hsu-Hwa Chang, assignors to Koninklijke Philips Electronics N.V
11 February 2003 (Class 600/459); filed 31 October 2000

In the realm of ultrasonic imaging, freehand imaging is a method to develop 3-D images in which the sonographer moves a 1-D array of transducers across a patient ''freehand'' and a specialized graphic processor attempts to warp together a 3-D image. One innovation that makes improved 3-D imaging possible is the use of location sensors externally mounted on a 1-D sound transducer to register the spatial location and orientation with respect to translation and angulation of acquired ultrasound images. The block diagram in the figure shows an ultrasound imaging system configured



PRIOR ART

for freehand scanning. A transducer connected to the ultrasound unit outputs and receives ultrasound signals under the control of the imaging unit so as to scan a patient in a known fashion. The transducer works in conjunction with location sensors that correlate the ultrasound echoes obtained by the transducer with the spatial location of the transducer. The gathered information gleaned from a transducer scanning is then ''stitched together'' to form a 3-D image. The novel aspect of this patent appears to be the combining of the separate cables into a single cable going to an integrated control unit.—DRR

6,520,281

43.80.Qf ELASTOMERIC ANTI-MICROBIAL STETHOSCOPE DIAPHRAGM

Richard J. Deslauriers *et al.*, assignors to Doctors Research Group
18 February 2003 (Class 181/131); filed 18 August 2000

This patent describes a replaceable and disposable elastomeric soft diaphragm for use with a stethoscope. The diaphragm (or cover) consists of a molded piece of elastomer that is sufficiently flexible to provide an airtight seal with a stethoscope and allow for easy application and removal. The disposability of the diaphragm is meant to prevent transmission of bacteria from one patient to the next during the use of the stethoscope.—DRR

6,520,914

43.80.Qf ULTRASONIC DENSITOMETER WITH PRE-INFLATED FLUID COUPLING MEMBRANES

Richard F. Morris *et al.*, assignors to Lunar Corporation
18 February 2003 (Class 600/449); filed 21 December 2000

This is an ultrasonic densitometer that uses ultrasound waves to measure bone integrity. The ultrasound is conducted into a human heel through

liquid-filled bladders. The preinflated bladders are arranged to form a cavity between them slightly smaller than the heel and the heel is slid in between the bladders. The sliding provides a wiping action that helps eliminate air trapped between the bladders and the foot. The shape and configuration of the bladders also help in the alignment of the foot with respect to the propagation axis.—DRR

6,520,918

43.80.Qf METHOD AND DEVICE FOR MEASURING SYSTOLIC AND DIASTOLIC BLOOD PRESSURE AND HEART RATE IN AN ENVIRONMENT WITH EXTREME LEVELS OF NOISE AND VIBRATIONS

Stergios Stergiopoulos and Amar C. Dhanantwari, assignors to Her Majesty the Queen in Right of Canada as represented by the Minister of National Defence
18 February 2003 (Class 600/490); filed in Canada 24 November 1999

In the field of auscultatory blood pressure monitoring, this apparatus removes noise and vibration effects from audible heart beat sounds. Blood pressure signals corresponding to the heartbeat are detected by a sensor placed on a patient near an artery. Another acoustic transducer is placed on the patient away from the artery to detect noise and vibrations. Pressure is applied to the artery thereby forcing the artery to close. The pressure is then reduced and, while reducing the pressure, the artery pressure signal and the two acoustic signals are relayed to a processing unit. The signal of the former acoustic sensor is then processed using an adaptive interference canceller algorithm with the signal from the second transducer acting as interferer. Heart beat pulses are determined from the processed signal. Relating the heart beat pulses to the pressure signal provides the systolic and diastolic blood pressure.—DRR

6,520,924

43.80.Qf AUTOMATIC DIAGNOSTIC APPARATUS WITH A STETHOSCOPE

Byung Hoon Lee, Kangnam-ku, Seoul, the Republic of Korea
18 February 2003 (Class 600/586); filed in the Republic of Korea 16 November 2000

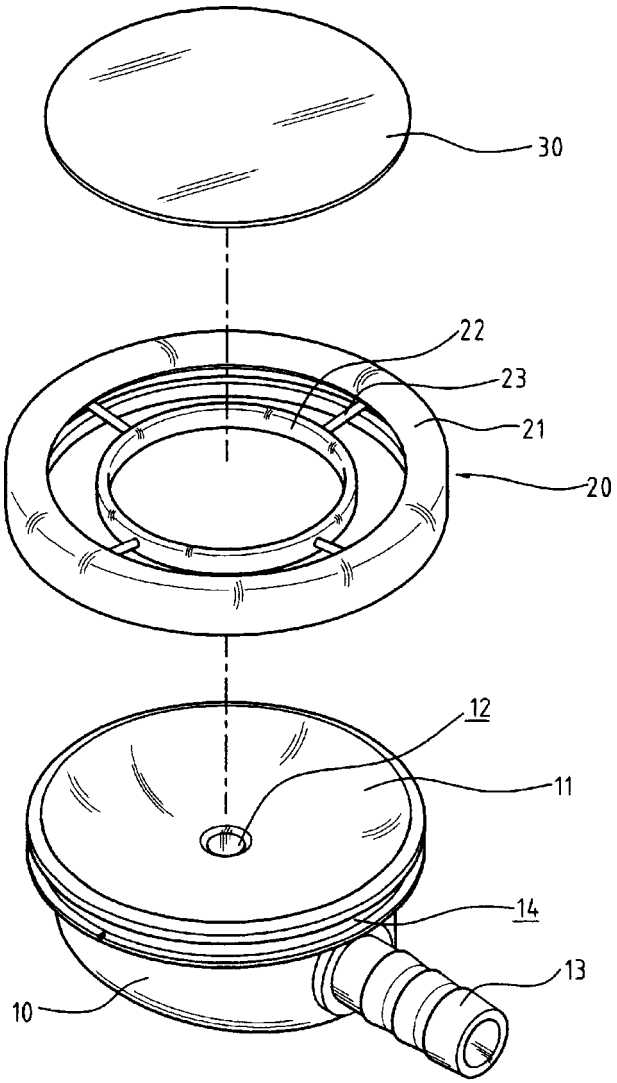
This is a computerized stethoscope attachment with the goal of disease identification. This is done by comparing the auscultated sounds with standard recorded data for cardiovascular and gastrovascular diseases. The waveforms of the auscultated stethoscope sounds are digitized and compared through a search mechanism with the standard data. The diseases are named and recorded. This information can also be displayed on appropriate monitors.—DRR

6,523,639

43.80.Qf RECEIVER STRUCTURE OF STETHOSCOPE

Woei-Kang Shieh, Taipei City, Taiwan, Province of China
25 February 2003 (Class 181/131); filed 5 September 2002

A receiver structure of a stethoscope, which can be switched back and forth between high- and low-frequency sounds, consists of a receiver body, a double hoop frame, and a membrane. A groove **14** is formed in the outer periphery of a sound-pickup cambered surface of the receiver body for buttoning up and assembling the double-hoop frame. The double hoop frame **20** consists of an outer hoop and an inner hoop, which are supported by at least two links **23**. A gap is formed between the outer hoop and the links for clamping and fixing the rim of the membrane. During use of this stethoscope, a large cavity would be formed and enclosed by the cambered surface, the outer hoop, and the membrane, which allows a user to hear low-frequency sounds via pickup duct **12**. A small cavity would be formed when the user depresses the receiver body harder to enable hearing high-frequency sounds.—DRR



6,524,244

43.80.Qf MEDICAL DIAGNOSTIC ULTRASOUND SYSTEM AND METHOD

Christopher B. Knell *et al.*, assignors to Ecton Incorporated
25 February 2003 (Class 600/437); filed 14 September 1999

The patent covers a medical diagnostic ultrasound system in which a non-real-time operating system is used to transfer image data. In a preferred embodiment, the system includes a central processing unit located outside of the received signal path to transfer image data. The non-real-time operating system can be used, e.g., in transferring image data from a hard disk to system memory for display, to transfer image data from system memory to the hard disk for storage, or to scroll image data to view a loop of image data.—DRR

6,524,249

43.80.Qf DOPPLER ULTRASOUND METHOD AND APPARATUS FOR MONITORING BLOOD FLOW AND DETECTING EMBOLI

Mark A. Moehring and Timothy R. Myers, assignors to Spentech, Incorporated
25 February 2003 (Class 600/438); filed 6 December 2000

This version of using Doppler shift phenomena to measure blood flow velocity provides an information display consisting of two simultaneously displayed graphics. One graphical display is a blood locator display that indicates locations along the axis of the ultrasonic beam at which blood flow is detected. The other graphical display constitutes a spectrogram that indicates velocities of the monitored blood flow at selected locations. The blood locator display may include a color region corresponding to the locations at which blood flow is detected. The intensity of the color may vary as a function of detected ultrasound signal amplitude or as a function of detected blood flow velocities. Embolus detection is also provided.—DRR

6,524,255

43.80.Qf ACOUSTIC COUPLING GUIDE FOR AN ULTRASONIC TRANSDUCER PROBE

Peter M. Pawluskiewicz, assignor to Koninklijke Philips Electronics N.V.
25 February 2003 (Class 600/459); filed 12 December 2000

Probes that are inserted into the body for scanning, such as transrectal and transvaginal probes, often necessitate the use of a coupling material between the transducer and the surface being examined. These probes may contain multiple transducers mounted in different positions at the distal end of the probe to obtain multiple scan planes. Coupling materials emulate body tissue and act as a standoff, which allows the transducer to focus on the surface being examined. A coupling guide is provided for controlling the size and shape of a condom-covered probe when the condom is inflated with a fluid such as water. The coupling guide can also be used to keep an organ stable while the organ is being scanned and to minimize discomfort to the patient.—DRR

6,514,203

43.80.Sh METHOD FOR ULTRASONIC CORONARY THROMBOLYSIS

Shmuel Bukshpan, assignor to Sonata Technologies Limited
4 February 2003 (Class 600/439); filed 12 February 2001

This involves a method and apparatus for achieving thrombolysis of coronary artery thromboses by means of ultrasonic energy. A thrombosis-specific ultrasound contrast agent is injected intravenously into a patient suspected of undergoing a coronary thrombosis. An array of ultrasound transducers within a pad placed on the patient's chest transmits ultrasound signals into the patient's thorax and then receives the reflected echoes. The reflected signals are analyzed to determine the ratio of second harmonic to first harmonic components in the signals. A high ratio indicates the presence of a thrombus in the field of interrogation. Time-of-flight data for each of the receiving transducers are used to compute the spatial location of the thrombus and a phased array ultrasonic transmission is then focused on the thrombus to achieve sonoparation.—DRR

6,524,251

43.80.Sh ULTRASONIC DEVICE FOR TISSUE ABLATION AND SHEATH FOR USE THEREWITH

Robert Rabiner and Bradley A. Hare, assignors to OmniSonics Medical Technologies, Incorporated
25 February 2003 (Class 600/439); filed 15 February 2001

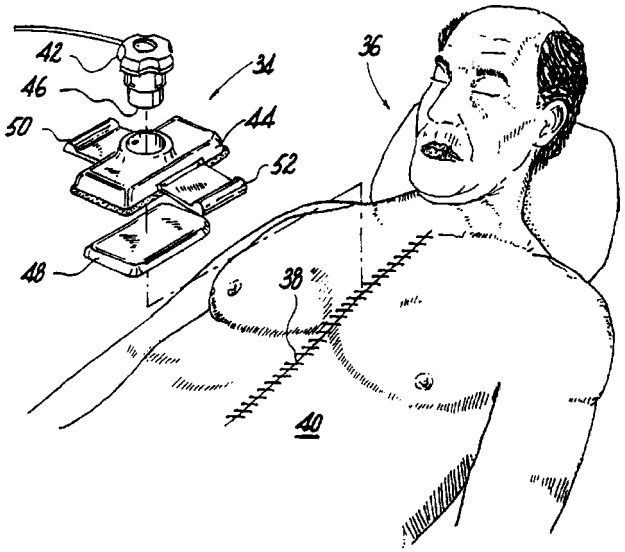
A shielded ultrasonic medical probe operates in a transverse mode to ablate and remove undesired tissue. The probe provides one or more acoustical sheaths for the purpose of controlling and focusing the energy emitted by the probe in a manner most suited to the planned medical procedure. The probe creates a cavitation area along its longitudinal length, thereby increasing the working surface of the probe.—DRR

6,524,261

43.80.Sh ULTRASOUND APPLICATION DEVICE FOR ACCELERATING STERNUM HEALING

Roger J. Talish *et al.*, assignors to Exogen, Incorporated
25 February 2003 (Class 601/2); filed 18 October 1999

This method for sternum healing of a patient includes the steps of positioning an ultrasound application device near a sternum and applying ultrasound to the area around the sternum to promote healing. The device consists of a transducer to generate ultrasound and a base for positioning the transducer near the sternum. A diverging lens may be positioned between the transducer and the sternum for acoustically diverging the ultrasound to cover



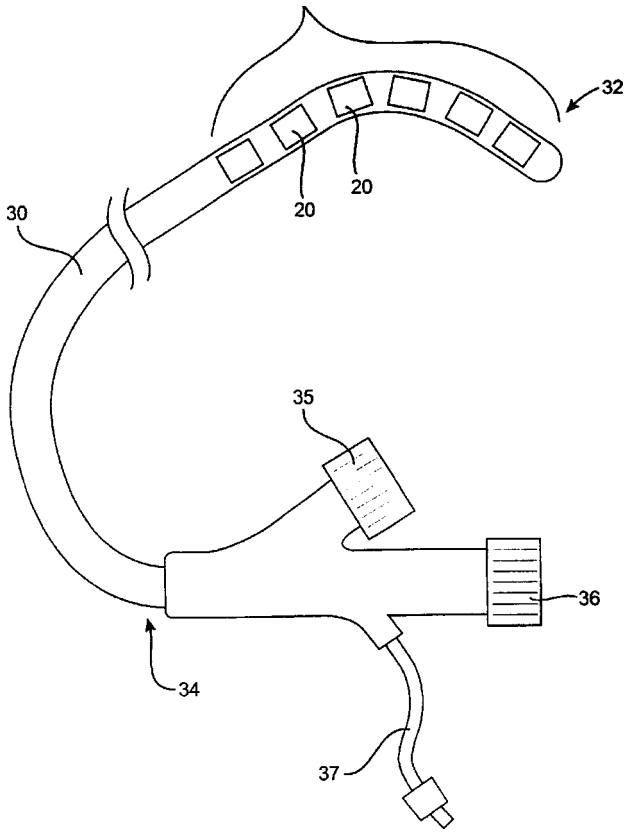
a wider area surrounding the sternum. A plurality of transducers may be included in a number of recesses of the base along the length of the sternum. A ring may also be used for positioning the transducers by securing itself around the neck of the patient. A metal strip that can be implanted within the patient adjacent to the sternum can be used to respond to signals from the sternum. A mesh implanted in the patient substantially adjacent to the sternum can also be used.—DRR

6,524,271

43.80.Sh THERAPEUTIC ULTRASOUND CATHETER FOR DELIVERING A UNIFORM ENERGY DOSE

Axel F. Brisken *et al.*, assignors to Pharmasonics, Incorporated
25 February 2003 (Class 604/22); filed 2 July 2001

This device provides a method for treating a target region in a body lumen by delivering a uniform dose of ultrasonic energy from the interior of



the lumen radially outward along the length of the lumen. A uniform dosage corresponds to ultrasonic energy producing a uniform biological effect around the circumference of the lumen. The purpose of the preferred embodiment is to provide a dosage such that the ultrasound energy received at any one point along the length of the lumen varies by no more than ± 6 dB from that received at any other point along the length.—DRR

6,514,202

43.80.Vj SYSTEMS AND METHODS FOR EVALUATING OBJECTS WITHIN AN ULTRASOUND IMAGE

Sorin Grunwald, assignor to SciMed Life Systems, Incorporated
4 February 2003 (Class 600/437); filed 23 April 2001

Statistics are used to distinguish between a scattering object and background. The statistics are calculated at selected regions and compared to a predetermined range of values that represent the object. If the values are within the range, the locations are marked to indicate that they are within the object. A border may then be drawn around the object and the area calculated.—RCW

6,514,205

43.80.Vj MEDICAL DIGITAL ULTRASONIC IMAGING APPARATUS CAPABLE OF STORING AND REUSING RADIO-FREQUENCY (RF) ULTRASOUND PULSE ECHOES

Min Hwa Lee and Tai Kyong Song, assignors to Medison Company, Limited
4 February 2003 (Class 600/443); filed in the Republic of Korea 9 February 1999

Pulse-echo rf signals from an array of transducer elements are stored and used to form an image. The stored signals are analyzed and the analysis is used to reprocess the stored signals and obtain an improved image.—RCW

6,514,208

43.80.Vj METHOD AND APPARATUS FOR POWER DOPPLER ULTRASOUND IMAGE ANALYSIS

Leopoldo C. Cancio *et al.*, assignors to The United States of America as represented by the Secretary of the Army
4 February 2003 (Class 600/454); filed 24 March 2000

Power Doppler ultrasound images of the kidney are analyzed to obtain a numerical value that correlates to blood flow in the kidney cortex.—RCW

6,515,657

43.80.Vj ULTRASONIC IMAGER

Claudio I. Zanelli, Menlo Park, California
4 February 2003 (Class 345/419); filed 11 February 2000

Cross-sectional images created from volumetric data and data from an interventional device such as a catheter are superimposed in one or more views by this imager. The interventional device position is updated continuously and superimposed on tissue images that may be updated less frequently. The system operates in a fluoroscopelike way or provides a near real-time projected image of the interventional device relative to an organ or tissue.—RCW

6,517,484

43.80.Vj ULTRASONIC IMAGING SYSTEM AND ASSOCIATED METHOD

Peter J. Wilk and Robert C. Stirbl, assignors to Wilk Patent Development Corporation
11 February 2003 (Class 600/437); filed 28 February 2000

This system includes subsystems for calibration, determination of transducer position, and formation of images. The calibration subsystem adjusts the other subsystems. For determination of position, sensors are considered as rigid bodies or as being contained in a flexible two-dimensional mat. Images may be viewed stereographically on a flat video screen or on a flexible video screen conforming to the outer surface of the subject.—RCW

6,517,485

43.80.Vj METHOD AND APPARATUS FOR PROVIDING REAL-TIME CALCULATION AND DISPLAY OF TISSUE DEFORMATION IN ULTRASOUND IMAGING

Hans Torp *et al.*, assignors to G.E. Vingmed Ultrasound AS
11 February 2003 (Class 600/438); filed 22 January 2002

A high frame rate of tissue velocity imaging or strain rate imaging is achieved by using the same ultrasonic pulses for tissue-based and Doppler-based images. The number of beams used in Doppler frames is increased to visualize tissue based only on Doppler. A sliding window is used in the processing.—RCW

6,517,486

43.80.Vj COMPOUNDING METHOD FOR REDUCING SPECKLE NOISE

Pai-Chi Li, assignor to Computed Ultrasound Global, Incorporated
11 February 2003 (Class 600/443); filed 16 August 2001

A reference image of an object is obtained and an external force is applied to deform the object. The deformed object is imaged in the same position using an estimated in-plane displacement to correct for motion. The reference and the corrected image are compounded to reduce speckle.—RCW

6,517,489

43.80.Vj METHOD AND APPARATUS FOR FORMING MEDICAL ULTRASOUND IMAGES

Patrick J. Phillips *et al.*, assignors to Acuson Corporation
11 February 2003 (Class 600/458); filed 19 April 2001

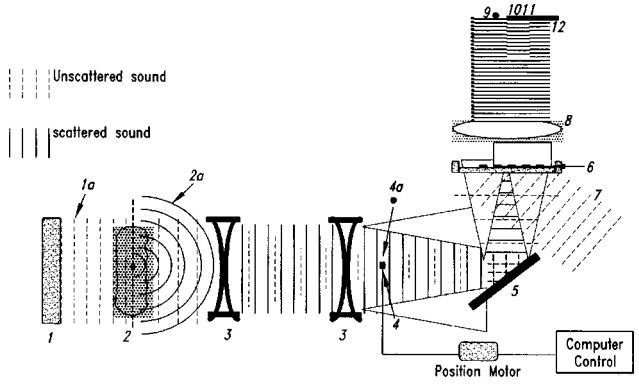
Essentially planar transmitted waves are used to improve contrast agent imaging by producing peak pressures that are relatively uniform in range. Depending on the type of contrast agent, echo frequencies of interest, and the strength of nonlinear response, multiple wavefronts can be produced at nearly the same time to increase peak pressures.—RCW

6,517,490

43.80.Vj APPARATUS AND PROCESS FOR ENHANCING IMAGING OF SUBTLE STRUCTURES

George F. Garlick, assignor to Advanced Diagnostics Systems, Incorporated
11 February 2003 (Class 600/459); filed 8 June 2000

Sound that diffracts, refracts, or scatters from the internal structure of



an object is selectively used and displayed in a holographic imaging system.—RCW

6,520,913

43.80.Vj SYSTEM FOR RAPIDLY CALCULATING EXPANSION IMAGES FROM HIGH-FREQUENCY ULTRASONIC ECHO SIGNALS

Andreas Pesavento and Helmut Ermert, assignors to Lorenz & Pesavento Ingenieurbüro für Informationstechnik
18 February 2003 (Class 600/438); filed in Germany 29 May 1998

Strain images are obtained rapidly by frequency-downshifting echo signals and iteratively calculating zero crossings.—RCW

6,520,915

43.80.Vj ULTRASOUND IMAGING SYSTEM WITH INTRINSIC DOPPLER CAPABILITY

Shengtz Lin and Hue Phung, assignors to U-Systems, Incorporated
18 February 2003 (Class 600/453); filed 24 November 2000

Phase is determined from b-mode frames and frame-to-frame changes in phase are processed to detect Doppler shift. The phase comes from baseband signal components. A threshold can be changed to eliminate low Doppler shifts. The flow image produced by this system has the same frame rate, spatial resolution, and coverage as the b-mode image.—RCW

6,520,916

43.80.Vj ULTRASOUND IMAGING SYSTEM AND METHOD FOR IMPLANTABLE AND INVASIVE DEVICES

Kenneth R. Brennen, assignor to Medtronic, Incorporated
18 February 2003 (Class 600/463); filed 2 August 2000

A guiding device such as a stylet or guidewire is coupled to a device that vibrates. The guiding device is inserted into the device to be visualized. In this configuration, longitudinal waves transferred to the device to be imaged produce a Doppler image of the implantable device.—RCW

6,524,247

43.80.Vj METHOD AND SYSTEM FOR ULTRASOUND IMAGING OF A BIOPSY NEEDLE

Danhua Zhao *et al.*, assignors to U-Systems, Incorporated
25 February 2003 (Class 600/437); filed 15 May 2001

Ultrasonic b-scan images obtained in the usual way are compounded with special b-scan frames having transmit and receive parameters modified to highlight reception of echoes from a biopsy needle. The beamwidth used for the special frames is wider in elevation than the corresponding beamwidth used to produce the other frames. The depth and orientation of a biopsy needle is determined so that imaging parameters can be adaptively changed.—RCW

6,524,252

43.80.Vj METHOD AND SYSTEM FOR GENERATING ULTRASOUND FRAMES WITH DECORRELATED SPECKLE PATTERNS AND GENERATING A COMPOUND ULTRASOUND IMAGE THEREFROM

Zengpin Yu *et al.*, assignors to U-Systems, Incorporated
25 February 2003 (Class 600/443); filed 24 November 2000

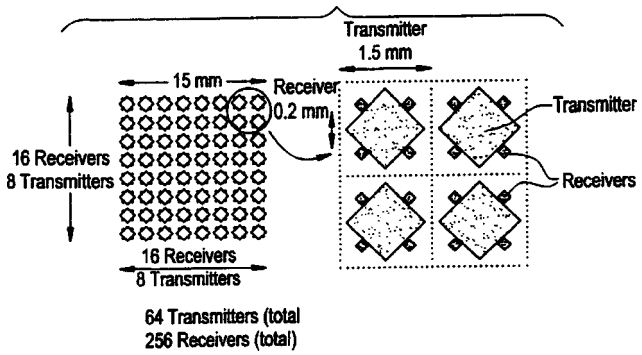
Successive sets of frames are obtained using different parameters that affect speckle. The parameters are selected to produce decorrelated speckle patterns. The frames are then compounded to produce the smoother image.—RCW

6,524,253

43.80.Vj VOLUMETRIC ULTRASOUND IMAGING WITH A THINNED ARRAY

Kenneth Abend, assignor to VueSonix Sensors Incorporated
25 February 2003 (Class 600/443); filed 20 March 2002

Each transmitter in this array illuminates a segment of a volume. Signals at receive elements are dynamically focused on intervals of the illumi-



nated segment. The spacing between the receivers is greater than one-half of the transmitted wavelength. The received beam pattern is formed so that positions of receiver grating lobes coincide with nulls of the transmit beam pattern to minimize the deleterious effects of grating lobes.—RCW

6,524,254

43.80.Vj ORTHOGONALLY RECONFIGURABLE INTEGRATED MATRIX ACOUSTICAL ARRAY

Kenneth R. Erikson, assignor to BAE Systems Information and Electronic Systems Integration, Incorporated

25 February 2003 (Class 600/447); filed 2 October 2001

The elements in this array can be switched in real time between two orthogonal multirow arrays of transducers. The multirow subarrays can be switched between vertical and horizontal strip configurations by integrated circuits attached to the subarrays. The integrated circuits perform a first level of transmit and receive beamforming for either horizontal or vertical scanning. Impedance matching in integrated circuits between transducers and cables improves signal transmission in cables from the array housing.—RCW

Primary microphone calibration system stability (L)

George S. K. Wong[a] and Lixue Wu
Acoustical Standards, Institute for National Measurement Standards, National Research Council, Ottawa, Ontario K1A 0R6, Canada

(Received 31 October 2002; revised 8 May 2003; accepted for publication 30 May 2003)

The stability of a primary microphone calibration system is investigated. In connection with the intercomparison on the calibration of type 4160 LS1P laboratory standard microphones involving the national laboratories of five countries, the pilot laboratory in Canada collected data over a period of 31 months. The measurements were made in a chamber with controlled environment at the reference condition of 23 °C, 101.325 kPa and 50% R.H. The microphone sensitivity levels were measured at seven frequencies from 125 to 8000 Hz. Over the above period the stability of the primary microphone calibration system was approximately 0.01 dB that included the stability of the transfer standard microphones.
[DOI: 10.1121/1.1593064]

PACS numbers: 43.38.Kb, 43.30.Yj [AJZ]

I. INTRODUCTION

During a relatively recent intercomparison on the calibration of laboratory standard microphones involving the national laboratories of five countries: Argentina (INTI), Brazil (INMETRO), Canada (NRC, INMS), Mexico (CENAM), and the United States (NIST), operating under the SIM (Sistema Interamericano de Metrologia) AUV. A-K1 key comparison designation, the Institute for National Measurement Standards of Canada served as the pilot laboratory. The transfer standards were a pair of Brüel & Kjær type 4160 LS1P laboratory standard microphones [serial no. (s/n) 907045 and 17434004]. A ''star'' delivery arrangement was adopted such that the pilot laboratory calibrated the microphones before and after the calibration by each of the participants. The microphones were ''hand carried'' to and from

TABLE I. Microphone sensitivity levels (in dB *re* 20 μV/Pa) measured over a period of 31 months. The mean sensitivity levels are the mean of the five measurements shown in columns 1–5.

Serial no.	Frequency (Hz)	1 Before NIST, Aug. 97	2 After NIST/ before CENAM Jan. 98	3 After CENAM/ before INMETRO Mar. 98	4 After INMETRO/ before INTI Dec. 98	5 After INTI Mar. 00	Mean sensitivity level (dB)
907045	125	−26.498	−26.498	−26.500	−26.505	−26.508	−26.502
	250	−26.518	−26.514	−26.521	−26.532	−26.530	−26.523
	500	−26.522	−26.520	−26.528	−26.533	−26.537	−26.528
	1000	−26.486	−26.481	−26.494	−26.496	−26.499	−26.491
	2000	−26.305	−26.301	−26.309	−26.315	−26.318	−26.310
	4000	−25.703	−25.701	−25.696	−25.706	−25.709	−25.703
	8000	−27.201	−27.191	−27.186	−27.176	−27.186	−27.188
1734004	125	−26.991	−26.978	−26.993	−26.989	−26.987	−26.988
	250	−27.011	−27.002	−27.007	−27.013	−27.014	−27.009
	500	−27.020	−27.005	−27.008	−27.025	−27.022	−27.016
	1000	−26.993	−26.981	−26.988	−26.993	−26.991	−26.989
	2000	−26.855	−26.845	−26.849	−26.859	−26.858	−26.853
	4000	−26.403	−26.398	−26.398	−26.404	−26.409	−26.402
	8000	−27.351	−27.353	−27.340	−27.347	−27.357	−27.350
907039	125	−26.558	−26.559	−26.557	−26.559	−26.563	−26.559
	250	−26.579	−26.577	−26.581	−26.578	−26.577	−26.578
	500	−26.580	−26.585	−26.582	−26.577	−26.576	−26.580
	1000	−26.542	−26.542	−26.541	−26.542	−26.542	−26.542
	2000	−26.376	−26.377	−26.377	−26.374	−26.379	−26.377
	4000	−25.847	−25.846	−25.844	−25.853	−25.855	−25.849
	8000	−27.284	−27.280	−27.284	−27.285	−27.301	−27.287

[a]Electronic mail: george.wong@nrc.ca

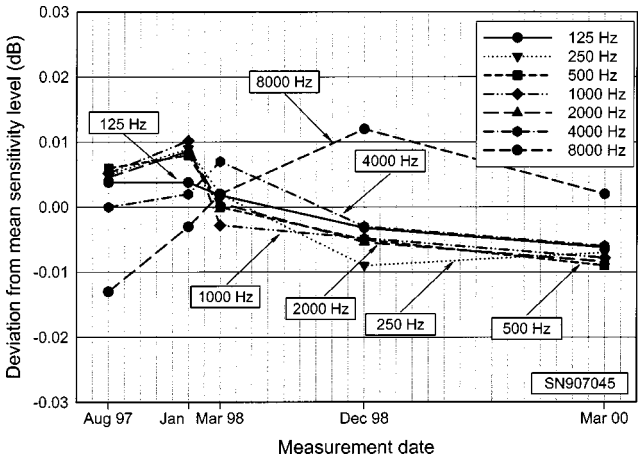


FIG. 1. Deviation from the mean sensitivity level of transfer standard microphone s/n 907045.

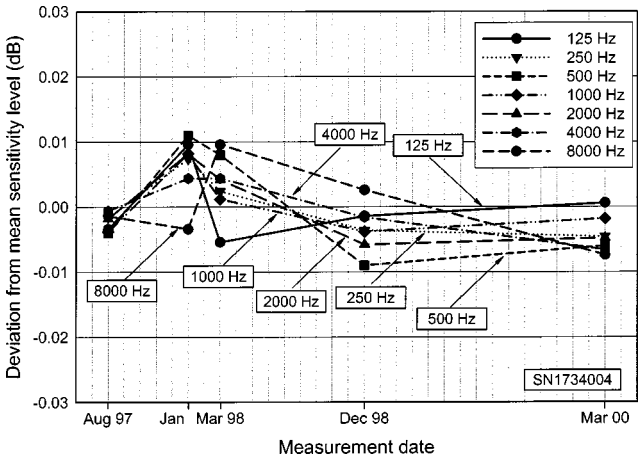


FIG. 2. Deviation from the mean sensitivity level of transfer standard microphone s/n 1734004.

the pilot laboratory. The pilot laboratory performed five sets of calibrations within a period of approximately 31 months, started in August 1997 and ended in March 2000.

II. CALIBRATION SYSTEM

The calibrations were performed in a chamber with controlled environment[1] at the reference condition of 23 °C, 101.325 kPa, and 50% R.H. The uncertainties of the environmental controls were 0.05 °C, 20 Pa, and 5% R.H., respectively. ISO standard air[2] was used for all measurements to avoid contamination with other gases, such as CO_2 that may influence the measurements. Since the reciprocity method requires three microphones, a third LS1P laboratory standard microphone (Brüel & Kjær type 4160, s/n 907039) was included to perform the calibration in accordance to the IEC international standard.[3] A commercial reciprocity calibration apparatus (Brüel & Kjær 4143EH4004) was used. Details of the instrumentation and measurement procedure for microphone calibration with the environmental controlled chamber had been described.[1]

III. MEASUREMENT RESULTS

In accordance with the protocol accepted by all the participants of the above SIM intercomparison of microphone calibration, the open-circuit sensitivity levels of the microphones were measured at seven frequencies: 125, 250, 500, 1000, 2000, 4000, and 8000 Hz. The microphone open-circuit sensitivity levels of five sets of measurements taken by the pilot laboratory are shown in Table I. Figures 1–3 show the deviations from the mean sensitivity level (the last column in Table I) of the three microphones, at each of the test frequencies, over the above period of 31 months. From Figs. 1 and 2 for transfer standard microphones s/n 907045 and s/n 1734004, respectively, the deviations are within 0.011 dB, except at 8000 Hz for microphone s/n 907045, at which the deviation is slightly larger at 0.013 dB. It is interesting to point out that for microphone s/n 907039, which remained at the pilot laboratory without the need to endure air transportation to and from the participating laboratories over the above period, the deviations shown in Fig. 3 were significantly smaller for frequencies below 8000 Hz and these deviations are within $+0.007$ and -0.006 dB. It should be noted that the deviations shown for the three microphones (Figs. 1–3) are the combined contributions from the sensitivity level stability of the microphones and the stability of the primary calibration system. Since the specifications for laboratory standard microphones[4] state that the type 4160 LSIP microphone has a long-term stability coefficient of <0.02 dB per year from 250 to 1000 Hz, and the manufacturer Brüel and Kjær indicated the stability of their LS1P microphones may have a stability coefficient of approximately 0.01 dB per year, one may conclude that the stability of the primary microphone calibration system of the pilot laboratory at INMS is excellent.

IV. CONCLUSIONS

The measurement of microphone open-circuit sensitivity levels of three LS1P laboratory standard microphones over a period of 31 months is discussed. The deviations of the sensitivity levels from the mean value are the combined contributions from the stability of the microphones and the stability of the primary calibration system. One may conclude that during the above period, the stability of the primary micro-

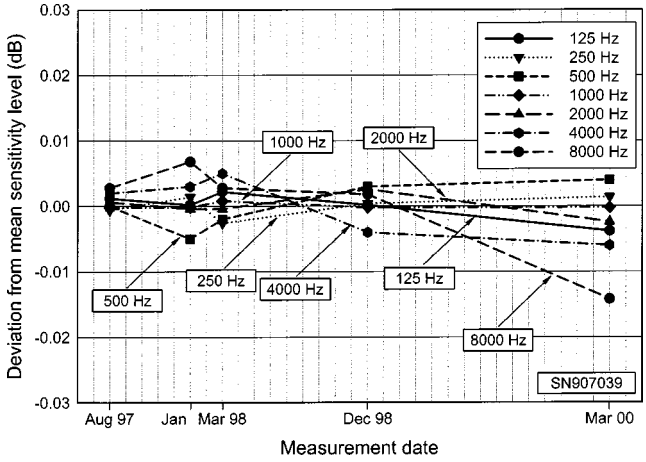


FIG. 3. Deviation from the mean sensitivity level of laboratory standard microphone s/n 907039.

phone calibration system is approximately 0.01 dB for measurements from 125 to 8000 Hz.

[1] G. S. K. Wong and L. Wu, ''Controlled environment for reciprocity calibration of laboratory standard microphones and measurement of sensitivity pressure correction,'' Metrologia **36**, 275–280 (1999).

[2] ISO 2533, *Standard Atmosphere* (International Organization for Standardization, 1975).

[3] IEC 1094-2 Ed.1 (1992), ''Measurement microphones Part 2: Primary method for pressure calibration of laboratory standard microphones by the reciprocity technique.''

[4] IEC 1094-1 Ed.2 (2000), ''Measurement microphones Part 1: Specifications for laboratory standard microphones.''

Effect of contralateral precursor type on the temporal effect in simultaneous masking with tone and noise maskers (L)

Sophie Savel and Sid P. Bacon
Psychoacoustics Laboratory, Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287-0102

(Received 19 January 2003; revised 8 May 2003; accepted 19 May 2003)

A sound (contralateral precursor) presented to the nontest ear prior to the onset of a masker and probe has been shown to reduce the temporal effect in simultaneous masking with noise maskers but not with tonal maskers. The present study examined this further. The probe was a 4.0-kHz tone. In experiment 1a, the masker was a 4.4-kHz tone and the precursor was a 4.4-kHz tone or an unmodulated (UM) or amplitude-modulated (AM) band of noise (4.4–8.0 kHz). In experiment 1b, the masker was a broadband noise and the precursor was a UM or an AM broadband noise. In both experiments the precursor consistently reduced the temporal effect for only one of the seven or eight subjects, regardless of precursor type. These largely negative results indicate that it may not be possible to use contralateral precursors to gain much insight into the mechanisms underlying temporal effects in simultaneous masking.
[DOI: 10.1121/1.1592162]

PACS numbers: 43.66.Dc, 43.66.Mk, 43.66.Ba [MRL]

I. INTRODUCTION

The threshold for a brief probe is often higher when it is presented at the beginning of a long masker than when it is presented later in time. This temporal effect has been observed with broadband noise and tonal maskers (e.g., Zwicker, 1965; Bacon and Viemeister, 1985). Although a variety of explanations has been proposed for this temporal effect (e.g., see Bacon and Smith, 1991), most recently it has been suggested that the temporal effect with broadband noise maskers could reflect the influence of the medial olivocochlear bundle on the outer hair cells (Turner and Doherty, 1997; Bacon and Liu, 2000; Zeng *et al.*, 2000). One line of supporting evidence comes from research with contralateral precursors. These are sounds that precede the masker and probe, but are presented to the nontest ear. Turner and Doherty (1997) found that a broadband precursor reduced the threshold at the onset of their broadband masker (and hence the temporal effect) by 1–11 dB in their eight (of ten) subjects with a sizable temporal effect. Bacon and Liu (2000) showed that the effectiveness of a contralateral precursor depended upon its spectral content, suggesting, among other things, that the precursor was not simply providing a timing cue for the onset of the probe. However, the contralateral precursor was effective in reducing the temporal effect with their broadband masker in only four of their 15 subjects.

It is unclear whether the same processing underlies the temporal effect with broadband and tonal maskers. To date, only one study has examined the effect of contralateral precursors on the temporal effect with tonal maskers. Bacon and Healy (2000) found that those precursors had no effect in the three subjects they tested. There are at least four possible explanations for this lack of an effect. The first is that Bacon and Healy did not test a sufficient number of subjects, given that large individual differences may be present (Bacon and Liu, 2000). The second is that the conditions were not adequate to demonstrate an effect. In particular, it could be that the possible role of the medial olivocochlear system in the psychophysical temporal effect will be observed only when probe levels are in a range where the basilar-membrane input–output functions are the most compressive (e.g., Ruggero *et al.*, 1997), that is, for levels above 40 dB SPL. In Bacon and Healy (2000), the probe level at threshold was around 35 dB SPL. The third explanation is that the tonal precursor did not produce a sufficient efferent response, given that tones are less likely to activate the efferent system than noise (Warren and Liberman, 1989). Finally, the fourth explanation is that the efferent system is simply not involved in the temporal effect with tonal maskers.

The purpose of the present study was to re-evaluate the effects of contralateral precursors on the temporal effect with tonal maskers. We extended previous work by (1) increasing the number of subjects; (2) using a shorter probe and smaller masker–probe frequency ratio to increase probe level; (3) comparing the effect of contralateral precursors for tone and noise maskers in the same group of subjects; and (4) employing a variety of precursor types, including modulated noise, which may be especially effective in eliciting an efferent response (Maison *et al.*, 1999).

II. METHODS

A. Apparatus and stimuli

All stimuli were digitally generated and produced at a 50-kHz sampling rate using a digital array processing card (TDT AP2) and digital-to-analog converter, or DAC (TDT DD1). The probe and masker were presented through one channel of the DAC. When a precursor was present, it was presented through the second channel. The output of each channel was low-pass filtered at 8 kHz (TDT FT6), attenuated (TDT PA4), and ultimately routed via a headphone buffer (TDT HB6) to a Sennheiser HD250 headphone.

The probe was a 4.0-kHz tone, 10 ms in duration, pre-

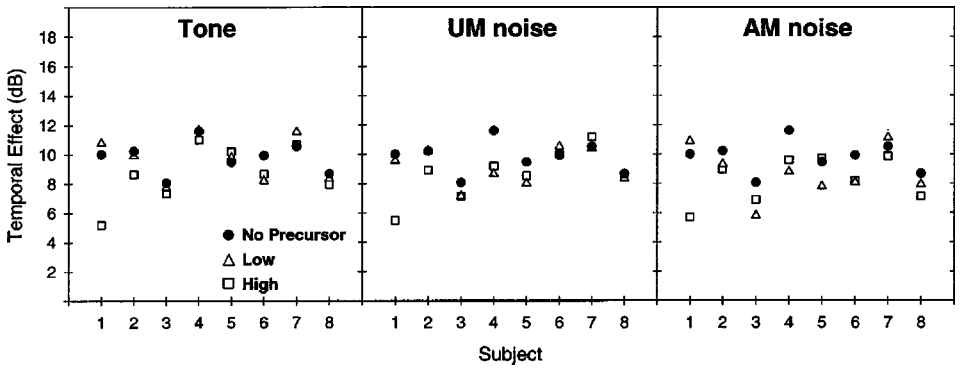


FIG. 1. Temporal effect (in dB) with a 4.4-kHz tonal masker obtained for each subject without a precursor (filled circles), with a high-level precursor (unfilled squares), and with a low-level precursor (unfilled triangles). Each of the three panels is for a different type of precursor (tone, UM noise, AM noise, respectively).

sented either at the beginning (0-ms delay) or at the temporal center (195-ms delay) of the 400-ms masker. In experiment 1a the masker was a 4.4-kHz tone presented at 80 dB SPL. The precursor was either a 4.4-kHz tone or a bandpass noise (4.4–8.0 kHz) that was unmodulated (UM) or amplitude modulated (AM). The level of the tonal precursor was 40 (low) or 80 (high) dB SPL. The spectrum level of the UM and AM precursors was −10 (low) or 30 (high) dB SPL. In experiment 1b the masker was a broadband noise (1–8000 Hz) presented at a spectrum level of 20 dB SPL. The precursor was either a UM or an AM broadband noise (1–8000 Hz), presented at a spectrum level of −20 (low) or 20 (high) dB SPL.

In both experiments the AM noise was modulated at a rate of 100 Hz and a depth of 100%, as these values have been shown to produce a maximum efferent effect (Maison *et al.*, 1999). The noises (masker and precursors) were generated in the frequency domain and transformed into the time domain via an inverse Fourier transform, using the SIGGEN software provided by TDT. A new sample of the noise was generated on each trial. The duration of the precursor (and masker) was 400 ms. The delay between the offset of the precursor and the onset of the masker was 0 ms. In the presence of a precursor, the probe was only presented at the beginning of the masker, as precursors generally have no effect on thresholds for a probe at the temporal center of a masker. All stimulus durations included 5-ms $\cos^2$ rise/fall times, and all durations and delays were determined from 0-voltage points. The level of the probe was varied adaptively via the array processor.

B. Procedure

The conditions were presented in random order. Subjects were tested in a single-walled, sound-attenuating booth located in a sound-insulated room. The masker and probe were presented to one ear, and the precursor (when present) was presented to the other ear. Thresholds were measured using an adaptive, three-interval, forced-choice procedure with a decision rule that estimates the 79.4%-correct point on the psychometric function (Levitt, 1971). A run started with the probe level 10–15 dB above the estimated threshold; it was decreased following three correct responses and increased following one incorrect response. The step size initially was 5 dB, but was decreased to 2 dB following the second reversal. Each run consisted of 12 reversals; the threshold estimate for that run was the mean level at the last 10 reversals. A run was included only if the standard deviation of the threshold estimate was 5 dB or less. Each threshold reported is the mean of at least three runs, obtained over 3 days. If the standard deviation of this mean exceeded 3 dB, an additional estimate was obtained and included in the mean. This continued until the standard deviation was less than 3 dB, or a total of six estimates was obtained and averaged. Over 95% (101/106) of the means presented here had standard deviations less than 3 dB.

C. Subjects

Eight individuals participated. They ranged in age from 20–29 years, and had thresholds of 15 dB HL or lower (ANSI, 1996) for octave frequencies from 0.5 to 8.0 kHz. Subjects had at least 2 h of practice prior to data collection. The practice consisted primarily of detecting a probe at masker onset, both with and without the contralateral precursor, as this was the most difficult condition. All subjects except S1 (author SS) were paid for their participation.

III. RESULTS

A. Experiment 1a: Tonal masker

The individual results for the tonal masker are shown in Fig. 1. The results are plotted in terms of the temporal effect (in dB)[1] obtained without a precursor (filled circles), with a high-level precursor (unfilled squares), and with a low-level precursor (unfilled triangles). Each of the three panels is for a different type of precursor (the data without a precursor are repeated across panels). In the absence of the precursor, the temporal effect ranged from 8 to about 12 dB, with a group mean of 9.8 dB. As shown in Fig. 1, there was no consistent effect of the precursor across subjects, nor was there a consistent effect of precursor type. Moreover, the effects were generally small. Indeed, the effect for S1 (for all three precursor types at the high level) was the only one in which the error bars (not shown) representing ±1 s.d. of the threshold estimates for the 0-ms delay (with and without a precursor) did not overlap. For this subject, the high-level precursor reduced the temporal effect by 5 dB.

B. Experiment 1b: Broadband noise masker

The individual results for the broadband noise masker are shown in Fig. 2. The temporal effect in the absence of a precursor (filled circles, repeated across panels) was 8–16

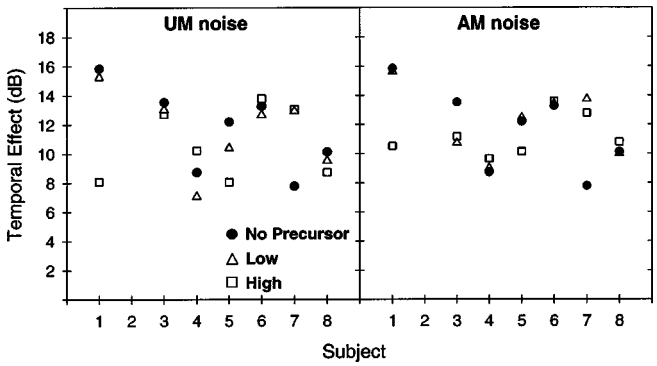


FIG. 2. Temporal effect (in dB) with a broadband noise masker obtained for each subject without a precursor (filled circles), with a high-level precursor (unfilled squares), and with a low-level precursor (unfilled triangles). The two panels are for the UM and AM noise precursor, respectively.

dB, with a group mean of 11.6 dB. There generally was no consistent effect of the precursor, except for S1 and S7. For S7, the precursor increased the temporal effect, presumably because the precursor served as a distraction. For S1, the precursor reduced the temporal effect. The reduction was about 5–7 dB, and only for the high-level precursor. For both subjects, the error bars (not shown) representing ± 1 s.d. of the threshold estimates for the 0-ms delay (with and without a precursor) did not overlap.

IV. DISCUSSION

One explanation for the temporal effect in simultaneous masking is that it reflects the time course of the medial olivocochlear system. A potentially fruitful way in which to test this explanation is to use contralateral precursors. The initial results of Turner and Doherty (1997) were particularly encouraging. They found that a contralateral precursor consistently reduced the temporal effect with a broadband masker in their eight subjects with a considerable temporal effect. Unfortunately, we have observed such a reduction in only 5 of 26 subjects in our laboratory using both broadband noise and tonal maskers (Bacon and Liu, 2000; Bacon and Healy, 2000; present study). It is unclear why we have been so much less successful in observing an effect of contralateral precursors. Our lack of success should not necessarily be taken as evidence that the efferent system is not involved in the temporal effect. Instead, it may simply reflect the fact that the medial olivocochlear system responds better to ipsilateral than contralateral stimulation (Liberman, 1988), and that there is considerable variability in efferent activity, in both animal (either under anesthesia or arousal) and human studies (Robertson and Gummer, 1985; Maison *et al.*, 1997). For example, larger contralateral evoked otoacoustic emission suppression was observed in professional musicians than in nonmusicians, suggesting that acoustic history could have an influence on the activity of the efferent system (Maison *et al.*, 1999). However, even if the temporal effect in simultaneous masking is mediated via the efferent system, our results suggest that the use of contralateral precursors may not be an especially fruitful way to study this processing.

ACKNOWLEDGMENTS

This research was supported by a grant from the National Institute on Deafness and Other Communication Disorders (NIDCD Grant No. DC01376). We thank Marjorie Leek and two anonymous reviewers for their helpful comments on an earlier version of this manuscript.

[1]The temporal effect is the difference between the thresholds at the 0-ms and 195-ms delays. Because precursors do not have an effect in long-delay conditions (Bacon and Smith, 1991; Turner and Doherty, 1997), the 195-ms delay was not run in the presence of the precursors. Thus, with precursors, the temporal effect represents the difference between the 0-ms delay with a precursor and the 195-ms delay without one.

ANSI (**1996**). ANSI S3.6-1996, ''Specifications for Audiometers'' (American National Standards Institute, New York).

Bacon, S. P., and Healy, E. W. (**2000**). ''Effects of ipsilateral and contralateral precursors on the temporal effect in simultaneous masking with pure tones,'' J. Acoust. Soc. Am. **107**, 1589–1597.

Bacon, S. P., and Liu, L. (**2000**). ''Effects of ipsilateral and contralateral precursors on overshoot,'' J. Acoust. Soc. Am. **108**, 1811–1818.

Bacon, S. P., and Smith, M. A. (**1991**). ''Spectral, intensive, and temporal factors influencing overshoot,'' Q. J. Exp. Psychol. **43A**(3), 373–399.

Bacon, S. P., and Viemeister, N. F. (**1985**). ''Simultaneous masking by gated and continuous sinusoidal maskers,'' J. Acoust. Soc. Am. **78**, 1220–1230.

Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Liberman, M. C. (**1988**). ''Response properties of cochlear efferent neurons: Monaural vs binaural stimulation and the effects of noise,'' J. Neurophysiol. **60**, 1779–1798.

Maison, S., Micheyl, C., and Collet, L. (**1999**). ''Sinusoidal amplitude modulation alters contralateral noise suppression of evoked otoacoustic emissions in humans,'' Neuroscience **91**, 133–138.

Maison, S., Micheyl, C., Chays, A., and Collet, L. (**1997**). ''Medial olivocochlear efferent system stabilizes active cochlear micromechanical properties in humans,'' Hear. Res. **113**, 89–98.

Roberston, D., and Gummer, M. (**1985**). ''Physiological and morphological characterization of efferent neurons in the guinea pig cochlea,'' Hear. Res. **20**, 63–77.

Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.

Turner, C. W., and Doherty, K. A. (**1997**). ''Temporal masking and the 'active process' in normal and hearing-impaired listeners,'' in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Mahwah, NJ), pp. 387–396.

Warren, III, E. H., and Liberman, M. C. (**1989**). ''Effects of contralateral sound on auditory-nerve responses. II. Dependence on stimulus variables,'' Hear. Res. **37**, 105–121.

Zeng, F. G., Martino, K. M., Linthicium, F. H., and Soli, S. D. (**2000**). ''Auditory perception in vestibular neurectomy subjects,'' Hear. Res. **142**, 102–112.

Zwicker, E. (**1965**). ''Temporal effects in simultaneous masking by white-noise bursts,'' J. Acoust. Soc. Am. **37**, 653–663.

Modeling the quasicompressional wave field of a rectangular transducer in a transversely isotropic solid

Dmitri Gridin[a)]
Centre for Waves and Fields, School of Engineering, South Bank University, 103 Borough Road, London SE1 0AA, United Kingdom

(Received 8 August 2002; revised 10 March 2003; accepted 13 March 2003)

A fast model for simulating the transient quasicompressional wave field of a rectangular ultrasonic transducer directly coupled to a transversely isotropic elastic half-space of general orientation is developed. The so-called two-tier asymptotic approach and the uniform stationary phase method are used to derive the high-frequency asymptotics of time-harmonic displacements. Then, transient fields are modeled by means of harmonic synthesis. In geometrical regions, the formulas involve elementary and inside boundary layers, well-known special functions (Fresnel integral and generalized Fresnel integral), and are applicable in the radiating near field. The asymptotics elucidate the physics in terms of various arrivals and give explicit dependence of the radiated waves upon model parameters. The asymptotic code is tested against a direct numerical solution. It is at least a thousand times faster but describes accurately both arrival times and amplitudes of various pulses radiated by the transducer. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1575745]

PACS numbers: 43.20.Bi, 43.20.Dk, 43.35.Zc [DEC]

I. INTRODUCTION

The problem of wave propagation in anisotropic media has attracted much attention due to its importance in many practical areas such as nondestructive evaluation (NDE), seismology, crystal optics, and magneto-hydrodynamics. In particular, many industrial materials exhibit anisotropy at typical ultrasonic inspection wavelengths. It is well known that such materials are difficult to inspect due the skewing, splitting, and focusing of ultrasonic beams.

Relatively few models of transducer wave fields in anisotropic solids have been developed so far due to the complexity of the problem. Most existing NDE models are based on full numerical schemes such as the boundary-integral equation method (Guo and Achenbach, 1995), Fourier-integral representation (Eriksson *et al.*, 2000), and the so-called elastodynamic finite integration technique (Langenberg *et al.*, 2000). Such models are extremely time-consuming and do not elucidate the physics of the problem. Several seminumerical approaches have also been developed. These include the point-source superposition (Spies, 1994; Lhémery *et al.*, 2000) and the Gaussian beam approach (Spies, 1999; Schmerr and Sedov, 2002). The latter method is very computationally efficient, but it relies on the paraxial approximation and thus is suitable for modeling only the main beam at distances not closer to the source than one-tenth of the near-field length.

As demonstrated in a number of recent publications (see Gridin and Fradkin, 1998, and references therein), high-frequency asymptotical methods may lead to fast and accurate codes for simulating transducer wave fields in isotropic solids; they also help to understand the underlying physics. This paper is the first step toward extending these models to anisotropic solids. In NDE, most frequently encountered anisotropic solids are transversely isotropic (TI); important examples are austenitic steel which is found, for instance, in claddings and welds of nuclear reactors, and fiber-reinforced composites. In this paper we concentrate primarily on TI solids.

The basis for our analysis is the so-called two-tier asymptotic approach proposed in Fradkin *et al.* (1998) for modeling the radiating near field of a uniform circular transducer directly coupled to an isotropic elastic half-space. The approach was applied to wave fields of rectangular transducers in isotropic solids in Gridin (1998). Below, we utilize a similar strategy to describe the radiating-near-quasicompressional (qP) wave field of a rectangular transducer directly coupled to a TI half-space. Note that the Gaussian beam approach has not yet been developed to deal with rectangular sources. The far-field expressions for circular and rectangular surface sources have been derived before by Spies (2002). It should be emphasized that the qP wave field considered in the present paper is not the complete transducer wave field; the latter should also include two types of quasishear (qS) fields. However, such wave fields are more difficult to model asymptotically, since the two types degenerate in certain directions and one of them may possess caustics. Although for a point-surface source these problems have been successfully dealt with in Gridin and Fradkin (2001), modeling the qS transducer wave fields is a more involved problem and will not be considered in this paper.

The paper is organized as follows: We start by describing the two-tier approach in Sec. III, and then derive the high-frequency asymptotics of the qP field of a rectangular transducer in Sec. IV. In Sec. V we describe an asymptotic

[a)]Present address: Department of Mathematics, Imperial College, London SW7 ABZ, United Kingdom; electronic mail: d.gridin@ic.ac.uk

and a direct numerical code for modeling pulse propagation, and carry out numerical tests.

II. FORMULATION OF THE PROBLEM

Let us consider the forced time-harmonic motion of a homogeneous TI elastic half-space caused by a direct contact rectangular transducer. The geometry of the problem is depicted in Fig. 1. We use the Cartesian coordinates (x_1,x_2,x_3), with the x_3 axis directed into the half-space, and the origin coinciding with the center of the loaded area. Without any loss of generality, the symmetry axis of the solid is chosen to lie in the (x_1,x_3) plane, and the angle between the axis and the x_3 axis is denoted by α. The transducer half-sides are l_1 and l_2, and the smallest angle between one of the sides and the x_1 axis is denoted by β.

As is well known (e.g., Payton, 1983), a TI solid may be characterized by density ρ and five independent elastic constants c'_{11}, c'_{12}, c'_{13}, c'_{33}, and c'_{44}, where Voigt's shortened notations are used (see, e.g., Auld, 1990). The elastic constants in (x_1,x_2,x_3) coordinates are denoted by c_{IJ} and are given in terms of c'_{IJ} in the Appendix.

The time-harmonic displacement field **u** inside the solid satisfies the equations of motion

$$(c_{ijkl}\partial_j\partial_l+\rho\omega^2\delta_{ik})u_k=0, \tag{1}$$

where ∂_j stands for $\partial/\partial x_j$, and ω is the circular frequency. The factor $\exp(-i\omega t)$ is implied but omitted here and everywhere. We choose the usual boundary conditions of prescribed tractions which are uniform over the loaded area T. In the case of a normal directly coupled transducer, these conditions are

$$\begin{aligned}&\sigma_{13}|_{x_3=0}=\sigma_{23}|_{x_3=0}=0,\\ &\sigma_{33}|_{x_3=0}=\begin{cases}-p_0, & (x_1,x_2)\in T;\\ 0, & \text{otherwise},\end{cases}\end{aligned} \tag{2}$$

where σ is the stress tensor. In the case of a shear transducer we assume for simplicity that shear tractions are aligned with one of the transducer sides, so that we have

$$\begin{aligned}&\sigma_{13}|_{x_3=0}=\begin{cases}-p_0\cos\beta, & (x_1,x_2)\in T;\\ 0, & \text{otherwise},\end{cases}\\ &\sigma_{23}|_{x_3=0}=\begin{cases}-p_0\sin\beta, & (x_1,x_2)\in T;\\ 0, & \text{otherwise},\end{cases}\\ &\sigma_{33}|_{x_3=0}=0.\end{aligned} \tag{3}$$

In Eqs. (2) and (3) p_0 is a constant and has dimension of pressure. Finally, we assume the radiation condition which allows only the outgoing waves.

III. THE TWO-TIER ASYMPTOTIC APPROACH

It is well-known that the field radiated by a transducer into a solid can be represented as an integral of the corresponding Green's function over the loaded area. This is the basis of the two-tier asymptotic approach introduced in Fradkin *et al.* (1998) for the description of the radiating near field of a uniform circular transducer directly coupled to an isotropic elastic half-space. The approach was applied to wave fields of normal rectangular transducers in isotropic solids in Gridin (1998). Below, we utilize similar strategy to describe qP wave fields in TI half-spaces.

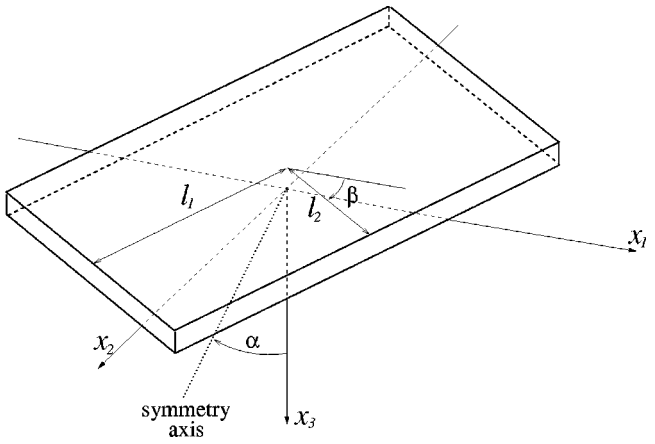


FIG. 1. The geometry of the problem.

The *first tier* involves finding the far-field asymptotics of the Green's tensor, that is solution of Lamb's problem of a point source acting on the surface of a half-space. For a TI half-space the problem has been solved in Gridin and Fradkin (2001). The leading term of the far-field asymptotic expansion of the qP wave produced by a point source acting at the origin can be written as

$$G_{ij}^{qP}(\omega,\mathbf{x})\simeq\frac{2\pi}{c'_{44}}D_{ij}(\hat{\mathbf{x}})\frac{e^{ik_0rf_1(\hat{\mathbf{x}})}}{r},\quad r\gg\lambda, \tag{4}$$

where $k_0=\omega/c_0$, $c_0=\sqrt{c'_{44}/\rho}$ is the shear wave speed along the symmetry axis, $\hat{\mathbf{x}}=\mathbf{x}/r$ is the unit vector pointing towards the observation point, and λ is a typical qP wavelength. Exact expressions for the directivity tensor D and the function f_1 can be found in Sec. 4(b) of Gridin and Fradkin (2001). We assume that the qP wavefront is a convex surface which does not possess any singular points; this case is most commonly encountered in practice.

The *second tier* involves integrating the asymptotics (4) over the loaded area T, so that we have

$$u_i^{qP(T)}(\omega,\mathbf{x})\simeq\frac{2\pi p_0}{c'_{44}}\int\int_T D_{i3}(\hat{\mathbf{d}})\frac{e^{ik_0|\mathbf{d}|f_1(\hat{\mathbf{d}})}}{|\mathbf{d}|}dx'_1\,dx'_2, \tag{5}$$

for a compressional transducer, and

$$u_i^{qP(T)}(\omega,\mathbf{x})\simeq\frac{2\pi p_0}{c'_{44}}\int\int_T[\cos\beta D_{i1}(\hat{\mathbf{d}})+\sin\beta D_{i2}(\hat{\mathbf{d}})]\frac{e^{ik_0|\mathbf{d}|f_1(\hat{\mathbf{d}})}}{|\mathbf{d}|}dx'_1\,dx'_2, \tag{6}$$

for a shear one, where we use the notations $\mathbf{d}=\mathbf{x}-\mathbf{x}'$, $\mathbf{x}'=(x'_1,x'_2,0)$ and $\hat{\mathbf{d}}=\mathbf{d}/|\mathbf{d}|$. Then, the high-frequency asymptotics of the integrals in (5) and (6) are evaluated. To simplify the presentation, below we concentrate on a compressional transducer, since analysis of Eq. (6) is similar to that of Eq. (5).

IV. THE HIGH-FREQUENCY ASYMPTOTICS

Under the near-zone condition

$$x_3 < [\min(l_1, l_2)]^2/\lambda, \tag{7}$$

the integrand in (5) contains a slowly varying amplitude and a rapidly oscillating exponential factor. It is well-known that main contributions to integrals of this type come from stationary points of phase functions, various types of critical boundary points, and singular points of amplitude functions (Bleistein and Handelman, 1986; Borovikov, 1994). Critical boundary points include points on smooth parts of the boundary where the derivative of the phase function along the boundary vanishes, as well as the boundary points such as corners, where analyticity of the boundary is violated.

Let us describe critical points of the integral in (5). To start with, the amplitude function has no singularities. The phase function, $|\mathbf{d}| f_1(\hat{\mathbf{d}})$, has one stationary point where it is at a minimum and which we denote by $\mathbf{x}^p = (x_1^p, x_2^p, 0)$. Point $\mathbf{x}^p$ is a critical point of the integral in (5) only if it lies within the integration domain, which is the case when the observation point is inside the main beam. For further convenience, let us introduce function χ^p by

$$\chi^p = \begin{cases} 1, & \mathbf{x}^p \in T; \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

Since the qP wavefront of a point source is convex, each edge (a straight segment of the loaded area) may contain no more than one edge critical point. Thus, there may be up to four edge critical points, depending on the observation point. It is useful to introduce points at which the phase function is at its minimum and which lie on straight lines containing the edges. We denote such points by $\mathbf{x}^e$ and, when necessary, use additional subscripts to indicate the edge. We also introduce χ^e, such that it is unity when $\mathbf{x}^e$ lies on the edge and zero when it lies at the same straight line but outside the edge. In general, each point $\mathbf{x}^e$ is to be found numerically.

Finally, the integration domain possesses four corner points, which we denote by $\mathbf{x}^c$. Let us now describe different types of waves which are contributions of the above critical points. Throughout, we use superscripts to distinguish the types.

A. Geometrical regions

The geometrical regions are so called because they may be modeled by using the geometrical elastodynamics, and the wave field inside them may be represented using a finite number of rays. When the radiated wave field is evaluated using asymptotic approximation of an integral of type (5), the wave field along each ray is a contribution of an isolated critical point.

1. Isolated stationary point: Plane wave

When integral (5) is evaluated asymptotically, the contribution of the isolated stationary point gives rise to a plane qP wave. However, rather than evaluating the double integral in (5) by the stationary phase method, below we find the exact expression for the wave.

Let us consider a load of infinite extent. Then, the symmetry of the problem allows us to rewrite the equations of motion (1) as a system of ordinary differential equations

$$\begin{aligned} &c_{55}\frac{\mathrm{d}^2 u_1}{\mathrm{d}x_3^2} + c_{35}\frac{\mathrm{d}^2 u_3}{\mathrm{d}x_3^2} + \rho\omega^2 u_1 = 0, \\ &c_{44}\frac{\mathrm{d}^2 u_2}{\mathrm{d}x_3^2} + \rho\omega^2 u_2 = 0, \\ &c_{35}\frac{\mathrm{d}^2 u_1}{\mathrm{d}x_3^2} + c_{33}\frac{\mathrm{d}^2 u_3}{\mathrm{d}x_3^2} + \rho\omega^2 u_3 = 0. \end{aligned} \tag{9}$$

Assuming for definiteness that the load exerts shear tractions along the x_1 axis, the boundary conditions become

$$\begin{aligned} &\left(c_{55}\frac{\mathrm{d}u_1}{\mathrm{d}x_3} + c_{35}\frac{\mathrm{d}u_3}{\mathrm{d}x_3}\right)\Bigg|_{x_3=0} = -p_0, \\ &c_{44}\frac{\mathrm{d}u_2}{\mathrm{d}x_3}\Bigg|_{x_3=0} = 0, \\ &\left(c_{35}\frac{\mathrm{d}u_1}{\mathrm{d}x_3} + c_{33}\frac{\mathrm{d}u_3}{\mathrm{d}x_3}\right)\Bigg|_{x_3=0} = 0. \end{aligned} \tag{10}$$

The solution of the boundary-value problem (9)–(10) can be found in the form

$$\mathbf{u} = \frac{p_0}{k_0 c'_{44}} \sum_{N=1}^{3} \mathbf{a}^N \mathrm{e}^{\mathrm{i}(\omega x_3/c_N + \pi/2)}, \tag{11}$$

where the qP, qSV, and qSH wave speeds are

$$\begin{aligned} &c_{1,2} = \left[\frac{1}{2\rho}\{c_{33} + c_{55} \pm \sqrt{(c_{33}-c_{55})^2 + 4c_{35}^2}\}\right]^{1/2}, \\ &c_3 = (c_{44}/\rho)^{1/2}. \end{aligned} \tag{12}$$

The dimensionless vector $\mathbf{a}^N$ is given by

$$\mathbf{a}^N = \mathrm{adj}\ M^N \left[\sum_{N'=1}^{3} B^{N'}\,\mathrm{adj}\ M^{N'}\right]^{-1} \mathbf{e}_1, \tag{13}$$

where $\mathbf{e}_1 = (1,0,0)^T$, adj M^N denotes the adjoint of matrix M^N, and matrices B^N and M^N can be found in the Appendix. When the boundary conditions specify shear tractions along the x_2 axis or else normal pressure, Eq. (13) should be modified by replacing $\mathbf{e}_1$ by $\mathbf{e}_2 = (0,1,0)^T$ or $\mathbf{e}_3 = (0,0,1)^T$, respectively.

According to the causality principle, the plane qP wave well inside the main qP beam is described by the $N=1$ term in Eq. (11). Thus, we have

$$\mathbf{u}^{qP(p)}(\omega, \mathbf{x}) = \chi^p \frac{p_0}{k_0 c'_{44}} \mathbf{a}^1 \mathrm{e}^{\mathrm{i}(\omega x_3/c_1 + \pi/2)} \equiv \chi^p \overline{\mathbf{u}}^{qP(p)}(\omega, \mathbf{x}), \tag{14}$$

where (p) stands for plane wave. The phase velocity of the plane qP wave is normal to the surface, while the group velocity is generally not; this leads to skewing of the main qP beam. This situation is illustrated in Fig. 2. The group velocity vector lies in the (x_1, x_3) plane, and the skewing angle ψ is given by

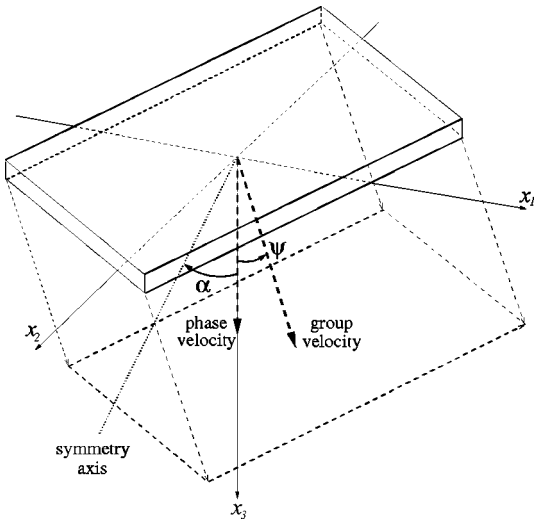


FIG. 2. Skewing of the main beam.

$$\psi = \tan^{-1}\frac{s_1'(\alpha)}{s_1(\alpha)}, \tag{15}$$

where s_1 is the qP slowness given by

$$s_1(\theta) = \{\tfrac{1}{2}[(\hat{c}_{11}'+1)\sin^2\theta + (\hat{c}_{33}'+1)\cos^2\theta + \sqrt{D}]\}^{-1/2}, \tag{16}$$

with

$$D = [(\hat{c}_{11}'-1)\sin^2\theta + (1-\hat{c}_{33}')\cos^2\theta]^2 + 4(\hat{c}_{13}'+1)^2\sin^2\theta\cos^2\theta. \tag{17}$$

[If ψ determined from (15) is positive (negative), the group velocity vector makes a blunt (acute) angle with the x_1 axis.]

Note that for a normal transducer $u_2^{qP(p)} = 0$ (because $a_2^1 = 0$ due to symmetry). However, in contrast to the isotropic case $u_1^{qP(p)} \neq 0$.

2. Isolated edge point: Edge wave

When integral (5) is evaluated asymptotically, the contribution of an isolated edge point gives rise to an edge qP wave. Utilizing Eq. (4.10) from Borovikov (1994), for the leading term of an edge wave we obtain

$$u_i^{qP(e)}(\omega,\mathbf{x}) \simeq \chi^e \frac{(2\pi)^{3/2} p_0}{k_0^{3/2} c_{44}'} \frac{D_{i3}(\hat{\mathbf{d}}^e)}{\sqrt{\partial_l^2(|\mathbf{d}|f_1)}\,\partial_n(|\mathbf{d}|f_1)} \times \frac{e^{i[k_0|\mathbf{d}^e|f_1(\hat{\mathbf{d}}^e)+3\pi/4]}}{|\mathbf{d}^e|} \equiv \chi^e \overline{u}_i^{qP(e)}(\omega,\mathbf{x}), \tag{18}$$

where $\mathbf{d}^e = \mathbf{x} - \mathbf{x}^e$, $\partial_l^2(|\mathbf{d}|f_1)$ denotes the second derivative of the phase function along the edge, and $\partial_n(|\mathbf{d}|f_1)$, the first derivative along the inward normal to the edge; both derivatives are taken at $\mathbf{x}^e$. The superscript (e) indicates the edge wave. Equation (18) demonstrates that the edge waves are asymptotically smaller (the order 1/2 smaller) than the plane wave.

3. Isolated corner point: Corner wave

The contribution to integral (5) of a corner point gives rise to a corner qP wave. Utilizing Eq. (4.12) from Borovikov (1994), for the leading term of the corner wave we obtain

$$u_i^{qP(c)}(\omega,\mathbf{x}) \simeq -\frac{2\pi p_0}{k_0^2 c_{44}'} \frac{D_{i3}(\hat{\mathbf{d}}^c)}{\partial_{n_1}(|\mathbf{d}|f_1)\,\partial_{n_2}(|\mathbf{d}|f_1)} \frac{e^{ik_0|\mathbf{d}^c|f_1(\hat{\mathbf{d}}^c)}}{|\mathbf{d}^c|}, \tag{19}$$

where $\mathbf{d}^c = \mathbf{x} - \mathbf{x}^c$, and $\partial_{n_1}(|\mathbf{d}|f_1)$ and $\partial_{n_2}(|\mathbf{d}|f_1)$ denote the first derivatives along the inward normals to the two edges joining at the corner; both derivatives are taken at $\mathbf{x}^c$. The superscript (c) indicates the corner wave. Equation (19) demonstrates that the corner waves are asymptotically smaller (the order 1 smaller) than the plane wave.

B. Boundary layers

The representation of the total qP field as a sum of the plane, edge, and corner waves is not applicable inside *boundary layers* between the geometrical zones. These are transitional regions surrounding lines or surfaces along which rays of different nature coalesce. Wave fields inside boundary layers are described in terms of asymptotic series which involve special functions rather than inverse powers of dimensionless frequency as ray series do. The size and shape of a boundary layer are both determined by the condition that the phase difference between rays of different nature is small (the usual criterion is that it is π). The higher the frequency is, the narrower the boundary layers. When the wave field is evaluated using asymptotic approximation of an integral of type (5), this situation corresponds to coalescence of two or more critical points.

1. Coalescence of the stationary and edge point: Edge penumbra

The first type of a penumbral boundary layer occurs as $\mathbf{x}^e \to \mathbf{x}^p$, i.e., when the phase difference between the plane and an edge wave tends to zero. Under this condition, expression (14) has a discontinuity and expression (18) tends to infinity. Mathematically, this situation corresponds to the case of coalescing stationary and edge critical point (Borovikov, 1994, Sec. 4.2.4). The leading term of the asymptotics which are applicable in this *edge* penumbral region involves the Fresnel integral

$$F(\xi) \simeq \frac{e^{-i\pi/4}}{\sqrt{\pi}} \int_{-\infty}^{\xi} e^{is^2} ds \tag{20}$$

[Borovikov, 1994, Eq. (6.1a)] and we obtain

$$\mathbf{u}^{qP(\mathrm{EP})} = \chi^e F[(2\chi^p - 1)\sqrt{k_0(|\mathbf{d}^e|f_1(\hat{\mathbf{d}}^e) - x_3/c_1)}]\,\overline{\mathbf{u}}^{qP(p)}, \tag{21}$$

where superscript (EP) indicates the edge penumbra. The size of the edge penumbral layer is determined by the condition that the phase difference between the plane and edge wave is π, that is, we have

$$k_0(|\mathbf{d}^e|f_1(\hat{\mathbf{d}}^e) - x_3/c_1) = \pi. \tag{22}$$

Since $u_2^{qP(p)}=0$, Eq. (21) gives $u_2^{qP(\mathrm{EP})}=0$, and therefore the next term in the asymptotic expansion for this component should be used. It coincides with the edge wave

$$u_2^{qP(\mathrm{EP})}=u_2^{qP(e)}. \tag{23}$$

2. Coalescence of a corner and edge point: Corner penumbra

The second type of penumbral boundary layer occurs as $\mathbf{x}^e\to\mathbf{x}^c$, i.e., when the phase difference between an edge and corner wave tends to zero. Under this condition, expression (18) has a discontinuity and expression (19) tends to infinity. Mathematically, the situation corresponds to the case of coalescing edge critical and corner point (Borovikov, 1994, Sec. 4.2.2). The leading term of the asymptotics which are applicable in this *corner* penumbral region also involves the Fresnel integral, and we have

$$\mathbf{u}^{qP(\mathrm{CP})}\simeq F[(2\chi^e-1)\sqrt{k_0(|\mathbf{d}^c|f_1(\hat{\mathbf{d}}^c)-|\mathbf{d}^e|f_1(\hat{\mathbf{d}}^e))}]\overline{\mathbf{u}}^{qP(e)}, \tag{24}$$

where superscript (CP) indicates the corner penumbra. The size of the corner penumbral layer is determined by the condition that the phase difference between the edge and corner wave is π, that is, we have

$$k_0(|\mathbf{d}^c|f_1(\hat{\mathbf{d}}^c)-|\mathbf{d}^e|f_1(\hat{\mathbf{d}}^e))=\pi. \tag{25}$$

3. Coalescence of the stationary and corner point: Edge-corner penumbra

The third type of penumbral boundary layer occurs when $\mathbf{x}^c\to\mathbf{x}^p$, i.e., when the phase difference between the plane and a corner wave tends to zero. This means that $\mathbf{x}^{e_1}\to\mathbf{x}^p$ and $\mathbf{x}^{e_2}\to\mathbf{x}^p$ as well, where $\mathbf{x}^{e_1}$ and $\mathbf{x}^{e_2}$ relate to the two edges joining at the corner. Under this condition expression (21) has a discontinuity and expression (24) tends to infinity. Mathematically, this situation corresponds to the case of coalescing stationary and corner points (Borovikov, 1994, Sec. 4.4). The leading term of the asymptotics which are applicable in this *edge-corner* penumbral region is expressed in terms of the generalized Fresnel integral

$$G(\eta,\nu)=\frac{\nu}{2\pi}\int_\eta^\infty \frac{e^{i(\xi^2+\nu^2)}}{\xi^2+\nu^2}d\xi \tag{26}$$

[Borovikov, 1994, Eq. (6.35)] and we obtain

$$\begin{aligned}\mathbf{u}^{qP(\mathrm{ECP})}\simeq\overline{\mathbf{u}}^{qP(p)}\{&\chi^p-G[-(2\chi^{e_1}-1)\sqrt{k_0(|\mathbf{d}^c|f_1(\hat{\mathbf{d}}^c)-|\mathbf{d}^{e_1}|f_1(\hat{\mathbf{d}}^{e_1}))},(2\chi^p-1)\sqrt{k_0(|\mathbf{d}^{e_1}|f_1(\hat{\mathbf{d}}^{e_1})-x_3/c_1)}]\\&-G[-(2\chi^{e_2}-1)\sqrt{k_0(|\mathbf{d}^c|f_1(\hat{\mathbf{d}}^c)-|\mathbf{d}^{e_2}|f_1(\hat{\mathbf{d}}^{e_2}))},(2\chi^p-1)\sqrt{k_0(|\mathbf{d}^{e_2}|f_1(\hat{\mathbf{d}}^{e_2})-x_3/c_1)}]\},\end{aligned} \tag{27}$$

where superscript (ECP) indicates the edge-corner penumbra. The size of the edge-corner penumbral layer is determined by the condition that the phase difference between the plane and corner wave is π, that is, we have

$$k_0(|\mathbf{d}^c|f_1(\hat{\mathbf{d}}^c)-x_3/c_1)=\pi. \tag{28}$$

Equation (27) gives $u_2^{qP(\mathrm{ECP})}=0$, and therefore we use the next term in the asymptotic expansion for this component

$$\begin{aligned}u_2^{qP(\mathrm{ECP})}=&F[(2\chi^{e_1}-1)\\&\times\sqrt{k_0(|\mathbf{d}^c|f_1(\hat{\mathbf{d}}^c)-|\mathbf{d}^{e_1}|f_1(\hat{\mathbf{d}}^{e_1}))}]\overline{u}_2^{qP(e_1)}\\&+F[(2\chi^{e_2}-1)\\&\times\sqrt{k_0(|\mathbf{d}^c|f_1(\hat{\mathbf{d}}^c)-|\mathbf{d}^{e_2}|f_1(\hat{\mathbf{d}}^{e_2}))}]\overline{u}_2^{qP(e_2)}.\end{aligned} \tag{29}$$

V. PULSE PROPAGATION

In practical applications, one is interested in the propagation of pulses rather than the time-harmonic displacement. Therefore, let us formulate the following initial-boundary value problem for the time-dependent displacement $\mathbf{U}(t,\mathbf{x})$. It satisfies the equations of motion

$$\partial_j\sigma_{ij}-\rho\partial_t^2U_i=0, \tag{30}$$

and is subject to the boundary conditions

$$\begin{aligned}&\sigma_{13}|_{x_3=0}=\sigma_{23}|_{x_3=0}=0,\\&\sigma_{33}|_{x_3=0}=\begin{cases}-P_0(t), & (x_1,x_2)\in T;\\ 0, & \text{otherwise},\end{cases}\end{aligned} \tag{31}$$

where $P_0(t)$ is a pressure input function. The half-space is assumed to be at rest prior to $t=0$, and the boundedness condition at infinity is imposed.

The displacement $\mathbf{U}(t,\mathbf{x})$ can be obtained from the time-harmonic displacement $\mathbf{u}(\omega,\mathbf{x})$ by using the inverse Fourier transform

$$\mathbf{U}(t,\mathbf{x})=\mathrm{Re}\frac{1}{\pi}\int_0^\infty \mathbf{u}(\omega,\mathbf{x})p_0(\omega)\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}\omega, \tag{32}$$

where $p_0(\omega)$ is the Fourier transform of $P_0(t)$. Alternatively, the convolution theorem gives

$$\mathbf{U}(t,\mathbf{x})=P_0(t)*\mathrm{Re}\frac{1}{\pi}\int_0^\infty \mathbf{u}(\omega,\mathbf{x})\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}\omega, \tag{33}$$

where the star denotes the time convolution.

In numerical comparisons below, again, having practical applications in mind, we consider the particle velocity $\mathbf{V}(t,\mathbf{x})$, that is, the first derivative of the displacement with respect to time. From Eqs. (32) and (33) we obtain

$$\mathbf{V}(t,\mathbf{x})=\mathrm{Im}\frac{1}{\pi}\int_0^\infty \mathbf{u}(\omega,\mathbf{x})p_0(\omega)\omega\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}\omega, \tag{34}$$

and

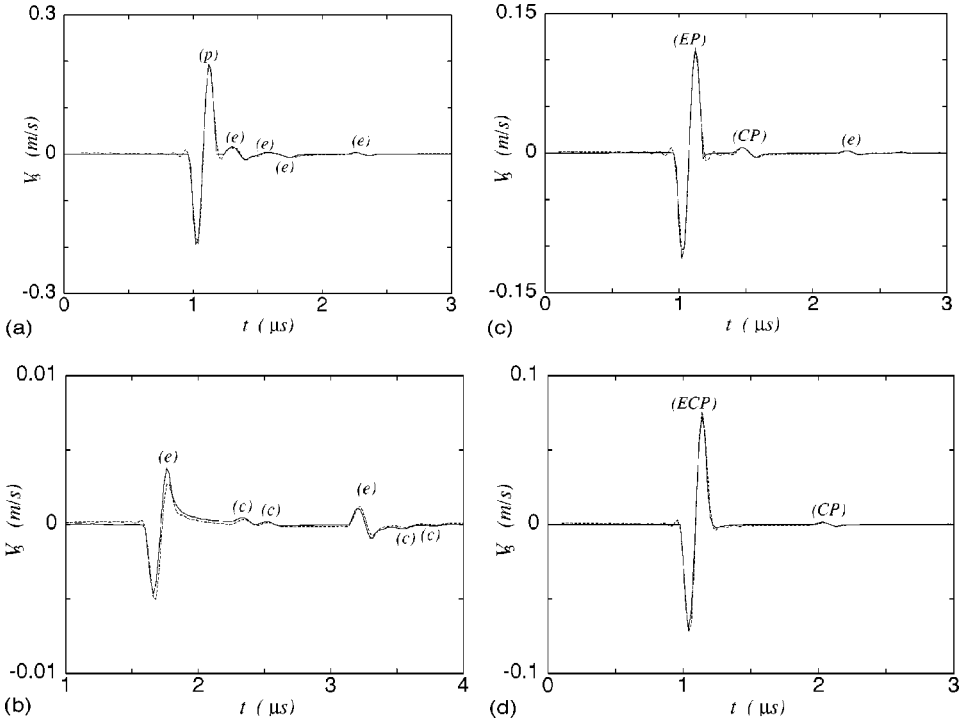


FIG. 3. Comparisons of the asymptotic (dashed lines) and direct numerical (solid lines) particle velocity component V_3 at various observation points; (a) $\mathbf{x}=(0,3,6)$ mm; (b) $\mathbf{x}=(10,7,6)$ mm; (c) $\mathbf{x}=(7,0,6)$ mm; (d) $\mathbf{x}=(-1,11,6)$ mm. Descriptors (p), (e), (c), (EP), (CP), and (ECP) indicate the plane, edge, corner, edge-penumbral, corner-penumbral and edge-corner-penumbral arrivals, respectively.

$$\mathbf{V}(t,\mathbf{x})=P_0'(t)*\mathrm{Re}\frac{1}{\pi}\int_0^\infty \mathbf{u}(\omega,\mathbf{x})\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}\omega. \tag{35}$$

A. Direct numerical solution

Let us first derive the integral representation of the time-harmonic transducer qP wave field. Applying the Fourier transform technique to the problem (1–2), we obtain

$$\mathbf{u}^{qP(T)}(\omega,\mathbf{x})$$
$$\simeq \frac{\mathrm{i}k_0}{c_{44}'}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{4\sin(k_0 l_1\hat{\xi}_\perp\sin\beta')\sin(k_0 l_2\hat{\xi}_\perp\cos\beta')}{k_0^2\hat{\xi}_\perp^2\sin\beta'\cos\beta'}$$
$$\times\hat{\mathbf{a}}^1(\hat{\xi}_1,\hat{\xi}_2)\nu_1(\hat{\xi}_1,\hat{\xi}_2)\mathrm{e}^{\mathrm{i}k_0 r f_1(\hat{\xi}_1,\hat{\xi}_2)}\mathrm{d}\hat{\xi}_1\,\mathrm{d}\hat{\xi}_2, \tag{36}$$

where we have introduced the notation

$$\beta'=\beta-\cos^{-1}(\hat{\xi}_1/\hat{\xi}_\perp),\quad \hat{\xi}_\perp=(\hat{\xi}_1^2+\hat{\xi}_2^2)^{1/2}. \tag{37}$$

The derivation of Eq. (36) is similar to that of Eq. (4.22) in Gridin and Fradkin (2001), and the only difference is the factor

$$\frac{4\sin(k_0 l_1\hat{\xi}_\perp\sin\beta')\sin(k_0 l_2\hat{\xi}_\perp\cos\beta')}{k_0^2\hat{\xi}_\perp^2\sin\beta'\cos\beta'}$$

in Eq. (36), which is due to the rectangular shape of the loaded area. [Explicit expressions for $\hat{\mathbf{a}}^1$ and f_1 and a description of ν_1 can be found in Gridin and Fradkin (2001).]

A numerical method based on the inverse Fourier transform of the double-integral representation (36) would be extremely time-consuming, since at high frequencies the integrands, whose evaluation requires considerable time in itself, are rapidly oscillating. Instead, we use the second approach based on the convolution equation (35). Substituting (36) into (35) and taking the integral with respect to ω, we arrive at

$$\mathbf{V}^{\mathrm{qP}(T)}(t,\mathbf{x})\simeq-\frac{c_0}{c_{44}'}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\hat{\mathbf{a}}^1(\hat{\xi}_1,\hat{\xi}_2)\chi^1(\hat{\xi}_1,\hat{\xi}_2)}{\hat{\xi}_\perp^2\sin\beta'\cos\beta'}$$
$$\times\{p_0(t-t_1)-p_0(t-t_2)-p_0(t-t_3)$$
$$+p_0(t-t_4)\}\mathrm{d}\hat{\xi}_1\,\mathrm{d}\hat{\xi}_2, \tag{38}$$

where the following notations have been introduced:

$$\begin{aligned}
t_1&=[rf_1(\hat{\xi}_1,\hat{\xi}_2)+l_1\hat{\xi}_\perp\sin\beta'+l_2\hat{\xi}_\perp\cos\beta']/c_0,\\
t_2&=[rf_1(\hat{\xi}_1,\hat{\xi}_2)-l_1\hat{\xi}_\perp\sin\beta'+l_2\hat{\xi}_\perp\cos\beta']/c_0,\\
t_3&=[rf_1(\hat{\xi}_1,\hat{\xi}_2)+l_1\hat{\xi}_\perp\sin\beta'-l_2\hat{\xi}_\perp\cos\beta']/c_0,\\
t_4&=[rf_1(\hat{\xi}_1,\hat{\xi}_2)-l_1\hat{\xi}_\perp\sin\beta'-l_2\hat{\xi}_\perp\cos\beta']/c_0.
\end{aligned} \tag{39}$$

The new integrand in Eq. (38) is not rapidly oscillating. Nevertheless, numerical evaluation of the double integral still takes a long time.

B. Numerical results

To carry out numerical tests we use the following material parameters: $\rho=7900\ \mathrm{kg\,m^{-3}}$, $c_{11}'=263$ GPa, $c_{12}'=98$ GPa, $c_{13}'=145$ GPa, $c_{33}'=216$ GPa, and $c_{44}'=129$ GPa (Ogilvy, 1986). The angle between the normal to the surface and the TI symmetry axis is $\alpha=30°$. The transducer half-sides are $l_1=10$ mm and $l_2=5$ mm, and we choose $\beta=30°$ as well. A wideband pulse containing high and intermediate frequencies is considered: $p_0(t)$ is one cycle of sine of frequency $f=5$ MHz and has amplitude $p_0=10$ MPa.

In Fig. 3 we compare outputs of the direct numerical and asymptotic codes. The direct numerical code is based on Eq. (38). The asymptotic code performs the fast Fourier transform (34) of the high-frequency asymptotic approximation derived in Sec. IV.

Figure 3(a) shows the train of pulses at point $\mathbf{x}=(0,3,6)$ mm, which is well inside the main beam. The first pulse is due to the plane wave and dominates. The other four pulses are due to the edge waves and are much smaller than the first. The corner pulses are too small to be discernible. Figure 3(b) shows the train of pulses at point $\mathbf{x}=(10,7,6)$ mm, which is well outside the main beam. The two largest pulses are due to the edge waves; the other four pulses are the corner arrivals. Note that the largest pulse is about 40 times smaller than the plane arrival inside the main beam; the corner arrivals are smaller still and in practical applications can be neglected.

Figure 3(c) shows the train of pulses at point $\mathbf{x}=(7,0,6)$ mm, which is close to the boundary of the main beam well away from the corners, that is inside an edge penumbra. The first, dominant pulse is due to the edge-penumbral arrival. The amplitude of the pulse is about half of the amplitude of the plane pulse; it would be half directly at the boundary of the main beam. The other discernible pulses are due to a corner-penumbral and edge arrival. Figure 3(d) shows the train of pulses at point $\mathbf{x}=(-1,11,6)$ mm, which is close to a corner of the main beam, that is inside an edge-corner penumbra. The first pulse is due to the edge-corner-penumbral arrival, with the amplitude of about one-third of the plane pulse; it would be one-fourth of the plane pulse directly at the corner of the main beam. The other discernible pulse is due to a corner-penumbral arrival.

Figure 3 demonstrates that the high-frequency asymptotics describe accurately both arrival times and amplitudes of various waves generated by the transducer. For the examples shown, the asymptotic code runs a few thousand times faster than the direct numerics, with 1 second being a typical run time on an HP9000/785 workstation and the HP FORTRAN 90 compiler being used.

VI. CONCLUSIONS

We have developed a computationally efficient model for simulating the transient qP wave field of a rectangular ultrasonic transducer directly coupled to a transversely isotropic elastic half-space of general orientation. The two-tier asymptotic approach and the uniform stationary phase method have been used to derive the high-frequency asymptotics of time-harmonic displacements. The formulas involve geometrical regions elementary and inside boundary layers, well-known special functions (Fresnel integral and generalized Fresnel integral). The results are applicable in the radiating near field, that is, the near field with the evanescent wave zone excluded. The far-field formulas are much simpler and have been obtained before in Spies (2002).

The asymptotic solutions elucidate the physics in terms of arrivals of the plane, edge, and corner waves, and give explicit dependence of the radiated waves upon model parameters. The edge waves are much smaller than the plane wave, and the corner waves are smaller still, but only they are present inside certain zones. The asymptotic code has been tested against a direct numerical scheme. It is at least a thousand times faster but describes accurately both arrival times and amplitudes of various qP pulses radiated by the transducer; thus, in contrast to the Gaussian beam approach, the method is applicable outside the main beam as well. The above results may be easily extended to other types of anisotropy, as long as the far-field asymptotics of the qP Green's tensor similar to Eq. (4) are established.

ACKNOWLEDGMENTS

The work has been carried out under the EPSRC Grant GR/M31552. I thank Professor L. Ju. Fradkin for helpful suggestions.

APPENDIX: ELASTIC CONSTANTS AND MATRICES M^N AND B^N

The elastic constants c_{IJ} in the coordinate system $(\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3)$ are obtained from the five independent elastic constants c'_{IJ} via simple transformations (see Auld, 1990, Chap. III, Sec. D). The nonzero constants are

$$
\begin{aligned}
&c_{11}=\cos^4\alpha c'_{11}+2\cos^2\alpha\sin^2\alpha c'_{13}+\sin^4\alpha c'_{33}+\sin^2 2\alpha c'_{44},\\
&c_{12}=\cos^2\alpha c'_{12}+\sin^2\alpha c'_{13},\\
&c_{13}=\cos^2\alpha\sin^2\alpha(c'_{11}+c'_{33})+(\sin^4\alpha+\cos^4\alpha)c'_{13}\\
&\qquad-\sin^2 2\alpha c'_{44},\\
&c_{15}=0.5\sin 2\alpha(\cos^2\alpha c'_{11}-\cos 2\alpha c'_{13}-\sin^2\alpha c'_{33})\\
&\qquad-\sin 2\alpha\cos 2\alpha c'_{44},\\
&c_{22}=c'_{11},\quad c_{23}=\sin^2\alpha c'_{12}+\cos^2\alpha c'_{13},\\
&c_{25}=0.5\sin 2\alpha(c'_{12}-c'_{13}),\\
&c_{33}=\sin^4\alpha c'_{11}+2\cos^2\alpha\sin^2\alpha c'_{13}+\cos^4\alpha c'_{33}+\sin^2 2\alpha c'_{44},\\
&c_{35}=0.5\sin 2\alpha(\sin^2\alpha c'_{11}+\cos 2\alpha c'_{13}-\cos^2\alpha c'_{33})\\
&\qquad+\sin 2\alpha\cos 2\alpha c'_{44},\\
&c_{44}=\cos^2\alpha c'_{44}+\sin^2\alpha c'_{66},\quad c_{46}=0.5\sin 2\alpha(c'_{66}-c'_{44}),\\
&c_{55}=0.25\sin^2 2\alpha(c'_{11}-2c'_{13}+c'_{33})+\cos^2 2\alpha c'_{44},\\
&c_{66}=\sin^2\alpha c'_{44}+\cos^2\alpha c'_{66},
\end{aligned}
\tag{A1}
$$

where $c'_{66}=(c'_{11}-c'_{12})/2$. In the body of the paper, we also use the dimensionless elastic constants

$$
\hat{c}_{IJ}=c_{IJ}/c'_{44},\quad \hat{c}'_{IJ}=c'_{IJ}/c'_{44}. \tag{A2}
$$

The matrices M^N and B^N are given by

$$
M^N=\begin{bmatrix}
-\hat{c}_{55}(\hat{c}_N)^{-2}+1 & 0 & -\hat{c}_{35}(\hat{c}_N)^{-2}\\
0 & -\hat{c}_{44}(\hat{c}_N)^{-2}+1 & 0\\
-\hat{c}_{35}(\hat{c}_N)^{-2} & 0 & -\hat{c}_{33}(\hat{c}_N)^{-2}+1
\end{bmatrix}, \tag{A3}
$$

and

$$B^N=\begin{bmatrix} \hat{c}_{55}(\hat{c}_N)^{-1} & 0 & \hat{c}_{35}(\hat{c}_N)^{-1} \\ 0 & \hat{c}_{44}(\hat{c}_N)^{-1} & 0 \\ \hat{c}_{35}(\hat{c}_N)^{-1} & 0 & \hat{c}_{33}(\hat{c}_N)^{-1} \end{bmatrix}, \tag{A4}$$

respectively, where we use dimensionless wave speeds

$$\hat{c}_N=c_N/c_0, \tag{A5}$$

and c_N are given by Eq. (12).

Auld, B. A. (**1990**). *Acoustic Fields and Waves in Solids*, 2nd ed. (Krieger, Melbourne, FL), Vol. I.

Bleistein, N., and Handelman, R. A. (**1986**). *Asymptotic Expansions of Integrals* (Dover, New York).

Borovikov, V. A. (**1994**). *Uniform Stationary Phase Method*, IEE Electromagnetic Series, Vol. 40 (IEE, London).

Eriksson, A. S., Mattsson, J., and Niklasson, A. J. (**2000**). ''Modeling of ultrasonic crack detection in anisotropic materials,'' NDT & E Int. **33**, 441–451.

Fradkin, L. J., Kiselev, A. P., and Krylova, E. (**1998**). ''The radiating near-field asymptotics of a time-harmonic circular normal ultrasonic transducer in an elastic half-space,'' J. Acoust. Soc. Am. **104**(3), 1178–1187.

Gridin, D. (**1998**). ''A fast method for simulating the propagation of pulses radiated by a rectangular normal transducer into an elastic half-space,'' J. Acoust. Soc. Am. **104**(6), 3199–3211.

Gridin, D., and Fradkin, L. J. (**1998**). ''High-frequency asymptotic description of pulses radiated by a circular normal transducer into an elastic half-space,'' J. Acoust. Soc. Am. **104**(6), 3190–3198.

Gridin, D., and Fradkin, L. J. (**2001**). ''The complete far-field asymptotic description of a point source acting on a transversely isotropic half-space,'' Proc. R. Soc. London, Ser. A **457**, 2675–2698.

Guo, Q. C., and Achenbach, J. D. (**1995**). ''Radiation of ultrasound into an anisotropic solid,'' Ultrasonics **33**, 449–456.

Langenberg, K. J., Hannemann, R., Kaczorowski, T., Marklein, R., Koehler, B., Schurig, C., and Walte, F. (**2000**). ''Application of modeling techniques for ultrasonic austenitic weld inspection,'' NDT & E Int. **33**, 465–480.

Lhémery, A., Calmon, P., Lecœur-Taibi, I., Raillon, R., and Paradis, L. (**2000**). ''Modeling tools for ultrasonic inspection of welds,'' NDT & E Int. **33**, 499–513.

Ogilvy, J. A. (**1986**). ''Ultrasonic beam profiles and beam propagation in an austenitic weld using a theoretical ray tracing model,'' Ultrasonics **24**, 337–347.

Payton, R. G. (**1983**). *Elastic Wave Propagation in Transversely Isotropic Media* (Martinus Nijhoff, The Hague).

Schmerr, L. W., and Sedov, A. (**2002**). ''Gaussian beam propagation in anisotropic, inhomogeneous elastic media,'' in *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 21A, edited by D. O. Thompson and D. E. Chimenti (American Institute of Physics, New York), pp. 123–129.

Spies, M. (**1994**). ''Transducer-modeling in general transversely isotropic media via point-source-synthesis: theory,'' J. Nondestruct. Eval. **13**, 85–99.

Spies, M. (**1999**). ''Transducer field modeling in anisotropic media by superposition of Gaussian base functions,'' J. Acoust. Soc. Am. **105**(2), 633–638.

Spies, M. (**2002**). ''Quantitative evaluation of defects in anisotropic media using the far-field radiation characteristics of vibrating sources,'' in *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 21A, edited by D. O. Thompson and D. E. Chimenti (American Institute of Physics, New York), pp. 107–114.

Elastic wave propagation and scattering in solids with uniaxially aligned cracks

Liyong Yang and Joseph A. Turner
Department of Engineering Mechanics, University of Nebraska-Lincoln, W317.4 Nebraska Hall, Lincoln, Nebraska 68588-0526

(Received 11 November 2002; revised 18 May 2003; accepted 19 May 2003)

In this article, elastic wave propagation and scattering in a solid medium permeated by uniaxially aligned penny-shaped microcracks are studied. The crack alignment refers to the case in which the unit normals of all cracks are randomly oriented within a plane of isotropy. The analysis is restricted to the limit of the noninteraction approximation among individual cracks. Explicit expressions for attenuations and wave speeds of the shear horizontal, quasilongitudinal, and quasishear vertical waves are obtained using stochastic wave theory in a generalized dyadic approach. The ensemble average elastic wave response is governed by the Dyson equation, which is solved in terms of the anisotropic elastic Green's dyadic. The analysis of expressions is limited to frequencies below the geometric optics limit. The resulting attenuations are investigated in terms of the directional, frequency, and damage dependence. In particular, the attenuations are simplified considerably within the low frequency Rayleigh regime. Finally, numerical results are presented and discussed in terms of the relevant dependent parameters.
[DOI: 10.1121/1.1592158]

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

I. INTRODUCTION

The investigation of wave propagation and scattering of elastic waves in damaged solids is of considerable interest to nondestructive evaluation and materials characterization, particularly for ultrasonic techniques. Analytical and experimental examinations of attenuation and wave speeds of ultrasonic waves in cracked solids provide a direct approach for the detection of material damage. Material responses, which are typically evaluated ultrasonically by the decrease in wave velocity or increase in wave attenuation, vary with microcracking changes. Both of these phenomena are caused by the stiffness degradation of the material by the cracks. Studies of elastic wave attenuation in cracked solid media have been studied for at least 30 years, since the work of Mal,[1,2] Piau,[3] Chatterjee *et al.*,[4] Hudson,[5] Martin,[6] Krenk,[7] and Martin and Wickham.[8] In addition, wave propagation in cracked solids has been reviewed by Zhang and Achenbach,[9] Zhang and Gross,[10,11] Smyshlyaev and Willis,[12] and Eriksson and Datta.[13] Previous research has primarily been focused on specific wave types and wave directions for either a single crack or distributed cracks, but a more comprehensive, general study has never been undertaken. In our previous article,[14] explicit general expressions of wave attenuations and wave speeds in a medium with damage from randomly distributed penny-shaped microcracks were derived. Under the assumption of statistical isotropy used in that work, the attenuation is independent of propagation direction. However, in the case of structural materials such as concrete, polycrystalline metals, fiber-reinforced composites, and others, those microcracks induced by directional loading or temperature are typically parallel to some direction. In this case, the effective media may acquire an anisotropy essentially due to the presence of such uniaxially aligned cracks. Thus, the scattering attenuation is a function of propagation direction. The analysis of this scattering attenuation is, therefore, more complicated than that of the isotropic case.

In this article, the framework used previously[14] is extended to study the attenuation of elastic waves in solids with uniaxially aligned cracks that are statistically homogeneous. Again, the microcracks are assumed to be noninteracting, penny-shaped cracks. Here, the unit normals of all cracks are assumed to be coplanar, but random within this plane of isotropy. Thus, the uniaxial symmetry direction is perpendicular to this plane. This case is different from the case of perfect crack alignment by all cracks, as discussed by Hudson,[5] for example. The effective elastic moduli of the medium that contains many penny-shaped cracks are reviewed by Nemat-Nasser,[15] Kachanov,[16] Krajcinovic,[17] and others. General wave propagation and scattering problems for anisotropic media are discussed by Stanke and Kino,[18] Ahmed and Thompson,[19] and Hirsekorn.[20,21] All those discussions involve a scattering integral with an isotropic Green's function. The use of an anisotropic Green's function for modeling the scattering in anisotropic media was investigated by Turner.[22] Here, this approach is employed as well to formulate the uniaxially aligned crack problem. In this way, the mean response is written in terms of the Dyson equation as discussed by Frisch[23] and Weaver.[24] The Dyson equation is solved in the spatial Fourier transform domain within the limits of the first-order smoothing approximation (FOSA), or Keller[25] approximation. A further approximation is also made which restricts the results to frequencies below the high-frequency geometric optics limit. The resulting attenuations are shown to be directional dependent, frequency dependent, and damage dependent for the shear horizontal, quasilongitudinal, and quasishear vertical waves. In particular, the angular dependence of the attenuations in the Rayleigh limit is obtained explicitly. Outside the Rayleigh limit,

simple expressions of the attenuations of the shear horizontal, quasilongitudinal, and quasishear vertical waves are derived in terms of integrations on the unit circle. Quantitative and qualitative comparisons with previous results by Zhang and Gross,[10,11] Zhang and Achenbach,[9] Eriksson and Datta,[13] Ahmed and Thompson,[19] and Turner[22] show that the more general, direct expressions derived here are reliable and comprehensive for practical applications of detecting damage from microcracks.

This article is organized in the following manner. The effective elastic properties of a solid containing uniaxially aligned cracks are discussed in Sec. II. The wave propagation and scattering model described by an anisotropic Green's function is included for completeness in Sec. III. The attenuations of each wave type through the uniaxially aligned cracks are given in a direct form subsequently in Sec. IV. In Sec. V, numerical results and discussion are presented. Finally, conclusions are presented in Sec. VI.

II. EFFECTIVE PROPERTIES IN SOLIDS WITH UNIAXIALLY ALIGNED CRACKS

The effective stiffness attributed to a single penny-shaped crack, which is located within an infinite, homogeneous, isotropic and elastic continuum is considered first. Estimating the effective elastic properties of a statistically homogeneous elastic solid which contains a large number of microcracks was investigated by Nemat-Nasser and Hori,[15] Kachanov,[16] and others. Using similar methods, the effective stiffness attributable to a single, penny-shaped crack of radius a in a unit volume, called the crack basis Green's function, is given in the form[14]

$$\mathbf{C}^{(s)}=2\mu\, d_i \mathbf{I}^i, \tag{1}$$

where the repeated index i denotes the summation convention over the range of $i=1-6$. It should be noted that this summation convention and range are used throughout this article. The coefficients d_i are given by

$$d_1=0,\quad d_2=\frac{16}{3}\frac{\nu^2(1-\nu)}{(1-2\nu)^2}a^3,\quad d_3=d_4=\frac{16}{3}\frac{\nu(1-\nu)}{1-2\nu}a^3,$$
$$d_5=\frac{32}{3}\frac{1-\nu}{2-\nu}a^3,\quad d_6=-\frac{16}{3}\frac{\nu(1-\nu)}{2-\nu}a^3. \tag{2}$$

The stiffness of a single crack is dependent on the unit normal $\hat{\mathbf{m}}$, which defines the crack orientation. This orientation is implicit in the tensors $\mathbf{I}$. These basis tensors are given in terms of unit vector $\hat{\mathbf{m}}$ and Kronecker delta function as follows.[26]

$$I^1_{ijmn}=\tfrac{1}{2}(\delta_{im}\delta_{jn}+\delta_{in}\delta_{jm}),\quad I^2_{ijmn}=\delta_{ij}\delta_{mn},$$
$$I^3_{ijmn}=\delta_{ij}\hat{m}_m\hat{m}_n,\quad I^4_{ijmn}=\hat{m}_i\hat{m}_j\delta_{mn},$$
$$I^6_{ijmn}=\hat{m}_i\hat{m}_j\hat{m}_m\hat{m}_n, \tag{3}$$
$$I^5_{ijmn}=\tfrac{1}{4}(\hat{m}_i\hat{m}_m\delta_{jn}+\hat{m}_i\hat{m}_n\delta_{jm}+\hat{m}_j\hat{m}_m\delta_{in}+\hat{m}_j\hat{m}_n\delta_{im}).$$

The ensemble average properties contributed by all cracks are considered next. The cracks are assumed to be embedded in an infinitely extended, homogeneous, isotropic and elastic three-dimensional continuum. The penny-shaped

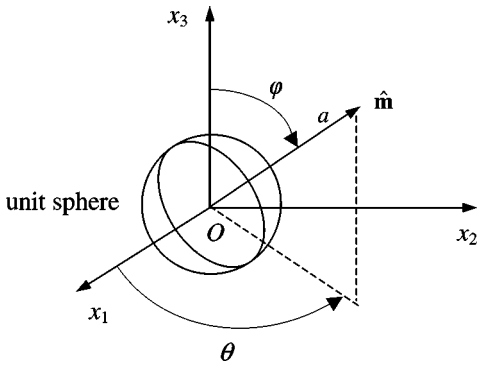


FIG. 1. Geometry of a penny-shaped crack.

crack is characterized by its radius a and two Euler angles θ and φ, which define the direction of the unit normal $\hat{\mathbf{m}}$ as shown in Fig. 1. It is also assumed that the microcracks do not interact with each other. Thus, the effective stiffness may be determined by integration over a continuous distribution of crack sizes and orientations. In general, the specific distribution of the crack radii and orientations is described by the probability density function $W(a,\theta,\varphi)$. In some cases, the microcrack radii and orientations are often correlated. Here, it is assumed that the microcrack radii and orientations are not correlated. As such, the density function may be separated into independent radius and orientation functions of the form

$$W(a,\theta,\varphi)=A(a)\zeta(\theta,\varphi). \tag{4}$$

It is also assumed that all microcracks are parallel to the x_3 axis ($\hat{\mathbf{n}}$ direction) with their unit normals (lying in the x_1-x_2 plane), having a random distribution, as shown in Fig. 2. In this situation, due to the symmetry about the x_1-x_2 plane, the average elastic properties are those of transverse isotropy, with the x_3 axis as the uniaxial symmetry axis. Here, the distribution of the microcracks is supposed to be dilute, and the distribution of the crack sizes is also assumed to be independent of their orientations. The crack orientation distribution function in Eq. (4), which implies that the orientation function $\zeta(\theta,\varphi)$ is independent of the angle θ, is then given by

$$\zeta(\theta,\varphi)=2\delta\left(\varphi-\frac{\pi}{2}\right). \tag{5}$$

Therefore, the effective continuum material properties caused by all microcracks per unit volume are weighted by

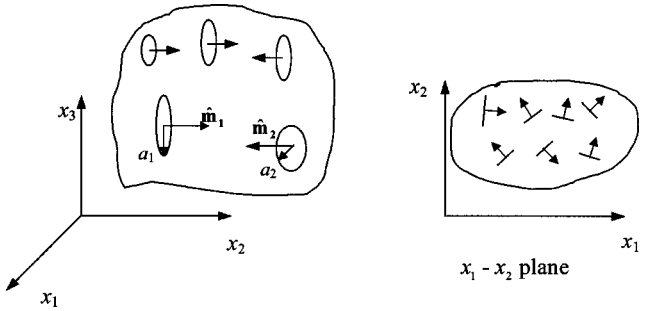


FIG. 2. The distribution of microcracks parallel to the x_3-axis.

the density function, Eq. (4), over the basis Green's function, Eq. (1), and are then given by

$$C^*_{ijmn}=\frac{\varepsilon}{2\pi}\int_0^{2\pi}\int_{-\pi/2}^{\pi/2}\delta\left(\varphi-\frac{\pi}{2}\right)C^{(s)}_{ijmn}(\theta,\varphi)\sin\varphi\, d\theta\, d\varphi. \tag{6}$$

In Eq. (6), the nondimensional microcrack density per unit volume is defined by

$$\varepsilon=N\langle a^3\rangle=\int_{a^-}^{a^+}A(a)a^3\, da, \tag{7}$$

where N is number of cracks per unit volume and the angular brackets denote the ensemble average. This damage density definition was introduced first by Walsh[27] for the case of a statistically isotropic distribution of penny-shaped microcracks. A more general form of the damage density is discussed by Budiansky.[28] The basis function $C^{(s)}_{ijmn}$ is expressed in Eq. (1). By integrating over the Euler angles in Eq. (6), the effective stiffness due to the distribution of uniaxially aligned penny-shaped microcracks is derived as

$$C^*_{ijmn}=Da_i\mathbf{I}^i, \tag{8}$$

where the coefficients D and a_i are

$$\begin{aligned}
&D=2\varepsilon\mu\frac{16}{3}\frac{1-\nu}{2-\nu},\quad a_1=1-\frac{\nu}{4},\\
&a_2=\frac{\nu(15-20\nu+4\nu^2)}{8(1-2\nu)^2},\quad a_3=a_4=\frac{\nu(-7+2\nu)}{8(1-2\nu)},\\
&a_5=1-\frac{\nu}{2},\quad a_6=-\frac{3\nu}{8}.
\end{aligned} \tag{9}$$

In the tensors $\mathbf{I}$ used in Eq. (8), the orientation is that of the symmetry direction $\hat{\mathbf{n}}$, rather than the direction $\hat{\mathbf{m}}$. It is hoped that this notation is not confusing to the reader. If the original undamaged state of the material is homogeneous and isotropic, the stiffness tensor is given in the standard form $\mathbf{C}^0=\lambda\mathbf{I}^2+2\mu\mathbf{I}^1$.

The ensemble effective stiffness is now redefined such that the average fluctuations are zero as done previously.[14] Such a procedure, while not necessary, is convenient for the calculation of material covariance and attenuation. The moduli are assumed to be spatially varying and of the form

$$\overline{\mathbf{C}}(\mathbf{x})=\overline{\mathbf{C}}^0+\delta\overline{\mathbf{C}}(\mathbf{x}), \tag{10}$$

where

$$\overline{\mathbf{C}}^0=\mathbf{C}^0-\mathbf{C}^*. \tag{11}$$

Thus, the average moduli have the form

$$\overline{\mathbf{C}}^0=\langle\overline{\mathbf{C}}(\mathbf{x})\rangle=\lambda_\perp\mathbf{I}^2+\mu_\perp\mathbf{I}^1+\Gamma_1(\mathbf{I}^3+\mathbf{I}^4)+\Gamma_2\mathbf{I}^5+\Gamma_3\mathbf{I}^6, \tag{12}$$

where the effective elastic constants are

$$\begin{aligned}
&\lambda_\perp=\lambda-\frac{D\nu(15-20\nu+4\nu^2)}{8(1-2\nu)^2},\quad \mu_\perp=\mu-\frac{D(4-\nu)}{8},\\
&\Gamma_1=\frac{D\nu(7-2\nu)}{8(1-2\nu)},\quad \Gamma_2=D\left(-1+\frac{\nu}{2}\right),\quad \Gamma_3=\frac{3D\nu}{8}.
\end{aligned} \tag{13}$$

The moduli fluctuations, which have zero mean, $\langle\delta\overline{C}\rangle=0$, are given by

$$\delta\overline{\mathbf{C}}=\mathbf{C}^*-\mathbf{C}^{(s)}H(\mathbf{x}), \tag{14}$$

and the function $H(\mathbf{x})$ is defined as

$$H(\mathbf{x})=\begin{cases}1 & \text{if } \mathbf{x}\in\mathbf{S}\\ 0 & \text{otherwise}\end{cases}, \tag{15}$$

where $\mathbf{S}$ denotes the space of all cracks.

In the next section, the fundamental elastodynamics of elastic wave propagation and scattering are introduced in terms of appropriate Green's dyadic. The formalism is developed for a transversely isotropic material. The mean response, which is expressed in terms of the Dyson equation, is discussed for the case of a transversely isotropic medium. The elastic modulus tensor is specified for this case and expressions of the attenuation for each wave type are given.

III. ELASTIC WAVE PROPAGATION AND SCATTERING MODEL

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

$$\begin{aligned}
&\{-\delta_{jm}\rho(\mathbf{x})\partial_t^2+\partial x_i\overline{C}_{ijmn}(\mathbf{x})\partial x_n\}G_{m\alpha}(\mathbf{x},\mathbf{x}';t)\\
&\quad=\delta_{j\alpha}\delta^3(\mathbf{x}-\mathbf{x}')\delta(t),
\end{aligned} \tag{16}$$

where $\delta^3(\mathbf{x}-\mathbf{x}')$ is the three-dimensional spatial delta function. The second order Green's dyadic, $G_{m\alpha}(\mathbf{x},\mathbf{x}';t)$, defines the response at location $\mathbf{x}$ in the mth direction to a unit impulse at location $\mathbf{x}'$ in the αth direction. The moduli are considered to vary spatially and density is assumed uniform throughout with units chosen such that density is unity. The moduli $\overline{\mathbf{C}}$ are assumed to be spatially heterogeneous and have the form in Eq. (10). The covariance of the moduli is characterized by an eighth-rank tensor

$$\langle\delta C_{ijmn}(\mathbf{x})\delta C_{\alpha\beta\gamma\delta}(\mathbf{y})\rangle=\Xi^{\alpha\beta\gamma\delta}_{ijmn}W(\mathbf{x}-\mathbf{y}). \tag{17}$$

The spatial and tensorial parts of the above covariance, Ξ and W, respectively, are assumed independent. The spatial correlation function W is also assumed a function of the difference between two vectors, $\mathbf{x}-\mathbf{y}$. This assumption implies that the medium is statistically homogeneous. However, the additional assumption of statistical isotropy as considered previously,[14] such that W is a function of $|\mathbf{x}-\mathbf{y}|$, is not made here.

The mean response, $\langle\mathbf{G}\rangle$, is governed by the Dyson equation which is given by[23,24]

$$\begin{aligned}
\langle G_{i\alpha}(\mathbf{x},\mathbf{x}')\rangle&=G^0_{i\alpha}(\mathbf{x},\mathbf{x}')+\int\int G^0_{i\beta}(\mathbf{x},\mathbf{y})M_{\beta j}(\mathbf{y},\mathbf{z})\\
&\quad\times\langle G_{j\alpha}(\mathbf{z},\mathbf{x}')\rangle d^3y\, d^3z.
\end{aligned} \tag{18}$$

In Eq. (18), the quantity $\mathbf{G}^0$ is the bare Green's dyadic. It defines the response of the medium without heterogeneities, namely, the solution of Eq. (16) with $\delta\overline{C}_{ijmn}(\mathbf{x})=0$. The second order tensor $\mathbf{M}$ is the mass or self-energy operator. The Dyson equation, Eq. (18), is easily solved in Fourier trans-

form space under the assumption of statistical homogeneity. The assumption of statistical homogeneity ensures that $\mathbf{G}^0$, $\mathbf{M}$ and $\langle\mathbf{G}\rangle$ are functions of a single wave vector in Fourier space. The Dyson equation is then transformed and solved to give the result for $\langle\mathbf{G}(\mathbf{p})\rangle$ of the form

$$\langle\mathbf{G}(\mathbf{p})\rangle=[\mathbf{G}^0(\mathbf{p})^{-1}-\tilde{\mathbf{M}}(\mathbf{p})]^{-1}. \tag{19}$$

Here, $\tilde{\mathbf{M}}$ is the spatial transform of the self-energy. An approximation of the self-energy $\mathbf{M}$ can be written as an expansion in powers of moduli fluctuations. To first order,[23,25] $\mathbf{M}$ is expressed as[24]

$$M_{\beta j}(\mathbf{y},\mathbf{z}) \approx\left\langle \frac{\partial}{\partial y_\alpha}\delta C_{\alpha\beta\gamma\delta}(\mathbf{y})\frac{\partial}{\partial y_\delta}G^0_{\gamma m}(\mathbf{y},\mathbf{z})\frac{\partial}{\partial z_i}\delta C_{ijmn}(\mathbf{z})\frac{\partial}{\partial z_n}\right\rangle. \tag{20}$$

Such an approximation is assumed valid if the fluctuations, $\delta\mathbf{C}$, are not too large. The components of $\tilde{\mathbf{M}}$, which are employed to calculate the phase velocity and attenuation of the wave modes, are discussed next. Further details of the scattering theory can be reviewed by the reader in the articles of Karal and Keller,[25] Frisch,[23] Stanke and Kino,[18] Weaver,[24] and Turner.[22]

The medium of uniaxially aligned cracks is considered to be transversely isotropic, a medium with a single symmetry axis defined here by the unit vector $\hat{\mathbf{n}}$. The fourth-rank elastic moduli tensor, $\overline{\mathbf{C}}$, in a transversely isotropic medium, given in Eq. (12), is written in terms of $\hat{\mathbf{n}}$ by

$$\begin{aligned}\overline{C}_{ijmn}=&\lambda_\perp\delta_{ij}\delta_{mn}+\mu_\perp(\delta_{im}\delta_{jn}+\delta_{in}\delta_{jm})+\Gamma_1(\delta_{ij}\hat{n}_m\hat{n}_n\\&+\delta_{mn}\hat{n}_i\hat{n}_j)+\Gamma_2(\delta_{im}\hat{n}_j\hat{n}_n+\delta_{in}\hat{n}_j\hat{n}_m+\delta_{jm}\hat{n}_i\hat{n}_n\\&+\delta_{jn}\hat{n}_i\hat{n}_m)+\Gamma_3\hat{n}_i\hat{n}_j\hat{n}_m\hat{n}_n.\end{aligned} \tag{21}$$

The above elastic constants are defined in Eqs. (13).

For propagation in the $\hat{\mathbf{p}}$ direction, the shear horizontal wave (SH) in a transversely isotropic medium is polarized in direction $\hat{\mathbf{u}}_1$, that is perpendicular to the plane defined by $\hat{\mathbf{p}}$ and $\hat{\mathbf{n}}$. The angle between the $\hat{\mathbf{p}}$ and $\hat{\mathbf{n}}$ is defined as Θ. The quasi-P and quasi-SV waves are polarized in directions defined by $\hat{\mathbf{u}}_2$ and $\hat{\mathbf{u}}_3$, respectively, both of which lie in the $\hat{\mathbf{p}}-\hat{\mathbf{n}}$ plane. It is noted that $\hat{\mathbf{u}}_1$, $\hat{\mathbf{u}}_2$, and $\hat{\mathbf{u}}_3$ form an orthonormal basis such that $\hat{\mathbf{u}}_3=\hat{\mathbf{u}}_1\times\hat{\mathbf{u}}_2$. The vector $\hat{\mathbf{u}}_2$ is directed at an angle ψ from the propagation direction $\hat{\mathbf{p}}$ (see Fig. 3). $\mathbf{G}^0$ may be diagonalized by using the directions $\hat{\mathbf{u}}_2$ and $\hat{\mathbf{u}}_3$, such that $\mathbf{I}-\hat{\mathbf{u}}_1\hat{\mathbf{u}}_1=\hat{\mathbf{u}}_2\hat{\mathbf{u}}_2+\hat{\mathbf{u}}_3\hat{\mathbf{u}}_3$.

Substituting the elastic stiffness tensor $\overline{\mathbf{C}}$ into the transform form of the equation of motion, Eq. (16), gives in direct notation[22]

$$\begin{aligned}\{&\hat{\mathbf{u}}_1\hat{\mathbf{u}}_1[\omega^2-p^2(\mu_\perp+\Gamma_2\cos^2\Theta)]+\hat{\mathbf{u}}_2\hat{\mathbf{u}}_2[\omega^2\\&-p^2(Q+P\cos^2\psi+R\cos^2(\Theta+\psi))]+\hat{\mathbf{u}}_3\hat{\mathbf{u}}_3[\omega^2\\&-p^2(Q+P\sin^2\psi+R\sin^2(\Theta+\psi))]\}\cdot\mathbf{G}^0(\mathbf{p})=\mathbf{I}.\end{aligned} \tag{22}$$

The quantities Q, P, and R in Eq. (22) are defined by

$$\begin{aligned}Q&=\mu_\perp+\Gamma_2(\hat{\mathbf{p}}\cdot\hat{\mathbf{n}})^2-(\Gamma_1+\Gamma_2)(1-(\hat{\mathbf{p}}\cdot\hat{\mathbf{n}})^2),\\P&=\lambda_\perp+\mu_\perp+\Gamma_1+\Gamma_2,\\R&=\Gamma_1+2\Gamma_2+\Gamma_3(\hat{\mathbf{p}}\cdot\hat{\mathbf{n}})^2.\end{aligned} \tag{23}$$

It should also be kept in mind that the vectors $\hat{\mathbf{u}}_2$ and $\hat{\mathbf{u}}_3$ are functions of the direction of propagation, $\hat{\mathbf{p}}$, relative to the uniaxial symmetry direction, $\hat{\mathbf{n}}$. This dependence, $\psi=\psi(\Theta)$ will remain implicit throughout.

The bare Green's dyadic is then presented in the form

$$\mathbf{G}^0(\mathbf{p})=g^0_{SH}(\mathbf{p})\hat{\mathbf{u}}_1\hat{\mathbf{u}}_1+g^0_{qP}(\mathbf{p})\hat{\mathbf{u}}_2\hat{\mathbf{u}}_2+g^0_{qSV}(\mathbf{p})\hat{\mathbf{u}}_3\hat{\mathbf{u}}_3, \tag{24}$$

where the dispersion relations for the bare response of the SH, qP, and qSV waves are given by

$$\begin{aligned}g^0_{SH}(\mathbf{p})&=[\omega^2-p^2(\mu_\perp+\Gamma_2\cos^2\Theta)]^{-1}\\&=[\omega^2-p^2c^2_{SH}]^{-1},\\g^0_{qP}(\mathbf{p})&=[\omega^2-p^2(Q+P\cos^2\psi+R\cos^2(\Theta+\psi))]^{-1}\\&=[\omega^2-p^2c^2_{qP}]^{-1},\\g^0_{qSV}(\mathbf{p})&=[\omega^2-p^2(Q+P\sin^2\psi+R\sin^2(\Theta+\psi))]^{-1}\\&=[\omega^2-p^2c^2_{qSV}]^{-1},\end{aligned} \tag{25}$$

with Q, P, and R defined in Eqs. (23).

The mean response, $\langle\mathbf{G}(\mathbf{p})\rangle$, is governed by the Dyson equation, Eq. (19). The solution of $\langle\mathbf{G}(\mathbf{p})\rangle$ is expressed in terms of $\mathbf{G}^0(\mathbf{p})$ and $\tilde{\mathbf{M}}(\mathbf{p})$. Similar to $\mathbf{G}^0$, the mean response $\langle\mathbf{G}(\mathbf{p})\rangle$ and self-energy $\tilde{\mathbf{M}}(\mathbf{p})$ may be written in terms of the orthonormal basis defined by $\hat{\mathbf{u}}_1$, $\hat{\mathbf{u}}_2$, and $\hat{\mathbf{u}}_3$ in the form

$$\begin{aligned}\langle\mathbf{G}(\mathbf{p})\rangle&=g_{SH}(\mathbf{p})\hat{\mathbf{u}}_1\hat{\mathbf{u}}_1+g_{qP}(\mathbf{p})\hat{\mathbf{u}}_2\hat{\mathbf{u}}_2+g_{qSV}(\mathbf{p})\hat{\mathbf{u}}_3\hat{\mathbf{u}}_3,\\\tilde{\mathbf{M}}(\mathbf{p})&=m_{SH}(\mathbf{p})\hat{\mathbf{u}}_1\hat{\mathbf{u}}_1+m_{qP}(\mathbf{p})\hat{\mathbf{u}}_2\hat{\mathbf{u}}_2+m_{qSV}(\mathbf{p})\hat{\mathbf{u}}_3\hat{\mathbf{u}}_3,\end{aligned} \tag{26}$$

where it is again noted that the propagation direction $\hat{\mathbf{p}}$ is implicit within the directions $\hat{\mathbf{u}}_1$, $\hat{\mathbf{u}}_2$, and $\hat{\mathbf{u}}_3$.

The dispersion relations for the mean response are then given by the solution of the Dyson equation, Eq. (19), as

$$g_\beta(\mathbf{p})=[g^0_\beta(\mathbf{p})^{-1}-m_\beta(\mathbf{p})]^{-1}=[\omega^2-p^2c^2_\beta-m_\beta(\mathbf{p})]^{-1}, \tag{27}$$

for each wave type, β. These are the expressions for the dispersion relation of the mean response, which defines the phase velocity and attenuation of each wave type from solution of

$$\omega^2-p^2c^2_\beta-m_\beta(\mathbf{p})=0, \tag{28}$$

for the wave vector $\mathbf{p}$. The attenuation of each wave type is given by the imaginary part of $\mathbf{p}$. Explicit expressions of the attenuation can be determined using an approximation valid below the high-frequency geometric optics limit ($m_\beta(\mathbf{p})\approx m_\beta(\omega/c_\beta\hat{\mathbf{p}})$).[18,24] This approximation allows the imaginary part of $\mathbf{p}$ to be calculated directly from Eq. (28). Thus, the attenuation of each wave type is given as

$$\alpha_\beta(\hat{\mathbf{p}})=-\frac{1}{2\omega c_\beta(\hat{\mathbf{p}})}\operatorname{Im}m_\beta\left(\frac{\omega}{c_\beta}\hat{\mathbf{p}}\right). \tag{29}$$

The attenuations for the three wave types, which are each defined in Eq. (29), are finally given in the general form[22]

$$\alpha_\beta(\hat{\mathbf{p}})=\frac{1}{c_\beta^3(\hat{\mathbf{p}})}\Bigg\{\frac{\pi}{4}\int d^2\hat{s}\frac{\omega^4}{c_{SH}^5(\hat{\mathbf{s}})}W\left(\frac{\omega}{c_\beta(\hat{\mathbf{p}})}\hat{\mathbf{p}}-\frac{\omega}{c_{SH}(\hat{\mathbf{s}})}\hat{\mathbf{s}}\right)\times\tilde{\Xi}_{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_1}^{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_1}+\frac{\pi}{4}\int d^2\hat{s}\frac{\omega^4}{c_{qP}^5(\hat{\mathbf{s}})}\times W\left(\frac{\omega}{c_\beta(\hat{\mathbf{p}})}\hat{\mathbf{p}}-\frac{\omega}{c_{qP}(\hat{\mathbf{s}})}\hat{\mathbf{s}}\right)\times\tilde{\Xi}_{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_2}^{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_2}+\frac{\pi}{4}\int d^2\hat{s}\frac{\omega^4}{c_{qSV}^5(\hat{\mathbf{s}})}\times W\left(\frac{\omega}{c_\beta(\hat{\mathbf{p}})}\hat{\mathbf{p}}-\frac{\omega}{c_{qSV}(\hat{\mathbf{s}})}\hat{\mathbf{s}}\right)\tilde{\Xi}_{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_3}^{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_3}\Bigg\},\tag{30}$$

where K is defined as the polarization for the wave type β (1, 2, or 3 for wave types SH, qP, and qSV, respectively). In Eq. (30), it can be seen that the integrals are over the unit sphere, which is defined by unit vector $\hat{\mathbf{s}}$. The direction $\hat{\mathbf{p}}$ defines the propagation direction, $\hat{\mathbf{s}}$ is the scattered direction, and $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ are defined as the polarization directions. The dependence of the vectors $\hat{\mathbf{u}}$ on $\hat{\mathbf{p}}$ and of $\hat{\mathbf{v}}$ on $\hat{\mathbf{s}}$ is implicit. The inner products on the covariance of the moduli fluctuations are given in terms of these four unit vectors. The argument of the correlation is the difference between the incoming and outgoing propagation directions.

It is clear that the above expressions of the attenuation for uniaxially aligned crack distributions are more complicated than those for a distribution of randomly oriented cracks as discussed previously.[14] They can be simplified to the forms given there in the case of statistical isotropy. This framework is the same as that given by Turner.[22] In the next section, the covariance and attenuation are specified.

IV. COVARIANCE AND ATTENUATION

The relevant inner products on the covariance of the effective moduli fluctuations are necessary for calculating the attenuations. The tensorial part of the covariance is represented by an eighth-rank tensor which is given explicitly by

$$\Xi(\mathbf{q})_{\cdots\hat{\mathbf{u}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}}^{\cdots\hat{\mathbf{u}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}}=\Xi(\mathbf{q})_{\alpha\beta\gamma\delta}^{ijmn}\hat{u}_\beta\hat{u}_m\hat{p}_\alpha\hat{p}_n\hat{s}_i\hat{s}_\delta\hat{v}_\gamma\hat{v}_j.\tag{31}$$

For the case of uniaxially aligned cracks, the covariance is dependent on the crack orientations $\hat{\mathbf{m}}$. To calculate the covariance, the following identities are needed

$$\begin{aligned}
&\langle\hat{m}_i\hat{m}_j\rangle=\tfrac{1}{2}\Delta_{ij},\\
&\langle\hat{m}_i\hat{m}_j\hat{m}_m\hat{m}_n\rangle=\tfrac{1}{8}(\Delta_{ij}\Delta_{mn}+\Delta_{im}\Delta_{jn}+\Delta_{in}\Delta_{jm}),\\
&\langle\hat{m}_i\hat{m}_j\hat{m}_m\hat{m}_n\hat{m}_\alpha\hat{m}_\beta\rangle\\
&\quad=\tfrac{1}{48}[\Delta_{ij}\Delta_{mn}\Delta_{\alpha\beta}+\text{all permutations}\\
&\qquad-15\text{ terms in all}],\\
&\langle\hat{m}_i\hat{m}_j\hat{m}_m\hat{m}_n\hat{m}_\alpha\hat{m}_\beta\hat{m}_\gamma\hat{m}_\delta\rangle\\
&\quad=\tfrac{1}{384}[\Delta_{ij}\Delta_{mn}\Delta_{\alpha\beta}\Delta_{\gamma\delta}+\text{all permutations}\\
&\qquad-105\text{ terms in all}],
\end{aligned}\tag{32}$$

where the brackets, $\langle\,\rangle$, denote an ensemble average, and $\Delta_{MN}=(\delta_{MN}-\hat{n}_M\hat{n}_N)$. The unit vector $\hat{\mathbf{n}}$ is the uniaxial symmetry axis. All averages of odd numbers of $\hat{\mathbf{m}}$'s are zero.

In addition to these tensorial averages, the average of the spatial part of the covariance must be determined. As discussed previously,[14] the necessary relation is given by $\langle H(\mathbf{x})H(\mathbf{y})\rangle=\varepsilon\,\Pr(r|0)$, where $\Pr(r|0)=(1-\varepsilon)W(r)+\varepsilon$, is defined as the conditional probability.[29] Due to the assumption of small damage density, the higher order terms may be neglected. Therefore, $\langle\delta C_{ijmn}(\mathbf{x})\delta C_{\alpha\beta\gamma\delta}(\mathbf{y})\rangle=\varepsilon W(r)\Xi_{ijmn}^{\alpha\beta\gamma\delta}$. Averaging over all crack orientations, the covariance is thus defined by

$$\Xi_{ijmn}^{\alpha\beta\gamma\delta}=\frac{1}{2\pi}\int_0^{2\pi}\int_{-\pi/2}^{\pi/2}(\overline{C}_{ijmn}^{(s)}\overline{C}_{\alpha\beta\gamma\delta}^{(s)})\,\delta\left(\varphi-\frac{\pi}{2}\right)\sin\varphi\,d\varphi\,d\theta,\tag{33}$$

where the definition of $\overline{\mathbf{C}}^{(s)}$ is given by Eq. (60) in Ref. 14. Substituting the identities of Eqs. (32) into Eq. (33), the generally compact form of Ξ is constructed in terms of Kronecker deltas and pairs of $\hat{\mathbf{n}}$'s. The general compact form of Ξ is not presented here due to brevity. The form of the attenuations given in Eqs. (30) is dependent on various inner products on the covariance tensor. The vectors $\hat{\mathbf{p}}$ and $\hat{\mathbf{s}}$, respectively, represent the incoming and outgoing propagation directions. The vectors $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ are vectors defining the polarization directions of the particular waves. These vectors are perpendicular to the plane defined by $\hat{\mathbf{s}}$ or $\hat{\mathbf{p}}$ and $\hat{\mathbf{n}}$ (for SH waves) or they lie in this plane (for qP and qSV).

Now the necessary inner products involved in determining the attenuations are calculated. The attenuations will vary angularly only within the plane defined by the propagation direction $\hat{\mathbf{p}}$ and the crack alignment direction $\hat{\mathbf{n}}$. Therefore, without loss of generality, a reference plane is defined as the $\hat{\mathbf{p}}$-$\hat{\mathbf{n}}$ plane (see Fig. 3). The following vectors are then defined with respect to a general $x_1x_2x_3$ coordinate system as

$$\begin{aligned}
&\hat{\mathbf{n}}=\hat{\mathbf{x}}_3,\quad \hat{\mathbf{p}}=\hat{\mathbf{x}}_2\sin\Theta+\hat{\mathbf{x}}_3\cos\Theta,\\
&\hat{\mathbf{s}}=\hat{\mathbf{x}}_1\sin\Theta'\cos\phi'+\hat{\mathbf{x}}_2\sin\Theta'\sin\phi'+\hat{\mathbf{x}}_3\cos\Theta'.
\end{aligned}\tag{34}$$

The polarization vectors are then defined with respect to these angles and ψ as discussed by Turner.[22] The angles γ and γ' used hereafter are defined by

$$\gamma=\Theta+\psi(\Theta),\quad \gamma'=\Theta'+\psi(\Theta').\tag{35}$$

These angles, γ and γ', define the orientation angle of the qP wave with respect to the $\hat{\mathbf{n}}$ direction, for the $\hat{\mathbf{p}}$ and $\hat{\mathbf{s}}$ directions, respectively.

Inserting these definitions of the relevant unit vectors into Eq. (33), the required inner products are reduced to a simple form. The inner products are:

for α_{SH}

$$\begin{aligned}
&\Xi_{\cdots\hat{\mathbf{u}}_1\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_1}^{\cdots\hat{\mathbf{u}}_1\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_1}=\sin^2\Theta\sin^2\Theta'[-\eta_1\cos^2\phi'\sin^2\phi'+\eta_2],\\
&\Xi_{\cdots\hat{\mathbf{u}}_1\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_2}^{\cdots\hat{\mathbf{u}}_1\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_2}=\sin^2\Theta\sin^2\Theta'\sin^2\gamma'[\eta_1\cos^2\phi'\sin^2\phi'+\eta_3],
\end{aligned}\tag{36}$$

$$\Xi_{\cdots\hat{\mathbf{u}}_1\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_3}^{\cdots\hat{\mathbf{u}}_1\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_3}=\sin^2\Theta\sin^2\Theta'\cos^2\gamma'[\eta_1\cos^2\phi'\sin^2\phi'+\eta_3],$$

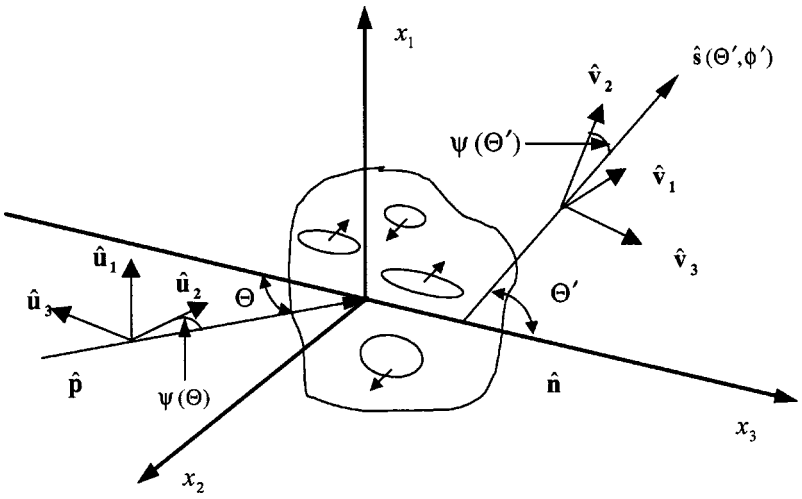


FIG. 3. Geometry for the propagation direction $\hat{\mathbf{p}}$, the scattered direction $\hat{\mathbf{s}}$, and the respective polarization directions $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ in the local coordinate system.

for α_{qP}

$$\Xi^{\cdots\hat{\mathbf{u}}_2\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_1}_{\cdots\hat{\mathbf{u}}_2\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_1}=\sin^2\Theta\sin^2\Theta'\sin^2\gamma[\eta_1\cos^2\phi'\sin^2\phi'+\eta_3],$$

$$\Xi^{\cdots\hat{\mathbf{u}}_2\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_2}_{\cdots\hat{\mathbf{u}}_2\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_2}=\sin^2\Theta\sin^2\Theta'\sin^2\gamma\times\sin^2\gamma'[\eta_1\sin^4\phi'+\eta_4\sin^2\phi'+\eta_5], \quad (37)$$

$$\Xi^{\cdots\hat{\mathbf{u}}_2\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_3}_{\cdots\hat{\mathbf{u}}_2\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_3}=\sin^2\Theta\sin^2\Theta'\sin^2\gamma\times\cos^2\gamma'[\eta_1\sin^4\phi'+\eta_4\sin^2\phi'+\eta_5],$$

and for α_{qSV}

$$\Xi^{\cdots\hat{\mathbf{u}}_3\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_1}_{\cdots\hat{\mathbf{u}}_3\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_1}=\sin^2\Theta\sin^2\Theta'\cos^2\gamma[\eta_1\cos^2\phi'\sin^2\phi'+\eta_3],$$

$$\Xi^{\cdots\hat{\mathbf{u}}_3\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_2}_{\cdots\hat{\mathbf{u}}_3\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_2}=\sin^2\Theta\sin^2\Theta'\cos^2\gamma\times\sin^2\gamma'[\eta_1\sin^4\phi'+\eta_4\sin^2\phi'+\eta_5], \quad (38)$$

$$\Xi^{\cdots\hat{\mathbf{u}}_3\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_3}_{\cdots\hat{\mathbf{u}}_3\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_3}=\sin^2\Theta\sin^2\Theta'\cos^2\gamma\times\cos^2\gamma'[\eta_1\sin^4\phi'+\eta_4\sin^2\phi'+\eta_5],$$

where γ and γ' are defined in Eq. (35). The coefficients η_i $(i=1,\ldots,5)$ and T_j $(j=1,\ldots,7)$, are given by

$$\eta_1=4T_3+16T_6+4T_7,\quad \eta_2=T_3+4T_6+4T_7,$$
$$\eta_3=T_4+4T_5+4T_6+6T_7,$$
$$\eta_4=4T_2+4T_4+32T_5+16T_6+16T_7, \quad (39)$$
$$\eta_5=T_1+4T_4+4T_7,$$

with

$$T_1=-\frac{\nu^2(80\nu^4-416\nu^3+472\nu^2+184\nu-235)M^2}{128(1-2\nu)^2},$$

$$T_2=-\frac{\nu(12\nu^3-28\nu^2+127\nu-184)M^2}{384},$$

$$T_3=\frac{\nu^2(1-2\nu)^2M^2}{384},$$

$$T_3=\frac{(\nu-4)(\nu-12)(1-2\nu)^2M^2}{384}, \quad (40)$$

$$T_4=\frac{\nu^2(10\nu-23)(2\nu-7)M^2}{384},$$

$$T_5=\frac{\nu(1-2\nu)(6\nu^2-31\nu+44)M^2}{384},$$

$$T_6=\frac{(3-2\nu)(1-2\nu)^2M^2}{96},$$

where the constant M is defined as $M=\mu(32/3)[1-\nu/(2-\nu)(1-2\nu)]$. The expressions given by Eqs. (36)–(38) are also directly related to the diffuse energy propagation, including backscatter.[30–32]

If the tensorial and spatial components of covariance are assumed to be independent, as discussed in Eq. (17), the spatial correlation function W is uncorrelated with the tensorial part. Here, the correlation function W is assumed to have an exponential form $W(r)=e^{-r/L}$, where L is the spatial correlation length, $L=2\langle a\rangle$. The limits of such an assumption have been reviewed by Stanke,[33] and Markov and Willis.[29]

Substituting the above inner products into Eqs. (30) and integrating over the azimuthal angle ϕ', the attenuations finally reduce to dimensionless forms

$$\alpha_{SH}(\Theta)L=x_{SH}^4\frac{\varepsilon}{2\rho^2\overline{c}_{SH}^4}r_{SH}^3(\Theta)\sin^2\Theta\left[I_{SH-SH}+I_{SH-qP}\left(\frac{\overline{c}_{SH}}{\overline{c}_{qP}}\right)^5+I_{SH-qSV}\left(\frac{\overline{c}_{SH}}{\overline{c}_{qSV}}\right)^5\right], \quad (41)$$

$$\alpha_{qP}(\Theta)L=x_{qP}^4\frac{\varepsilon}{2\rho^2\overline{c}_{qP}^4}r_{qP}^3(\Theta)\sin^2\Theta\sin^2\gamma\times\left[I_{qP-SH}\left(\frac{\overline{c}_{qP}}{\overline{c}_{SH}}\right)^5+I_{qP-qP}+I_{qP-qSV}\left(\frac{\overline{c}_{qP}}{\overline{c}_{qSV}}\right)^5\right], \quad (42)$$

$$\alpha_{qSV}(\Theta)L = x_{qSV}^4 \frac{\varepsilon}{2\rho^2 \overline{c}_{qSV}^4} r_{qSV}^3(\Theta)\sin^2\Theta\cos^2\gamma \times\left[I_{qSV-SH}\left(\frac{\overline{c}_{qSV}}{\overline{c}_{SH}}\right)^5 + I_{qSV-qP}\left(\frac{\overline{c}_{qSV}}{\overline{c}_{qP}}\right)^5 + I_{qSV-qSV}\right], \tag{43}$$

with the density, ρ, now included in the general form. The terms denoted by $I_{\beta-\gamma}$ within the square brackets represent integrals defined by

$$I_{SH-SH} = \int_0^\pi \left[\frac{\eta_1(6X_{SH-SH}^2 - Y_{SH-SH}^2)}{2Y_{SH-SH}^2} + \frac{\eta_1(2X_{SH-SH}Y_{SH-SH}^2 - 3X_{SH-SH}^2)}{Y_{SH-SH}^4(X_{SH-SH}^2 - Y_{SH-SH}^2)^{1/2}} \times \frac{\eta_2 X_{SH-SH}}{(X_{SH-SH}^2 - Y_{SH-SH}^2)^{3/2}} \right] r_{SH}^5(\Theta')\sin^3\Theta'\, d\Theta',$$

$$I_{SH-\alpha} = \int_0^\pi \left[\frac{\eta_1(Y_{SH-\alpha}^2 - 6X_{SH-\alpha}^2)}{2Y_{SH-\alpha}^2} + \frac{\eta_1(3X_{SH-\alpha}^2 - 2X_{SH-\alpha}Y_{SH-\alpha}^2)}{Y_{SH-\alpha}^4(X_{SH-\alpha}^2 - Y_{SH-\alpha}^2)^{1/2}} \times \frac{\eta_3 X_{SH-\alpha}}{(X_{SH-\alpha}^2 - Y_{SH-\alpha}^2)^{3/2}} \right] r_\alpha^5(\Theta')\Pi_\alpha \sin^3\Theta'\, d\Theta', \tag{44}$$

and

$$I_{\alpha-SH} = \int_0^\pi \left[\frac{\eta_1(Y_{\alpha-SH}^2 - 6X_{\alpha-SH}^2)}{2Y_{\alpha-SH}^2} + \frac{\eta_1(3X_{\alpha-SH}^2 - 2X_{\alpha-SH}Y_{\alpha-SH}^2)}{Y_{\alpha-SH}^4(X_{\alpha-SH}^2 - Y_{\alpha-SH}^2)^{1/2}} \times \frac{\eta_2 X_{\alpha-SH}}{(X_{\alpha-SH}^2 - Y_{\alpha-SH}^2)^{3/2}} \right] r_{SH}^5(\Theta')\sin^3\Theta'\, d\Theta',$$

$$I_{\delta-\alpha} = \int_0^\pi \left[\frac{\eta_1(6X_{\delta-\alpha}^2 + Y_{\delta-\alpha}^2) + 2\eta_4 Y_{\delta-\alpha}^2}{2Y_{\delta-\alpha}^4} + \frac{\eta_1(4X_{\delta-\alpha}^3 Y_{\delta-\alpha}^2 - 3X_{\delta-\alpha}^5) + \eta_4 Y_{\delta-\alpha}^2(2X_{\delta-\alpha}Y_{\delta-\alpha}^2 - X_{\delta-\alpha}^3)}{Y_{\delta-\alpha}^4(X_{\delta-\alpha}^2 - Y_{\delta-\alpha}^2)^{3/2}} + \frac{\eta_5 X_{\delta-\alpha}Y_{\delta-\alpha}^4}{Y_{\delta-\alpha}^4(X_{\delta-\alpha}^2 - Y_{\delta-\alpha}^2)^{3/2}} \right] r_\alpha^5(\Theta')\Pi_\alpha \sin^3\Theta'\, d\Theta', \tag{45}$$

with

$$X_{\beta-\gamma} = 1 + x_\beta^2 r_\beta^2(\Theta) + x_\gamma^2 r_\gamma^2(\Theta') - 2x_\beta x_\gamma r_\beta(\Theta) r_\gamma(\Theta')\cos\Theta\cos\Theta',$$
$$Y_{\beta-\gamma} = 2x_\beta x_\gamma r_\beta(\Theta) r_\gamma(\Theta')\sin\Theta\sin\Theta', \tag{46}$$

for the different wave types, β and γ. The subscripts δ and α denote either the qP or qSV wave type, and the notation $\Pi_{qP} = \sin^2(\Theta' + \psi(\Theta'))$, $\Pi_{qSV} = \cos^2(\Theta' + \psi(\Theta'))$ is used. In Eqs. (41)–(46), the angular averaged wave speeds are defined as $\overline{c}_\beta = \frac{1}{2}\int_0^\pi c_\beta(\Theta)\sin\Theta\, d\Theta$, for each wave type, β. Three nondimensional frequencies are then defined as $x_\beta = \omega L/\overline{c}_\beta$ and the slowness surface for each wave type is defined by the dimensionless quantity $r_\beta(\Theta) = \overline{c}_\beta / c_\beta(\Theta)$. Equations (41)–(45) are the primary results of this article.

In the long wavelength Rayleigh limit, $x_\beta \ll 1$ and these integrals become independent of incident direction and frequency. Therefore, they reduce to a much simpler form as

$$I_{SH-SH} = \int_0^\pi \left(-\frac{\eta_1}{8} + \eta_2 \right) r_{SH}^5(\Theta')\sin^3\Theta'\, d\Theta',$$
$$I_{SH-\alpha} = \int_0^\pi \left(\frac{\eta_1}{8} + \eta_3 \right) r_\alpha^5(\Theta')\Pi_\alpha \sin^3\Theta'\, d\Theta', \tag{47}$$

and

$$I_{\alpha-SH} = \int_0^\pi \left(\frac{\eta_1}{8} + \eta_3 \right) r_{SH}^5(\Theta')\sin^3\Theta'\, d\Theta',$$
$$I_{\delta-\alpha} = \int_0^\pi \left(\frac{3\eta_1}{8} + \frac{\eta_4}{2} + \eta_5 \right) r_\alpha^5(\Theta')\Pi_\alpha \sin^3\Theta'\, d\Theta' \tag{48}$$

for all outgoing wave types. In the Rayleigh limit, the angular dependence of the attenuation is explicitly seen. In the subsequent section, example numerical results and discussion are presented.

V. EXAMPLE RESULTS

Numerical results are now presented for a specific case, in which the observed anisotropy of the cracked material is essentially due to the presence of the uniaxially aligned cracks. The material properties of the uncracked medium used are Young's modulus $E = 2.0\times10^{11}$ Pa, Poisson's ratio $\nu = 0.30$, and density $\rho = 7850$ kg/m^3. Using the dispersion relations given in Eqs. (25), the slowness surfaces calculated for different damage densities, $\varepsilon = 0$ and 0.1, are shown in Fig. 4. The normalized effective wave velocity, $c_\beta(\varepsilon)/c_\beta(\varepsilon = 0)$, of each wave type is presented in Fig. 5. The effective velocities decrease with increasing damage density ε within the considered frequency range. The reduction of velocity of SH and qP wave due to the presence of the uniaxially aligned cracks is a maximum at $\Theta = 90°$, it becomes smaller as Θ decreases, and the reduction reaches a minimum at Θ

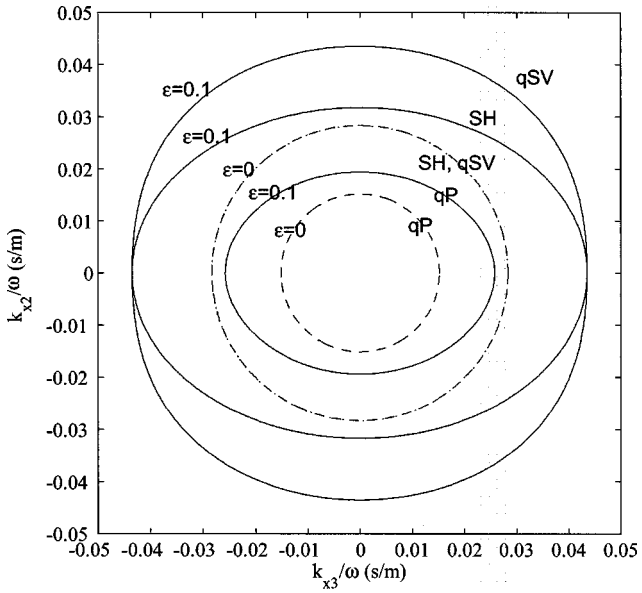


FIG. 4. Slowness surfaces for damage densities $\varepsilon = 0, 0.1$.

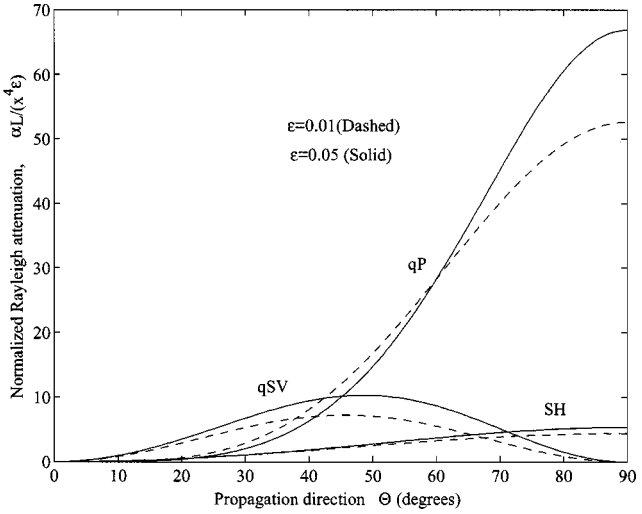


FIG. 6. Rayleigh limit as a function of direction for the SH, qP, and qSV waves for damage density $\varepsilon = 0.01$ (dashed) and $\varepsilon = 0.05$ (solid). The dimensionless attenuation αL has been normalized by the fourth power of the dimensionless frequency and damage density for the respective wave type: $\alpha_{SH}L/(x_{SH}^4\varepsilon)$, $\alpha_{qP}L/(x_{qP}^4\varepsilon)$, and $\alpha_{qSV}L/(x_{qSV}^4\varepsilon)$.

$=0°$, though the changes are not substantial. These results are in basic agreement with those of Zhang and Gross[10] and Eriksson and Datta.[13] The qSV wave velocity is seen to have a greater reduction at $\Theta = 45°$ than at $\Theta = 0°$ and 90°.

In the Rayleigh limit, the attenuations simplify considerably since the integrals reduce to those given by Eqs. (47) and (48). The attenuation depends on the fourth power of frequency in the Rayleigh regime. Thus, the angular Rayleigh attenuation results shown in Fig. 6 are given in a general form of, $\alpha L/(x^4\varepsilon)$, for each wave type. In Fig. 6, the SH and qP waves are observed to have their maxima at $\Theta = 90°$—perpendicular to the crack alignment direction $\hat{\mathbf{n}}$. The qSV wave is observed to have zero attenuation for propagation along the symmetry axis ($\Theta = 0°$) and perpendicular to it ($\Theta = 90°$). All wave types have zero attenuation along the symmetry axis, because the material properties do not vary in that direction. Those results are qualitatively the same as previous work.[19,21,22] Zhang and Gross[10] comment that their attenuation results are not zero for propagation along the symmetry axis. They speculate that the attenuation arises from Poisson effects. However, such a comparison is difficult to make since the focus of their work was at much higher frequencies. An additional feature observed for the qSV wave in Fig. 6, is the asymmetry that develops as ε increases. This peak is around $\Theta = 45°$, but shifts slightly as ε increases from 0.01 to 0.05.

Finally, using Eqs. (41)–(43), attenuation results are given in terms of the single dimensionless frequency $x_{SH} = \omega L/\overline{c}_{SH}$. Outside the Rayleigh regime, the attenuations were calculated using the complete integrals, Eqs. (44) and (45), by numerical integration. In Fig. 7, the normalized SH wave attenuation, α_{SH}/k_{SH}, is presented as a function of propagation direction for three different damage densities at frequency $x_{SH} = 1.0$. The attenuation for propagation perpendicular to the crack alignment direction is seen to increase more quickly than for other directions as the damage in-

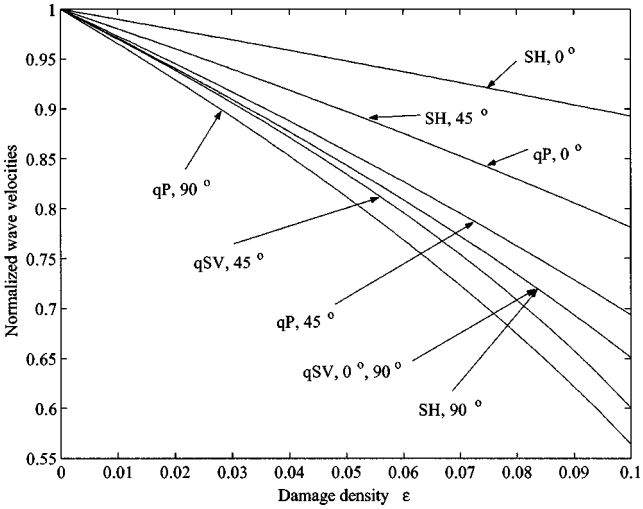


FIG. 5. Wave velocity of each wave type normalized to the undamaged wave speed, $c_\beta(\varepsilon)/c_\beta$, versus damage density ε at $\Theta = 0°$, 45° and 90°.

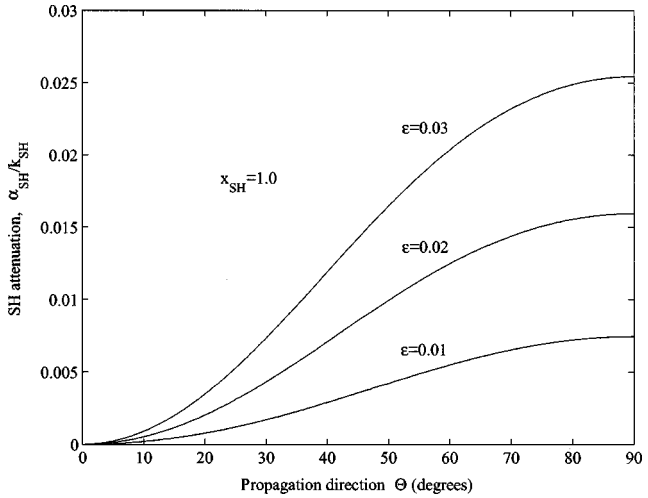


FIG. 7. Angular dependence of the normalized SH attenuation, α_{SH}/k_{SH} for various damage densities ε at frequency $x_{SH} = 1.0$.

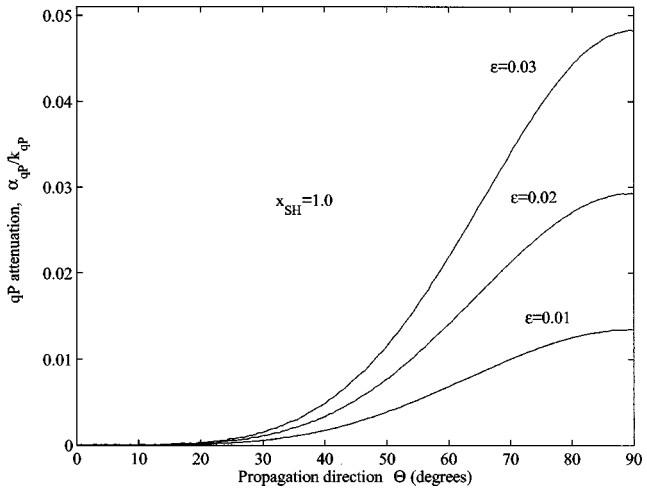


FIG. 8. Angular dependence of the normalized qP attenuation, α_{qP}/k_{qP} for various damage densities ε at frequency $x_{SH}=1.0$.

creases. The results for the normalized qP attenuation, α_{qP}/k_{qP}, are shown in Fig. 8. These results display similar behavior as the SH attenuation in terms of the change with angle and damage. Analogous results have been observed in textured polycrystals by Hirsekorn,[21] Ahmed and Thompson,[19] and Turner.[22] In Fig. 9, the normalized qSV attenuation, α_{qSV}/k_{qSV}, is presented at various damage densities for frequency $x_{SH}=1.0$. The attenuation for propagation at $\Theta=0°$ and 90° is zero as discussed above. For propagation at $\Theta=45°$, the attenuation is the largest. In addition, it is seen that the peak of maximum attenuation shifts as the damage increases, although this shift is not significant. The direction of maximum α_{qSV} is dependent upon both frequency and damage. This shift is thought to be the result of the induced anisotropy from the cracks as shown in the slowness plots in Fig. 4 as speculated elsewhere.[22] However, further investigation is necessary to determine the precise reason for this peak shift.

VI. CONCLUSIONS

In this article, elastic wave propagation and scattering have been examined for media with uniaxially aligned cracks. These cracks have unit normals that are randomly oriented within a plane of isotropy. The ensemble average elastic wave response is governed by the Dyson equation which is solved within the limits of the first-order smoothing approximation. The general Green's dyadic for a transversely isotropic medium was employed to derive expressions of the attenuation of the shear horizontal, quasilongitudinal and quasishear vertical waves. This dyadic approach is convenient to make the results coordinate free. Thus, the final forms of the attenuations for the three wave types were given directly by simple compact expressions involving integrations over the unit circle. In particular, the integrals are simplified considerably in the Rayleigh regime. The general attenuations for each wave type are dependent on frequency, wave velocity, wave direction and damage density. Finally, numerical results show how the attenuations and the effective wave velocity of each wave type are affected by those parameters. The general formulation is also directly related to other types of elastic wave scattering such as backscatter. The simple form of the results makes them particularly useful for nondestructive testing and materials characterization research. However, the neglect of mutual interactions among the microcracks may have a large influence for the scattering effects. This analysis will be investigated in subsequent work.

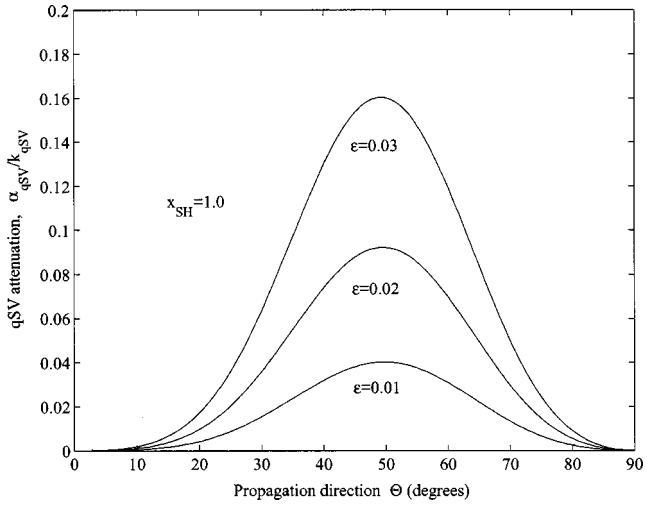


FIG. 9. Angular dependence of the normalized qSV attenuation, α_{qSV}/k_{qSV} for various damage densities ε at frequency $x_{SH}=1.0$.

ACKNOWLEDGMENTS

The National Science Foundation (Grants No. CMS-9978707 and INT-0089548) and the National Bridge Research Organization (NaBRO) are gratefully acknowledged by the authors for supporting this work.

[1] A. K. Mal, "Interaction of elastic waves with a penny-shaped crack," Int. J. Eng. Sci. **8**, 381–388 (1968).
[2] A. K. Mal, "A note on the low frequency diffraction of elastic waves by a Griffith crack," Int. J. Eng. Sci. **10**, 609–612 (1972).
[3] M. Piau, "Attenuation of a plane compressional wave by a random distribution of thin circular cracks," Int. J. Eng. Sci. **17**, 151–167 (1979).
[4] A. Chatterjee, A. K. Mal, L. Knopoff, and J. Hudson, "Attenuation of elastic waves in a cracked, fluid-saturated solid," Math. Proc. Cambridge Philos. Soc. **88**, 547–561 (1980).
[5] J. A. Hudson, "Wave speeds and attenuation of elastic waves in material containing cracks," Geophys. J. R. Astron. Soc. **64**, 133–150 (1981).
[6] P. A. Martin, "Diffraction of elastic waves by a penny-shaped crack," Proc. R. Soc. London, Ser. A **378**, 263–285 (1981).
[7] S. Krenk and H. Schmidt, "Elastic wave scattering by a circular crack," Philos. Trans. R. Soc. London, Ser. A **308**, 167–198 (1982).
[8] P. A. Martin and G. R. Wickham, "Diffraction of elastic waves by a penny-shaped crack: Analytical and numerical results," Proc. R. Soc. London, Ser. A **390**, 91–129 (1983).
[9] C. Zhang and J. Achenbach, "Effective wave velocity and attenuation in a material with distributed penny-shaped cracks," Int. J. Solids Struct. **27**, 751–767 (1991).
[10] C. Zhang and D. Gross, "Wave attenuation and dispersion in randomly cracked solids-I. Slit cracks," Int. J. Eng. Sci. **31**, 841–858 (1993).
[11] C. Zhang and D. Gross, "Wave attenuation and dispersion in randomly cracked solids-II. Penny-shaped cracks," Int. J. Eng. Sci. **31**, 859–872 (1993).
[12] V. P. Smyshlyaev and J. R. Willis, "Linear and nonlinear scattering of elastic waves by microcracks," J. Mech. Phys. Solids **42**, 585–610 (1994).
[13] A. S. Eriksson, A. Bostrom, and S. K. Datta, "Ultrasonic wave propagation through a cracked solid," Wave Motion **22**, 297–310 (1995).
[14] L. Yang and J. A. Turner, "Scattering of elastic waves in damaged media," J. Acoust. Soc. Am. **113**, 2992–3000 (2003).
[15] S. Nemat-Nasser and M. Hori, *Micromechanics: Overall Properties of*

Heterogeneous Materials (North-Holland, Amsterdam, 1993).
[16] M. Kachanov, ''Elastic solids with many cracks and related problems,'' Adv. Appl. Mech. **29**, 259–445 (1993).
[17] D. Krajcinovic, *Damage Mechanics* (North-Holland, Amsterdam, 1996).
[18] F. E. Stanke and G. S. Kino, ''A unified theory for elastic wave propagation in polycrystalline materials,'' J. Acoust. Soc. Am. **75**, 665–681 (1984).
[19] S. Ahmed and R. B. Thompson, ''Propagation of elastic waves in equiaxed stainless-steel polycrystals with aligned [001] axes,'' J. Acoust. Soc. Am. **99**, 2086–2096 (1996).
[20] S. Hirsekorn, ''The scattering of ultrasonic waves in polycrystalline materials with texture,'' J. Acoust. Soc. Am. **77**, 832–843 (1985).
[21] S. Hirsekorn, ''Directional dependence of ultrasonic propagation in textured polycrystals,'' J. Acoust. Soc. Am. **79**, 1269–1279 (1986).
[22] J. A. Turner, ''Elastic wave propagation and scattering in heterogeneous, anisotropic media: Textured polycrystalline materials,'' J. Acoust. Soc. Am. **106**, 541–552 (1999).
[23] U. Frisch, ''Wave propagation in random media,'' in *Probabilistic Methods in Applied Mathematics*, edited by A. T. Barucha-Reid (Academic, New York, 1968), Vol. 1, pp. 75–198.
[24] R. L. Weaver, ''Diffusion of ultrasound in polycrystals,'' J. Mech. Phys. Solids **38**, 55–86 (1990).
[25] F. C. Karal and J. B. Keller, ''Elastic, electromagnetic, and other waves in a random medium,'' J. Math. Phys. **5**, 537–547 (1964).
[26] I. Kunin, *Elastic Media with Microstructure II* (Springer-Verlag, Berlin, 1983).
[27] J. Walsh, ''The effect of cracks on compressibility of rocks,'' J. Geophys. Res. **70**, 381–389 (1965).
[28] B. Budiansky and R. J. O'Connell, ''Elastic moduli of a cracked solid,'' Int. J. Solids Struct. **12**, 81–97 (1976).
[29] K. Z. Markov and J. R. Willis, ''On the two-point correlation function for dispersions of nonoverlapping spheres,'' Math. Methods Appl. Sci. **8**, 359–377 (1998).
[30] J. A. Turner and R. L. Weaver, ''Radiative transfer and multiple scattering of diffuse ultrasound in polycrystalline media,'' J. Acoust. Soc. Am. **96**, 3675–3683 (1994).
[31] J. A. Turner and R. L. Weaver, ''Time dependence of multiply scattered diffuse ultrasound in polycrystalline media,'' J. Acoust. Soc. Am. **97**, 2639–2644 (1995).
[32] J. A. Turner and R. L. Weaver, ''Ultrasonic radiative transfer theory: Effects of a fluid-solid interface,'' J. Acoust. Soc. Am. **98**, 2801–2808 (1995).
[33] F. E. Stanke, ''Spatial autocorrelation functions for calculations of effective propagation constants in polycrystalline materials,'' J. Acoust. Soc. Am. **80**, 1479–1485 (1986).

Interaction of a bounded ultrasonic beam with a thin inclusion inside a plate

Sigfried Vanaverbeke[a] and Oswald Leroy
Interdisciplinary Research Center (IRC), KULAK, Etienne Sabbelaan 53, 8500 Kortrijk, Belgium

Gennady Shkerdin[b]
Department of Electronics and Information Processing (ETRO), Pleinlaan 2, B-1050, Brussels, Belgium

(Received 25 March 2002; revised 23 April 2003; accepted 23 April 2003)

A theoretical study of the reflection of a two-dimensional Gaussian ultrasonic beam, incident at a Lamb angle of a plate containing a thin rectangular inclusion at an arbitrary position, is presented on the basis radiation mode theory. The inclusion is parallel to the plate surface and its thickness is assumed to be much smaller than the ultrasonic wavelength. It is shown that the amplitude and phase of the reflected beam profile can be used for accurate inclusion characterization. However, this only holds for certain internal positions of the inclusion and for material combinations that do not strongly perturb the excitation of Lamb waves in the plate. When these conditions are satisfied, it is possible to define the Lamb waves and the associated experimental conditions for which good estimates can be obtained of the position of the beginning point of the inclusion as well as of the length and the thickness of the inclusion. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1587149]

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

I. INTRODUCTION

In the field of acoustics, numerous attempts have been made to study the scattering of ultrasound caused by various kinds of discontinuities and imperfections in solid and fluid media. The interaction of ultrasonic waves with inclusions is a problem of considerable interest in the field of nondestructive characterization. In the past, many papers were published about wave scattering from solid and fluid inclusions in unbounded media.[1–4] Rhodes and Sachse[1] showed that the diameter and longitudinal wave speed of a cylindrical solid inclusion can be approximately determined from the measurement of the arrival times of scattered ultrasonic pulses. Ben-Menahem *et al.*[2] considered wave scattering from an anisotropic inclusion in an isotropic solid matrix and demonstrated that observations of the spatial patterns of scattered wave amplitudes can be inverted to determine the elastic constants of an inclusion. The detection of voids or inclusions inside a plate has been studied by Wu *et al.*[5] using transient elastic waves generated by a steel ball impact. Teng[6] has used a finite element method to investigate three-dimensional wave scattering from an inclusion in a solid half-space with an external source at its free surface. An important contribution is the recent work of Maslov *et al.*[7] In this paper, a new ultrasonic technique based on acoustic microscopy has been developed. This method allows to distinguish small spherical voids from small spherical elastic and rigid inclusions using phase information extracted from reflected ultrasonic echoes. It was shown that conventional acoustic techniques using only amplitude information and ignoring the phase information do not allow to make this distinction.

Devolder *et al.*[8–10] elaborated an acousto-optic ndt technique which also uses phase information as an essential diagnostic tool. The acousto-optic ndt technique involves insonifying a solid sample by a bounded ultrasonic beam incident at one of its critical angles. In this way, Rayleigh or Lamb waves are generated inside the specimen and the reflected field is split up in a specular and a nonspecular component which are 180° out of phase.[11,12] In the presence of a coating, a defect or an adhesion problem, a significant phase shift appears in the nonspecular lobe.[10] This phase shift can be determined without disturbing the sound field by first measuring the amplitude and phase modulation of a diffracted laser beam in the optical near field behind the reflected sound beam.[8–10] The amplitude and phase of the reflected beam can be reconstructed from these measurements using methods described in detail by Blomme and Leroy and Windels and Leroy.[13–15] It was shown experimentally that the thickness of thin opaque coatings on a solid substrate can be accurately measured using the acousto-optic ndt technique[8] and the method has been verified by various theoretical models.

In a similar way, the acousto-optic ndt technique can be used to detect inclusions inside a plate by choosing a Lamb angle of the plate and measuring the phase variations in the reflected beam caused by the interaction of the excited Lamb mode with the inclusion. Because of the complicated scattering and mode conversion phenomena that arise when a bounded beam insonifies an inclusion, radiation mode theory (RMT) is most appropriate for this kind of problems.[16,17]

Briers *et al.*[18] made use of RMT to analyze the interaction of a Stoneley wave with a rectangular inclusion located near the surface of a liquid/solid interface. This topic was investigated experimentally by Chamuel.[19] The reflection of

[a]Research Assistant for the Flemish Institute for the Encouragement of the Scientific and Technological Research in Industry (IWT).
[b]Permanent address: Institute for Radio Engineering and Electronics, Russian Academy of Sciences, Moscow, Russia.

a Gaussian beam from the same liquid/solid configuration was also studied recently by Briers *et al.*[20] It was shown that the phase of the reflected profile is sensitive to the presence of an inclusion. However, a straightforward method to determine both the length and the thickness of the inclusion did not emerge from their analysis. Moreover, the above-mentioned RMT model dealt with a solid halfspace and has not yet been extended to study bounded beam diffraction on a *plate* containing an inclusion either located at the surface or at the interior of the plate. The purpose of this paper is to fill this gap by considering the most general case: scattering of a bounded Gaussian beam by a rectangular inclusion located at an arbitrary position in the plate.

The outline of the paper is as follows. In Sec. II, the interaction of a bounded beam with a rectangular inclusion is treated using RMT. In order to implement this configuration in our model, the plate with the inclusion is divided in three substructures. Overlap integrals are defined and a final analytical formula is derived. Section III is devoted to numerical simulations. The influence of the length, the thickness, the depth and the material parameters of an inclusion on the amplitude and phase of the reflected field in a Lamb angle of the intact plate is studied. Finally, our conclusions are presented in sec. IV.

II. INTERACTION OF A BOUNDED BEAM WITH A RECTANGULAR INCLUSION

In order to gain a more profound insight into the procedure described in this section, it is useful to recall the basic structure of RMT. RMT is essentially based on the construction of a complete and orthogonal set of acoustic modes in which an arbitrary acoustic field within the considered configuration can be expanded.[21–23] The whole spectrum of acoustic modes within the structure can be regarded as a discrete spectrum of eigenmodes and a continuous spectrum of radiation modes.[21–23] In order to define the concept of a radiation mode, one usually introduces an acoustic wave source located at infinity.[21] Each plane wave emitted by this source is reflected and transmitted through the interfaces between the different media in the considered structure. The combination of incident, reflected and transmitted waves is called a radiation mode. An eigenmode corresponds to a resonance solution of the structure which does not reradiate into the coupling medium (e.g., the two Stoneley modes of a plate[23]). It is important to remark that according to this definition, leaky Rayleigh and Lamb waves are not eigenmodes, because they can be decomposed into radiation modes of the structure. The complete and orthogonal set of acoustic modes for a given structure is constructed by means of suitable orthogonalization procedures.[21]

In order to handle complicated structures, it is often necessary to divide the entire structure into several substructures. The radiation modes and eigenmodes are developed for each of these substructures and the acoustic field in each structure is obtained by requiring the continuity of the displacements and stresses at the vertical interfaces between the substructures. It should be emphasized that RMT has been used extensively in the past to solve a large number of acoustic scattering problems on discontinuous structures.[21–24] The interested reader is referred to earlier papers for a thorough description of the fundamentals and applications of RMT in acoustics.[21–24]

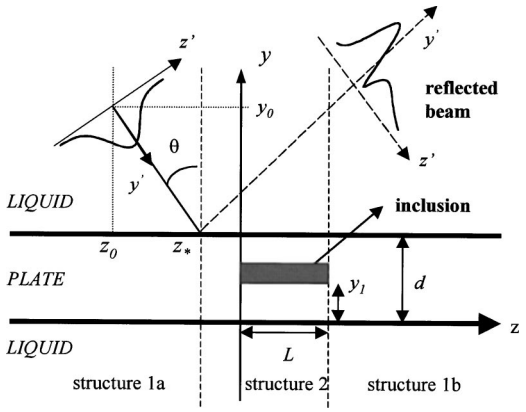


FIG. 1. Reflection of a Gaussian beam from an inclusion in a solid plate. Geometrical configuration and definition of the parameters.

The configuration of the problem considered in this paper is illustrated in Fig. 1. We consider an isotropic solid plate of thickness d immersed in water. The plate contains a rectangular inclusion with length L and thickness d_1 which is assumed to be parallel to the surface of the plate. The plate has density ρ_1, dilatational velocity v_{d1} and shear velocity v_{s1}; the material constants of the inclusion and the liquid are ρ_2, v_{d2}, v_{s2} and ρ, v, respectively. The corresponding Lamé constants will be denoted by λ_1, μ_1 for the plate, λ_2, μ_2 for the inclusion and λ for the liquid. We assume perfect contact between the inclusion and the plate. The plate is insonified from the surrounding liquid by means of a Gaussian ultrasonic beam with incidence angle θ and circular frequency $\omega=2\pi f$. The center of the beam is located at (y_0,z_0). Only two-dimensional configurations will be considered in this paper. In the coordinate system shown in Fig. 1, this means that the incident beam is bounded in the (y,z) plane and assumed to be uniform in the direction orthogonal to the plane of incidence. Our goal is to derive expressions for the reflected sound field with the aim of investigating the influence of the inclusion on the reflected amplitude and phase distribution.

In order to apply RMT to this problem, the configuration is divided into three substructures (see Fig. 1): structure 1a, a liquid/solid/liquid structure ($z\leqslant 0$); structure 1b, also a liquid/solid/liquid structure ($z\geqslant L$) and structure 2 ($0\leqslant z\leqslant L$), a liquid/multilayer/liquid structure. In each substructure, the acoustic field is represented by the orthogonal set of acoustic modes of the considered structure. Structure 2 can be considered as a multilayer which contains three layers. The material constants of the first and the third layer are identical to the material constants of the plate and the second layer is located at the position of the rectangular inclusion. The internal position, the material constants and the thickness of the inclusion can be changed by simply adjusting the properties of the second layer in structure 2.

In the context of RMT, the displacement field in structure 1a is described by[23]

$$\vec{u}_i(y,z)=\sum_{n=1,2}\int_0^k C_i^n(k_{y,i})\vec{u}_1^n(k_{y,i},y,z)dk_{y,i}, \qquad (1)$$

in which the expansion coefficients of the incident acoustic field in structure 1a are denoted by C_i^n (the subscript i refers to the incident field) and $\vec{u}_1^n$ are the displacement fields of the symmetric $(n=1)$ and the antisymmetric $(n=2)$ radiation modes of a liquid/solid/liquid structure. Summation is taken over these two types of radiation modes. The expressions for the radiation modes of a liquid/solid/liquid structure can be found in Ref. 23. The acoustic field in structure 2 is represented by a similar expression,

$$\vec{u}_2(y,z)=\sum_{m=1,2}\int_0^k C_2^m(k_{y,2})\vec{u}_2^m(k_{y,2},y,z)dk_{y,2}, \tag{2}$$

where $\vec{u}_2^m$ are the displacement fields of the addition $(m=1)$ and subtraction $(m=2)$ modes of a multilayer structure with expansion coefficients C_2^m. The radiation modes for the acoustic field in structure 2 are a special case of the radiation modes for a viscoelastic multilayer structure derived by Vandeputte *et al.*[24] Because we do not intend to study the influence of viscoelasticity in this paper, we have used the expressions of Vandeputte *et al.* with the values of the absorption parameters set equal to zero.

Finally, the acoustic field in the second liquid/solid/liquid structure (structure 1b) is given by

$$\vec{u}_1(y,z)=\sum_{p=1,2}\int_0^k C_1^p(k_{y,1})\vec{u}_1^p(k_{y,1},y,z)dk_{y,1}, \tag{3}$$

and the corresponding expansion coefficients are denoted by C_1^p. Because structure 1b is the same as structure 1a, its displacement field is expanded in the same set of radiation modes used to describe the incident field in structure 1a. By requiring that the inclusion is sufficiently thin compared to the ultrasonic wavelength, the reflected acoustic waves at the vertical boundaries of the inclusion can be neglected.

The integration variables in Eqs. (1)–(3) are defined as

$$k_{y,i}=(k^2-k_{z,i}^2)^{1/2} \quad \text{and} \quad k_{y,\alpha}=(k^2-k_{z,\alpha}^2)^{1/2} \quad (\alpha=1,2), \tag{4}$$

where k is the longitudinal wave number in the liquid and $k_{z,i}$ and $k_{z,\alpha}$ are the z components of the wave vectors of the radiation modes that appear in the mode expansions. The corresponding y components are denoted by $k_{y,i}$ and $k_{y,\alpha}$, respectively. The integrals in Eqs. (1)–(3) are only calculated within the interval $(0,k)$. This means that only propagating radiation modes are used in the mode expansions and evanescent modes are neglected.[21] Similar mode expansions can be written down for the components of the stress tensors in structures 1a, 2, and 1b. For example, the stress tensor components in structure 2 are given by

$$T_{ij,2}(y,z)=\sum_{n=1,2}\int_0^k C_2^n(k_{y,2})T_{ij,2}^n(k_{y,2},y,z)dk_{y,2}. \tag{5}$$

Note that Eqs. (1)–(5) are similar to the corresponding equations in Ref. 20, but because we now consider a solid *plate* instead of a half-space, we have to include summations over two different types of radiation modes.

For the numerical simulations reported in this paper, the reflected acoustic field has been calculated at a sufficiently large distance from the plate surface, so that the reflected beam is restricted to structure 1b. In this way, the displacement field of the reflected beam is fully described by Eq. (3). It is also important to remark that the contributions of the Stoneley modes have been neglected in Eqs. (1)–(3). This approximation is valid if the distance of the source of the incident sound field to the plate surface is larger than the penetration depth of the Stoneley modes in the liquid surrounding the plate.

The expansion coefficients C_i^n of the incident sound field in structure 1a can be determined if the incident field is known in a plane perpendicular to the surface of the plate.[21,23] In the intrinsic coordinate system (y',z') shown in Fig. 1, the magnitude of the acoustic displacement field of a Gaussian beam in the plane $z'=0$ is given by the following expression:

$$u_i(y',z'=0)=u_0e^{-(y'/w)^2}, \tag{6}$$

where u_0 is an arbitrary constant and w is the half beamwidth. The methodology for determining the coefficients C_i^n is summarized in Appendix A.

The unknown expansion coefficients C_1^p and C_2^m in structure 2 and structure 1b can be determined by imposing boundary conditions (continuity of displacement and stresses) along the division planes at $z=0$ and $z=L$. We obtain the following solution for the coefficients C_2^m:

$$C_2^m(k_{y,2})=\frac{1}{2Q_r(k_{y,2})}\sum_{n=1,2}\int_0^k C_i^n(k_{y,i}) \times[I_{12}^{n,m}(k_{y,i},k_{y,2})+K_{12}^{n,m}(k_{y,i},k_{y,2})]dk_{y,i}, \tag{7}$$

with $I_{12}^{n,m}$ and $K_{12}^{n,m}$ the overlap integrals,[21]

$$I_{12}^{n,m}(k_{y,i},k_{y,2})=\int_{-\infty}^{+\infty}[T_{yz,1}^n(k_{y,i})u_{y,2}^{m*}(k_{y,2}) - T_{zz,2}^{n*}(k_{y,2})u_{z,1}^m(k_{y,i})]_{z=0}\,dy, \tag{8}$$

$$K_{12}^{n,m}(k_{y,i},k_{y,2})=\int_{-\infty}^{+\infty}[T_{zz,1}^n(k_{y,i})u_{z,2}^{m*}(k_{y,2}) - T_{yz,2}^{m*}(k_{y,2})u_{y,1}^n(k_{y,i})]_{z=0}\,dy. \tag{9}$$

These integrals describe the coupling of a radiation mode of type "n" in structure 1a with a radiation mode of type "m" in structure 2 in the plane $z=0$. Applying boundary conditions at $z=L$ leads to the following expressions for the expansion coefficients C_1^n of the acoustic field in structure 1b:

$$C_1^p(k_{y,1})=\frac{1}{2Q_r(k_{y,1})}\sum_{m=1,2}\int_0^k C_2^m(k_{y,2})[I_{21}^{m,p}(k_{y,2},k_{y,1}) + K_{21}^{m,p}(k_{y,2},k_{y,1})]e^{i(k_{z,2}-k_{z,1})L}\,dk_{y,2}. \tag{10}$$

The overlap integrals $I_{21}^{m,p}$ and $K_{21}^{m,p}$ describe the interaction between a radiation mode of type "m" in structure 2 and a radiation mode of type "p" in structure 1b at $z=L$. The following symmetry properties hold if the wave vector component $k_{y,i}$ in Eqs. (8) and (9) is replaced by $k_{y,1}$:

$$I_{21}^{m,p}=-(K_{12}^{p,m})^*, \quad K_{21}^{m,p}=-(I_{12}^{p,m})^*. \tag{11}$$

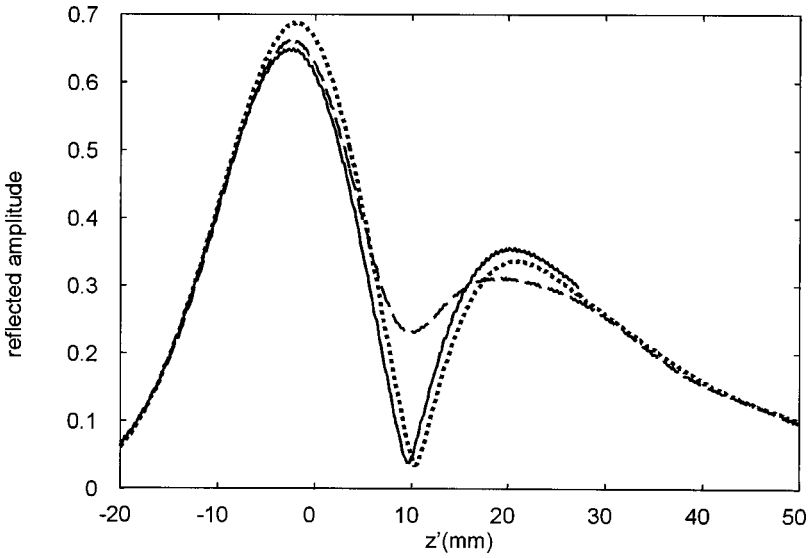


FIG. 2. Amplitude distribution of the reflected beam for different values of z_*; $y_1=0.75$ mm, $L=15$ mm, $d_1=25\ \mu$m. Continuous line, $z_*=-30$ mm; dashed line, $z_*=0$ mm; dotted line, $z_*=15$ mm.

Substitution of Eqs. (7) in Eqs. (10) finally results in the following expressions for the expansion coefficients C_1^p in structure 1b:

$$C_1^p(k_{y,1}) = -\frac{1}{4Q_r(k_{y,1})}\sum_{m,n=1,2}\int_0^k\int_0^k \frac{dk_{y,2}\,dk_{y,i}}{Q_r(k_{y,2})} C_i^n(k_{y,i}) \times[I_{12}^{n,m}(k_{y,i},k_{y,2}) + K_{12}^{n,m}(k_{y,i},k_{y,2}) \times[I_{12}^{p,m*}(k_{y,1},k_{y,2}) + K_{12}^{p,m*}(k_{y,1},k_{y,2})]e^{i(k_{z,2}-k_{z,1})L}. \quad (12)$$

The evaluation of the overlap integrals $I_{12}^{n,m}$, $K_{12}^{n,m}$ is explained in more detail in Appendix B. By substituting these expansion coefficients in Eq. (3), we can calculate the amplitude and phase distribution of the reflected beam in structure 1b.

III. NUMERICAL RESULTS

We now present numerical simulations obtained by the theoretical approach of RMT. We consider a water loaded steel plate of thickness $d=1.5$ mm with a brass inclusion in its interior. The ultrasonic frequency is 4 MHz, $w=12$ mm, and $y_0=150$ mm. The incidence angle corresponds to the first antisymmetric (A1) Lamb mode of the plate without inclusion ($\theta=24.96°$). Unless otherwise mentioned, this setup and the values of the corresponding parameters will not be changed throughout this section. The material parameters for water, steel and brass are listed in Table I. We have also included the parameters for lead which will be used later on in the discussion.

TABLE I. Values of the material parameters used in the numerical calculations.

Material	Density (kg/m^3)	Shear velocity (m/s)	Longitudinal velocity (m/s)
Water	1000		1480
Steel	7900	3100	5790
Brass	8600	2260	4660
Lead	11 200	700	2200

We first consider a configuration for which the depth of the inclusion is known. The interior position y_1 of the inclusion could for example be determined by an independent pulse-echo experiment. We choose $y_1=0.75$ mm and consider an inclusion with $L=15$ mm and $d_1=25\ \mu$m. Our aim is to study the influence of the z-coordinate of the central position (marked as z_* in Fig. 1) of the incident beam on the reflected Lamb profile. Therefore we move the incident beam parallel to the surface of the plate along the $y=y_0$ axis. As a result, z_* varies from negative to positive values. In this way we can simulate a scan over the region of the plate where the inclusion is located. We investigate the influence of different values of z_* on the reflected beam profiles in the phase plane at $y'=164$ mm. A number of resulting amplitude profiles are shown in Fig. 2.

If the meeting point is far in front of the beginning point of the inclusion, the incident beam does not feel the presence of the inclusion and the reflected beam profile is a well-known Lamb profile.[11] This behavior is illustrated by the continuous line ($z_*=-30$ mm). The reflected beam consists

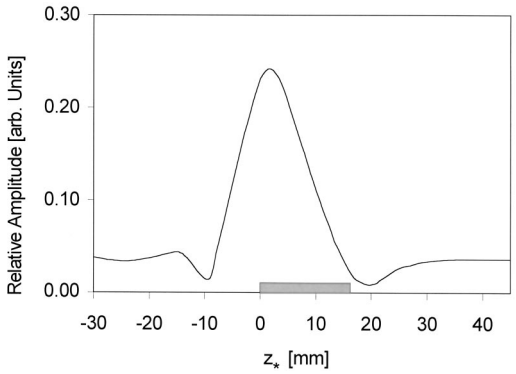


FIG. 3. Simulation of a scan performed by translating the incident beam along the plate. The amplitude of the local minimum between the specular and the nonspecular lobe of the reflected beam is plotted as a function of z_* ($y_1=0.75$ mm, $L=15$ mm, $d_1=25\ \mu$m). The position of the inclusion is indicated by a rectangle.

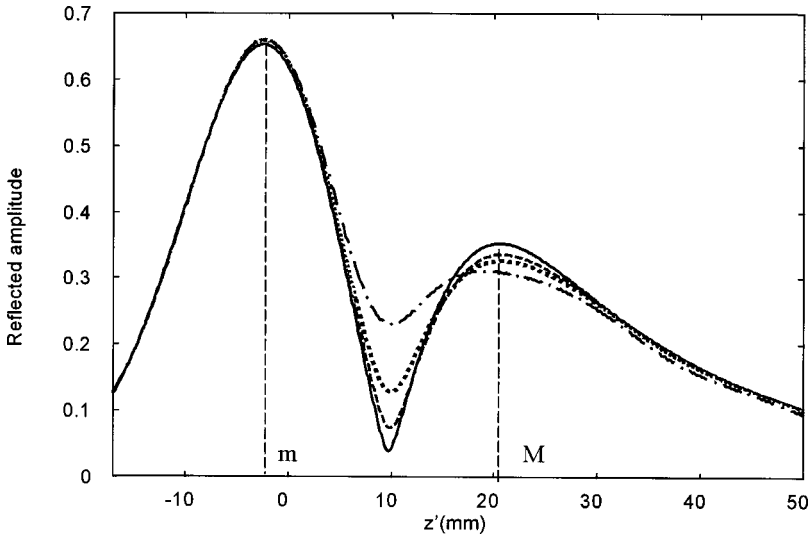


FIG. 4. Amplitude distribution of the reflected beam for different values of the thickness of the inclusion ($y_1 = 0.75$ mm, $L = 15$ mm, $z_* = 0$ mm). Continuous line, no inclusion; dashed line, $d_1 = 10\,\mu$m; dotted line, $d_1 = 15\,\mu$m; dashed–dotted line, $d_1 = 25\,\mu$m. The position of the maxima of the specular and the nonspecular lobe are indicated by *m* and *M*, respectively.

of a specular and a nonspecular lobe separated by a clear minimum in the sound intensity. For later reference, we will call the relative position of this minimum in this case the ''undisturbed'' minimum. The dashed and dotted lines illustrate the profiles when the meeting point coincides with the beginning point ($z_* = 0$ mm) and the end point ($z_* = 15$ mm) of the inclusion, respectively. The amplitude of the reflected beam in the neighborhood of the undisturbed minimum position is clearly sensitive to the presence of the inclusion. The sound amplitude at the position of the minimum increases strongly if the meeting point coincides with the beginning point of the inclusion but falls back almost to its undisturbed level at the second edge of the inclusion. This result suggests that useful information on the location and size of the inclusion can be obtained from amplitude measurements in the reflected beam profile.

In order to investigate in more detail the behavior of the amplitude at the position of the ''undisturbed'' minimum, we have plotted the amplitude at this point in the reflected beam profile as a function of z_* in Fig. 3. It is not surprising to see that the incident beam starts to feel the presence of the defect at $z_* \approx -15$ mm because the half beamwidth is close to the length of the inclusion ($w = 12$ mm). The amplitude increases towards a maximum at a z_* location which almost coincides with the beginning point of the inclusion and subsequently declines over a distance which is approximately equal to its length. Figure 3 clearly shows that semiquantitative estimates of the position of the beginning point and the length of an inclusion can be obtained by scanning the surface of the plate and measuring the sound intensity at the minimum between the two lobes of the reflected beam. The position of this minimum can be readily found by scanning the profile of the reflected beam on a region of the sample with no inclusions. We do not consider the transmitted beam profiles separately, because we found that the presence of the inclusion does not induce appreciable amplitude changes in the transmitted sound field.

Figure 4 illustrates the reflected beam profiles when the meeting point coincides with the beginning point of the inclusion and the thickness is varied. The reflected beam profiles are plotted for $d_1 = 10\,\mu$m (dashed line), $d_1 = 15\,\mu$m (dotted line), and $d_1 = 25\,\mu$m (dashed–dotted line) with $L = 15$ mm. The continuous line shows the reflected profile on a plate with no inclusion. We observe that the amplitude of the minimum increases monotonically with growing thickness of the inclusion. Variations in the length of the inclusion

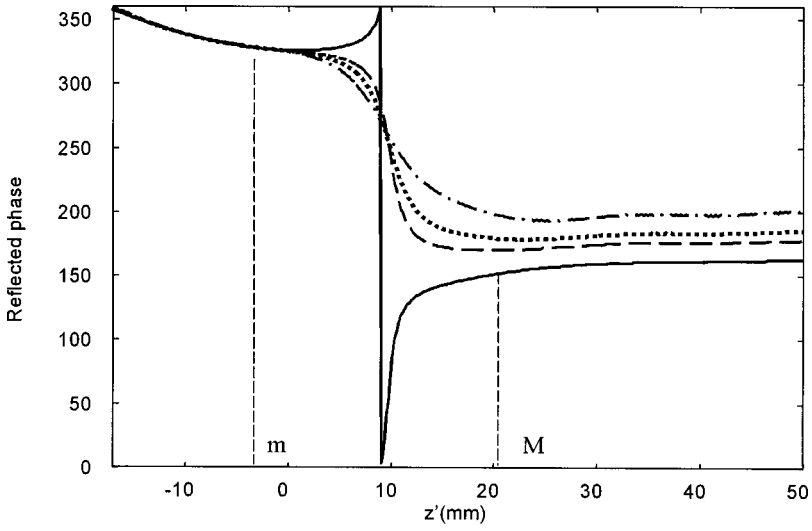


FIG. 5. Phase distribution of the reflected beam for as a function of the thickness of the inclusion ($y_1 = 0.75$ mm, $L = 15$ mm, $z_* = 0$ mm). Continuous line, no inclusion; dashed line, $d_1 = 10\,\mu$m; dotted line, $d_1 = 15\,\mu$m; dashed–dotted line, $d_1 = 25\,\mu$m. The position of the maxima of the specular and the nonspecular lobe are indicated by *m* and *M*, respectively.

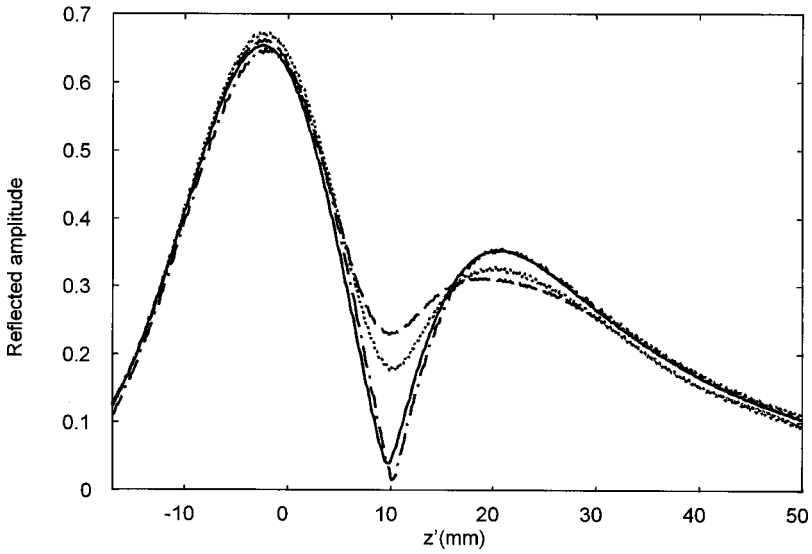


FIG. 6. Amplitude distribution of the reflected beam for different internal positions of the inclusion ($L=15$ mm, $d_1=25\ \mu$m, $z_*=0$ mm). Continuous line, no inclusion; dashed line, $y_1=0.75$ mm; dotted line, $y_1=1.25$ mm; dashed–dotted line, $y_1=1.475$ mm.

at fixed thickness have a similar effect which is not shown separately in this paper. This analysis means that the amplitude of the minimum between the two lobes of the reflected beam is proportional to the surface area ($d_1.L$) of the inclusion. If the length of the inclusion is first estimated by performing a scan as suggested above, the thickness can be estimated by comparing the recorded sound intensity with the numerical predictions obtained by RMT for an inclusion with the appropriate length.

Figure 5 shows the phase distributions which correspond to the amplitude profiles presented in Fig. 4. The position of the maxima of the specular lobe and the nonspecular lobe in the case of a plate with no inclusions have been indicated by *m* and *M* in Figs. 4 and 5, respectively. The position of the maxima of the two lobes in the reflected beam is almost unaffected by the presence of the inclusion. These two points can therefore serve as useful reference points in the reflected beam profile. Figure 5 shows that the phase at the maximum of the specular lobe is almost unaffected by the presence of the inclusion. On the other hand, the phase at the maximum of the nonspecular lobe changes significantly. The *phase difference* between the maximum of the specular lobe and the maximum of the nonspecular lobe gradually becomes smaller when the thickness of the inclusion increases. This phase difference thus provides another method to determine the thickness of the inclusion using the acousto-optic ndt technique mentioned in the introduction.

The feasibility of accurate inclusion characterization as illustrated in the preceding paragraphs strongly depends on the internal position and the material parameters of the inclusion. Figures 6 and 7 illustrate the influence of the internal position of the inclusion on the amplitude and phase distribution of the reflected beam. The parameters of the inclusion are the same as in Fig. 2, but we now compare the reflected beam profiles for several internal positions of the inclusion: $y_1=0.75$ mm (dashed lines), $y_1=1.25$ mm (dotted lines), and $y_1=1.475$ mm (dashed–dotted lines). The central axis of the incident beam reaches the plate at $z_*=0$ mm. The continuous curves represent the amplitude and phase distribution for a plate without inclusion. Figures 6 and 7 show that the amplitude and phase changes caused by the inclusion are strongly reduced when the inclusion is located at the plate surface. Significantly larger effects are seen when it is located in the bulk of the plate. The symmetry properties of

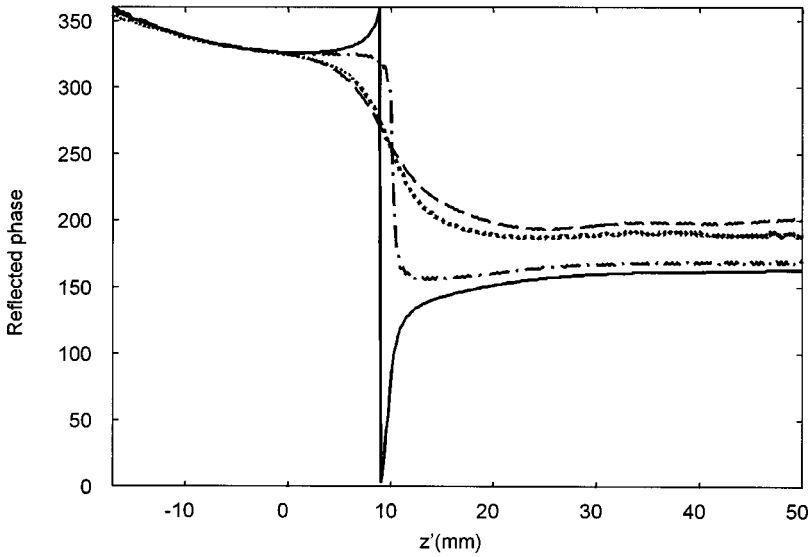


FIG. 7. Phase distribution of the reflected beam for different internal positions of the inclusion ($L=15$ mm, $d_1=25\ \mu$m, $z_*=0$ mm). Continuous line, no inclusion; dashed line, $y_1=0.75$ mm; dotted line, $y_1=1.25$ mm; dashed–dotted line, $y_1=1.475$ mm.

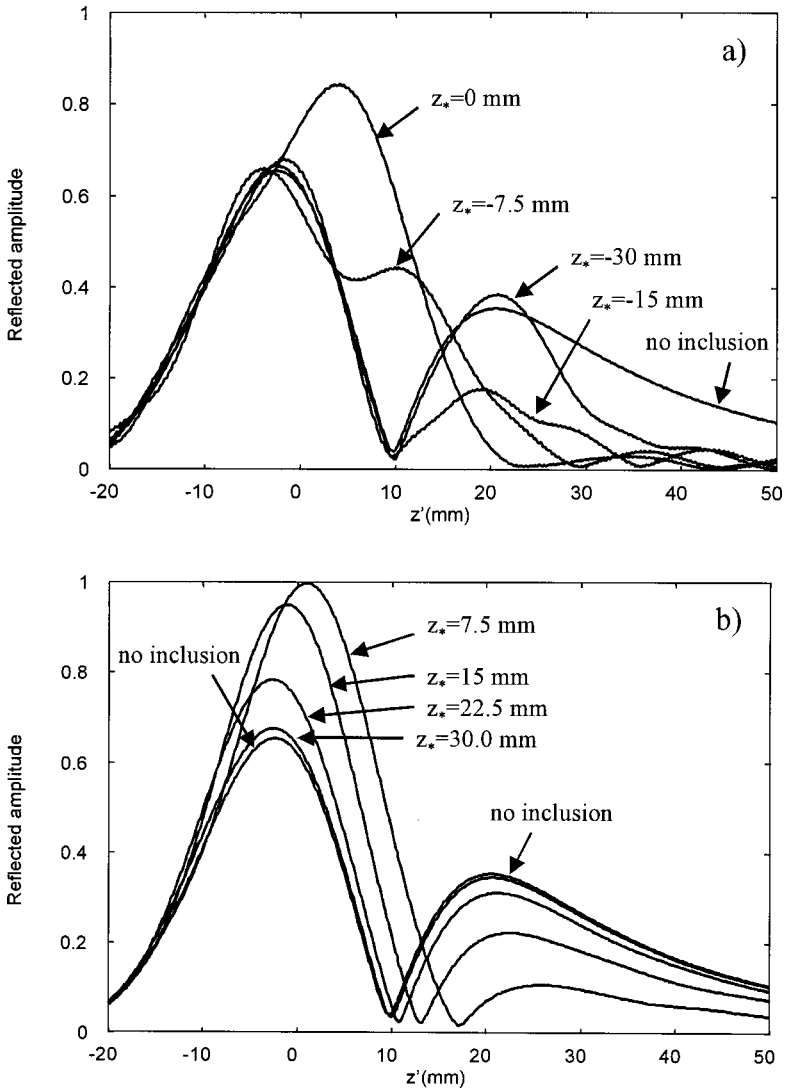


FIG. 8. Reflected amplitude distribution for a steel plate without inclusion as well as at different values of z_* on a steel plate with a lead inclusion ($L=15$ mm, $d_1=25\ \mu$m, $y_1=0.75$ mm). (a) Reflected profiles for $z_*\leqslant 0$, (b) reflected profiles for $z_*>0$.

Lamb waves lead to a reflection symmetry with respect to the midplane of the plate. Apart from this ambiguity, the amplitude and phase changes are generally not a unique function of the interior position of an inclusion. It is therefore not possible to separate the effects of the position and the dimension of an inclusion and to determine all these parameters individually from measurements in a single Lamb angle. It is possible that the degeneracy can be removed by combining several Lamb angles and working at different frequencies. This is an area that we are currently exploring in greater detail.

The influence of the material parameters is illustrated in Fig. 8. We consider a lead/steel configuration with $L=15$ mm, $d_1=25\ \mu$m, $y_1=0.75$ mm, and $\theta=24.96°$ [the Lamb angle of the antisymmetric (A1) mode of a steel plate without inclusion]. Figure 8 shows the reflected beam profile for a steel plate without inclusion as well as the reflected profiles at different values of z_* in the case of a steel plate with a lead inclusion. We observe that the nonspecular lobe of the reflected beam is strongly suppressed as the central axis of the incident beam approaches the beginning point of the inclusion [Fig. 8(a)]. The reflected beam profile at the location of the beginning point is almost specular and has no clear features. The nonspecular lobe subsequently regains strength for positive z_* locations representing insonification on top of and behind the inclusion [Fig. 8(b)]. The reflected profile returns to the Lamb profile of an intact plate at $z_*\approx 30$ mm. This behavior of the reflected beam can be explained by the fact that the resonant excitation of the Lamb wave in the plate is completely disturbed because of the large impedance mismatch between lead and steel. Figure 8 shows that it is difficult to obtain accurate estimates of the location of the beginning point and the dimension of the inclusion because the amplitude changes of the reflected beam occur over a distance which is considerably larger than the length of the inclusion and because useful reference points in the reflected beam profile cannot be easily identified. Accurate characterization of inclusions is therefore only possible for material combinations for which the excitation of Lamb waves is only slightly perturbed.

IV. CONCLUSIONS

Radiation mode theory has been used to develop a theoretical model for the reflection of a Gaussian beam incident at a Lamb angle on a solid plate containing a thin rectangular inclusion. The inclusion is assumed to be parallel to the plate surface. We have shown that the amplitude and phase of the

reflected beam profile can be used for accurate inclusion characterization, but only for suitable internal positions of the inclusion and for material combinations that do not strongly perturb the excitation of Lamb waves in the plate. When these conditions are satisfied, it is possible to obtain good estimates of the position of the beginning point of the inclusion as well as the length and the thickness of the inclusion. Without prior knowledge of the internal position of an inclusion, its internal position and dimensions cannot be unambiguously determined using measurements performed in a single Lamb angle of the plate. The inversion of the geometrical and material parameters of an inclusion from measurements performed in several Lamb angles remains a topic for further research.

ACKNOWLEDGMENTS

The first author (S.V.) would like to thank the Flemish Institute for the Encouragement of the Scientific and Technological Research in Industry (IWT) for their financial support. The authors thank Professor K. Van Den Abeele for useful comments which helped to improve the contents of the paper.

APPENDIX A

For well-bounded beams ($w \gg \lambda$, λ being the wavelength in the liquid), z component of the displacement field of the incident beam in the plane $z=z_0$ can be obtained by simply projecting Eq. (6) on the $z=z_0$ axis:

$$u_{i,z}(y,z=z_0)=u_0 \sin(\theta) e^{[(y-y_0)^2 \sin^2(\theta)/w^2]} e^{-ik\cos(\theta)(y-y_0)}. \tag{A1}$$

The coefficients $C_i^n(k_{y,i})$ can be found by evaluating the following integrals:[21,23]

$$C_i^n(k_{y,i})=-\frac{\int_{-\infty}^{+\infty}[u_{i,z}(y,z=z_0)T_{zz,l}^{n}{}^*(k_{y,i})]}{Q_r(k_{y,i})}, \tag{A2}$$

where the function $Q_r(k_{y,i})$ is given by

$$Q_r(k_{y,i})=i\pi\lambda k^2 k_{z,i}|A|^2, \tag{A3}$$

with A an arbitrary normalization constant and $k_{z,i}$ defined in accordance with the relations (5). $T_{zz,l}^{n}{}^*$ is the stress-tensor component of a fixed radiation mode of type n in structure 1a. Substituting Eq. (A1) in Eq. (A2) and assuming that $y_0 > 3w$ results in the following expressions for the coefficients C_i^n:

$$\begin{aligned} C_i^n(k_{y,i})=\frac{u_0 w k^2}{\sqrt{\pi}k_{z,i}}(&e^{ik_{y,i}(y_0-d)}e^{-i\varphi_n(k_{y,i})} \\ &\times e^{-[w^2(k\cos(\theta)-k_{y,i})^2/\sin^2(\theta)]}-e^{-ik_{y,i}(y_0-d)} \\ &\times e^{i\varphi_n(k_{y,i})}e^{-[w^2(k\cos(\theta)+k_{y,i})^2/\sin^2(\theta)]}), \end{aligned} \tag{A4}$$

where the functions $\varphi_n(k_{y,i})$ $(n=1,2)$ represent the phase of the type 1 and type 2 radiation modes in structure 1a. These phases were derived in Ref. 23.

APPENDIX B

The overlap integrals $I_{12}^{n,m}$ and $K_{12}^{n,m}$ can be evaluated by splitting the integration interval in three parts: in the liquid below the plate ($-\infty<y<0$), inside the plate ($0\leqslant y\leqslant d$), and in the liquid above the plate ($d<y<\infty$). The resulting expressions can be written as

$$\begin{aligned} I_{12}^{n,m}(k_{y,i},k_{y,2})=|A|^2 k_{z,i}\lambda k^2[&\alpha_{n,m}(k_{y,i},k_{y,2})\pi\delta(k_{y,i}-k_{y,2}) \\ &+I_e^{n,m}(k_{y,i},k_{y,2})+I_s^{n,m}(k_{y,i},k_{y,2})], \end{aligned} \tag{B1}$$

$$\begin{aligned} K_{12}^{n,m}(k_{y,i},k_{y,2})=|A|^2 k_{z,2}\lambda k^2[&\alpha_{n,m}(k_{y,i},k_{y,2})\pi\delta(k_{y,i}-k_{y,2}) \\ &+I_e^{n,m}(k_{y,i},k_{y,2})+K_s^{n,m}(k_{y,i},k_{y,2})]. \end{aligned} \tag{B2}$$

The contributions of the overlap integrals over the interior of the plate are contained in the functions $I_s^{n,m}(k_{y,i},k_{y,2})$ and $K_s^{n,m}(k_{y,i},k_{y,2})$. The analytical form of these integrals in very complicated and too cumbersome to be listed explicitly. The integrals cannot be evaluated in closed form and were computed numerically.

The values of the functions $I_e^{n,m}$ are determined as follows:

$$\begin{aligned} I_e^{1,1}(k_{y,i},k_{y,2})=\frac{1+e^{i\psi(k_{y,2})}}{4}\Bigg[&iP\left(\frac{1}{k_{y,2}-k_{y,i}}\right) \\ &\times(e^{i(\chi(k_{y,2})+\varphi_1(k_{y,i}))}-e^{-i\varphi_1(k_{y,i})}) \\ &+\frac{i(e^{i\varphi_1(k_{y,i})}-e^{i(\chi(k_{y,2})-\varphi_1(k_{y,i}))})}{k_{y,i}+k_{y,2}}\Bigg], \end{aligned} \tag{B3}$$

$$\begin{aligned} I_e^{2,2}(k_{y,i},k_{y,2})=\frac{1+e^{i\psi(k_{y,2})}}{4}\Bigg[&iP\left(\frac{1}{k_{y,2}-k_{y,i}}\right) \\ &\times(e^{i(\chi_1(k_{y,2})+\varphi_2(k_{y,i}))}-e^{-i\varphi_2(k_{y,i})}) \\ &+\frac{i(e^{i\varphi_2(k_{y,i})}-e^{i(\chi_1(k_{y,2})-\varphi_2(k_{y,i}))})}{k_{y,i}+k_{y,2}}\Bigg], \end{aligned} \tag{B4}$$

$$\begin{aligned} I_e^{1,2}(k_{y,i},k_{y,2})=\frac{e^{i\psi(k_{y,2})}-1}{4}\Bigg[&-iP\left(\frac{1}{k_{y,2}-k_{y,i}}\right) \\ &\times(e^{i(\chi_1(k_{y,2})+\varphi_1(k_{y,i}))}+e^{-i\varphi_1(k_{y,i})}) \\ &+\frac{i(e^{i\varphi_1(k_{y,i})}+e^{i(\chi_1(k_{y,2})-\varphi_1(k_{y,i}))})}{k_{y,i}+k_{y,2}}\Bigg], \end{aligned} \tag{B5}$$

$$\begin{aligned} I_e^{2,1}(k_{y,i},k_{y,2})=\frac{e^{i\psi(k_{y,2})}-1}{4}\Bigg[&-iP\left(\frac{1}{k_{y,2}-k_{y,i}}\right) \\ &\times(e^{i(\chi(k_{y,2})+\varphi_2(k_{y,i}))}+e^{-i\varphi_2(k_{y,i})}) \\ &+\frac{i(e^{i\varphi_2(k_{y,i})}+e^{i(\chi(k_{y,2})-\varphi_2(k_{y,i}))})}{k_{y,i}+k_{y,2}}\Bigg], \end{aligned} \tag{B6}$$

where P stands for the principal value function and the angles χ_1, χ, and ψ are defined as

$$e^{i\chi}=-\frac{2(\beta_0+\beta_{01})-K_1\sqrt{1+\left(\frac{K}{K_1}\right)^2}\left(\frac{\lambda k^2}{k_{y,2}}\beta_1-\frac{k_{y,2}}{\lambda k^2}\beta_2\right)+iH}{K^2+K_1^2+4}, \tag{B7}$$

$$e^{i\chi_1}=\frac{2(\beta_0+\beta_{01})+K_1\sqrt{1+\left(\frac{K}{K_1}\right)^2}\left(\frac{\lambda k^2}{k_{y,2}}\beta_1-\frac{k_{y,2}}{\lambda k^2}\beta_2\right)+iH}{K^2+K_1^2+4}, \tag{B8}$$

$$e^{i\psi}=\frac{1+i\frac{K}{K_1}}{\sqrt{1+\left(\frac{K}{K_1}\right)^2}}, \tag{B9}$$

with

$$H=2\frac{\lambda k^2}{k_{y,2}}\beta_1-2\frac{k_{y,2}}{\lambda k^2}\beta_2+K_1\sqrt{1+\left(\frac{K}{K_1}\right)^2}(\beta_1+\beta_{01}), \tag{B10}$$

$$K=\beta_0-\beta_{01}, \tag{B11}$$

and

$$K_1=\frac{\lambda k^2}{k_{y,2}}\beta_1+\frac{k_{y,2}}{\lambda k^2}\beta_2. \tag{B12}$$

The quantities β_1, β_2, β_0, and β_{01} are given by

$$\beta_1=A_{21}-\frac{A_{43}A_{21}}{A_{41}}, \tag{B13}$$

$$\beta_2=A_{32}-\frac{A_{42}A_{31}}{A_{41}}, \tag{B14}$$

$$\beta_0=A_{33}-\frac{A_{43}A_{31}}{A_{41}}, \tag{B15}$$

$$\beta_{01}=A_{22}-\frac{A_{21}A_{43}}{A_{41}}, \tag{B16}$$

where A_{ij} $(i,j=1,4)$ are the elements of the transfer matrix[25] of a multilayer structure. The functions $\alpha_{1,1}(k_{y,i},k_{y,2})$ and $\alpha_{1,2}(k_{y,i},k_{y,2})$ are defined as

$$\alpha_{1,1}(k_{y,i},k_{y,2})=\frac{(1+e^{i\psi(k_{y,2})})(e^{-i\varphi_1(k_{y,i})}+e^{i(\chi(k_{y,2})+\varphi_1(k_{y,i}))})}{4}, \tag{B17}$$

$$\alpha_{1,2}(k_{y,i},k_{y,2})=\frac{(e^{i\psi(k_{y,i})}-1)(e^{-i\varphi_1(k_{y,i})}+e^{i(\chi_1(k_{y,2})+\varphi_1(k_{y,i}))})}{4}, \tag{B18}$$

and the functions $\alpha_{2,1}$ and $\alpha_{2,2}$ are identical to $\alpha_{1,2}$ and $\alpha_{1,1}$ apart from the following substitutions: $\varphi_1\rightarrow\varphi_2$, $\chi_1\rightarrow\chi$ and $\varphi_1\rightarrow\varphi_2$, $\chi_1\rightarrow\chi$, respectively.

[1] D. Rhodes and W. Sachse, "Arrival times of scattered ultrasonic signals from a solid inclusion in an elastic material," J. Acoust. Soc. Am. **65**, 1116–1120 (1979).

[2] A. Ben-Menahem and R. L. Gibson, "Scattering of elastic waves by localized anisotropic inclusions," J. Acoust. Soc. Am. **87**, 2300–2309 (1990).

[3] L. Flax and H. Überall, "Resonant scattering of elastic waves from spherical solid inclusions," J. Acoust. Soc. Am. **67**, 1432–1441 (1980).

[4] B. Peterson, V. K. Varadan, and V. V. Varadan, "Scattering of elastic waves by a fluid inclusion," J. Acoust. Soc. Am. **73**, 1487–1493 (1983).

[5] T.-T. Wu and J.-H. Hong, "Application of transient elastic waves to the non-destructive evaluation of plate structure with cavity or inclusion," J. Acoust. Soc. Am. **94**, 1453–1460 (1993).

[6] Y.-C. Teng, "Three-dimensional finite element analysis of waves in an acoustic medium with inclusion," J. Acoust. Soc. Am. **86**, 414–422 (1989).

[7] K. L. Maslov, T. Kundu, and O. I. Lobkis, "Acoustic microscopy for spherical inclusion characterization," J. Acoust. Soc. Am. **100**, 80–85 (1996).

[8] S. Devolder, M. Wevers, P. Demeester, and O. Leroy, "Thin layer measurements based on the acousto-optic ndt-technique," Appl. Phys. Lett. **68**, 1732–1734 (1996).

[9] M. Wevers, S. Devolder, P. Demeester, and O. Leroy, "Acousto-optic technique: A new non-destructive technique to evaluate thin layered structures," Appl. Phys. Lett. **66**, 1466–1468 (1995).

[10] S. Devolder, "A new acousto-optic nondestructive technique for the quality control of layered materials," Ph.D. thesis, K. U. Leuven Campus, Kortrijk, 1996.

[11] L. Pitts and T. Plona, "Theory of non-specular reflection effects for an ultrasonic beam incident on a solid plate in a liquid," IEEE Trans. Sonics Ultrason. **SU-24(2)**, 101–109 (1977).

[12] H. L. Bertoni and T. Tamir, "Unified theory of Rayleigh-angle phenomena for acoustic beams at liquid-solid interfaces," Appl. Phys. **2**, 157–172 (1973).

[13] E. Blomme and O. Leroy, "Plane wave analysis of the near field of light diffracted by ultrasound," J. Acoust. Soc. Am. **91**, 1474–1483 (1992).

[14] F. Windels and O. Leroy, "Acousto-optical sound pulse analysis under Raman–Nath conditions and beyond," Acust. Acta Acust. **87**, 184–190 (2001).

[15] F. Windels and O. Leroy, "The acousto-optical interaction of narrow laser beams under Raman–Nath conditions," J. Opt. A, Pure Appl. Opt. **3**, S12–S17 (2001).

[16] F. Windels, S. Vanaverbeke, and O. Leroy, "Thin coating characterization by Rayleigh waves: An analytical model based on normal-mode theory," J. Acoust. Soc. Am. **110**, 1349–1359 (2001).

[17] A. Neubrand and P. Hess, "Laser generation and detection of surface acoustic waves: elastic properties of surface layers," J. Appl. Phys. **71**, 227–238 (1992).

[18] R. Briers, O. Leroy, and G. N. Shkerdin, "Influence of an inclusion on the phase of a Stoneley wave," J. Acoust. Soc. Am. **101**, 1358–1364 (1997).

[19] J. R. Chamuel, "Ultrasonic studies of transient seismo-acoustic waves in bonded solids and liquid/solid interfaces," Sonoquest Report No. JRC-34-91, 1991, National Technical Information Document No. AD-A243441.

[20] R. Briers, O. Leroy, and G. N. Shkerdin, "Bounded beam interaction with thin inclusions: Characterization by phase differences at Rayleigh angle incidence," J. Acoust. Soc. Am. **108**, 1622–1630 (2000).

[21] R. Briers, O. Leroy, G. N. Shkerdin, and Y. Gulyaev, "Mode theory as a framework for the investigation of the generation of a Stoneley wave at a liquid/solid interface," J. Acoust. Soc. Am. **95**, 1953–1966 (1994).

[22] J. Vandeputte, O. Leroy, R. Briers, and G. N. Shkerdin, "Extension of the

mode method for viscoelastic media and focused ultrasonic beams,'' J. Acoust. Soc. Am. **108**, 1614–1621 (2000).
[23] R. Briers, O. Leroy, and G. N. Shkerdin, ''Conversion of a Stoneley wave at the extremity of a fluid loaded plate,'' J. Acoust. Soc. Am. **101**, 1347–1357 (1997).
[24] J. Vandeputte, G. N. Shkerdin, and O. Leroy, ''Radiation mode model for multilayered structures,'' in *Mechanical Waves for Composite Structures Evaluation*, edited by D. A. Sotiropoulos (Kluwer Academic, Dordrecht, The Netherlands, 2001).
[25] L. Brekhovskikh, *Waves in Layered Media* (Academic, San Diego, 1980).

The reflection of the fundamental torsional mode from cracks and notches in pipes

A. Demma, P. Cawley, and M. Lowe
Department of Mechanical Engineering, Imperial College, London SW7 2BX, United Kingdom

A. G. Roosenbrand
Shell Global Solutions, P.O. Box 38000, 1030 BN Amsterdam, The Netherlands

(Received 1 May 2002; accepted for publication 12 April 2003)

A quantitative study of the reflection of the $T(0,1)$ mode from defects in pipes in the frequency range 10–300 kHz has been carried out, finite element predictions being validated by experiments on selected cases. Both cracklike defects with zero axial extent and notches with varying axial extents have been considered. The results show that the reflection coefficient from axisymmetric cracks increases monotonically with depth at all frequencies and increases with frequency at any given depth. In the frequency range of interest there is no mode conversion at axisymmetric defects. With nonaxisymmetric cracks, the reflection coefficient is a roughly linear function of the circumferential extent of the defect at relatively high frequencies, the reflection coefficient at low circumferential extents falling below the linear prediction at lower frequencies. With nonaxisymmetric defects, mode conversion to the $F(1,2)$ mode is generally seen, and at lower frequencies the $F(1,3)$ mode is also produced. The depth and circumferential extent are the parameters controlling the reflection from cracks; when notches having finite axial extent, rather than cracks, are considered, interference between the reflections from the start and the end of the notch causes a periodic variation of the reflection coefficient as a function of the axial extent of the notch. The results have been explained in terms of the wave-number-defect size product, ka. Low frequency scattering behavior is seen when $ka<0.1$, high frequency scattering characteristics being seen when $ka>1$. [DOI: 10.1121/1.1582439]

PACS numbers: 43.20.Fn, 43.20.Mv [DEC]

I. INTRODUCTION

The presence of defects in pipelines is a major concern in the oil and chemical industries and NDT techniques are required to assess the integrity of pipes in service. Conventional ultrasonic testing such as local thickness gauging uses bulk waves and only tests the region of pipe below the transducer. An alternative ultrasonic technique is to use guided waves for long-range inspection. This can be done by using a pulse-echo arrangement from a single location on a pipe. Using this configuration waves propagate along the length of the pipe in both directions from the point where the wave is excited and all of the reflected signals are detected and analyzed. It is therefore important to quantify the reflection coefficient from defects of different size and shape.

The interaction of cylindrically guided waves with discontinuities in the geometry of the waveguide is a topic that has stimulated a great deal of interest.[1–10] The ability of guided waves to locate cracks and notches in pipes has been documented.[1–5] Furthermore the effect of defect size on the reflection and transmission characteristics has been investigated by many researchers.[6–11] There are some exact closed-form analytical solutions to elastodynamic scattering problems in the case of bulk waves. Closed form solutions for scattering of bulk waves can be obtained at high frequencies or at low frequencies[12] but in the midfrequency range numerical methods are necessary.

The guided wave scattering problem presents an enhanced complexity of the solution compared with bulk waves due to the presence of at least two modes at any given frequency. The large number of possible wave modes in a pipe is illustrated in the group velocity dispersion curves for a 3 inch schedule 40 steel pipe in Fig. 1. The software DISPERSE[13] has been used to trace the dispersion curves. In order to refer to different modes in cylindrical systems consistently, we use here a modified version of the system used by Silk and Bainton[3] which tracks the modes by their type, their circumferential order and their consecutive order. The modes are labeled $L(0,n)$, $T(0,n)$, and $F(m,n)$ and they, respectively, refer to longitudinal, torsional, and flexural modes. The first index m gives the harmonic number of circumferential variation and the second index n is a counter variable. Figure 1 shows that the modes travel at different speeds and are dispersive so that the original wave packet is distorted as it travels through the given structure. Much of the effort in flaw detection in pipes has been concentrated on the generation of a single mode in order to reduce the difficulty of dealing with many modes.[14] In a long range test it is also convenient to use a mode in a nondispersive region,[15] though a new technique which enables compensation for the signal spread caused by dispersion in a long-range guided wave testing is now being studied.[16]

Most of the previous investigators studied the effect of defects with small axial extent on the reflection of longitudinal waves.[7–9,17] Quantitative studies of the reflection due to notches[8] examined the correlation between the circumferential and through-thickness extent of notches and the amplitude of the reflected mode when a longitudinal $L(0,2)$ mode

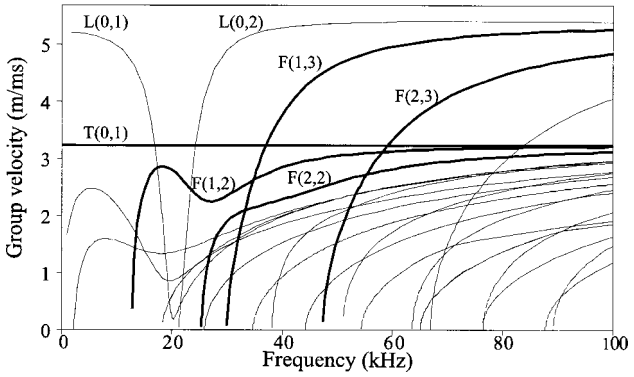


FIG. 1. Group velocity dispersion curves for 3 inch diameter pipe. Modes of interest are shown in bold.

is generated in the waveguide. Mode-conversion studies[7] when the $L(0,2)$ mode is launched in the test pipe enabled discrimination between axially symmetric reflectors such as circumferential welds and nonaxially symmetric defects. Guided wave tuning of nonaxisymmetric waves has also been used for defect detection in tubing.[11] Recently, work on crack characterization using guided circumferential waves has been published.[18] Guided circumferential waves have the limitation that they do not propagate along the length of the pipe so the range of the inspection is limited to a small percentage of the pipe.

In principle both axisymmetric and nonaxisymmetric modes can be used for long range inspection. Axisymmetric modes are in general preferable because they are easier to excite and have relatively simple acoustic fields. Initial practical testing was done using the longitudinal $L(0,2)$ mode.[7–9] However, more recent testing has employed the torsional mode.[10] This has the advantage that, in contrast to the $L(0,2)$ mode, the $T(0,1)$ mode propagation characteristics are not affected by the presence of liquid in the pipe and there is no other axially symmetric torsional mode in the frequency range, so axially symmetric torsional excitation will only excite the $T(0,1)$ mode, whereas when the $L(0,2)$ mode is used, the transducer system must be carefully designed to suppress the $L(0,1)$ mode.

No information is currently available on the reflection and mode conversion characteristics when a torsional mode is incident at a defect. The aim of this paper is to determine the reflection coefficients from cracks and notches of varying depth, circumferential and axial extent when the $T(0,1)$ mode is travelling in the pipe. An experimental and modelling study on the reflection caused by a series of defect geometries in both 3 inch and 24 inch pipes is presented. It is hoped later to generalize the findings to other pipe sizes.

II. GUIDED MODE PROPERTIES

Wave propagation properties in pipes are extremely complicated. Figure 1 shows the group velocity dispersion curves for a 3 inch, schedule 40 steel pipe. They were calculated using the program DISPERSE,[13] developed at Imperial College, and they show the velocity of propagation of the wave packets. The curves shown in bold are the modes which were used for the study. The dispersion curves of the

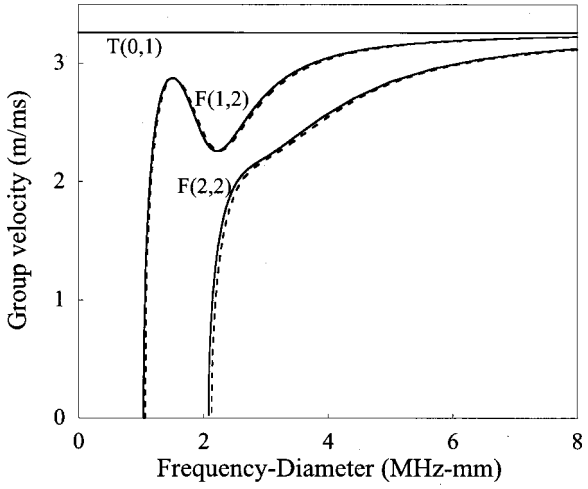


FIG. 2. Group velocity dispersion curves for $T(0,1)$, $F(1,2)$, and $F(2,2)$ modes in 24 inch (solid lines) and 3 inch pipes (dashed lines) as a function of frequency-diameter product.

$F(1,2)$ and $F(2,2)$ modes scale approximately with the frequency-diameter product in the frequency region of interest and the $T(0,1)$ velocity curve remains constant at all frequency values. This is shown in Fig. 2 which plots the group velocities of these modes as a function of frequency-diameter product for 3 inch (5.5 mm wall thickness) and 24 inch (20 mm wall thickness) pipes. The small difference between the curves is due to the fact that the diameter-to-thickness ratio is different for the two pipe sizes. This study considered an incident $T(0,1)$ mode in all cases, but mode conversion to the nonsymmetric modes could occur at nonaxisymmetric features. The frequency regions under examination were 10–50 kHz and 40–100 kHz for the 24 inch and the 3 inch pipes, respectively. The lowest frequencies of the range were chosen to be the frequencies usually used in practical testing (45–65 kHz for the 3 inch pipe and 10–20 kHz for the 24 inch pipe). It was also of interest to study the sensitivity of the $T(0,1)$ mode at higher frequencies.

The mode shape of the torsional $T(0,1)$ mode is not frequency dependent and it is completely nondispersive at all frequencies; its group velocity is the bulk shear velocity. No other torsional mode is present in the frequency range which was used for both finite element models and experiments. The mode shape of the $T(0,1)$ mode in a 3 inch pipe is shown in Fig. 3. This shows the profile of the tangential displacement through the thickness of the pipe wall. No axial or radial displacements are present in this mode. It can be seen that the tangential displacements are approximately constant through the wall thickness, indicating that defects will be detectable anywhere in the cross section of the pipe and there may also be potential for defect sizing. Since the $T(0,1)$ mode shape is also constant with frequency, pure mode excitation was obtained by simply imposing the mode shape at the center frequency.[19] The torsional mode in the 24 inch pipe has a very similar mode shape so all the comments about the $T(0,1)$ mode in the 3 inch pipe are also valid for the 24 inch pipe.

When an axisymmetric mode is incident at an axisymmetric defect, only the axisymmetric propagating modes ex-

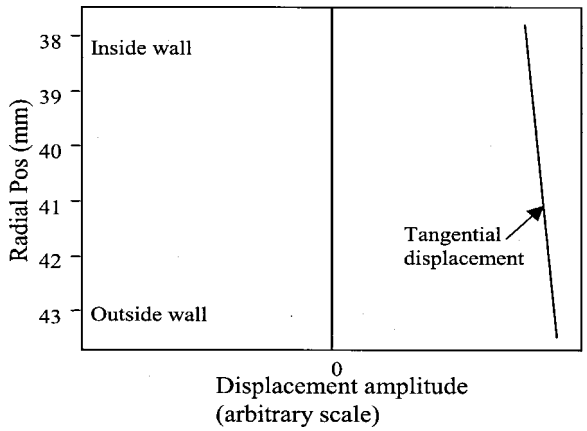


FIG. 3. $T(0,1)$ mode shape in a 3 inch pipe at 45 kHz. Radial and axial displacements are zero.

isting at the frequency of interest will contribute to the reflected field. Since we worked below the $T(0,2)$ cutoff frequency, there was no mode conversion at an axisymmetric defect and $T(0,1)$ was simply reflected or transmitted past the defect. However $T(0,1)$ can convert to the $F(1,2)$, $F(1,3)$, and $F(2,2)$ modes when an asymmetric feature exists in the pipe. The conversion is dependent on the similarity in particle motion of the incident and mode converted modes. Figures 4(a)–4(c) show the mode shapes of the $F(1,2)$, $F(1,3)$, and $F(2,2)$ modes, respectively, at 45 kHz. The $F(1,2)$ and $F(2,2)$ modes have dominant tangential displacement so the conversion to these modes from $T(0,1)$ is strong. Their torsional behavior becomes even more dominant at higher frequencies as illustrated in Figs. 5(a) and (b) which show the mode shapes at 100 kHz. Both tangential and longitudinal displacements are significant in the case of the $F(1,3)$ mode at 45 kHz so this mode has both torsional and extensional behavior at this frequency. The torsional behavior becomes less significant for this mode as the frequency increases [see the mode shape for the $F(1,3)$ mode at 100 kHz in Fig. 5(c)]. It is also important to understand the mode shape characteristics at low frequency; Fig. 5(d) shows that the $F(1,2)$ mode shape at 25 kHz is characterized by longitudinal and tangential displacements of the same order.

III. EXPERIMENTAL SETUP

Laboratory experiments were performed on a set of 3 inch schedule 40 steel pipes to determine the reflection sensitivity of the torsional $T(0,1)$ mode to a series of notches with different dimensions. The specimens which were used were 3.1 meter long pipes and the notches were machined at 2.3 meters from end *A* using a 3.2 mm diameter slot drill cutter (see Fig. 6). The pipe was tested horizontally; it was radially clamped at end *B* using 8 bolts evenly distributed around the circumference and rested on a steel support positioned approximately at the midpoint of the pipe. The support produced a negligible reflection of the ultrasonic signal; the reflection from end *B* was not monitored so the nature of the clamping there was unimportant. The cutter was aligned along a radius of the pipe so the circumferential extent of the cut was changed by simply rotating the pipe about its axis of symmetry. Two rings of 16 dry coupled transducers were clamped to the pipe at 0.9 meters from end *A* and left in place during the cutting so all of the experiments on one pipe were done without removing the transducers. The use of two rings enabled us to distinguish the direction of propagation of the reflected signal (left or right with respect to the ring position).[10] Using piezoelectric transducers equally spaced to produce tangential displacement, only torsional modes are generated.[10] In order to generate only axially symmetric

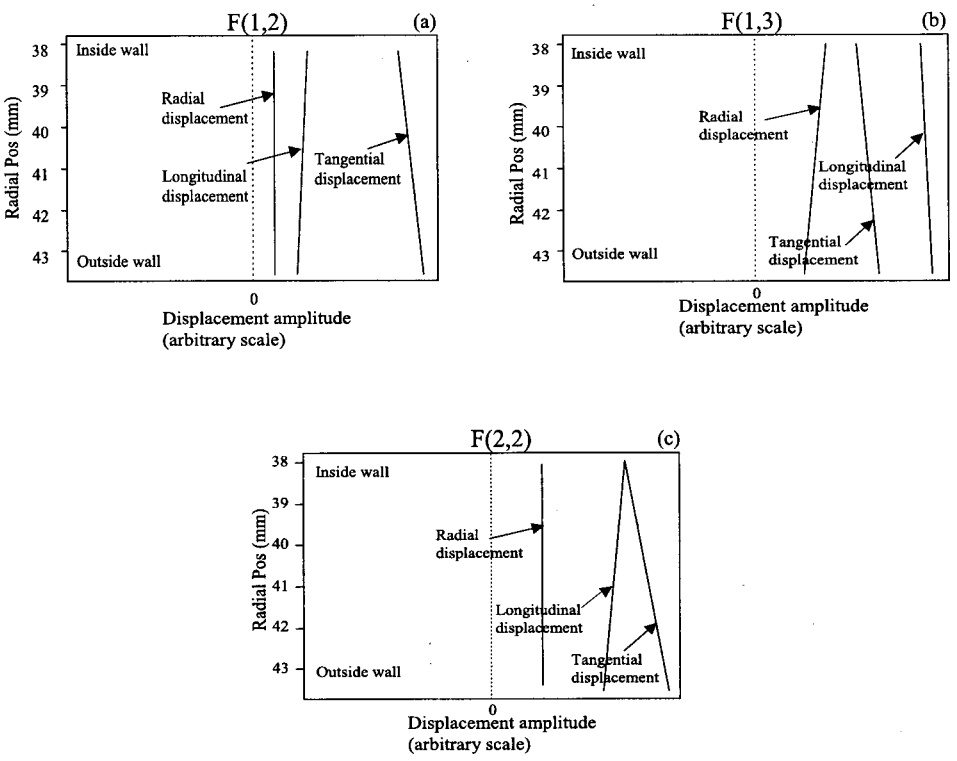


FIG. 4. Displacement mode shapes in a 3 inch pipe at 45 kHz for $F(1,2)$ (a), $F(1,3)$ (b), and $F(2,2)$ mode (c).

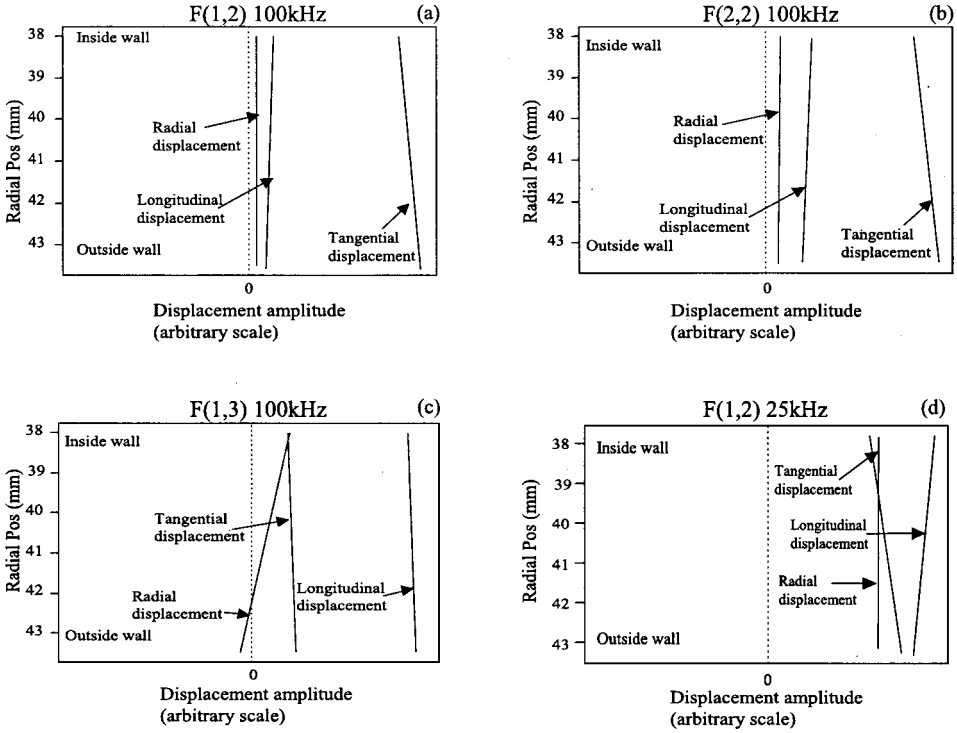


FIG. 5. Displacement mode shapes in a 3 inch pipe at 100 kHz for $F(1,2)$ (a), $F(2,2)$ (b), and $F(1,3)$ mode (c) and at 25 kHz for $F(1,2)$ mode (d).

modes the number of elements in the ring has to be greater than n where $F(n,1)$ is the highest order flexural mode whose cutoff frequency is within the bandwidth of the excitation signal.[20] A Guided Ultrasonics Ltd. Wavemaker 16 instrument was used to generate a six cycle Hanning-windowed toneburst at many frequency values from 40 to 70 kHz. The same instrument was also used for signal reception. In order to obtain a reference measurement, reflections from end A were recorded before any machining was done. By doing that we verified that the vibration induced from the cutting left the contact characteristics of the transducer rings on the pipe unchanged. Experiments were conducted on four separate pipes, using the following notches: (a) a through-thickness notch with axial extent equal to about 3.5 mm and varying circumferential extent; (b) a notch extending over the full circumference of the pipe with axial extent equal to about 3.5 mm and varying notch depth; (c) a notch extending over 25% of the circumference of the pipe with axial extent equal to about 3.5 mm and varying notch depth; (d) a notch extending over the full circumference of the pipe with notch depth equal to 20% of the thickness and varying axial extent.

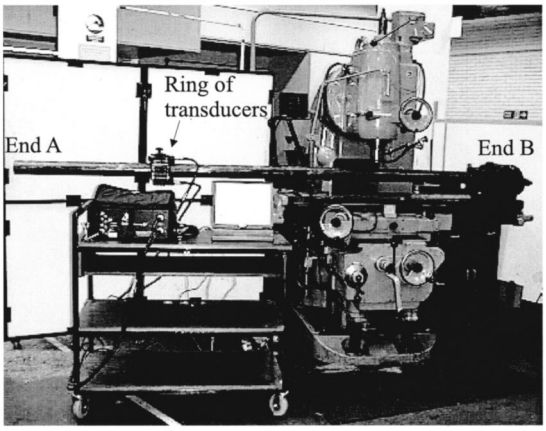


FIG. 6. Experimental setup.

IV. FINITE ELEMENT MODELS

The finite element (FE) method has been extensively and successfully used to study the interaction between guided waves and defects in structures.[7–9,15,21–25] In general a three-dimensional solid model is required to perform a numerical analysis of the interaction between guided waves and discrete defects. However 3D models are computationally expensive so when possible we use simplified models.[7] In some cases it is possible to study a specific 3D model by using a combination of two 2D models. Here we considered three types of models:

(1) *Membrane*: This is a three-dimensional analysis in which the pipe wall is represented by membrane elements. These allow defects removing a fraction of the pipe circumference to be modelled, but the defects remove the full wall thickness. The axial extent of the defect can be varied.

(2) *Axisymmetric*: This is a two-dimensional analysis which can deal with defects which remove part of the wall thickness, but are axially symmetric and so cover the whole pipe circumference. The axial extent of the defect can be varied.

(3) *Three dimensional*: Previous work[7–9] had indicated that the $L(0,2)$ reflection coefficient from a defect was linearly proportional to its circumferential extent. The reflectivity of a defect of a given depth and circumferential extent could therefore be predicted from results of the axisymmetric analysis by obtaining the reflection coefficient for an axisymmetric defect having the depth of interest and scaling by the fraction of the circumference covered by the defect. The three-dimensional models were used to demonstrate the va-

lidity of making predictions using combinations of results from the different types of two-dimensional model when the torsional $T(0,1)$ mode is incident.

In each case a length of the pipe was modelled and a notch was introduced at some distance along it. The input wave signal was excited in the model by prescribing time-varying displacements at one end of the pipe. Narrow band signals were used, typically composed of 10 cycles of the desired center frequency modulated by a Gaussian window. The tangential displacements were then monitored at a location between the excitation end and the notch, thereby detecting both the incident wave on its way to the notch and the reflected wave returning. The reflection ratio was calculated by dividing the amplitude of the reflected mode by the amplitude of the incident mode in the frequency domain.

In both axisymmetric and 3D models the excitation was achieved by prescribing the displacements to match the exact mode shape at the end of the pipe. The displacement distribution through the thickness was derived using Disperse.[13] In the membrane model the torsional wave was excited by prescribing the displacement all around one circumference; displacement distribution through thickness is not possible using this model. The notches were introduced either by disconnecting adjacent elements or by removing elements.

A. Membrane models (full depth, part circumference)

For the membrane models the geometry of the pipes was discretized using membrane finite elements. This type of element describes a two-dimensional rectangular region and it models the stresses which lie in the plane: the two orthogonal in-plane direct stresses and the in-plane shear stress. It assumes plane stress conditions, therefore incorporating the capacity of the plate to change thickness when the in-plane stresses are applied. Although it is a two-dimensional element by shape and (local) behavior, it may be used in a three-dimensional model, meaning simply that it can be orientated arbitrarily in a three-dimensional space. In a three-dimensional model its nodes have displacement degrees of freedom in the three coordinate directions and it has mass which participates in any coordinate direction.

The justification for using membrane elements is the nature of the mode shapes of all of the modes which were studied. As already seen in Figs. 3–5, the pipe wall shows extensional or torsional behavior for all of the modes (especially at high frequency). Therefore the modes of interest could be modelled reasonably well by using the membrane elements, each element lying at the center of the pipe wall. It is not possible to use membrane models in order to simulate modes in which there is local bending of the pipe wall, nor to model part-through notches.

Half of the circumferential extent of a length of pipe was modelled as in previous work with the $L(0,2)$ mode[7,8] but in this case we assumed one plane of antisymmetry. A wave was excited at the end of the pipe, a defect was simulated and the signals were monitored on a line between the excitation and the defect. A mesh of identically sized linear (four noded) quadrilateral membrane elements was used. The models of the 24 inch pipe represented a 3.6 m length, using 900 elements along the length and 40 elements around the circumference. As a result each element was about 4 mm long and 24 mm along the circumference. The models of the 3 inch pipe also represented a 3.6 m length, using 900 elements along the length and 40 elements around the circumference. Hence each element was about 4 mm long and 3 mm in the circumferential direction. Convergence studies indicated that the meshes gave reliable results. The zero-length notches were modelled by disconnecting adjacent elements; thus although the elements on each side of the notch had nodes at coincident locations, they were not connected. The notches with nonzero lengths were modelled by removing elements from the model and were therefore rectangular in shape. The notches were introduced adjoining one of the planes of antisymmetry so that the geometric model described half of the pipe and half of the notch, and the center of the notch was therefore at the plane of antisymmetry.

The excitation of the axially symmetric modes (circumferential order 0 modes) was achieved by prescribing identical tangential displacement histories at all of the nodes at the end of the pipe. On reception of the multimode reflected signal, the order 0 modes were extracted simply by adding the tangential displacements at all of the monitoring nodes around the circumference of the pipe. For the order 1 modes, a phase delay of $\theta/2\pi$ was added to each signal before summing them,[7,8] where θ is the angular distance from the center of the notch. Thus a separate processing calculation was performed in order to extract the amplitude of each type of mode. In principle three order 0 modes $[L(0,1),L(0,2),T(0,1)]$ and three order 1 modes $[F(1,1),F(1,2),F(1,3)]$ can exist in the high frequency range. However $L(0,1)$ and $L(0,2)$ are both longitudinal and are not excited, and $F(1,1)$ does not propagate in the membrane model because it is characterized by through-wall bending behavior. In most cases the processing already described was therefore sufficient to identify separately the remaining $T(0,1)$, $F(1,2)$, and $F(1,3)$ modes. At some frequencies the $F(1,2)$ and $F(1,3)$ modes have very similar group velocity so it was not possible to separate them in the time domain. Since the signals were rather narrow-band, the processing could reasonably have been performed directly on the raw time records. However, for better accuracy, the calculations were performed in the frequency domain.

Typical time records from the simulations are shown in Figs. 7(a) and 7(b). Both of these records are for the 24 inch pipe with a notch which extends around 25% of the circumference, and for the $T(0,1)$ mode incident at 50 kHz; the difference between them is in the processing. Figure 7(a) shows the signal with processing for order 0. This shows clearly the incident signal on its way towards the notch and then the reflected $T(0,1)$ mode. Figure 7(b) shows the signal when the same raw results are processed to extract the order 1 modes. This now shows the reflection of the $F(1,2)$ mode. Also, the incident mode has vanished in this plot because it is order 0 and is therefore not detected by the order 1 processing.

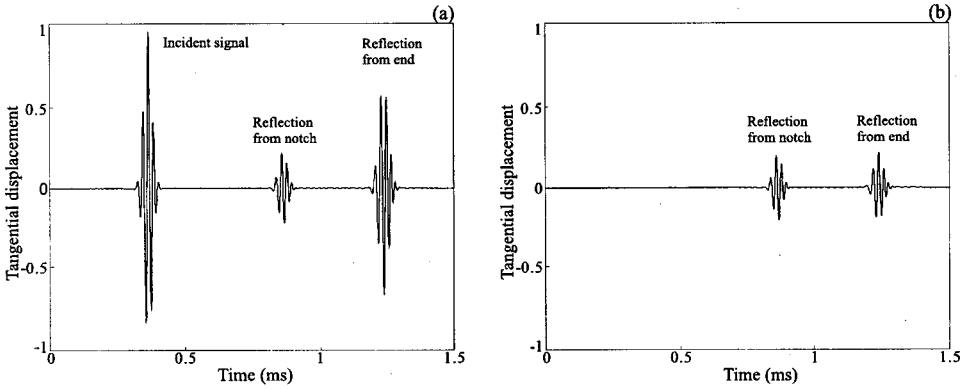


FIG. 7. Predicted time record for membrane model of 24 inch pipe with notch extending around 25% of circumference and $T(0,1)$ mode incident. Results processed to show order 0 modes (a) and order 1 modes (b).

B. Axisymmetric models (full circumference, part depth)

The axisymmetric elements describe a two-dimensional region representing a radial–axial section through an axially symmetric structure. Thus in application to the pipe models they represent a section through the pipe wall on a plane which contains the axis of the pipe. They model the three in-plane stresses (the radial and axial direct stresses and the radial-axial shear stress) and also the circumferential stress. Thus radial displacements of the elements are correctly coupled to circumferential stresses. It follows that any geometric shape which is modelled must be axially symmetric. Therefore, although it is possible to model notches which extend part-way through the pipe wall (unlike the membrane models which only represent through-wall notches), any notch is assumed to extend around the full circumference. Since the geometry is axially symmetric, there is no mode conversion between modes of different circumferential orders, so only modes of the same order as the excitation signal can propagate. In our models only the $T(0,1)$ torsional mode can be reflected from the axisymmetric notch because this is the only torsional mode existing in the working frequency range.

The excitation, monitoring and notch locations were arranged along the pipe similarly to the membrane models discussed earlier. Identically sized linear (four noded) quadrilateral axisymmetric elements were used. The wall thickness of the 24 inch and 3 inch pipes were taken to be 20 mm and 5.5 mm, respectively. The models of the 24 inch pipe represented a 2.4 m length, using 1000 elements along the length and 10 elements through the thickness. Therefore each element was 2.4 mm long and 2 mm in the thickness direction. The models of the 3 inch pipe with the $T(0,1)$ mode represented a 2.4 m length, using 2000 elements along the length and 10 elements through the thickness. Therefore each element was 1.2 mm long and 0.55 mm in the thickness direction. Again convergence studies showed that this discretization was satisfactory. Two other FE models of the 3 inch pipe with an 80% defect depth were run at 150 kHz and 240 kHz center frequency in order to obtain the reflection coefficient at high frequencies. A finer mesh was needed in the 240 kHz case according to the practical rule of having at least eight elements per wavelength of the smallest wavelength of the excitation signal. The ratio between the frequency spectrum of the reflected signal and the spectrum of the input was then derived for each FE run. Using four FE models with different center frequencies (40 kHz, 100 kHz, 150 kHz, and 240 kHz) we covered the frequency bandwidth from 15 kHz to 270 kHz. Higher frequencies were not considered because of the presence of the second torsional mode $T(0,2)$ [the cutoff frequency for the $T(0,2)$ torsional mode is 300 kHz in a 3 inch pipe].

The effect of the axial extent of the notch on the reflection coefficient of an axisymmetric notch was also studied. Several FE models with 20% defect depth and different axial extent have been studied. The notches were created in the FE simulations by simply removing elements in the mesh. When the axial extent was long enough, the reflection of the start of the notch was separated from the reflection from the end of the notch.

C. 3D models (part circumference, part depth)

The three-dimensional modelling used 8-node ''brick'' elements to discretize a length of pipe fully in three dimensions. The brick elements are cuboid in shape and are used to represent, piecewise, the whole volume of the structure. Potentially this offers the advantage of being able to model notches which have limited extent both in the circumferential and the through-wall directions. However, there is a heavy computational penalty because of the large numbers of elements which are necessary for three-dimensional models. We here modelled a pipe with a part-circumference and part-thickness notch. The purpose was to compare the reflection coefficient obtained with the one derived by combining the results of membrane and axisymmetric models in order to confirm the assumption which has been made that the reflectivity from such notches can be inferred from the combined results of the membrane and axisymmetric models.

Half of the circumferential extent of a 2.4 m length of the 3 inch pipe was modelled, assuming one plane of antisymmetry. The excitation, notch and monitoring locations were arranged similarly to the membrane models. The pipe was modelled with 600 elements along the length, 40 elements around the circumference, and three elements through the 5.5 mm wall thickness. The $T(0,1)$ mode at 100 kHz was excited and monitored. The excitation was achieved by prescribing the displacements to match the exact mode shape at the end of the pipe. The $T(0,1)$ mode was monitored by adding the circumferential displacements at all of the nodes

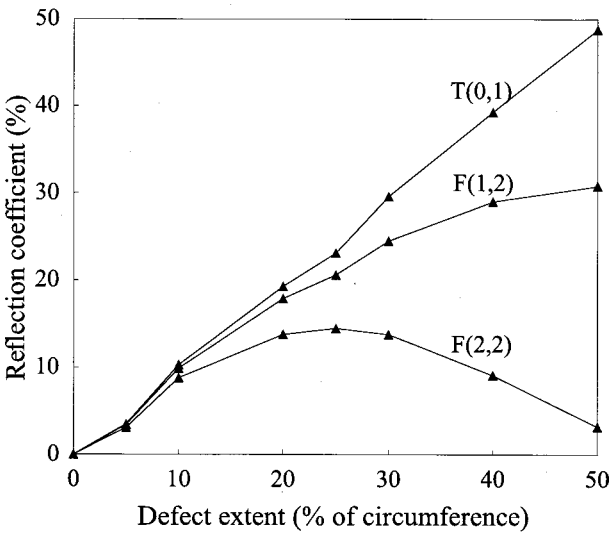


FIG. 8. Variation of reflection ratio with defect circumferential extent for zero axial length, full wall thickness defect. Results are from membrane model with $T(0,1)$ incident in 3 inch pipe at 100 kHz.

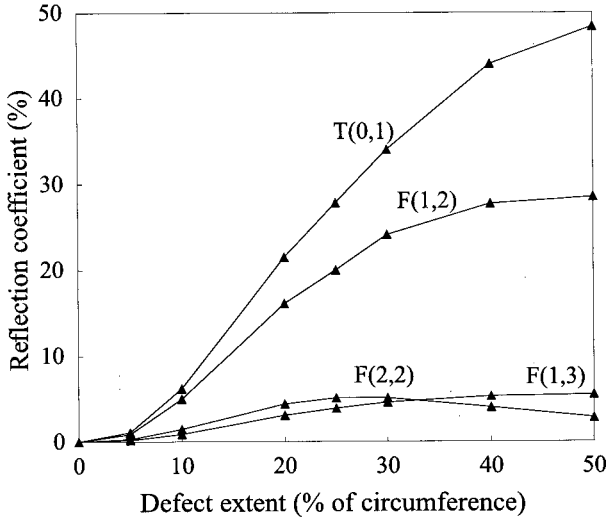


FIG. 9. Variation of reflection ratio with defect circumferential extent for zero axial length, full wall thickness defect. Results are from membrane model with $T(0,1)$ incident in 3 inch pipe at 45 kHz.

around one circumference at the monitoring location.

Before tackling the problem of the notch with limited extent in both circumferential and depth directions, some analyses of an axially symmetric part-depth notch and a through-wall part-circumference notch were studied using the 3D model and compared with the related 2D models. The results from the 3D mesh compared well with those obtained from the 2D models, so validating the discretization.

A 3D model of a part-depth (33%), part-circumference (25%) notch was analyzed using 100 kHz $T(0,1)$ excitation. A notch axial length of 180% of the wavelength ($\lambda=32.6$ mm) was chosen, so the defect was a square patch (60 mm$\times$60 mm) 1.8 mm deep.

V. STUDY OF PARAMETERS AFFECTING REFLECTION AND MODE CONVERSION

The numerical results are presented below. We study separately the influence of parameters such as pipe dimension, frequency of the excitation, circumferential extent of the defect, depth of the defect and axial extent of the notch. The section is divided into three sections in which we describe the results for the three different FE models.

A. Through thickness defects with zero axial extent (membrane model)

Using the results of the membrane models we studied the effect of the circumferential extent of a defect on the backscattering of the torsional wave at different frequencies. We first present the results of the simulation for the 3 inch pipe case and later offer a comparison with the 24 inch case.

Practical testing of 3 inch pipes using torsional excitation is usually performed at relatively low frequency (45–65 kHz).[10] We start by presenting the results at higher frequency (100 kHz) because of the simplicity of this case and we will focus later on the frequencies used in practical testing. The reflection coefficient on the 3 inch pipe with $T(0,1)$ input at 100 kHz is shown in Fig. 8, the computed points being joined by straight lines.

The $T(0,1)$ reflection coefficient is approximately linear with circumferential extent, the $F(1,2)$ reflection coefficient has a maximum at 50% of the circumferential extent, and the $F(2,2)$ has a maximum at 25% of the circumferential extent and a minimum corresponding to 50% of the circumferential extent. The results obtained in this case are very similar to the results obtained in previous work on the $L(0,2)$ reflection coefficient in a 3 inch pipe at 70 kHz.[7] No significant presence of the $F(1,3)$ mode is noticeable at this frequency.

The variation of the $T(0,1)$ reflection coefficient with circumferential extent at 100 kHz shown in Fig. 8 indicates some divergence from the overall linear trend; this is probably due to numerical problems. In principle a finer mesh could be used to investigate this but the model which was used here is close to the limits of the available resources so this has not been pursued. In any case the extent of the scatter is small in absolute terms, and is not important from the point of view of practical application.

The interpretation of the 3 inch FE simulation at lower frequencies is more complicated than at high frequency because of changes in the mode shapes of the relevant modes. In the intermediate frequency range (from 30 to 70 kHz) $F(1,3)$ is seen in the reflected signals along with the other modes. This is because there is significant circumferential motion in its mode shape over this frequency range as shown in Fig. 4(b), whereas this displacement component is smaller at higher frequency [see Fig. 5(c)]. Figure 9 shows the results at a frequency of 45 kHz. At all circumferential extents the amplitude of the $F(1,3)$ reflection is about 20% of that of the $F(1,2)$ mode. Figure 10 shows the $T(0,1)$, $F(1,2)$, and $F(1,3)$ reflection coefficients at frequencies from 25 kHz to 100 kHz for the 25% circumferential extent case. As expected from the mode shapes, the ratio between the $F(1,3)$ and $F(1,2)$ reflection coefficients becomes smaller at higher frequencies. At low frequency (25 kHz) the $T(0,1)$ reflection coefficient is small. Figure 10 also shows that the $F(1,2)$ mode reflection coefficient at low frequency is very small compared with the other results at higher frequencies. This is

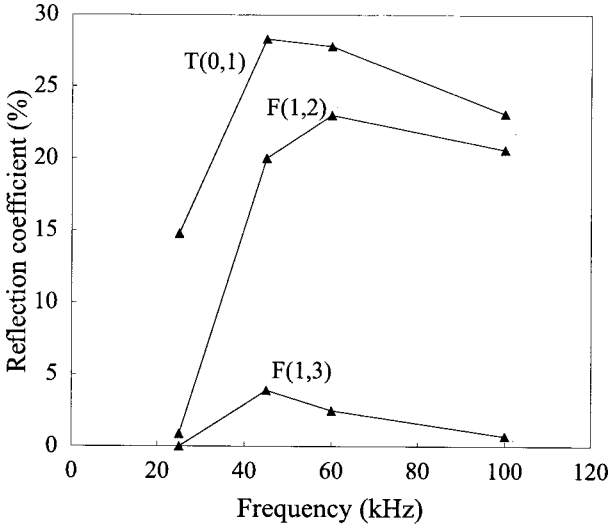


FIG. 10. Variation of reflection ratio with frequency for $T(0,1)$ mode input in 3 inch pipe using membrane model with 25% notch circumferential extent.

due again to the $F(1,2)$ mode shape at low frequency. From Fig. 5(d) it can be seen that at 25 kHz the $F(1,2)$ mode has a longitudinal displacement bigger than the tangential displacement so this mode does not have predominantly torsional behavior at low frequency.

In the case of a 24 inch pipe the general behavior of the reflection coefficient seen in the 3 inch pipe at 100 kHz case is confirmed (see Fig. 11). No presence of the $F(1,3)$ mode was observable in the 24 inch FE simulations that were run.

B. Part thickness axisymmetric defects (axisymmetric model)

The results of the axisymmetric FE modelling are plotted in Figs. 12–14. Since the defect is axisymmetric, there is no mode conversion between the axisymmetric (T) modes and the flexural (F) modes. Figure 12 shows the trend of the $T(0,1)$ reflection coefficient for both 3 inch and 24 inch

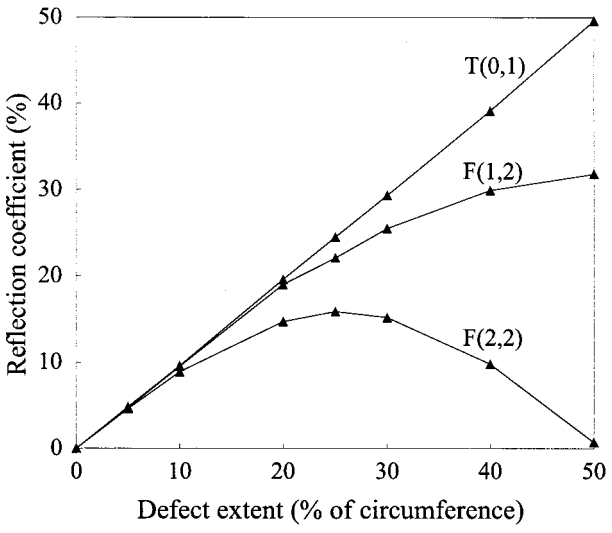


FIG. 11. Variation of reflection ratio with defect circumferential extent for zero axial length, full wall thickness defect. Results are from membrane model with $T(0,1)$ incident in 24 inch pipe at 50 kHz.

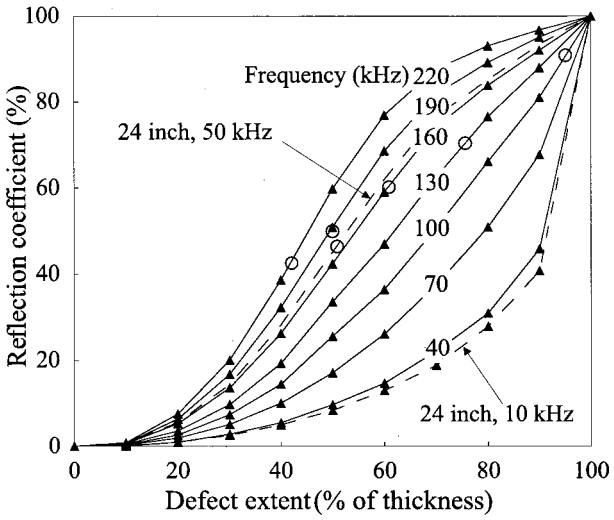


FIG. 12. Variation of reflection ratio with defect depth for zero axial length at various frequencies, axisymmetric defect. Results are from axisymmetric model with $T(0,1)$ incident in 3 inch (solid lines) and 24 inch pipes (dashed lines). The empty circles indicate the depth value for which $ka=1$ at each frequency.

pipes as a function of defect depth at different excitation frequencies. If we consider one pipe dimension at one frequency value we observe that the reflection coefficient increases with defect depth in a nonlinear manner. The shape of the curves is similar to that predicted for a 3 inch pipe at 70 kHz using the $L(0,2)$ mode in earlier published work.[8] If we now consider one pipe dimension at different frequencies of excitation, we can see that as the frequency decreases, the curves become increasingly concave, the reflectivity at low defect depths decreasing markedly. It therefore becomes more difficult to detect shallower defects as the test frequency is decreased. It is also of interest to observe that when the frequency-thickness product is similar, the reflection is similar. Specifically, the reflection coefficients for a 3 inch pipe with 5.5 mm wall thickness at 40 kHz (frequency-

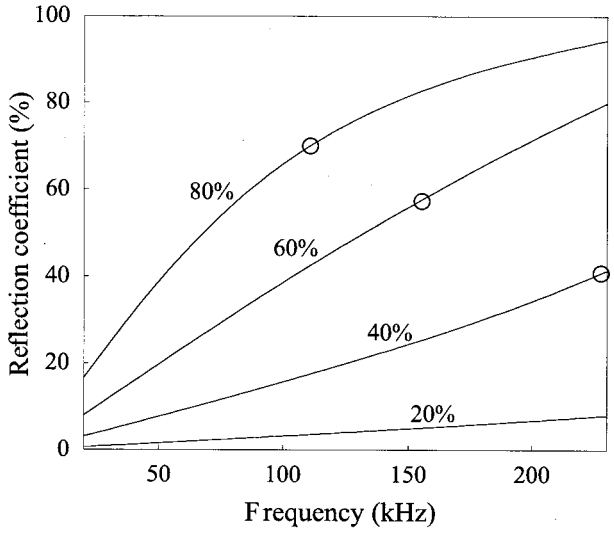


FIG. 13. Variation of reflection ratio with frequency for zero axial length at various defect depths, axisymmetric defect. Results are from axisymmetric model with $T(0,1)$ incident in 3 inch pipe. The empty circles indicate the frequency at which $ka=1$ for each defect depth.

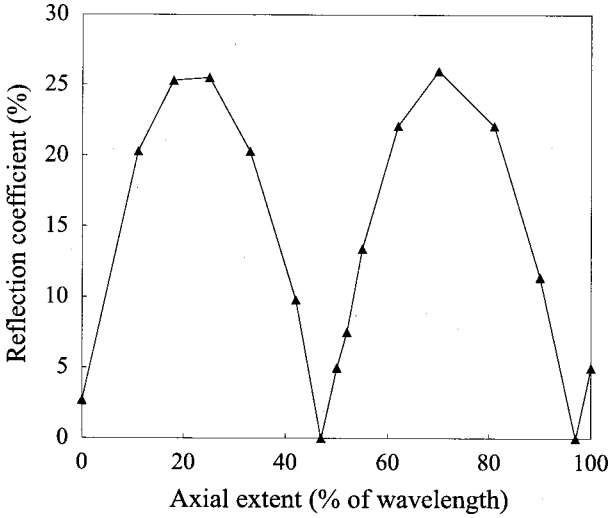


FIG. 14. Variation of reflection ratio with axial extent when there is an axisymmetric defect with 20% thickness depth. Results are from axisymmetric model with $T(0,1)$ incident in 3 inch pipe at 100 kHz.

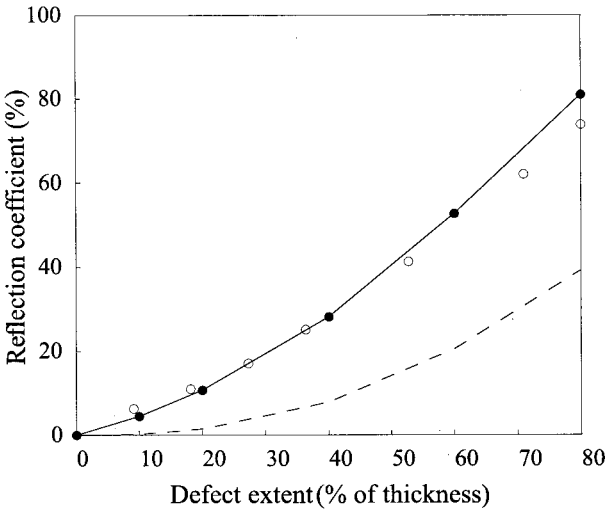


FIG. 15. Variation of reflection ratio in incident mode with defect depth for 3 mm axial length, axisymmetric defect. FE (–●–) and experimental (○) results are for $T(0,1)$ incident in 3 inch pipe at 55 kHz. The crack case (zero axial extent) is also displayed for comparison (dashed line).

thickness product 220 kHz mm) are similar to the reflection coefficients for a 24 inch pipe with 20 mm wall thickness at 10 kHz (frequency-thickness product 200 kHz mm). The same observation is valid at high frequency-thickness values; the 3 inch pipe at 160 kHz (880 kHz mm) has a reflection coefficient similar to that of the 24 inch pipe at 50 kHz (1000 kHz mm). Figure 13 shows the reflection coefficient in the frequency domain when a 3 inch pipe with various depth notches is excited with the torsional $T(0,1)$ mode. The reflection coefficient increases with frequency at all depth values. In the 80% depth case the rate of change decreases as the frequency increases.

The effect of the axial extent of the defect on the reflectivity of the $T(0,1)$ mode in a 3 inch pipe is plotted in Fig. 14. The reflection coefficient has a maximum at an axial extent of about a quarter wavelength and a minimum at about half-wavelength. Similar behavior has been reported for the s_0 mode in a plate.[25]

C. 3D models

The results of the 3D analysis are given in Table I. The results demonstrate excellent agreement between the predictions of the 2D axisymmetric analysis with those of the 3D analysis; for the case of an axisymmetric, 33% depth defect, the predicted reflection coefficients are 23.5% and 23.7% with the 2D axisymmetric and 3D models, respectively. Similarly good agreement is shown between the membrane and 3D models for the case of a through wall defect with 25% circumferential extent, the predicted reflection being 21.1% with both models.

Finally, Table I shows the results for the 3D analysis of a notch with 25% circumferential extent, 33% depth and axial extent of 180% of the wavelength. The 3D analysis predicts a reflection ratio of 5.2% for this case. This is closely matched by the multiplication of the axisymmetric and membrane results which give 5.0%. This confirms the validity of combining the axisymmetric and membrane results to predict the 3D case.

VI. EXPERIMENTAL VALIDATION OF NUMERICAL MODELLING

A series of experiments was carried out in order to validate the results obtained from the FE modelling. From previous modelling it was clear that the axial extent can be an important parameter affecting the reflection coefficient from notches, so when simulating the laboratory test cases using FE models it was necessary to introduce a notch with the precise axial extent used in the experiments. Part-thickness, part-circumference defects were modelled using a combination of 2D axisymmetric and 2D membrane models.

Figure 15 shows the measured and predicted reflection coefficients at 55 kHz for a series of axisymmetric notches with 3 mm axial extent in which the notch depth was varied, together with the prediction for a crack (zero axial extent).

TABLE I. Comparison of $T(0,1)$ reflection ratio in a 3 inch pipe at 100 kHz from 3D model with combined results from axisymmetric and membrane models.

Axial extent [% of λ(32.6 mm)]	180	180	180
Circumferential extent (%)	100	25	25
Through-wall extent (%)	33	100	33
Reflection ratio (%): Axisymmetric model (2D)	23.5		
Reflection ratio (%): Membrane model (2D)		21.1	
Reflection ratio (%): 3D model	23.7	21.1	5.2
Reflection ratio predicted by combining 2D analyses			5.0

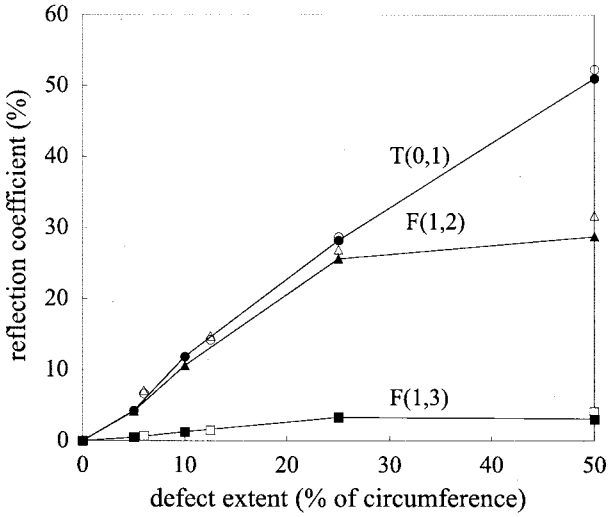


FIG. 16. Variation of reflection ratio with circumferential extent for 3 mm axial length, through thickness defect. FE (lines with solid symbols) and experimental (empty symbols) results are for $T(0,1)$ incident in a 3 inch pipe at 55 kHz.

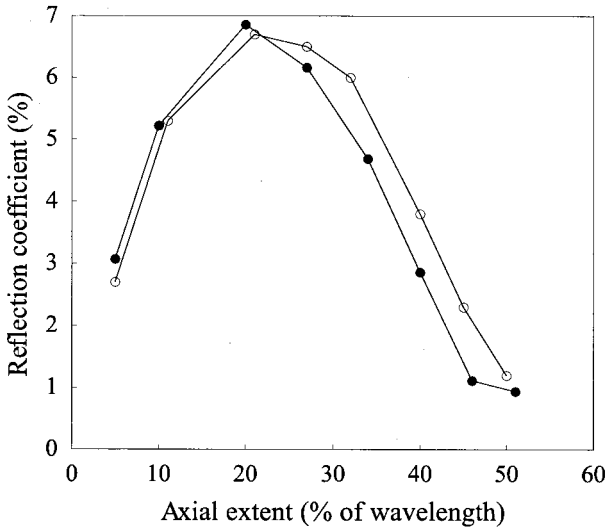


FIG. 18. Variation of reflection ratio in incident mode with axial extent for a 20% depth, 25% circumferential extent defect. FE (●) and experimental results (○) are for $T(0,1)$ incident in 3 inch pipe at 55 kHz.

The experimental measurements and finite element predictions can be seen to agree well. The reflection coefficients for the notch case are higher than those for the crack case. Moreover the concavity of the curve is reduced with respect to the crack case.

Figure 16 shows a comparison between experimental and FE results in the case of through-thickness notches with 3 mm axial extent and varying circumferential extent when the $T(0,1)$ mode is excited at 55 kHz. Again there is good agreement between the experimental and modeling results.

Figure 17 shows the dependence of the reflection coefficient on the frequency in the case of a through-thickness defect with 25% circumferential extent and axial extent equal to 3 mm. The reflection coefficient for $T(0,1)$ decreases when the frequency is increased from 50 kHz to 70 kHz in both experiments and predictions. Good agreement between experiments and finite elements is also found in this case.

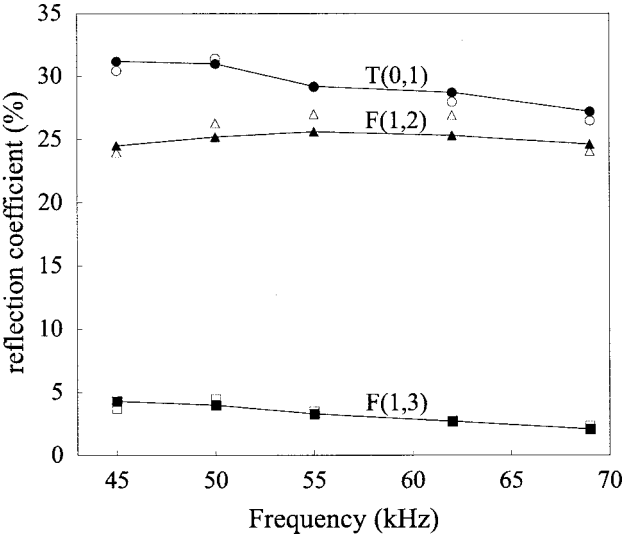


FIG. 17. Variation of reflection ratio with frequency for 3 mm axial length, through thickness defect extending over the 25% of the circumference of a 3 inch pipe. Both FE (lines with solid symbols) and experimental results (empty symbols) are displayed.

Figure 18 shows the behavior of the reflection coefficient when the axial extent is changed. The results are derived for a 20% depth defect extending over 25% of the circumference at 55 kHz center frequency excitation. The form of the experimental results agrees with the predictions, the peak of the reflection coefficient occurring when the axial extent is about a quarter wavelength and the minimum at about half wavelength.

VII. DISCUSSION

A defect in a solid body represents a scatterer for elastic waves and so in principle it can be detected and characterized by its effect on an incident pulse of ultrasonic wave motion. Elastodynamic scattering problems can be solved in closed form in the high frequency regime using either the Kirchoff approximation or the geometrical diffraction theory and at low frequency applying the Rayleigh approximation.[12] In the midfrequency range these methods are not valid because the wavelength of the incident pulse is of the same order as the characteristic dimension of the defect and preferable alternatives are numerical methods such as finite element, boundary element or finite difference. Scattering of bulk waves from a crack in an infinite medium has been studied by many investigators and much of this work has been reviewed by Kraut[26] and Datta.[27] Elastodynamic ray theory has been thoroughly studied by Achenbach *et al.*[28] in order to construct scattering fields generated by cracks.

The interaction of Lamb waves with cracks and notches in plate structures is a topic that has received a great deal of interest in recent years.[9,24,25,29–31] Wave scattering of guided elastic waves from cracks and notches in isotropic hollow cylinders has also been investigated.[6–9,11,17,18,32] The scattering of guided waves is more difficult than the bulk wave case as a consequence of the higher number of modes to be considered and the complexity of the mode shapes. When a

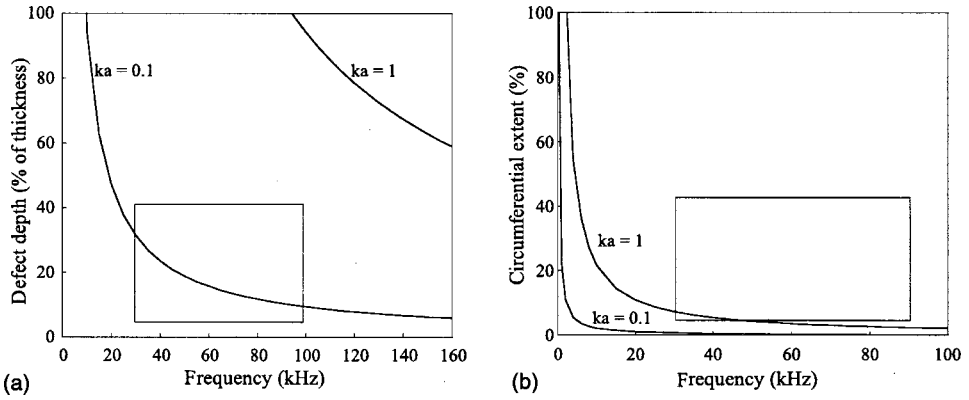


FIG. 19. Scattering regime regions in the case of the axisymmetric defect in 3 inch pipe with 5.5 mm wall thickness (a) and through thickness defect (b). The boxes indicate the practical testing regions.

single mode in a waveguide encounters the defect there is coupling between the incident mode and all of the propagating and nonpropagating modes of the waveguide.[33] If the location of the detector of the wave is placed far enough from the defect, only the propagating modes will have significant amplitude. In fact only a few of the propagating modes contain a significant amount of scattered energy so the problem can be further simplified by only considering these modes.[7]

A. Analysis of the sensitivity of torsional mode to cracks (zero axial extent)

It is well known that three scattering regimes can exist depending on the dimension of the scatterer in relation to the wavelength of the ultrasonic wave.[12] A parameter frequently used to define and distinguish the scattering regimes is ka where a is the characteristic dimension of the defect and k is the wave number defined by

$$k=\frac{2\pi f}{V}, \tag{1}$$

where f is the frequency and V is the phase velocity. In the specific case of a torsional wave, the velocity is constant with frequency so the wave number simply increases linearly with frequency. When $ka<0.1$ the problem can be approximated using a low frequency or quasistatic approach. At values for which $ka>1$ it is possible to use a high frequency approximation (however $ka=1$ cannot be considered as a strict boundary between the different regimes). For part-thickness, axisymmetric defects the characteristic length is the defect depth, while for through-thickness, part-circumference defects the characteristic length is the circumferential extent. Figures 19(a) and (b) show the scattering regimes for axisymmetric and through thickness cracks in the case of a 3 inch pipe with 5.5 mm wall thickness. In both figures a rectangular region which is important from the point of view of practical testing is also highlighted. It is evident that in the case of axisymmetric cracks we are in the low–intermediate frequency regime, whereas in the case of through thickness defects we are in the intermediate–high frequency regime.

1. Discussion on through thickness cracks

The motivation of separately studying the reflection characteristics of the through-thickness, part-circumference, and the part-depth axisymmetric defects is the difference in geometry and scattering regime of the two classes of crack. The reflection coefficients of the 24 inch pipe with $T(0,1)$ input at 10 kHz and 50 kHz as a function of circumferential extent are shown in Fig. 20. The circumferential extent is expressed as a percentage of the pipe circumference (1980 mm). The $T(0,1)$ reflection coefficients are very similar at the two frequencies and they increase roughly linearly with circumferential extent, especially at the higher circumferential extents (corresponding to higher ka). A somewhat smaller reflection is obtained for a small defect (5% of circumference) at low frequency (10 kHz). The same behavior is noticeable in the 3 inch case.

Let us first give an explanation of how the frequency of the test influences the reflection coefficient from through thickness cracks and we will then explain why the dimension of the defect is an important parameter. Lowe *et al.*[7,8] showed that the $L(0,2)$ reflection coefficient varies roughly linearly with respect to the circumferential extent of the defect. They presented a calculation in order to discover whether the reflection coefficients could be estimated from a simple assumption of the opening profile of the crack. The profile of the opening displacement of the FE nodes along

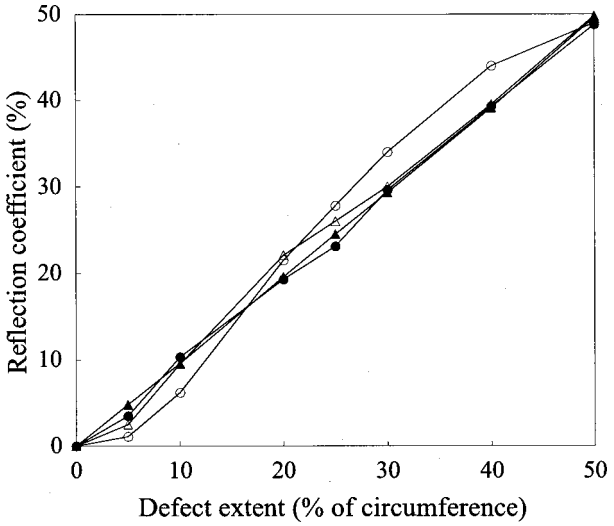


FIG. 20. Variation of reflection ratio in incident mode with defect circumferential extent for zero axial length, full wall thickness defect. Results are from membrane model with $T(0,1)$ incident in 24 inch pipe at 10 kHz (empty triangle) and 50 kHz (solid triangle) and in 3 inch pipe at 45 kHz (empty circle) and 100 kHz (solid circle).

the crack face was almost constant in the case of the $L(0,2)$ mode at 70 kHz incident on a crack. Clearly when the crack opening displacement (COD) is constant along the defect, the reflection coefficient of the axisymmetric mode is equal to the extent of the crack as a fraction of the circumference, leading to a linear behavior of the reflection coefficient as a function of circumferential extent. As they explained, the COD is approximately constant over the crack face when the wavelength of the incoming wave is sensibly smaller than the circumferential extent of the crack (i.e., $ka>1$). Under such conditions the crack face simply behaves as a free surface which reflects the incoming wave. When ka diminishes, the crack opening displacement can no longer be considered as constant. The variation of the COD with frequency was observed by Lowe in the case of a s_0 wave incident on a crack in a plate.[9] He showed that the COD changes with frequency and there is a transition between the high frequency COD and the quasistatic COD. From that analysis it is clear that only at relatively high frequency ($ka>1$) can the COD reasonably be approximated by a constant displacement along the face of the crack.

It is possible to interpret the results at 5% of the circumferential extent for both the 3 inch pipe at 45 kHz and the 24 inch pipe at 10 kHz (see Fig. 20) by saying that the ka value in these cases is smaller than 1, therefore the reflection coefficient is smaller than its linear approximation. Another explanation of the results at low frequency and small circumferential extent can be proposed. The COD is zero at the boundary points where the crack starts and ends; consequently in the regions of the crack near these boundary points the COD goes from zero to a finite value. When the defect is big enough in terms of circumferential extent, the COD might be considered to be constant and the linear approximation for the reflection coefficient would be valid. However, when the defect is relatively small, the boundary effects described would dominate the COD. In this case the COD cannot be considered as constant over the length of the crack and the reflection coefficient will not be satisfactorily approximated by a linear fit.

Another issue to be tackled is the understanding of mode conversion. Following the analysis of Ditri[6] and Lowe *et al.*,[7] the strength of conversion to each mode by a circumferential crack may be estimated from the degree to which the crack opening profile matches the stresses in the mode. Let us first consider the simple case in which only one of the modes of a given circumferential order existing at the frequency of interest has a stress profile similar to the incident mode. In this case it is possible to evaluate the reflection coefficient of this mode converted mode by using a spatial Fourier decomposition of the displacement around the circumference at the location of the crack. As shown by Lowe *et al.*[7] this spatial Fourier transform gives the excitation strengths of the mode converted waves. This is applicable for example to the $F(1,2)$ mode in a 3 inch pipe when the $T(0,1)$ mode is incident at 100 kHz. When more than one mode (or none of the modes) with a given circumferential order has a mode shape similar to the stress profile of the incident mode, the simple spatial Fourier decomposition is no longer valid. Using the spatial Fourier decomposition and supposing that only one of the possible order two modes is involved in the mode conversion, the value of reflection coefficient at a circumferential extent of 50% would be zero, as explained by Lowe *et al.*[7] In fact in all of the modelled cases in this paper this value is not zero [see $F(2,2)$ at 50% circumferential extent in Figs. 8, 9, and 11]. This is due to the fact that more than one order two mode is involved in the mode conversion phenomenon.

2. Discussion on axisymmetric cracks

In the case of axisymmetric cracks the scattering regime in which we are interested is the low-intermediate frequency shown in Fig. 19(b). In Fig. 13 the amplitude reflection coefficient for a series of crack depths is plotted as a function of frequency, the $ka=1$ points being indicated with empty circles. We can notice that the curve of reflection coefficient versus frequency is roughly linear until it approaches the value $ka=1$. It is also interesting to notice that at $ka=1$ the reflection coefficient is approximately equal to the percentage depth of the defect. We then see a small change in behavior going from the intermediate to the high frequency scattering regime. The reflection coefficient curve is less steep at high frequency which can be explained by saying that the reflection coefficient tends to a value which is the asymptotic value at high frequency. This behavior has been encountered in previous work on scattering from cracks in plates.[9,25] Unfortunately in our case it was not possible to increase the test frequency in order to derive the value of the reflection coefficient at very high frequency because another torsional mode appears at 300 kHz [$T(0,2)$].

In Fig. 12 we showed the amplitude reflection coefficient of the $T(0,1)$ mode for axisymmetric cracks as a function of defect depth at different frequency values. The value at which the high frequency scattering regime is reached is indicated in the figure by an empty circle. Around $ka=1$ the curve changes its shape from convex to concave. This behavior can again be attributed to a change in scattering regime.

However, the reflection coefficient is not only ka dependent. Figure 12 shows that if we consider one pipe size with two defect depths and the same ka value, a deep defect at low frequency produces larger reflection coefficient than a shallow defect at high frequency. For example, the reflection coefficient for an 80% depth defect at 40 kHz is 31% whereas that from a 20% defect depth at 160 kHz is 5.3%. This is probably because the geometry of the system, and hence the stress distribution, is very different for shallow and deep defects; the reflection coefficient would be dependent on ka only at small defect depths (e.g., below 10%). However, it has not been possible to check this because of limitations on the finite element mesh.

B. Effect of axial extent

Figure 14 shows the amplitude reflection coefficient for a series of axisymmetric notches with 20% defect depth and different axial extent at 100 kHz. The notch case is shown schematically in Fig. 21. The simplest case to be considered is that in which the notch has an axial extent which is long enough to separate in time the first reflection coming from

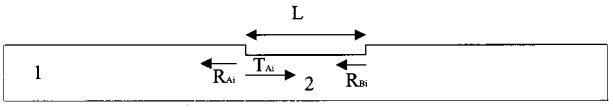


FIG. 21. Example notch case to explain reflection and transmission characteristics at the start and at the end of the notch.

the start of the notch and the second reflection coming from the end of the notch. The reflection coefficient from a step down in a pipe in which we removed 20% of the thickness is plotted in Fig. 22 as a function of frequency. It was simply obtained by dividing the FFT of the reflected signal by the FFT of the input signal. It is clear that this reflection is almost independent of frequency. No mode conversion is present in this case since the defect is axisymmetric. Furthermore the $T(0,1)$ mode is not dispersive so there is no change in velocity due to the change in the thickness of the pipe wall. If we consider that the tangential displacement is almost constant through the thickness and that the radius of the pipe is very much bigger than the thickness, we can approximate the value of reflection coefficient obtained at a step in a pipe by using the formula

$$R=\frac{1-\alpha}{1+\alpha} \tag{2}$$

in which $\alpha=A_2/A_1$, where A_1 is the cross-section area before the notch and A_2 is the cross section at the notch location. The signal arriving at the start of the notch has a component which is reflected and a component which is transmitted. The same phenomenon is repeated at the end of the notch. Let us now consider all of the reflections inside the notch (back and forth):

$$R_{A1}=\frac{1-\alpha}{1+\alpha}, \tag{3}$$

$$T_{A1}=\frac{2}{1+\alpha}, \tag{4}$$

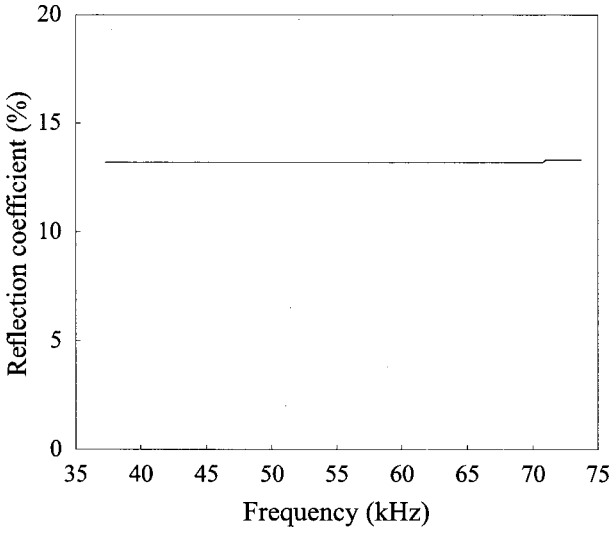


FIG. 22. Variation of reflection ratio with frequency for a step of 20% of the thickness of the pipe. Results are for $T(0,1)$ incident in 3 inch pipe using signal with center frequency of 55 kHz.

$$R_{B1}=\frac{\alpha-1}{1+\alpha}, \tag{5}$$

where R_{A1} is the reflection coefficient from the start of the notch, T_{A1} is the transmission coefficient past the start of the notch and R_{B1} is the signal reflection coefficient from the end of the notch. The reflection described by (5) is time delayed by L/V compared to that of (3) where L is the axial extent of the notch and V is the velocity of the torsional wave, therefore the total phase shift between the two reflections is $\pi+2fL\pi/V$, where the addition of π takes into account the sign reversal in the numerator between (3) and (5), and f is the frequency of the wave. The transmission coefficient at the end of the notch, T_{B1}, is

$$T_{B1}=\frac{2\alpha}{1+\alpha} \tag{6}$$

also with time delay L/V compared to T_{A1}. Similarly, the second reflection from the start of the notch, R_{A2} is

$$R_{A2}=T_{A1}\cdot R_{B1}\cdot T_{B1}, \tag{7}$$

with time delay $2L/V$;

$$R_{A3}=T_{A1}\cdot (R_{B1})^3\cdot T_{B1}, \tag{8}$$

with time delay $4L/V$;

$$R_{Aj}=T_{A1}\cdot (R_{B1})^{2j-3}\cdot T_{B1}, \tag{9}$$

with time delay $j\cdot 2\cdot L/V$. The total reflection coefficient is given by

$$R_{\mathrm{TOT}}=R_{A1}+R_{A2}+\cdots+R_{Aj}. \tag{10}$$

The total reflection coefficient has a maximum when L is equal to a quarter wavelength ($\lambda/4$) and a minimum when L is equal to $\lambda/2$. The maximum is about twice the value of the reflection from a step down of the same depth of the notch. The FE predictions of Fig. 14 and the experiments of Fig. 18 confirm that the maximum is obtained at about $\lambda/4$ and the minimum at about $\lambda/2$. In practice, the reflection maxima and minima are not at axial extents of exactly 25% and 50% of the wavelength. This is because the phase delay at a transition between two different thicknesses is not zero or π as predicted from equations (3) and (5), respectively. It should be stressed here that such a large increase in reflectivity when the defect length is a quarter wavelength long is unlikely in practice since real defects would not have a sharp, rectangular profile. Nevertheless, maxima and minima can occur in practical inspection and it is wise to test at more than one frequency.

VIII. CONCLUSIONS

A quantitative study of the reflection of the $T(0,1)$ mode from defects in pipes in the frequency range 10–300 kHz has been carried out, finite element predictions being validated by experiments on selected cases. Both cracklike defects with zero axial extent and notches with varying axial extents have been considered. The predictions have largely been done on part-thickness, axisymmetric defects and on full-

wall-thickness, part-circumference defects, both of which can be modelled with two-dimensional analyses. However, it has also been shown that it is reasonable to use a combination of these two approaches to predict the reflection from part-thickness, part-circumference defects.

It has been shown that the reflection coefficient from axisymmetric cracks increases monotonically with depth at all frequencies and increases with frequency at a given depth. In the frequency range of interest, $T(0,1)$ is the only propagating axisymmetric mode so there is no mode conversion at axisymmetric defects. With nonaxisymmetric cracks, the reflection coefficient is a roughly linear function of the circumferential extent of the defect at relatively high frequencies, the reflection coefficient at low circumferential extents falling below the linear prediction at lower frequencies. With nonaxisymmetric defects, mode conversion to the $F(1,2)$ mode occurs at all but the lowest frequencies, the amplitude of the mode converted signal being a maximum when the circumferential extent is 50%; at low circumferential extents, the amplitudes of the mode converted and direct reflections are similar. Some mode conversion to $F(1,3)$ as well as $F(1,2)$ is seen at lower frequencies.

The depth and circumferential extent of the defect are the parameters controlling the reflection from cracks; when notches having finite axial extent, rather than cracks, are considered, interference between the reflections from the start and the end of the notch causes a periodic variation of the reflection coefficient as a function of the axial extent of the notch, maxima occurring when the notch width is 25% and 75% of the wavelength, and minima appearing when the width is zero (the crack case), 50% or 100% of the wavelength. This paper has only considered square sided defects; real defects will not have such regular shapes so the interference effects will be less severe. However, some frequency dependence is likely to be seen.

The results have been explained in terms of the wavenumber-defect size product, ka. Low frequency scattering behavior is seen when $ka<0.1$, high frequency scattering characteristics being seen when $ka>1$.

The results demonstrate that the torsional wave is attractive for the practical testing of pipes. The sensitivity to defects is improved at higher frequencies, though this conflicts with the desire to use lower frequencies to increase the propagation range, and hence the length of pipe that can be inspected in a single test. When testing for defects that may have significant axial extent, it is wise to test at more than one frequency in order to avoid missing defects due to destructive interference of the reflections from the two ends of the defect.

The 3 inch and 24 inch pipes have been considered in this paper. It would be valuable to generalize the results to other pipe sizes and crack dimensions without the need to analyze each specific case. This will be the subject of a future paper.

[1] R. Thompson, G. Alers, and M. Tennison, "Application of direct electromagnetic Lamb wave generation to gas pipeline inspection," Proceedings of the 1971 IEEE Ultrasonic Symposium, (IEEE, New York, 1972), pp. 91–94.

[2] W. Mohr and P. Holler, "On inspection of thin-walled tubes for transverse and longitudinal flaws by guided ultrasonic waves," IEEE Trans. Sonics Ultrason. **SU-23**, 369–378 (1976).

[3] M. Silk and K. Bainton, "The propagation in metal tubing of ultrasonic wave modes equivalent to Lamb waves," Ultrasonics **17**, 11–19 (1979).

[4] M. Brook, T. Ngoc, and J. Eder, "Ultrasonic inspection of steam generator tubing by cylindrical guided waves," in *Review of Progress in Quantitative NDE*, edited by D. Thompson and D. Chimenti (Plenum, New York, 1990), pp. 243–249.

[5] J. Ditri, J. Rose, and A. Pilarski, "Generation of guided waves in hollow cylinders by wedge and comb type transducers," in *Review of Progress in Quantitative NDE*, edited by D. Thompson and D. Chimenti (Plenum, New York, 1993), pp. 211–218.

[6] J. Ditri, "Utilization of guided elastic waves for the characterization of circumferential cracks in hollow cylinders," J. Acoust. Soc. Am. **96**, 3769–3775 (1994).

[7] M. Lowe, D. Alleyne, and P. Cawley, "The mode conversion of a guided wave by a part-circumferential notch in a pipe," J. Appl. Mech. **65**, 649–656 (1998).

[8] D. Alleyne, M. Lowe, and P. Cawley, "The reflection of guided waves from circumferential notches in pipes," J. Appl. Mech. **65**, 635–641 (1998).

[9] M. Lowe, "Characteristics of the reflection of lamb waves from defects in plates and pipes," in *Review of Progress in Quantitative NDE*, edited by D. Thompson and D. Chimenti (Plenum, New York, 1998), p. 113.

[10] D. Alleyne, B. Pavlakovic, M. Lowe, and P. Cawley, "Rapid, long range inspection of chemical plant pipework using guided waves," Insight **43**, 93–96, 101 (2001).

[11] H. Shin and J. Rose, "Guided wave tuning principles for defect detection in tubing," J. Nondestruct. Eval. **17**, 27–36 (1998).

[12] G. Kino, *Acoustic Waves: Devices, Imaging and Analogue Signal Processing* (Prentice-Hall, New Jersey, 1987).

[13] B. Pavlakovic, M. Lowe, D. Alleyne, and P. Cawley, "DISPERSE: A general purpose program for creating dispersion curves," in *Review of Progress in Quantitative NDE*, edited by D. Thompson and D. Chimenti (Plenum, New York, 1997), Vol. 16, pp. 185–192.

[14] D. Alleyne and P. Cawley, "Long range propagation of Lamb waves in chemical plant pipework," Mater. Eval. **55**, 504–508 (1997).

[15] D. Alleyne and P. Cawley, "The interaction of Lamb waves with defects," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 381–397 (1992).

[16] P. Wilcox, M. Lowe, and P. Cawley, "A signal processing technique to remove the effect of dispersion from guided wave signals," in *Review of Progress in Quantitative NDE*, edited by D. Thompson and D. Chimenti (American Institute of Physics, New York, 2001), Vol. 20, pp. 555–562.

[17] W. Zhuang, A. Shah, and S. Datta, "Axisymmetric guided wave scattering by cracks in welded steel pipes," J. Pressure Vessel Technol. **119**, 401–406 (1997).

[18] C. Valle, M. Niethammer, J. Qu, and L. Jacobs, "Crack characterization using guided circumferential waves," J. Acoust. Soc. Am. **110**, 1282–1290 (2001).

[19] B. Pavlakovic, "Leaky guided ultrasonic waves in NDT," Ph.D. thesis, University of London, 1998.

[20] D. Alleyne and P. Cawley, "The excitation of Lamb waves in pipes using dry-coupled piezoelectric transducers," J. Nondestruct. Eval. **15**, 11–20 (1996).

[21] D. Alleyne, "The nondestructive testing of plates using ultrasonic lamb waves," Ph.D. thesis, University of London, 1991.

[22] M. Koshiba, H. Hasegawa, and M. Suzuki, "Finite-element solution of horizontally polarized shear wave scattering in an elastic plate," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **34**, 461–466 (1987).

[23] F. Moser, L. Jacobs, and J. Qu, "Modelling elastic wave propagation in waveguides with finite element method," NDT & E Int. **33**, 225–234 (1999).

[24] N. Kishore, I. Sridhar, and N. Iyengar, "Finite element modelling of the scattering of ultrasonic waves by isolated flaws," NDT & E Int. **33**, 297–305 (2000).

[25] M. Lowe and O. Diligent, "Low frequency reflection characteristics of the S0 Lamb wave from a rectangular notch in a plate," J. Acoust. Soc. Am. **111**, 64–74 (2002).

[26] E. Kraut, "Review of theories of scattering of elastic waves by cracks," IEEE Trans. Sonics Ultrason. **SU-23**, 162 (1972).

[27] S. Datta, "Scattering of elastic waves," Mechanics Today **4**, 149–208 (1978).

[28] J. Achenbach, A. Gautesen, and H. McMaken, *Ray Methods for Waves in Elastic Solids* (Pitman Books Limited, New York, 1982).
[29] D. Alleyne and P. Cawley, ''The interaction of Lamb waves with defects,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 381–397 (1992).
[30] Y. Wang and J. Shen, ''Scattering of elastic waves by a crack in a isotropic plate,'' Ultrasonics **35**, 451–457 (1997).
[31] J. Wang, K. Lam, and G. Liu, ''Wave scattering of interior vertical crack in plates and detection of the crack,'' Eng. Fract. Mech. **59**, 1–6 (1998).
[32] H. Bai, A. Shah, N. Popplewell, and S. Datta, ''Scattering of guided waves by circumferential cracks in steel pipes,'' J. Appl. Mech. **68**, 619–631 (2001).
[33] B. Auld, *Acoustic Fields and Waves in Solids* (Krieger, Malabar, FL, 1990), Vol. 2.

A Born scatterer in an acoustical waveguide

D. N. Ghosh Roy
Center for Space Physics, Boston University, Boston, Massachusetts 02215

Gregory J. Orris
Naval Research Laboratory, Washington, DC 20375

(Received 24 May 2002; revised 24 April 2003; accepted 19 May 2003)

Scattering from a Born inhomogeneity in a homogeneous, acoustical waveguide is considered and results compared to the corresponding scattering in a homogeneous, unbounded medium. It is found that the Ewald sphere in the unbounded medium case is replaced by a Ewald ''strip'' in a waveguide, the strip consisting of many individual Ewald spheres embedded in a ball the radius of which is twice that of an individual sphere. The physics behind the Ewald strip is discussed along with the implications of waveguide Born data. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1590974]

PACS numbers: 43.20.Fn, 43.60.Pt, 43.20.Mv [LLT]

I. INTRODUCTION

In this paper, we consider the problem of scattering from a compactly supported *inhomogeneity* occupying a region of space Ω in a homogeneous ambient medium Ω^0. Ω, $\Omega^0 \subset \mathbb{R}^3$. The scattering of the field (e.g., velocity potential) ϕ is governed by the following boundary value problem:

$$[\Delta + k^2(\mathbf{r})]\phi(\mathbf{r}) = 0 \quad \mathbf{r} \in \Omega, \tag{1}$$

$$[\Delta + k_0^2]\phi_0(\mathbf{r}) = 0 \quad \mathbf{r} \in \Omega^0, \tag{2}$$

$$\phi = \phi_0 \quad \mathbf{r} \in \partial\Omega, \tag{3}$$

and

$$\partial_n \phi = \partial_n \phi_0 \quad \mathbf{r} \in \partial\Omega, \tag{4}$$

in addition to Sommerfeld's radiation condition at infinity. In Eq. (1), the space-dependent wave number $k(\mathbf{r})$ describes the inhomogeneity, and k_0 in Eq. (2) is the wave number of the homogeneous ambient medium Ω^0. For example, for the *variable velocity* problem, in which density is held constant throughout the scatterer, one has[1]

$$k^2(\mathbf{r}) = k_0^2\left[\frac{c_0^2}{c^2(\mathbf{r})}\right] = k_0^2 n^2(\mathbf{r}),$$

$c(\mathbf{r})$ being the speed of sound in the inhomogeneity and $n(\mathbf{r})$ its index of refraction. Δ is the three-dimensional Laplacian operator, $\partial\Omega$ is the boundary of the scatterer, and ∂_n denotes the normal derivative on the boundary. $\partial\Omega$ is assumed to be smooth. Moreover, ϕ is the total field, that is

$$\phi = \phi^i + \phi^s.$$

ϕ^s is the scattered field, and similarly for ϕ_0. ϕ^i is the incident plane-wave field. The boundary conditions (3) and (4) are characteristics of an inhomogeneity as opposed to an *obstacle*.[2] For an obstacle, the normal derivative undergoes a jump discontinuity across $\partial\Omega$.

By defining the scattering function $V(\mathbf{r})$ as

$$V(\mathbf{r}) = k_0^2(1 - n^2(\mathbf{r})),$$

the scattered field ϕ^s is described by the well-known *Lippman–Schwinger equation*, and can be written as[3]

$$\phi^s(\mathbf{r}) = -\int_\Omega d^3r' G_0(\mathbf{r},\mathbf{r}';k_0) V(\mathbf{r}')\phi(\mathbf{r}'), \tag{5}$$

where $G_0(\mathbf{r},\mathbf{r}';k_0)$ is the free-space Green's function satisfying

$$\Delta G_0(\mathbf{r},\mathbf{r}';k_0) + k_0^2 G_0(\mathbf{r},\mathbf{r}';k_0) = -\delta^3(\mathbf{r}-\mathbf{r}'),$$

namely

$$G_0(\mathbf{r},\mathbf{r}',k_0) = G_0(|\mathbf{r}-\mathbf{r}'|;k_0) = \frac{e^{ik_0|\mathbf{r}-\mathbf{r}'}}{4\pi|\mathbf{r}-\mathbf{r}'|}.$$

In the limit as $|\mathbf{r}| \to \infty$, the free-space Green's function becomes

$$\lim_{r\to\infty} G_0(\mathbf{r},\mathbf{r}';k_0) = \frac{e^{ik_0 r}}{4\pi r} e^{-ik_0 \hat{\mathbf{r}}\cdot\mathbf{r}'}, \tag{6}$$

where $\hat{\mathbf{r}}$ is the unit vector in the direction of the vector $\mathbf{r}$ and r denotes its magnitude, i.e., $r = |\mathbf{r}|$.

Consider that the scatterer Ω is placed in an infinite medium and a plane wave is incident upon it. That is, $\Omega^0 \in \mathbb{R}^3 \backslash \overline{\Omega}$, $\phi^i(\mathbf{r}) = e^{ik_0\hat{\mathbf{k}}\cdot\mathbf{r}}$, and the radiation condition at infinity is

$$\lim_{r\to\infty} r\left(\frac{\partial\phi^s}{\partial r} - ik_0\phi^s\right) = o\left(\frac{1}{r^2}\right),$$

uniformly in $\hat{\mathbf{r}} \in S^2$ the unit sphere in $\mathbb{R}^3$. The corresponding scattering geometry is shown in Fig. 1. Using the large r limit of Green's function given in Eq. (6), the scattered field in the asymptotic limit becomes

$$\phi_\infty(\mathbf{r}) = \frac{e^{ik_0 r}}{4\pi r}\phi_\infty(\hat{\mathbf{r}};\hat{\mathbf{k}}) + \mathcal{O}\left(\frac{1}{r^2}\right). \tag{7}$$

$\phi_\infty(\hat{\mathbf{r}},\hat{\mathbf{k}})$ in Eq. (7) is the *scattering amplitude*. If the field $\phi(\mathbf{r}')$ under the integral in the Lippman–Schwinger equation (5) is replaced by the incident field, $\phi^i(\mathbf{r}')$, then the scattering amplitude $\phi_\infty(\hat{\mathbf{r}};\hat{\mathbf{k}})$ reduces to

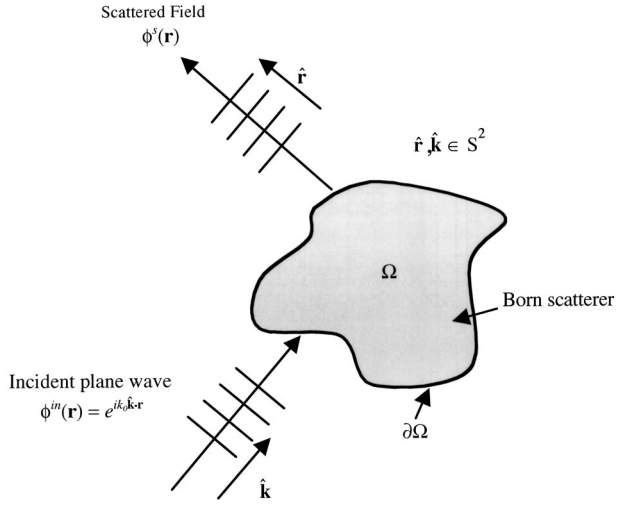


FIG. 1. Scattering of a plane wave from an object in an infinite homogeneous medium. $\hat{\mathbf{k}}$ is the unit vector along the direction of the incident wave and $\hat{\mathbf{r}}$ is the unit vector along the direction of scattering. $\hat{\mathbf{r}}$ and $\hat{\mathbf{k}} \in S^2$.

$$\phi_\infty^B(\hat{\mathbf{r}};\hat{\mathbf{k}}) = -\int_\Omega d^3r' e^{-ik_0(\hat{\mathbf{r}}-\hat{\mathbf{k}})\cdot\mathbf{r}'} V(\mathbf{r}'), \tag{8}$$

where in fact $\phi_\infty^B(\hat{\mathbf{r}},\hat{\mathbf{k}})$ is actually $\phi_\infty^B(\hat{\mathbf{r}}-\hat{\mathbf{k}})$. Equation (8) is the celebrated *Born approximation*.[4,5] This is explicitly indicated by the superscript "*B*" on ϕ_∞. The approximation holds under the condition that the integral operator T_{LS} of the Lippman–Schwinger equation satisfies the bound[6]

$$\|T_{\mathrm{LS}}\|_\infty \leqslant \frac{(k_0 R_s)^2}{2}\left\| 1 - \frac{c_0^2}{c^2}\right\|_\infty,$$

in the supremum norm, where R_S is the scale size of the scattering region. This is the well-known *weak scattering approximation*.

Equation (8) shows that the scattering amplitude under the Born approximation is the Fourier transform of the function $V(\mathbf{r})$ describing the inhomogeneity, the conjugate wave number being $\tilde{V}(\hat{\mathbf{r}}-\hat{\mathbf{k}})$. The Fourier transform of a function $f(\mathbf{r})$ was defined as

$$\tilde{f}(\boldsymbol{\zeta}) = \int_{\mathbb{R}^3} f(\mathbf{r}) e^{-i\boldsymbol{\zeta}\cdot\mathbf{r}} d^3r.$$

It is also clear that for a specified incident wave, the Fourier transform, $\tilde{V}(\hat{\mathbf{r}}-\hat{\mathbf{k}})$ of the unknown function $V(\mathbf{r})$ is obtained only on a sphere of radius k_0, the so-called *Ewald sphere* (S_E in Fig. 2). An excellent discussion of Ewald spheres was given by Langenberg.[5] Any reconstruction based on this information alone would necessarily be nonunique.

Next, we replace the infinite medium by a waveguide, and place the Born scatterer Ω in the guide. It is assumed that the scatterer function $V(\mathbf{r}) \in \mathbf{L}^2(\Omega)$, i.e., $V(\mathbf{r})$ is square integrable. $V(\mathbf{r})$ is also considered to be of short range compared to the height H of the waveguide. In other words, it decays rapidly enough so as to vanish on the waveguide boundaries. It is further assumed that the scatterer is located at a sufficient depth in the waveguide such that the smallest sphere circumscribing it does not intersect the plane-bounding surfaces of the guide.

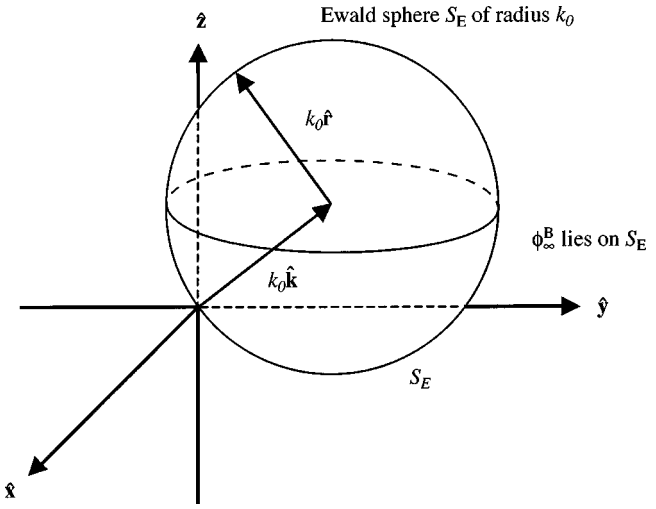


FIG. 2. A Ewald sphere of radius k_0 is shown in this figure. For scattering in an unbounded homogeneous medium under the Born approximation and for a fixed incidence, the scattering amplitude data reside on the sphere.

The objective of this paper is to determine what happens to the Ewald sphere picture of Fig. 2, and the Born data $\phi_\infty^B(\hat{\mathbf{r}}-\hat{\mathbf{k}})$ as the infinite medium is replaced by a waveguide. In particular, attention is focused on comparing the Born data in a waveguide with that in an unbounded medium as to their relative information content in the Fourier domain. No attempt is made here in solving the inverse reconstruction problem. Our approach is heuristic and physical rather than rigorous and mathematical.

II. BORN SCATTERING IN A WAVEGUIDE

The geometry of scattering from a Born object in a waveguide of depth H is shown in Fig. 3. The upper surface of the waveguide is assumed to be a pressure-release surface (i.e., satisfies a Dirichlet boundary condition), whereas the bottom surface is chosen to be rigid (a Neumann boundary). The origin of the coordinate system is assumed to be inside the scatterer which is located at a depth d from the pressure release surface. A time-harmonic source is situated at $\mathbf{r}_0$ and a receiver at $\mathbf{r}$ inside the waveguide. Cylindrical coordinates are used throughout. These are shown in Fig. 4. The vertical z axis is positive in the downward direction, as shown in the figure. $\mathbf{r}_0 = (\mathbf{r}_{0\perp}, z_0)$, $\mathbf{r} = (\mathbf{r}_\perp, z)$, and $\mathbf{r}' = (\mathbf{r}'_\perp, z')$. $\mathbf{r}'$ is a point of the scatterer, and the subscript $\perp$ denotes the projection on the horizontal plane through the center O. $\hat{\mathbf{r}}_{0\perp}$ and $\hat{\mathbf{r}}_\perp$ are unit vectors along $\mathbf{r}_{0\perp}$ and $\mathbf{r}_\perp$, respectively. z_0, z, and z' are the vertical components of $\mathbf{r}_0$, $\mathbf{r}$, and $\mathbf{r}'$, respectively.

The wave field ϕ in the waveguide is determined from the following problem, namely:

$$[\Delta + k_0^2]\phi(\mathbf{r}) = \delta^3(\mathbf{r}-\mathbf{r}_0), \quad \mathbf{r} \in \Omega^0$$

$$\phi = 0, \quad z = -d,$$

$$\partial_n \phi = 0, \quad z = H - d$$

$$\phi_+ = \phi_- \quad \text{on} \quad \partial\Omega,$$

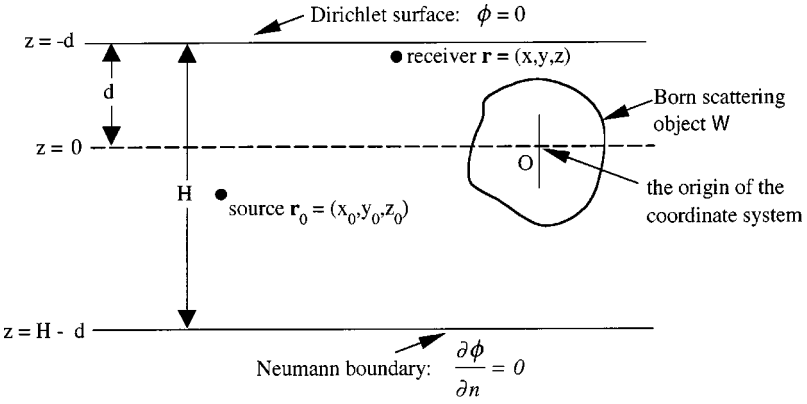


FIG. 3. A Born scatterer Ω in an acoustical waveguide of height H. The field ϕ at the upper boundary obeys Dirichlet conditions, and at the lower boundary, Neumann boundary conditions. The center of the coordinate system is inside the scatterer and located at a depth d from the upper boundary.

and

$$\partial_n \phi_+ = \partial_n \phi_- \quad \text{on} \quad \partial\Omega.$$

$\Omega^0 = \mathbb{R}^3_H \backslash \overline{\Omega}$, where $\mathbb{R}^3_H = \{\mathbf{r} = (\mathbf{r}_\perp, z), \mathbf{r}_\perp \in \mathbb{R}^2, 0 \leqslant z \leqslant H\}$. ϕ_+, ϕ_- denote the wave fields in the exterior and interior of $\partial\Omega$, respectively. It is assumed that $\phi \in C^2(\mathbb{R}^3_H \backslash \overline{\Omega}) \cap C(\mathbb{R}^3_H \backslash \Omega)$. The radiation condition must be satisfied as $|\mathbf{r}| \to \infty$. It is known that the above boundary value problem has a unique solution.[3]

Let $R(\mathbf{r}') \in \mathbf{L}^2(\Omega)$ describe the scatterer inhomogeneity. Using R emphasizes the fact that in the Born approximation, the scattering is essentially a reflection of the incident field at every point $\mathbf{r}'$ of the inhomogeneity, the strength of the reflection at that point being $R(\mathbf{r}')$. Then, in the Born approximation, the Lippmann–Schwinger equation (5) determines the scattered field at $\mathbf{r} \in \mathbb{R}^3_H \backslash \overline{\Omega}$ to be

$$\phi^{sB}(\mathbf{r}) = \int_\Omega d^3 r' G(\mathbf{r}', \mathbf{r}_0) R(\mathbf{r}') G(\mathbf{r}, \mathbf{r}'). \tag{9}$$

It is further assumed that the scattered fields are being observed on a cylindrical surface of constant radius R_s with its center at the origin O.

The Green's function between two position vectors $\mathbf{r}_a$ and $\mathbf{r}_b$ in a waveguide with a Dirichlet top and a Neumann bottom boundary is known to be[7]

$$G(\mathbf{r}_a|\mathbf{r}_b) = \frac{i}{\sqrt{2H}} \sum_\ell \psi_\ell(z_a) \psi_\ell(z_b) H_0^{(1)}(k_\ell |\mathbf{r}_{a\perp} - \mathbf{r}_{b\perp}|), \tag{10}$$

where $H_0^{(1)}$ is the Hankel function of the first kind of order zero. The source term in the equation for $G(\mathbf{r}_a|\mathbf{r}_b)$ was taken to be a point source in an axially symmetric environment

$$S(\mathbf{r}, \mathbf{r}_S) = -\frac{1}{2\pi r} \delta(r - r_s) \delta(z - z_s),$$

where the subscript "s" indicates the source position, and the radial Green's function is

$$G(\mathbf{r}_a|\mathbf{r}_b) = \frac{i}{4} H_0^{(1)}(k_l |\mathbf{r}_{a\perp} - \mathbf{r}_{b\perp}|).$$

The functions $\psi_\ell(z)$ are the modal wave functions given by

$$\psi_\ell(z) = \sqrt{\frac{2}{H}} \sin(\lambda_\ell (z + d)),$$

where

$$\gamma_\ell = \frac{\ell + \frac{1}{2}}{H} \pi, \quad \ell = 1,2,...,$$

and

$$k_\ell = \left[k_0^2 - \frac{1}{H^2} \left(\ell + \frac{1}{2} \right)^2 \pi^2 \right]^{1/2}.$$

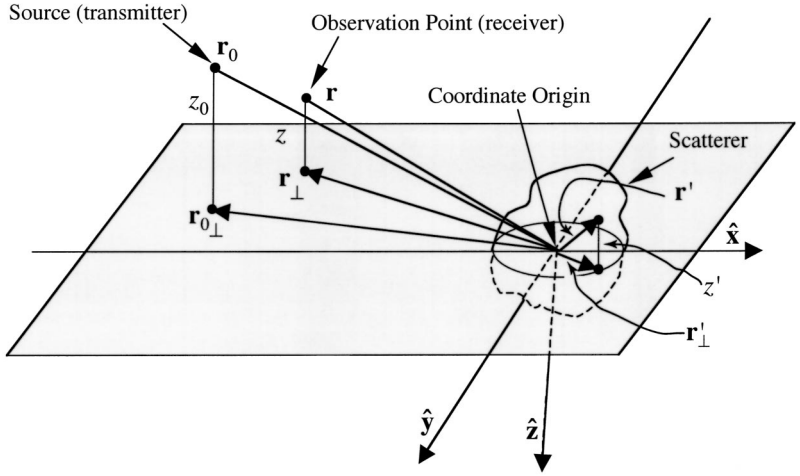


FIG. 4. The geometry of scattering in a waveguide showing the cylindrical coordinate system with origin inside the scatterer.

The wave number k_0 is that of the monochromatic point source. Moreover, we have the important relation

$$\gamma_\ell^2 + k_\ell^2 = k_0^2. \tag{11}$$

ψ_ℓ is valid to within a multiplicative factor. However, a resonance will be allowed in the guide so that γ_l can equal the free-space wave number k_0. Moreover, since the index $\ell = 1,2,...,N$, no confusion with the free-space wave number k_0 can arrise by using the letter k to indicate the horizontal wave number also. Taking the asymptotic limit $|\mathbf{r}_{a\perp}| \gg |\mathbf{r}_{b\perp}|$ in the Hankel function $H_0^{(1)}(k_\ell|\mathbf{r}_{a\perp} - \mathbf{r}_{b\perp}|)$ in Eq. (10), Green's function in this limit becomes

$$G(\mathbf{r}_a|\mathbf{r}_b) = \sum_\ell \psi_\ell(z_a)\psi_\ell(z_b)\frac{e^{ik_\ell|\mathbf{r}_{a\perp}|}}{\sqrt{k_\ell|\mathbf{r}_{a\perp}|}} e^{-ik_\ell \mathbf{r}_{b\perp}\cdot\hat{\mathbf{r}}_{a\perp}}, \tag{12}$$

Using the asymptotic form of Green's function of Eq. (12) in the scattering equation (9), the far-field limit of the wave field ϕ_∞^B in the Born approximation is obtained as

$$\begin{aligned}\phi_\infty^B(\xi) = {} & \frac{i}{\sqrt{2H}} e^{-i(\pi/4)} \sum_\ell \sum_m \sin(\gamma_\ell(z_0+d)) \\ & \times \sin(\gamma_m(z+d)) \left[\frac{e^{ik_\ell|\mathbf{r}_{0\perp}|}}{\sqrt{k_\ell|\mathbf{r}_{0\perp}|}} \frac{e^{ik_m|\mathbf{r}_\perp|}}{\sqrt{k_m|\mathbf{r}_\perp|}}\right] \\ & \times \int_\Omega d^3r' R(\mathbf{r}') \sin[\gamma_\ell(z'+d)] \\ & \times \sin[\gamma_m(z'+d)] e^{-i(k_\ell\hat{\mathbf{r}}_{0\perp} + k_m\hat{\mathbf{r}}_\perp)\cdot\mathbf{r}'_\perp},\end{aligned} \tag{13}$$

where now $|\mathbf{r}_\perp| = R_s$ in Eq. (13), and the asymptotic form for the Hankel function $H_0^{(1)}(\lambda r)$ was used, namely

$$\lim_{r\to\infty} H_0^{(1)}(\lambda r) \approx \sqrt{\frac{2}{\pi r\lambda}} e^{i(\lambda r - \pi/4)}.$$

Moreover, $\xi = (\hat{\mathbf{r}}_{0\perp}, \hat{\mathbf{r}}_\perp; z)$. From now on, the constant factor $(i/\sqrt{wH})\exp(-i\pi/4)$ in front of the summation in Eq. (13) will be written simply as iC. Note that for each index m

$$\sqrt{r}\left(\frac{\partial\phi_\infty^B}{\partial r} - ik_m\phi_\infty^B\right) \to 0,$$

as $|\mathbf{r}_\perp| \to \infty$. The radiation condition at infinity is thus satisfied. Equation (13) can be rewritten as

$$\begin{aligned}\phi_\infty^B(\xi) = {} & iC\sum_\ell \sum_m \mathcal{F}_{\ell m}(\boldsymbol{\xi}) \int_\Omega d^3r' R(\mathbf{r}') \sin[\gamma_\ell(z'+d)] \\ & \times \sin[\gamma_m(z'+d)] e^{-i(k_\ell\hat{\mathbf{r}}_{0\perp} + k_m\hat{\mathbf{r}}_\perp)\cdot\mathbf{r}'_\perp},\end{aligned} \tag{14}$$

where $\boldsymbol{\xi} = (\gamma_\ell, \gamma_m, k_\ell, k_m, z_0, z, R_s, |\mathbf{r}_{0\perp}|, d)$ is a vector containing various parameters in its argument, and the terms outside the integral are absorbed in the factor $\mathcal{F}_{\ell m}(\xi)$.

Next, we break up the standing-wave parts of Eq. (14), that is, the terms involving the sine functions, in their plane-wave components. This results in the expression

$$\begin{aligned}&\sin(\gamma_\ell(z'+d))\sin(\gamma_m(z'+d)) e^{-i(k_\ell\hat{\mathbf{r}}_{0\perp} + k_m\hat{\mathbf{r}}_\perp)\cdot\mathbf{r}'_\perp} \\ &\quad = e^{i\gamma_{\ell m}^+ d} e^{i\mathbf{r}'\cdot(\mathbf{v}_{0\ell} + \mathbf{v}_m)} + e^{-i\gamma_{\ell m}^+ d} e^{i\mathbf{r}'\cdot(\mathbf{w}_{0\ell} + \mathbf{w}_m)} \\ &\qquad - e^{i\gamma_{\ell m}^- d} e^{i\mathbf{r}'\cdot(\mathbf{v}_{0\ell} + \mathbf{w}_m)} - e^{i\gamma_{\ell m}^- d} e^{i\mathbf{r}'\cdot(\mathbf{w}_{0\ell} + \mathbf{v}_m)},\end{aligned} \tag{15}$$

where

$$\gamma_{\ell m}^\pm = \gamma_\ell \pm \gamma_m, \quad \mathbf{v}_n = \gamma_n\hat{\mathbf{z}} - k_n\hat{\mathbf{r}}_\perp,$$

$$\mathbf{v}_{0n} = \gamma_n\hat{\mathbf{z}} - k_n\hat{\mathbf{r}}_{0\perp}, \quad \mathbf{w}_n = -\gamma_n\hat{\mathbf{z}} - k_n\hat{\mathbf{r}}_\perp,$$

$$\mathbf{w}_{0n} = -\gamma_n\hat{z} - k_n\hat{\mathbf{r}}_{0\perp}.$$

Let $\tilde{R}(\boldsymbol{\zeta})$ be the Fourier transform of $R(\mathbf{r})$, $\boldsymbol{\zeta}$ being the Fourier space wave vector. Then, according to Eq. (15), for fixed values of the indices ℓ, m, and fixed source and receiver positions, the wave vectors $\boldsymbol{\zeta}$ are determined by the vectors $\mathbf{v}_{0\ell} + \mathbf{v}_m$, $\mathbf{v}_{0\ell} + \mathbf{w}_m$, $\mathbf{w}_{0\ell} + \mathbf{w}_m$, and $\mathbf{w}_{0\ell} + \mathbf{v}_m$. Assume that the scattering data are acquired over a unit circle around the Born object, that is, $\hat{\mathbf{r}}_\perp$ is allowed to vary over S^1. Then, the Fourier transform, $\tilde{R}(\boldsymbol{\zeta})$, is obtained in that volume of the Fourier domain which is swept out as the mode indices ℓ and m vary over their allowed values, and the unit vector $\hat{\mathbf{r}}_\perp$ sweeps out the unit circle S^1.

Consider, for example, the vector $\mathbf{v}_{0\ell} + \mathbf{v}_m$. Now

$$\mathbf{v}_{0\ell} + \mathbf{v}_m = (\gamma_\ell\hat{\mathbf{z}} - k_\ell\hat{\mathbf{r}}_{0\perp}) + (\gamma_m\hat{\mathbf{z}} - k_m\hat{\mathbf{r}}_\perp),$$

and recall that the unit vector $\hat{\mathbf{r}}_{0\perp}$ is fixed. Note also that if the energy flux due to evanescent waves is absent in the scattered field (see the Discussion section), then $\gamma_{\ell(m)}$ is strictly positive, whereas $k_{\ell(m)}$ is semi-positive. The latter includes situations in which the height H of the waveguide is such that a resonance can be allowed. Most importantly, however, is the relation $\gamma_n^2 + k_n^2 = k_0^2$ as was given in Eq. (11). This leads to the fact that $|\mathbf{v}_n| = |\mathbf{w}_n| = k_0$. Let $(\gamma_\ell\hat{\mathbf{z}} - k_\ell\hat{\mathbf{r}}_{0\perp})$ be considered the "incident" wave, analogous to the vector $\hat{\mathbf{k}}_0$ in the case of an infinite medium (see Fig. 2). The corresponding "scattered" vector is $(\gamma_m\hat{\mathbf{z}} - k_m\hat{\mathbf{r}}_{0\perp})$, the base of which is at the tip of the incident vector $(\gamma_\ell\hat{\mathbf{z}} - k_\ell\hat{\mathbf{r}}_\perp)$. Again, this is analogous to the addition of the scattered vector $k_0\hat{\mathbf{r}}$ to the incident vector $k_0\hat{\mathbf{k}}$ in Fig. 2 for an unbounded medium. Clearly, the direction of the scattered vector depends upon the index m. Consequently, as m takes on its allowed values, the vector $(\gamma_m\hat{\mathbf{z}} - k_m\hat{\mathbf{r}}_\perp)$ sweeps out an arc which is centered at $(\gamma_\ell\hat{\mathbf{z}} - k_\ell\hat{\mathbf{r}}_{0\perp})$ even though the direction $\hat{\mathbf{r}}$ remains unchanged. This is unlike the infinite medium situation, where $k_0\hat{\mathbf{r}}$ sweeps out a circle of radius k_0 (in the plan of the incident direction) with its center at $k_0\hat{\mathbf{k}}$ (Fig. 2) only as $\hat{\mathbf{r}}$ is varied over the unit circle S^1. The variation of the vector $\mathbf{v}_{0\ell} + \mathbf{v}_m$ with m, $\hat{\mathbf{r}}_\perp$ being held fixed, is shown in Fig. 5(a) which is drawn in the plane of the incident direction $(\gamma_\ell\hat{\mathbf{z}} - k_\ell\hat{\mathbf{r}}_{0\perp})$. Without a loss in generality, it was assumed that $\ell = L$, in which case $k_L = 0$ and $\gamma_L = k_0$ (see Refs. 6 and 7, and Discussion below). The vector $(\gamma_m\hat{\mathbf{z}} - k_m\hat{\mathbf{r}}_\perp)$ is seen to sweep out a first quadrantal circular arc AB of radius k_0, the center of the arc being at $(0,k_0)$. The $\boldsymbol{\zeta}$-space picture corresponding to the full Eq. (15) is shown in the same Fig. 5(a), where $\ell = L$, $\hat{\mathbf{r}}_\perp$ is held constant, and m varies over all its indices. The various quadrantal arcs OB, OC, and CD in this

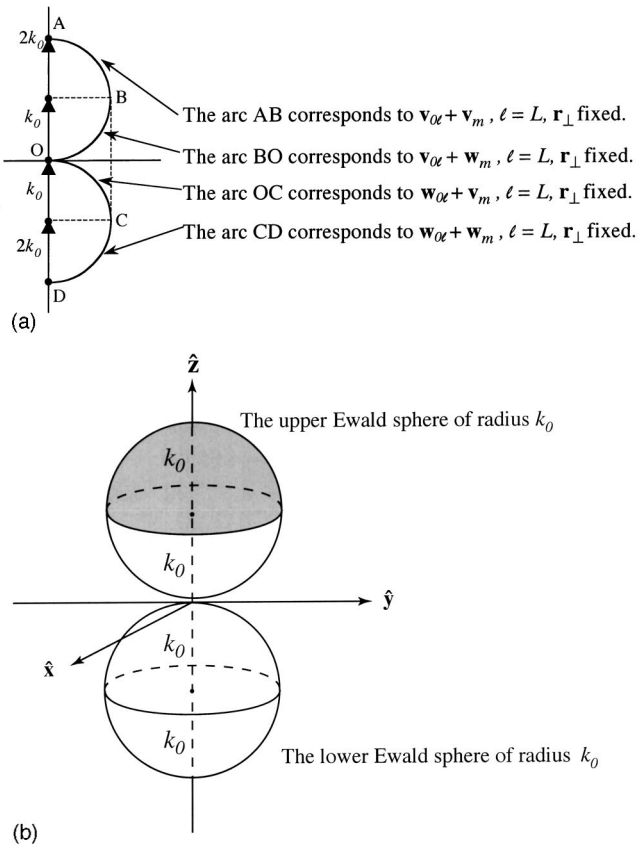


FIG. 5. (a) The $\boldsymbol{\zeta}$-space picture for $\ell = L$, i.e., $\gamma_L = k_0$, $\hat{\mathbf{r}}_\perp$ fixed, and the index m varying over its allowed values. (b) Same as (a), but after $\mathbf{r}_\perp$ sweeps through the unit circle S^1.

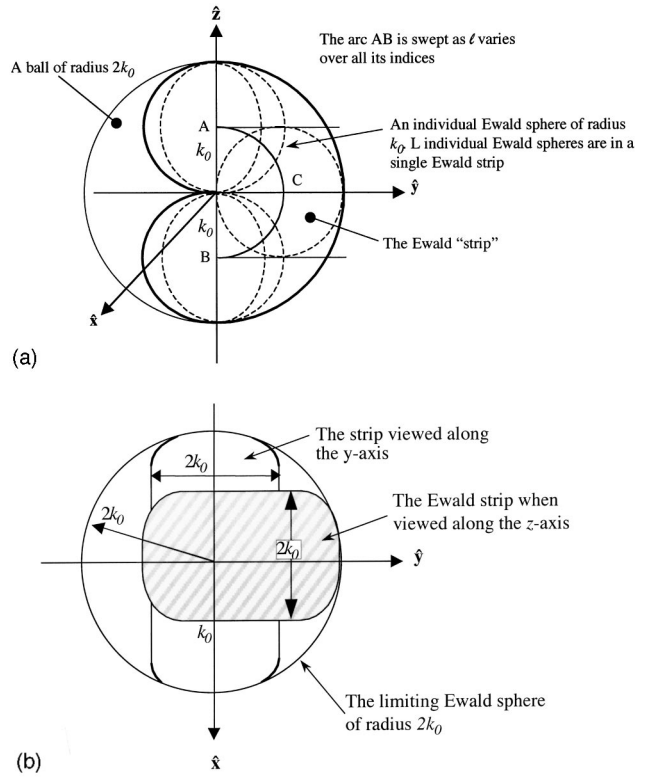


FIG. 6. Figure 5(b) after the index ℓ is varied over all allowed values. The figure demonstrates how a Ewald sphere for an unbounded, homogeneous medium is transformed into a Ewald strip in a waveguide. The Ewald strips looking along the y-and z axes. The strip width is $2k_0$ and is situated in a ball of the same radius, the so-called "limiting Ewald sphere."

figure are generated by $\mathbf{v}_{0\ell}+\mathbf{w}_m$, $\mathbf{w}_{0\ell}+\mathbf{v}_m$, and by $\mathbf{w}_{0\ell}+\mathbf{w}_m$, respectively.

In the next step, $\hat{\mathbf{r}}_\perp$ is allowed to sweep over the unit circle S^1, all other considerations remaining the same as in Fig. 5(a). As $\hat{\mathbf{r}}_\perp$ varies over S^1, each arc sweeps out a hemisphere with base perpendicular to the $\hat{\mathbf{k}}_0$ direction. Thus, for example, the arc AB sweeps out the upper hemisphere of the upper Ewald sphere S_U, shown shaded in Fig. 5(b). The result is Fig. 5(b). Two Ewald spheres, each of radius k_0, result and these are tangent to each other at the origin. The doubling of the Ewald spheres is a consequence of the fact that both an upgoing and a downgoing standing wave are simultaneously present in the waveguide. Figure 5(b) is perhaps the closest analogy of the waveguide picture to that in an infinite medium.

The final picture is obtained by letting the index ℓ take all its permissible values. Figure 5(b) then results in Fig. 6(a). As ℓ varies, the upper center point A in Fig. 6(a) describes an arc of radius k_0 through the point C terminating at the lower center B, and vice versa. At the same time, the upper Ewald sphere S_U rides downward on the arc BCA, terminating at the lower Ewald sphere S_L, and similarly, the lower Ewald sphere S_L rides upward on the arc ABC terminating at the upper Ewald sphere S_U. We then have what may be called a *Ewald strip* in a waveguide instead of a single Ewald sphere in an infinite ambience. It should be pointed out that it is not a strip in the usual sense of the word, that is, a region of space between two planes. Cut centrally, its section would resemble a cardioid instead of a rectangle. The Ewald strip is thus a cardioid-like strip or a swath of a "limiting Ewald ball." Also note that the radius of the ball in which the strip is embedded is twice that of a single Ewald sphere; this is the so-called "limiting Ewald sphere," well known in the field of x-ray crystallography. Figure 6(b) shows this strip schematically as it should appear when viewed along the y- and the z axes, respectively. Moreover, just as the Ewald sphere for an infinite medium is a single sphere centered at the tip of the vector $k_0\hat{\mathbf{k}}$, so is the Ewald strip for a waveguide centered around the arc ABC in Fig. 6(a).

Let us note that the Ewald strip is not filled up in a continuous manner. In other words, there exists no subset of the strip that is dense in it. Since indices ℓ and m vary discretely, an individual sphere as well as the strip itself are filled only discretely. More precisely, a Ewald strip in a waveguide consists of exactly L individual Ewald spheres, corresponding to L values of the index ℓ, and each of these L spheres, in its turn, consists of Fourier vectors which lie on M number of circles on the sphere, M being the maximum value of the index m. This discrete nature of filling of an individual Ewald sphere is shown in Fig. 7. This is in sharp contrast to the infinite medium case, in which the sphere is swept by the scattered vector in a continuous fashion. The Ewald strip in a waveguide would thus consist of LM number of circles which will contain the information about the

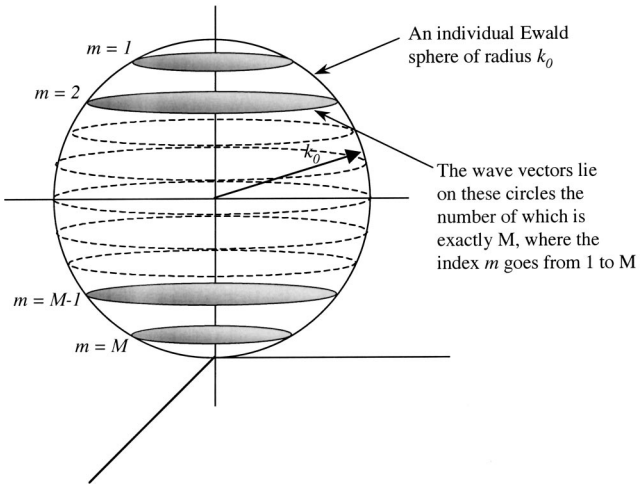


FIG. 7. The discrete nature of the scattering data on an individual Ewald sphere of an Ewald strip. The figure demonstrates that the data on an individual Ewald strip are distributed discretely on M circles corresponding M modes of the index m.

scatterer in the Fourier space. The crucial point is that unlike an infinite medium, one obtains information of the Fourier transform of a Born inhomogeneity inside the Ewald strip if the scatterer is placed in a waveguide.

The effect can be appreciated further if only backscattering is considered. The corresponding picture is essentially a section through the Ewald strip of Fig. 6(a), in which $\hat{\mathbf{r}}_\perp$ is taken to be $-\hat{\mathbf{r}}_{0\perp}$. Now, for a fixed incident wave, the backscattered Born data in an infinite medium consist only of a single point on the Ewald sphere. Comparatively, the backscattered far-field Born data in a waveguide are significantly richer, containing regions of the Ewald strip.

In view of the above discussions, the far-field limit of Eq. (14) in a waveguide under the Born approximation can be written as

$$\phi_\infty^B(\mathbf{r}) = iC\sum_{j=1}^{4}\sum_{\ell}\sum_{m}\mathcal{F}_{\ell m}(\boldsymbol{\xi})\psi_{\ell m}^j(d)\tilde{R}^j(\boldsymbol{\zeta}_{\ell m}), \tag{16}$$

where $\tilde{R}$ represents the Fourier transform of $R(\mathbf{r}')$, and $\boldsymbol{\zeta}_{\ell m}$ are the Fourier frequencies which lie in the Ewald strip in the manner described in previous paragraphs. The quantity $\phi_{\ell m}^j$ consists of $e^{\pm i\gamma_{\ell m}^{\pm}d}$ terms in Eq. (15). This means that each ℓm Fourier component, $\tilde{R}^j(\boldsymbol{\zeta}_{\ell m})$, of the Born object $R(\mathbf{r}')$, is weighted by its corresponding ℓm phase factor $\psi_{\ell m}^j$. Equation (16) is the waveguide analog of Eq. (8) for an infinite medium.

Let us consider an azimuthally symmetric Born scatterer in free space (not in a waveguide) which is axially symmetric. That is, if a spherical coordinate system is considered and the axis of symmetry is taken to be the $\hat{\mathbf{z}}$ axis, then the inhomogeneity is spherically symmetric and is distributed independently of the azimuthal angle φ. The Born scattering amplitude in this case is given by

$$\phi_\infty^B(\hat{\mathbf{r}},\hat{\mathbf{k}}) = 2\pi\int_0^\infty r^2\,dr\int_0^\pi \sin\theta\,d\theta\,R(r,\theta) \times J_0(k_0 r\sin\Theta\sin\theta)e^{-ik_0 r\cos\Theta\cos\theta}, \tag{17}$$

where Θ is the polar angle of the vector $\hat{\mathbf{r}}-\hat{\mathbf{k}}$. Equation (17) gives the Fourier transform $\tilde{R}(\boldsymbol{\zeta})=\tilde{R}(k_0,\Theta)$, and shows again that the Fourier information is on the Ewald sphere of radius k_0. In a waveguide under axial symmetry, on the other hand, we have

$$\begin{aligned}\int_\Omega d^3r'R(\mathbf{r}')e^{i\hat{\mathbf{r}}'\cdot[(\gamma_\ell\hat{\mathbf{z}}-k_\ell\hat{\mathbf{r}}_{0\perp})+(\gamma_m\hat{\mathbf{z}}-k_m\hat{\mathbf{r}}_\perp)]}&\\ =\int_0^{2\pi}d\theta'\int_{-\infty}^{\infty}dz'\int_0^\infty r'_\perp\,dr'_\perp\,R(\mathbf{r}'_\perp,z')&\\ \times e^{i(\gamma_\ell+\gamma_m)z'}e^{-i\mathbf{r}'_\perp\cdot(k_m\hat{\mathbf{r}}_\perp-k_\ell\hat{\mathbf{r}}_{0\perp})}&\\ =\int_{-\infty}^{\infty}dz'\int_0^\infty r'_\perp\,dr'_\perp\,R(r'_\perp,z')J_0(r'_\perp\sqrt{k_m^2+k_\ell^2})&\\ \times e^{-i(\gamma_\ell+\gamma_m)z'}&\\ =\tilde{R}(\sqrt{k_m^2+k_\ell^2},\gamma_\ell+\gamma_m).&\end{aligned} \tag{18}$$

The expression (18) was written for $\hat{\mathbf{r}}_\perp\perp\hat{\mathbf{r}}_{0\perp}$. Generally, $\sqrt{k_m^2+k_\ell^2}$ is meant to be replaced by $|k_m\hat{\mathbf{r}}_\perp-k_\ell\hat{\mathbf{r}}_{0\perp}|$. Using Eq. (18) in Eqs. (15) and (16), the scattered field in the far-field limit under cylindrical symmetry in the Born approximation reduces to

$$\begin{aligned}\phi_\infty^B(\hat{\mathbf{r}},\hat{\mathbf{k}})=\sum_\ell\sum_m\mathcal{F}_{\ell m}(\boldsymbol{\xi})\cdot\{&e^{i\gamma_{\ell m}^+d}\tilde{R}(\sqrt{k_m^2+k_\ell^2},\gamma_\ell+\gamma_m)\\ &+e^{-i\gamma_{\ell m}^+d}\tilde{R}(\sqrt{k_m^2+k_\ell^2},-(\gamma_\ell+\gamma_m))\}\\ \times\{&e^{i\gamma_{\ell m}^-d}\tilde{R}(\sqrt{k_m^2+k_\ell^2},\gamma_\ell-\gamma_m)\\ &+e^{-i\gamma_{\ell m}^-d}\tilde{R}(\sqrt{k_m^2+k_\ell^2},-(\gamma_\ell-\gamma_m))\}.\end{aligned} \tag{19}$$

A comparison between Eqs. (17) and (18) shows that the fixed wave number k_0 appearing in Eq. (17) for an infinite medium is replaced by the variable quantity $\sqrt{k_m^2+k_\ell}$ (or more generally $|k_m\hat{\mathbf{r}}_\perp-k_\ell\hat{\mathbf{r}}_{0\perp}|$) in Eq. (18) in a waveguide. This implies that the Born data in the case of a waveguide contain spatial frequency information of the scatterer not only inside a Ewald sphere, but in a Ewald strip for a waveguide. Considerations of only the normal modes ($\ell=m$) in Eq. (18) show that the Born data contain wave vectors both in and out of the $\boldsymbol{\zeta}_\perp$ plane in ζ space.

III. DISCUSSION

In the above, it was assumed that the height H of the waveguide was such that a resonance was allowed to occur in which case k_ℓ was zero. The Ewald strip in Fig. 6(a) corresponded to this case. If H does not admit a resonance, no generality is lost save the fact that the strip will be restricted to lie between an upward and a downward vector, the tip angle (from the vertical) of each corresponding to the maximum value of the index ℓ. In general, the incident wave vector, $\mathbf{v}_{0\perp}$, would have a small angle θ given by

$$\theta=\tan^{-1}\left[\frac{\sqrt{k_0^2-(L+\frac{1}{2})^2\pi^2/H^2}}{(L+\frac{1}{2})\pi/H}\right],$$

with respect to the z axis. This means the the arc ACB in Fig. 6(a) would subtend an angle of $(\pi-2\theta)$ at the origin instead of π when resonance is allowed.

Figure 6(a) shows the Ewald strip for a single source which is positioned at $\mathbf{r}_0$. The strip lies in one hemisphere of a ball B of radius $2k_0$, the so-called ''limiting Ewald sphere'' mentioned earlier. By placing the source at the coordinates $(-\mathbf{r}_{0\perp}, z_0)$, the strip can be made to fall in the other hemisphere. It is also evident that in contrast to scattering in an infinite medium, the ball B can be filled (discretely) with a relatively handful of source locations. Let us note that for a single source, but for all observation points, the reconstruction of the inhomogeneity from the Born data would be nonunique both in an unbounded medium as well as in a waveguide. There is, however, a difference between the two cases. In the former, the nonuniqueness is a consequence of the data being obtained on a spherical surface which is of measure zero in $\mathbb{R}^3$, as was discussed in the Introduction. In the case of a waveguide, however, the nonuniqueness is due to ''incomplete data.'' In order to further appreciate this point, consider again a spherically or cylindrically symmetric scatterer. If the scattering takes place in an unbounded medium, the data are on the surface (in the Fourier space) of measure zero regardless of the symmetry. The situation alters significantly if the scatterer is in a waveguide. In the latter case, the spectral information is obtained inside the Ewald strip residing in one of the hemispheres of the ball B. However, since the object is angularly symmetrical, the data can be obtained in the other hemisphere simply by consideration of symmetry. The nonuniqueness that results in this case then cannot be due to a Lebesgue space consisting of functions which are defined to within some equivalence class, but rather owing to the data not being complete. In other words, the nonuniqueness derives from the ill-posedness of the problem due to an incomplete data set. Suppose that the number of modes is large and an interpolation scheme can, in principle, be designed in order to obtain the missing data (as, for example, in the fashion of tomographic reconstruction from incomplete projection data). Then, a reconstruction can at least be attempted from the waveguide data. This is, however, not the case if the medium is unbounded. Next, consider that the object does not have an angular symmetry, but two sources are incident from opposite directions. In an unbounded medium, the nonuniqueness due to the data lying on the surface of measure zero still holds in this case. In contrast, the Ewald strip in a waveguide now falls into both hemispheres of the ball B, but would still give rise to nonuniqueness resulting from ill-posedness. As was mentioned in the Introduction, to outline the procedures for reconstruction of the Born scatterer from the waveguide data is not the objective of this paper, which is concerned with the information content of the scattering amplitude in a waveguide in comparison to that in an unbounded medium. The discussion of reconstruction is postponed for later.

The relatively richer information content of the scattering amplitude in a waveguide is, of course, owing to the fact that waves are trapped in the guide. In order to see this more clearly, consider the scattering integral of Eq. (14). The ℓm component of the integral takes the form

$$I^{\infty}_{\ell m}(\mathbf{r}) = \int_{-d}^{H-d} dz'\, \phi_{\infty}(\mathbf{r};\hat{\mathbf{r}}_0, k_m, k_\ell; z') e^{i\Phi_{\ell m} z'}, \tag{20}$$

where $\Phi_{\ell m}$ are the various phases described in Eq. (15), and the superscript ∞ in $I^{\infty}_{\ell m}$ indicates that far-field limit is being considered. Now by applying *Rellich's lemma*, and noting that the scattering solutions of the Helmholtz equation are analytic functions in all their variables, it follows from Eq. (20) that[3]

$$I_{\ell m}(\mathbf{r}) = \int_{-d}^{H-d} dz'\, \phi^{s}(\mathbf{r};\hat{\mathbf{r}}_0, k_m, k_\ell; z') e^{i\Phi_{\ell m} z'}. \tag{21}$$

The scattered field $\phi^s(\mathbf{r};\hat{\mathbf{r}}_0, k_m, k_\ell; z')$ in Eq. (21) or $\phi^{\infty}(\mathbf{r};\hat{\mathbf{r}}_0, k_m, k_\ell; z')$ in Eq. (20) are the scattered field and its far-field limit, respectively, that result when a cylindrical wave is incident upon the scatterer with wavefronts propagating along the planes $z'=$ constant. In these equations, the integration of these fields over z', after being multiplied by the one-dimensional plane wave $e^{i\Phi_{\ell m} z'}$, demonstrates that the scattered wave functions are trapped. In other words, the ''aperture'' of the scattering configuration is wider in a waveguide than in an unbounded medium. This results in the information content in a scattering amplitude data set being richer in a waveguide relative to that in an unbounded medium.

Analogous observations are known to exist in time-reversal focusing in a waveguide, and in scattering in an unbounded environment containing randomly distributed inhomogeneities. In the former, the presence of boundaries increases the apparent aperture of a time-reversal mirror (TRM) (see, for example, the review article by Fink and Prada, Ref. 8). The multipath reverberations give rise to virtual images so that to an observer located at the source, the aperture of the TRM appears to be increased. Similar considerations also apply to the analysis of the information content of the scattered data from an object embedded in random media. Here, multiple scattering from the randomly distributed inhomogeneities results in waves arriving at the detector which would otherwise be lost if the environment was homogeneous and unbounded.[9,10]

It was assumed that the mode frequencies γ_ℓ and k_ℓ were positive numbers with L as the upper limit of ℓ. In other words, evanescent waves were considered to be absent in the far-field data. This can be justified as follows. In the frequency domain in which we are working here, the total flux over all frequencies leaving through the boundary $\partial\Omega$ of the scatterer can be written as

$$\begin{aligned}\frac{\partial W}{\partial t}(\phi) &= \int_{-\infty}^{\infty} d\omega \int_{\partial\Omega} d\Gamma(-i\omega\rho_0)\phi\frac{\partial \phi^*}{\partial n} \\ &= -(ic_0\rho_0)\int_{-\infty}^{\infty} dk \int_{\partial\Omega} d\Gamma\, k\phi\frac{\partial \phi}{\partial n}.\end{aligned} \tag{22}$$

Equation (22) can be obtained by first calculating the time-integrated energy flux using the time-dependent propagation equation, namely

$$\left[\Delta - \frac{1}{c^2(\mathbf{r})}\frac{\partial^2}{\partial t^2}\right]\phi(\mathbf{r},t) = 0, \quad \mathbf{r}\in\Omega^0,$$

and then applying Parseval's relation to write it in terms of the Fourier domain wave function. Now, the field ϕ in the interior of the scatterer Ω in the Lippmann–Schwinger equation is simply the incident field if the Born approximation is invoked. According to Eq. (22), if no evanescent wave energy is present in the incident field, it cannot be present in the far-field either. Since the point source at $\mathbf{r}_0$ is in the far field, no energy flux due to evanescent waves can be expected to be present in the scattered wave in the far field.

ACKNOWLEDGMENTS

This work supported by the Naval Research Laboratory and the Office of Naval Research.

[1] M. Cheney and J. Rose, "Three-dimensional inverse scattering for the classical wave equation with variable speed," in Review of progress in quantitative nondestructive evaluation edited by D. O. Thomson and D. E. Chimenti (Plenum Press, New York, 1986), Vol. 5, pp. 1–5.
[2] L. S. Couchman, D. N. G. Roy, and J. Warner, "Inverse Neumann obstacle problem," J. Acoust. Soc. Am. **104**, 2615–2621 (1998).
[3] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory* (Springer, Berlin, 1992).
[4] R. H. T. Bates, V. A. Smith, and R. D. Murch, "Managable multidimensional inverse scattering theory," Phys. Rep. **201**(4), 185–277 (1991).
[5] K. J. Langenberg, "Applied inverse problems for acoustic, electromagnetic and elastic wave scattering," in *Basic Methods of Tomography and Inverse Problems*, edited by P. C. Sabatier (Hilger, Bristol, 1987), pp. 127–470.
[6] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems* (Springer, Berlin, 1996).
[7] F. Ingenito, "Scattering from an object in a stratified medium," J. Acoust. Soc. Am. **82**, 2051–2059 (1987).
[8] M. Fink and C. Prada, "Acoustic time-reversal mirrors," Inverse Probl. **17**, R1–R38 (2001).
[9] D. N. G. Roy and G. Yoon, "Double-passage recovery of an object embedded in a random medium," Opt. Lett. **17**, 553–555 (1992).
[10] T. Mavroidis and J. C. Dainty, "Imaging after double passage through a random screen," Opt. Lett. **15**, 857–859 (1990).

Influence of planar cracks in plates on reflected and transmitted fields of Gaussian acoustic beams

J. Vandeputte, O. Leroy, and G. Shkerdin[a)]
Interdisciplinary Research Center, K. U. Leuven Campus Kortrijk, B-8500 Kortrijk, Belgium

(Received 20 July 2001; revised 10 April 2003; accepted 23 April 2003)

The scattering of a Gaussian beam by a rectangular crack is studied by means of the radiation mode model. The orthogonal set of modes for a L/S/V structure is derived, and a general mode decomposition formula is developed. The influence of a crack on the reflected and transmitted fields of a bounded beam, incident at critical angles of the L/S/L- and the L/S/V structure, is studied and physical interpretations are given. From those interpretations, the starting point of the crack is deduced along with the length and distance to the surface. Effects on focused beams are considered as well. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1587148]

PACS numbers: 43.20.Gp [DEC]

I. INTRODUCTION

Many attempts were made in the past to study the scattering of acoustic waves by all kinds of discontinuities and imperfections. A particular interest can be addressed to a crack because of its possible danger to grow and cause failure of the material. In practice, ultrasonic methods for nondestructive testing (NDT) of material are playing a role of increasing importance. The aim is, in the first place, to detect a flaw in the material. But, afterwards, a good understanding and interpretation of numerical results can lead to an accurate characterization of the defect, such as its location, orientation, size, shape, COD, etc. Many shapes of cracks were tackled in the last half century. A rectangular crack was studied by Guan and Norris.[1] This study was motivated by its possible existence in composite materials and also because of corner effects. Corners are cinematically more stiff than other parts of the void. They showed that it is impossible to distinguish the shape of a crack based on low-frequency scattered data. An idealized case of a rectangular crack is a planar void with zero thickness, as was studied by Visscher.[2] Of course, in reality the crack is opened with an extent greater than the induced amplitude of vibration of the opposing crack faces. Note that when crack faces touch each other while insonifying, nonlinear effects are involved.[2,3] Others investigated the scattering by penny-shaped cracks, spherical cavities,[3–5] elliptic, fluid-filled, partly closed cracks, etc.

Most of the mathematical models which formalize the crack problem are based on integral representations. These express the displacement corresponding to the scattered field in terms of the crack-opening displacement (COD).[1,2,3,6–8] Auld's formula, which gives the change of Γ, i.e., $\delta\Gamma$, due to the scattering by an imperfection was used in Refs. 4 and 7. Γ denotes the ratio of the received signal strength over the incident signal strength. Sharp peaks of $|\delta\Gamma|$ in $\delta\Gamma$ graphs correspond to resonance effects of the thin layer between a horizontal crack face and the free surface.[7] Usually, normally incident plane waves are considered to handle the crack problem. Surface waves are chosen for inspection of cracks near the surface (surface-breaking cracks or subsurface cracks). Obliquely incident plane waves were considered in Refs. 1, 4, and 7. The varying angle of incidence is an extra parameter in order to characterize the crack. In the present paper, the scattering of an obliquely incident, bounded Gaussian beam by a planar crack is handled. This is the most general case in the sense that we deal with a Gaussian beam instead of a plane wave and that we consider oblique incidence instead of normal incidence. The mathematical model is the radiation mode model.[9,10]

In Sec. II, the orthogonal set of radiation modes for a liquid/solid/vacuum structure (L/S/V) is derived. Referring to the radiation modes of a liquid/solid/liquid structure (L/S/L) is done in Sec. III. In the fourth section, the bounded beam scattering by a rectangular crack is treated. The plate with the void is divided into three substructures in order to implement the above-mentioned structures into our model. Overlap integrals are defined and a final analytic formula is derived. Section V is devoted to numerical calculations. The influence of several parameters, such as depth of the crack, length of the crack, beginning of the crack, etc. on reflected and transmitted fields is shown and physically explained. Also, focused Gaussian beams are considered. Finally, conclusions are drawn in Sec. VI.

II. THE ORTHOGONAL SET OF ACOUSTIC MODES FOR A LIQUID/SOLID/VACUUM STRUCTURE (L/S/V)

A. The continuous spectrum of radiation modes

Consider an isotropic plate of thickness d, bordered at the upper side by a liquid and at the lower side by vacuum (see Fig. 1). We suppose a located-at-infinity ($y=+\infty$) acoustic source from which emerges plane waves, as was done previously.[9–11] These plane waves reflect upon and transmit through the plate. In total, six different potential functions are involved

$$\Phi_I = A\exp[i(-k_y y + k_z z - \omega t)],$$

$$\Phi_R = AR\exp[i(k_y y + k_z z - \omega t)],$$

[a)]Permanent address: Institute of Radio Engineering and Electronics, Russian Academy of Sciences, Moscow, Russia.

$$\Phi_l^{\pm} = AW_l^{\pm} \exp[i(\pm k_{ly}y + k_z z - \omega t)],$$

$$\boldsymbol{\Psi}_t^{\pm} = AW_t^{\pm} \exp[i(\pm k_{ty}y + k_z z - \omega t)]\mathbf{e}_x, \tag{1}$$

with the wave numbers in the liquid and solid

$$k_y^2 + k_z^2 = k^2, \quad k_{ly}^2 + k_z^2 = k_l^2, \quad k_{ty}^2 + k_z^2 = k_l^2. \tag{2}$$

A denotes the amplitude of the incident wave, ω the angular frequency, R, $W_l^{\pm}$, and $W_t^{\pm}$ the longitudinal and transversal reflection and transmission coefficients, respectively, as shown in Fig. 1. Requirements of continuity of the normal displacements and normal and tangential stress components at $y=d$, and of the stresses to be zero at $y=0$, leads to the following reflection and transmission coefficients (for a more detailed derivation we refer the reader to Ref. 12):

$$R = \frac{2\left(\frac{\rho_1}{\rho}\right)\left(1-2\frac{k_z^2}{k_t^2}\right)\left(2-2\cos(k_{ty}d)\cos(k_{ly}d)+\left(\frac{b}{a}+\frac{a}{b}\right)\sin(k_{ty}d)\sin(k_{ty}d)\right)}{D}$$
$$-\frac{2i\frac{k_{ly}}{k_y}\left(\frac{1}{1-(2k_z^2/k_t^2)}\right)\left(\cos(k_{ty}d)\sin(k_{ly}d)+\left(\frac{b}{a}\right)\cos(k_{ly}d)\sin(k_{ty}d)\right)}{D}, \tag{3}$$

$$W_l^{\pm} = \frac{2\left(1-e^{\pm k_{ly}d}\left(\cos(k_{ty}d)\pm i\left(\frac{b}{a}\right)\sin(k_{ty}d)\right)\right)}{D} \tag{4}$$

$$W_t^{\pm} = \pm\frac{4b\left(1-e^{\pm k_{ty}d}\left(\cos(k_{ly}d)\pm i\left(\frac{b}{a}\right)\sin(k_{ly}d)\right)\right)}{D}, \tag{5}$$

with

$$D = 2\left(\frac{\rho_1}{\rho}\right)\left(1-2\frac{k_z^2}{k_l^2}\right)\left(2-2\cos(k_{ty}d)\cos(k_{ly}d)\right.$$
$$\left.+\left(\frac{b}{a}+\frac{a}{b}\right)\sin(k_{ty}d)\sin(k_{ly}d)\right)+2i\frac{k_{ly}}{k_y}\left(\frac{1}{1-\frac{2k_z^2}{k_t^2}}\right)$$
$$\times\left(\cos(k_{ty}d)\sin(k_{ly}d)+\left(\frac{b}{a}\right)\cos(k_{ly}d)\sin(k_{ry}d)\right), \tag{6}$$

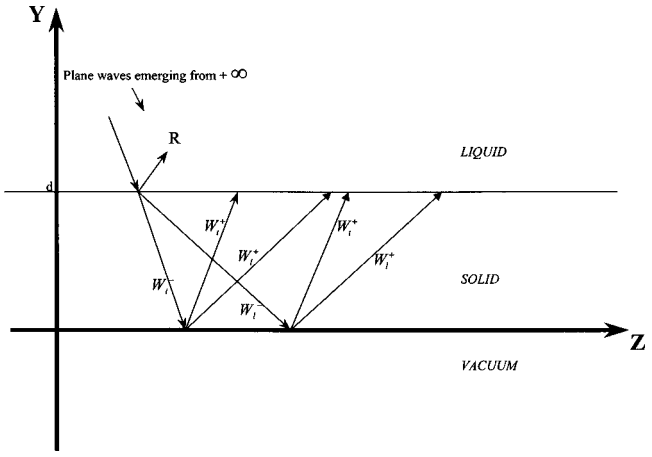


FIG. 1. Liquid/solid/vacuum structure.

where $a = k_z k_{ly}/(k_t^2 - 2k_z^2)$, $b = (k_t^2 - 2k_z^2)/4k_z k_{1y}$, (7)

ρ and ρ_1, the densities of the liquid and the solid, respectively.

As a result, the displacement components can be derived by means of

$$\mathbf{u} = \text{grad}(\Phi_I + \Phi_R), \quad \text{(in the liquid)} \tag{8}$$

$$\mathbf{u} = \text{grad}(\Phi_l^+ + \Phi_l^-) + \nabla\times(\boldsymbol{\Psi}_t^+ + \boldsymbol{\Psi}_t^-), \quad \text{(in the solid)} \tag{9}$$

and leads to

for $y>d$: (components in the liquid)

$$u_z = Bk_z\cos(k_y(y-d)-\varphi)e^{ik_z z - i\omega t},$$
$$u_y = Bik_y\sin(k_y(y-d)-\varphi)e^{ik_z z - i\omega t}. \tag{10}$$

for $0<y<d$: (components in the solid)

$$u_z = \frac{2B}{|D|}\left[\begin{array}{c}(\cos(k_{ly}d)-\cos(k_{ty}d))(k_z\cos(k_{ly}y)+2bk_{ty}\cos(k_{ty}y))\\ +\left(\sin(k_{ly}d)+\frac{b}{a}\sin(k_{ty}d)\right)(k_z\sin(k_{ly}y)-2ak_{ty}\sin(k_{ty}y))\end{array}\right]e^{ik_z z-i\omega t}$$

$$u_y = \frac{2iB}{|D|}\left[\begin{array}{c}(\cos(k_{ly}d)-\cos(k_{ty}d))(k_{ly}\sin(k_{ly}y)-2bk_z\sin(k_{ty}y))\\ -\left(\sin(k_{ly}d)+\frac{b}{a}\sin(k_{ty}d)\right)(k_{ly}\cos(k_{ly}y)+2ak_z\cos(k_{ty}y))\end{array}\right]e^{ik_z z-i\omega t}, \tag{11}$$

with $B=1/k^2$

$$\sin(\varphi)=\frac{2k_{ly}\left(\cos(k_{ty}d)\sin(k_{ly}d)+\frac{b}{a}\cos(k_{ly}d)\sin(k_{ty}d)\right)}{k_{ly}\left(1-\frac{2k_z^2}{k_t^2}\right)|D|}, \tag{12}$$

$$\cos(\varphi)=\frac{2\frac{\rho_1}{\rho}k_{ly}\left(1-\frac{2k_z^2}{k_t^2}\right)\left(2-\cos(k_{ty}d)\cos(k_{ly}d)+\left(\frac{b}{a}+\frac{a}{b}\right)\sin(k_{ty}d)\sin(k_{ty}d)\right)}{|D|}. \tag{13}$$

B. The discrete spectrum of eigenmodes

The considered structure (see Fig. 1) supports only one Stoneley mode. It is easy to derive the displacement components.

For $y>d$: (components in the liquid)

$$u_z = B_S k_{sz} e^{-k_{sy}(y-d)+ik_{sz}z},$$
$$u_y = iB_S k_{sy} e^{-k_{sy}(y-d)+ik_{sz}z}. \tag{14}$$

For $y<d$: (components in the solid)

$$u_z = -B_S k_{sy}\left[\frac{k_{sz}}{k_{sly}}\left(1-\frac{2k_{sz}^2}{k_t^2}\right)e^{k_{sly}(y-d)} + \frac{2k_{sz}k_{sty}}{k_t^2}e^{k_{sty}(y-d)}\right]e^{ik_{sz}z},$$
$$u_y = iB_s k_{sy}\left[\left(1-\frac{2k_{sz}^2}{k_t^2}\right)e^{k_{sly}(y-d)} + \frac{2k_{sz}^2}{k_t^2}e^{k_{sty}(y-d)}\right]e^{ik_{sz}z}, \tag{15}$$
$$B_S = 1/k^2.$$

"k_{sz}" is a solution of the characteristic dispersion equation for a L/S/V structure

$$4\frac{\rho_1}{\rho}k_y k_z^2\left[\left(k_z-\frac{k_t^2}{2k_z}\right)^2-k_{ly}k_{ty}\right]+k_{ly}k_t^4=0, \tag{16}$$

and

$$k_{sy}^2+k_{sz}^2=k^2,\quad k_{sly}^2+k_{sz}^2=k_l^2,\quad k_{sty}^2+k_{sz}^2=k_t^2. \tag{17}$$

So, the complete set of modes for the L/S/V structure consists of the continuous spectrum of radiation modes and one Stoneley mode.

III. THE ORTHOGONAL SET OF ACOUSTIC MODES FOR A LIQUID/SOLID/LIQUID STRUCTURE (L/S/L)

The formulas for the radiation modes and the eigenmodes for this structure were handled extensively in the past,[10,11] and will not be repeated here. The interested reader is referred to these earlier publications.[9–11]

IV. BOUNDED BEAM SCATTERING FROM A RECTANGULAR CRACK

When an incident Gaussian beam reaches a plate with a rectangular crack, the acoustic field will be strongly disturbed. Reflected and transmitted fields will give us a lot of information on the shape and localization of the crack. Consider an incident Gaussian beam as is sketched in Fig. 2. The plate has a thickness d, the crack has a thickness $c-g$, and a length L_1. We suppose the crack to be parallel to the plate side.

In order to handle this problem with the mode model, we divide the plate into three substructures: structure 1a, an L/S/L structure ($z\leqslant 0$); structure 1b, also an L/S/L structure

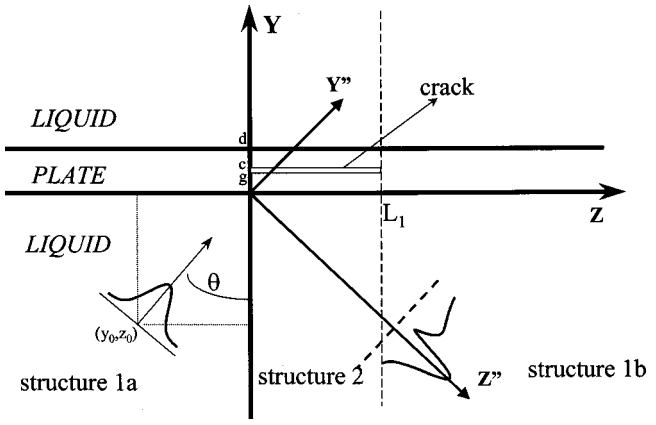


FIG. 2. Crack problem: division in substructures.

($z \geqslant L_1$); and structure 2, an L/S/V/S/L structure ($0 \leqslant z \leqslant L_1$) (see Fig. 2). In each substructure, the acoustic field can be represented by the orthogonal set of modes of the considered structure. Substructure 2 can be considered as the combination of two L/S/V structures. The radiation modes of such a structure were already derived in Sec. II. Formulas related to the upper L/S/V region ($y > c$) will be indicated by a "+" superscript, those of the lower part ($y < g$) by a "−" superscript. So, in region 1a the displacement field is given[10,11] as

$$\mathbf{u}_i(y,z) = \sum_{n=1,2} \int_0^k C_i^{(n)}(k_{y,i}) \mathbf{u}_1^{(n)}(k_{y,i};y,z)\, dk_{y,i}. \tag{18}$$

Explicit formulas for $u_1^{(n)}$ can be found in Ref. 10. We neglect reflection from $[g,c]$ at $z=0$ by requiring that the crack is sufficiently thin (such that the side-wall reflection is negligible in comparison with the total reflected field by the crack). In structure 2 we have

$$\mathbf{u}_2(y,z) = \int_0^k C_2^{\pm}(k_{y,2}) \mathbf{u}_2^{\pm}(k_{y,2};y,z)\, dk_{y,2}. \tag{19}$$

And, finally, in structure 1b

$$\mathbf{u}_1(y,z) = \sum_{\rho=1,2} \int_0^k C_1^{(p)}(k_{y,1}) \mathbf{u}_1^{(p)}(k_{y,1};y,z)\, dk_{y,1}. \tag{20}$$

Note that the integration is only taken from 0 up to k, so the local modes[9] are negligible (we only study reflected fields at a certain distance away from the plate) and are not taken into account (local modes correspond to $k_{y,1} > k$). Also, the Stoneley modes can be omitted; they have expansion coefficients almost equal to zero [the Stoneley modes are a second part of the decomposition of $\mathbf{u}_1$; see Eq. (6) in Ref. 11]. Analogous formulas can be written for the stress tensor components.

The determination of the expansion coefficient $C_i^{(n)}$ was shown in Ref. 10. So, we still have to calculate the unknown coefficients C_2^+, C_2^-, $C_1^{(1)}$, and $C_1^{(2)}$. We briefly repeat the method to calculate these coefficients as was described extensively previously (Ref. 13). There, we introduced the overlap integrals in order to simplify the long-winded expressions. Imposing boundary conditions at $z=0$ and $z=L_1$ (continuity of the normal and tangential displacement and stress components) enables us to calculate these coefficients. At $z=0$, we define next overlap integrals

$$I_{12}^{n+}(k_{y,1};k_{y,2}) = \int_c^{+\infty} [T_{yz}^{(n)}(k_{y,1};y,0) u_{2y}^{(+)*}(k_{y,2};y,0) - u_{1z}^{(n)}(k_{y,1};y,0) T_{zz}^{(+)*}(k_{y,2};y,0)]\, dy \quad (n=1,2), \tag{21}$$

$$K_{12}^{n+}(k_{y,1};k_{y,2}) = \int_c^{+\infty} [T_{zz}^{(n)}(k_{y,1};y,0) u_{2z}^{(+)*}(k_{y,2};y,0) - u_{1y}^{(n)}(k_{y,1};y,0) T_{yz}^{(+)*}(k_{y,2};y,0)]\, dy \quad (n=1,2). \tag{22}$$

"$k_{y,1}$" stands for the y component of the wave vector in the liquid in structure 1a, $k_{y,2}$ in the liquid of structure 2. "T" refers to the stress tensor components, "*" is the complex conjugate. These integrals describe the coupling between a radiation mode of type "n" in structure 1a, with a radiation mode of type "+" in structure 2.[13] Analogous integrals can be defined in order to describe the interaction with the "−" modes of structure 2: I_{12}^{n-} and K_{12}^{n-} ($n=1,2$). The integration interval is then $[-\infty,g]$ instead of [c, $+\infty$]. Analogously, at $z=L_1$ we define the overlap integrals $I_{21}^{\pm p}$ and $K_{21}^{\pm p}$ which describe the coupling between a radiation mode of type "±" of structure 2 with a radiation mode of type "p" of structure 1b.[13] It is easy to show that

$$I_{21}^{\pm p} = -(K_{12}^{p\pm})^*, \quad K_{21}^{\pm p} = -(I_{12}^{p\pm})^*. \tag{23}$$

For each overlap integral, we split the integration interval into two parts: one in the solid, and one in the liquid. For instance, I_{12}^{1+} (with integration interval [c, $+\infty$]) splits into $I_{12,S}^{1+}$ (integrating over $[c,d]$) and $I_{12,L}^{1+}$ (integrating over $[d, +\infty]$). By means of these integrals, the boundary conditions and the mode orthogonality relations,[9–11,13] we achieve, after very hard and rigorous calculations, the next expressions for the expansion coefficients

$$C_2^{\pm}(k_{y,2}) = C_i^{(1)}(k_{y,2}) \cos[\varphi^{\pm}(k_{y,2}) - \varphi(k_{y,2})] \pm C_i^{(2)}(k_{y,2}) \cos[\varphi^{\pm}(k_{y,2}) - \psi(k_{y,2})] + \frac{1}{\pi k_{z,2}} \int_0^k [C_i^{(1)}(k_{y,1}) A_1^{\pm}(k_{y,1};k_{y,z}) + C_i^{(2)}(k_{y,1}) A_2^{\pm}(k_{y,1};k_{y,2})]\, dk_{y,1} \tag{24}$$

$$C_2^{(1,2)}(k_{y,1}) = \tfrac{1}{2}(C_2^{\pm}(k_{y,1}) \cos(\varphi^{+}(k_{y,1}) - M(k_{y,1})) \pm C_2^{-}(k_{y,1}) \cos(\varphi^{-}(k_{y,1}) - M(k_{y,1}))) + \frac{1}{2\pi k_{z,1}} \int_0^k [C_2^{+}(k_{y,2}) A_{1,2}^{+}(k_{y,1};k_{y,2}) + C_2^{-}(k_{y,2}) A_{1,2}^{-}(k_{y,1};k_{y,2})] e^{i(k_{z,2}-k_{z,1})L_1} dk_{y,2}, \tag{25}$$

$$M(k_{y,1}) = \varphi(k_{y,1}) \quad \text{for } C_2^{(1)}(k_{y,2})$$

with

$$M(k_{y,1}) = \psi(k_{y,1}) \quad \text{for } C_2^{(2)}(k_{y,2}).$$

φ and ψ were already defined in the past; see Ref. 10, Eqs. (25) and (32).

$$A_1^{\pm}(k_{y,1};k_{y,2}) = \frac{k_{z,1}+k_{z,2}}{2} \tilde{I}_{12,0}^{1\pm}(k_{y,1};k_{y,2}) + \frac{2\mu k_{y,1} \sin(\varphi(k_{y,1}))}{\lambda k^2 |D^{\pm}|} \times [\pm k_{z,1} \tilde{I}_{12,s}^{1\pm} + k_{z,2} \tilde{K}_{12,s}^{1\pm}], \tag{26}$$

$$A_2^{\pm}(k_{y,1};k_{y,2}) = \pm\frac{k_{z,1}+k_{z,2}}{2}\tilde{I}_{12,0}^{2\pm}(k_{y,1};k_{y,2}) + \frac{2\mu k_{y,1}\sin(\psi(k_{y,1}))}{\lambda k^2|D^{\pm}|} \times [k_{z,1}\tilde{I}_{12,s}^{2\pm} \mp k_{z,2}\tilde{K}_{12,s}^{1\pm}], \quad (27)$$

and $D^{\pm}$ equals D as given in formula (6), with "d" replaced by "d-c" for D^{+} and replaced by "g" for D^{-}. λ is the Lamé constant of the liquid, λ_1 and μ are the Lamé constants of the solid plate. $K_{z,1}$ and $k_{z,2}$ denote the z component of the wave vector in the liquid of, respectively, structure 1 and structure 2. Formulas for $\sin(\varphi^{\pm})$, $\cos(\varphi^{\pm})$, $\tilde{I}_{12,0}^{1\pm}$, and $\tilde{I}_{12,0}^{2\pm}$ are given in the Appendix. Detailed formulas for $\tilde{I}_{12,s}^{1,2\pm}$ and $\tilde{K}_{12,s}^{1,2\pm}$ are too long-winded and will not be given in this paper. If analytical expressions are wished, the author(s) can be contacted.

V. NUMERICAL RESULTS

In this section, numerical illustrations are given of the scattering of a bounded Gaussian beam by a rectangular crack. The results are analyzed and physically explained. A steel plate is considered, which is immersed in water. The acoustic velocity v in water is 1460 m/s; those in the plate: $v_1 = 5720$ m/s (longitudinal), $v_t = 3160$ m/s (transversal). The densities are, respectively, $\rho = 1000$ kg/m^3 (water), $\rho_1 = 7800$ kg/m^3 (steel plate). The frequency is 4 Mhz, the plate thickness $d = 3$ mm, and $y_0 = -10$ cm (y_0 denotes the y coordinate of the position of the transducer). The beam half-width L is 1 cm. These parameters and values of variables are valid through the whole section, unless otherwise mentioned.

A. Influence of a crack at a known depth on the reflected Lamb profile of the L/S/V structure

In this paragraph, we assume the depth of the crack to be known. This can be done by an independent pulse–echo experiment or by the surface wave method as described in Ref. 8 for subsurface cracks. The crack is also assumed to be sufficiently long, so it makes sense to talk about resonance modes of the L/S/V structure. We select a critical angle, 11.84°, of this structure which can be calculated by setting Eq. (6) equal to zero. This is not a critical angle of the adjacent L/S/L structure (in order to avoid confusion between Lamb profiles of the L/S/L structure and those of the L/S/V structure). We will study the influence of the z coordinate of the position of the incident beam on the reflected Lamb profile of the L/S/V structure. Therefore, we move the incident beam parallel to the plate (along the $y = y_0$ axis). As a result, the meeting point of the center of the beam with the z-axis varies. This meeting point will be denoted by z^* (see Fig. 3). Letting vary z^* from positive to negative values, which corresponds to a shift of the transducer, the bounded beam is partially incident on the L/S/L structure and partially on the L/S/V structure. As a consequence, the resonant character of the reflected field will partly have disappeared. Therefore, we study the effect of moving the transducer on the shifted component of the reflected Lamb profile. This is illustrated in Fig. 4. Figure 4(a) plots the reflected amplitude field at the phase plane $z'' = 15$ cm for different values of z^*. We remark that the shifted component is stretched out over a longer distance (in comparison with the reflected Lamb profile of a L/S/L structure). This is due to the reradiation which is completely oriented to the lower liquid. It's clear that the lower the value of z^*, the less the resonance condition is fulfilled, the smaller the nonspecular component is in amplitude. For

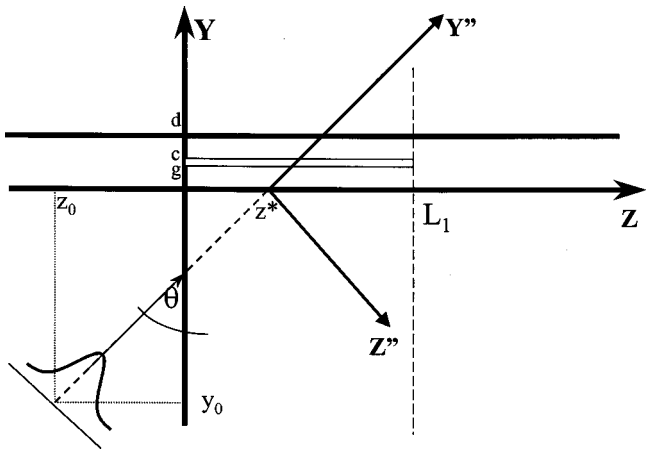


FIG. 3. Crack in plate: configuration, definition of z^*.

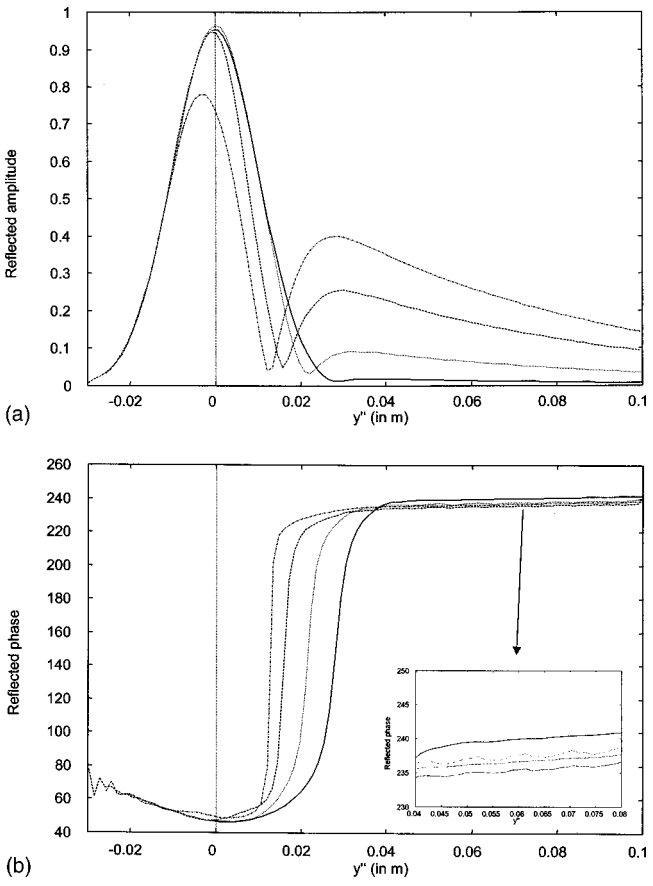


FIG. 4. (a) Reflected amplitude at $z'' = 15$ cm for different values of z^*. The dash-dot line represents $z^* = 1$ cm; the dashed line: $z^* = 0$; the dotted line: $z^* = -1$ cm; and the full line: $z^* = -3$ cm ($g = c = 0.6$ mm). (b) Reflected phase profile at $z'' = 15$ cm for different values of z^*. The dash-dot line represents $z^* = 1$ cm; the dashed line: $z^* = 0$ cm; the dotted line: $z^* = -1$ cm; and the full line: $z^* = -3$ cm ($g = c = 0.6$ mm).

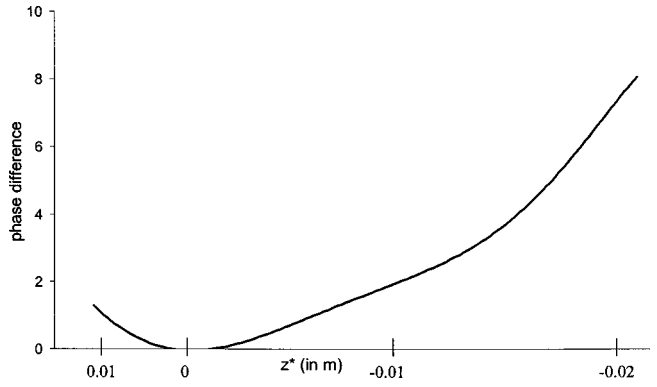


FIG. 5. Phase difference as a function of z^* in the reflected field, caused by the appearance of a crack.

$z^*=1$ cm, the whole Gaussian beam detects a L/S/V structure and reradiates much more than for $z^*=0$ or negative values, where for $z^*=-3$ cm the beam in incident on a L/S/L structure and thus outside the Lamb angle for that structure.

In Fig. 4(b), the corresponding phase curves are given. In the right-hand corner a detail is given for $y'' \in [0.04:0.08]$. The phase effect of moving the transducer (varying z^*) is more sensitive and will be used to estimate the starting position of the crack. In Fig. 5, the dependence of the phase effect, at the position where the maximum of the reflected amplitude field in the shifted component occurs, on z^* is investigated. With phase effect we mean the phase difference between the phase in case of $z^* \neq 0$ and the phase in case of $z^*=0$. Analyzing the curve in Fig. 5, we observe a minimum at $z^*=0$. As a result, considering phase effects by moving the transducer enables us to locate the starting point of the crack, namely where the phase reaches a minimum.

B. Influence of a crack on the reflected Lamb profile of the L/S/L structure

In this paragraph an alternative method is given in order to localize the starting point of the crack. Here, Lamb modes of the plate (L/S/L) will be used. So, in this case the depth of the crack (value of c and g) has not to be known *a priori*. The angle of incidence is chosen as 20.40°, which is the S2 Lamb mode. Again, we suppose that the transducer can be moved along the $y=y_0$ axis as was done in Sec. IV A. Far away from the crack, in the sense that $z^* \ll 0$, the reflected field will be the well-known reflected Lamb profile[11,14,15] (with its unshifted and shifted component, and its phase jump). The more we approach the crack (z^* comes closer to 0), the more the crack discontinuity will disturb the smooth Lamb profile. This will be an indication of an irregularity in the plate.

Consequently, reflected amplitude and phase curves are calculated for different values of z^*. This is done in Figs. 6(a) and (b), at the phase plane $z''=26$ cm. The crack was located at $c=g=0.89$ mm. We remark very clearly sudden falloffs in the reradiated lobe, which reveals the presence of the crack. The position after the last peak where the amplitude reaches 1/3 of the value of that last peak corresponds exactly to the projected z^*

$$\tilde{y}'' = -z^* \cos(\theta). \tag{28}$$

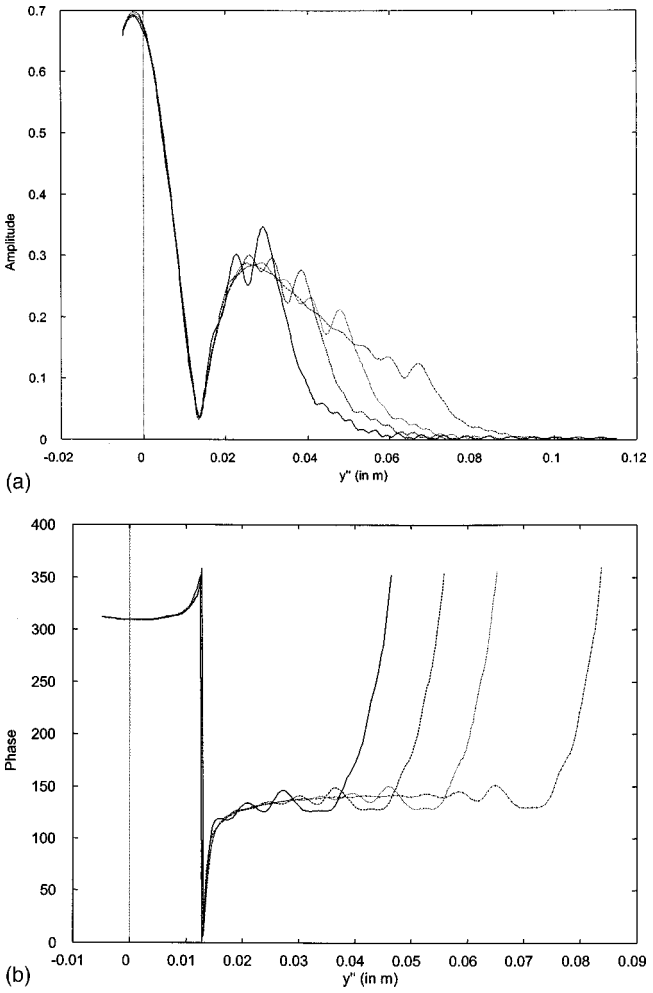


FIG. 6. (a) Reflected amplitude profile at $z''=26$ cm for different values of z^*. The full line represents $z^*=-4$ cm; the dashed line: $z^*=-5$ cm; the dotted line: $z^*=-6$ cm; and the dash-dot line: $z^*=-8$ cm ($g=c=0.89$ mm). (b) Reflected phase profile at $z''=26$ cm for different values of z^*. The full line represents $z^*=-4$ cm; the dashed line: $z^*=-5$ cm; the dotted line: $z^*=-6$ cm; and the dash-dot line: $z^*=-8$ cm ($g=c=0.89$ mm).

In Fig. 6(b), the corresponding phase curves are given. Here, we observe a sudden steep increase after a relatively plane part. This happens at the point $\tilde{y}''$.

In Fig. 7, the crack is placed at another depth, $c=g=1.5$ mm, in the middle of the plate. Looking at the reflected shifted beam component, we notice three parts: a smooth part, which is caused by the reradiation of the plate (L/S/L structure); a second part, due to the distortion of the crack, and a third part which exists of several adjacent peaks. These are caused by multiple reflection in the small thin plate bordered by the crack face and the liquid. The closer z^* to the crack, the faster the distortion in the second lobe occurs. The resonance condition of the incident beam for the L/S/L structure is not fulfilled anymore in the L/S/V structure, and as a result there is no constructive interference anymore of the induced partial waves which explains the various peaks.

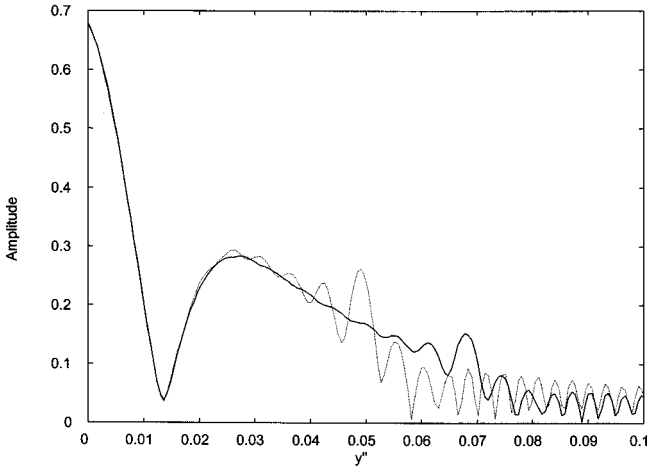


FIG. 7. Reflected amplitude profile at $z''=26$ cm for different values of z^*. The full line represents $z^*=-8$ cm; the dotted line: $z^*=-6$ cm ($c=g=1.5$ mm).

In Fig. 8, another Lamb angle of the L/S/L structure is considered. We chose 25.67°, which is the S1 mode and locate the crack at $c=g=0.5$ mm. Again, the effect of approaching the crack by shifting the transducer parallel to the plate on the reflected amplitude field is investigated. We consider a crack located at $c=g=0.5$ mm. The effect of the occurring crack is reflected in the amplitude field by very steep falloffs. After the falloff no acoustic energy is reflected anymore; no multiple reflection peaks in the L/S/V plate are present. These only occur for thicker L/S/V plates (higher value of c). The position $\tilde{y}$, as defined in (28), corresponds exactly to the position where the amplitude reaches 40% of the last maximum.

So, by means of resonance modes of the L/S/L structure, the starting point of the crack can be localized by amplitude and phase analysis. When the crack is situated quite at the bottom side of the plate (small value of c), the amplitude at $\tilde{z}$ equals a constant fraction of the last maximum amplitude. This constant fraction seems to be very close to $\sin(\theta)$, with θ the angle of incidence.

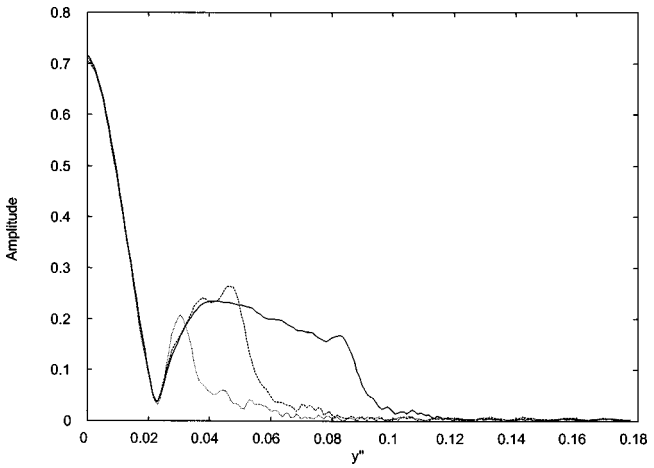


FIG. 8. Reflected amplitude profile at $z''=26$ cm for different values of z^*. The full line represents $z^*=-10$ cm; the dashed line: $z^*=-6$ cm; the dotted line: $z^*=-4$ cm ($c=g=0.5$ mm, $\theta=25.67°$).

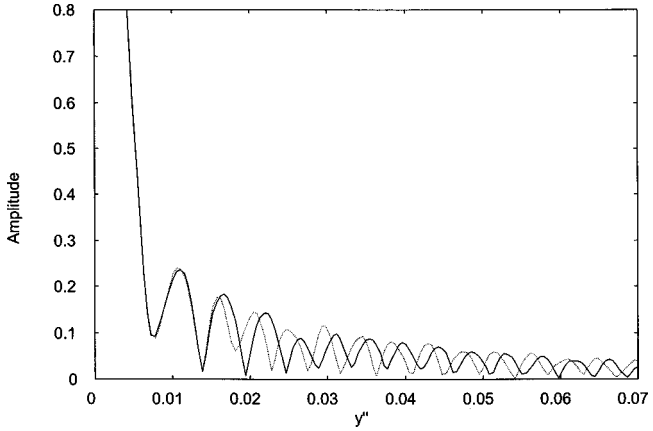


FIG. 9. Reflected amplitude field at $z''=10$ cm of a focused Gaussian beam. The full lines represents $z^*=-3$ cm; the dotted line: $z^*=-2$ cm ($c=g=2$ mm, $\theta=26°$).

C. Influence of a crack on the reflected and transmitted field of a focused Gaussian beam

The description of a focused Gaussian beam in the mode method was done previously in Ref. 11. It comes down to attributing a complex value to the halfwidth L. In our examples we consider Gaussian beams with focus on the plate. The angle of incidence is 26°, which is not a critical angle. In Fig. 9 the reflected amplitude field of a focused beam is plotted at $z''=10$ cm for different values of z^*. It's clear that the multiple reflection peaks of the L/S/L structure (e.g., $y'' \in [0;0.025]$ for the full line) suddenly change to the multiple reflected peaks of the L/S/V structure (e.g., $y'' \in [0.03;0.07]$ for the full line). The shape is changing and the distance between two adjacent peaks is changing. This is an indication that the wave has reached the crack. Moreover, the distance between two peaks of the L/S/V plate can be used to estimate the depth of the crack (by projecting the wave vectors in the plate onto the Z axis). The transmitted fields are shown in Fig. 10 at the plane $y=8$ cm. We didn't choose a phase plane because only amplitude fields will be investigated. The same observations can be made as for the reflected field. The only advantage to focus the beam on the

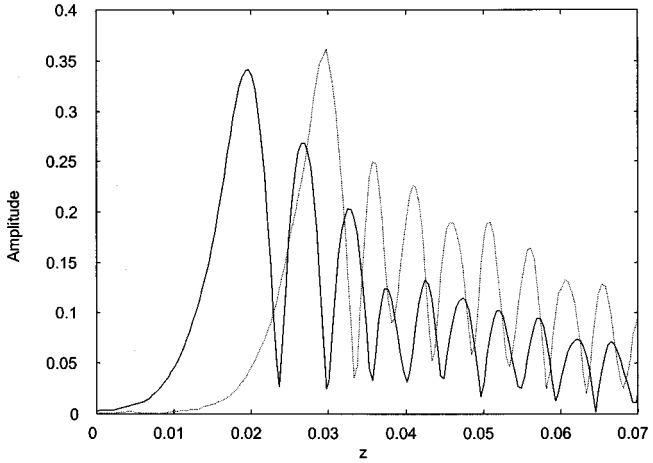
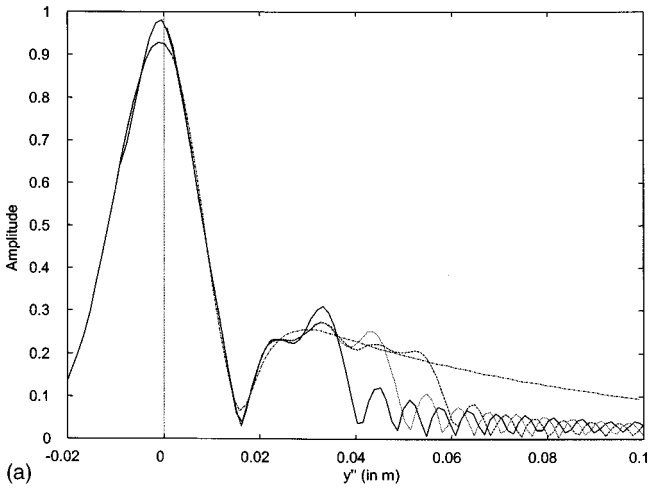


FIG. 10. Transmitted amplitude field at $y=8$ cm of a focused Gaussian beam. The full lines represents $z^*=-3$ cm; the dotted line: $z^*=-1$ cm ($c=g=1$ mm, $\theta=26°$).



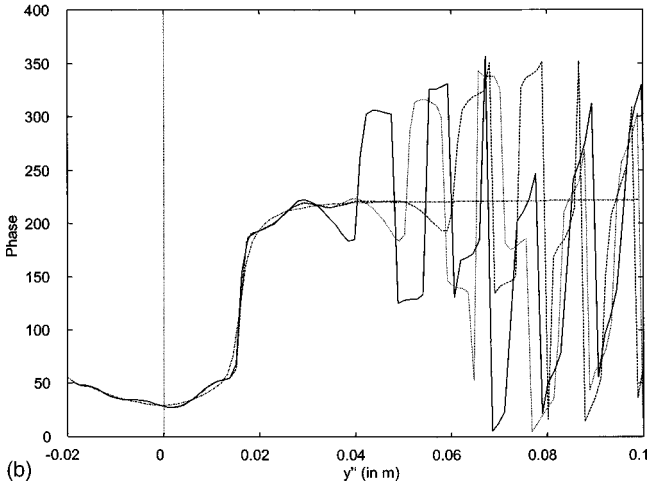


FIG. 11. (a) Reflected amplitude at $z''=30$ cm. The dash-dot lines: $L_1=\infty$; the dashed lines: $L_1=6$ cm; the dotted line: $L_1=5$ cm; the full lines; $L_1=4$ cm. (b) Reflected phase profiles at $z''=30$ cm. The dash-dot lines: $L_1=\infty$; the dashed lines: $L_1=6$ cm; the dotted line: $L_1=5$ cm; the full lines: $L_1=4$ cm.

plate seems to be the opportunity to estimate the depth of a crack very accurately by insonifying at a randomly chosen angle.

D. Influence of the length of a crack on the reflected amplitude and phase field

Assuming the starting position at $z=0$ and the depth of the crack given by $g=c=0.6$ mm, we analyze how the length of a crack is affecting the nonspecular reflected field of a Gaussian beam, incident at a critical angle $\theta=11.84°$ of the L/S/V structure. We chose this angle to be not close to a resonance mode of the L/S/L plate, otherwise it would be difficult to distinguish between Lamb profiles of the L/S/V- and the L/S/L structures, respectively. We assume z^* equal to zero, so the central axis of the beam reaches the plate at the origin of the (Y,Z)-coordinate system.

We study reflected amplitude and phase profiles at $z''=30$ cm. We consider different values of L_1, the length of the crack, and analyze the effects on the reflected curves. Figure 11(a) shows the reflected amplitude field, Fig. 11(b) the reflected phase profile. Looking at the amplitude curves,

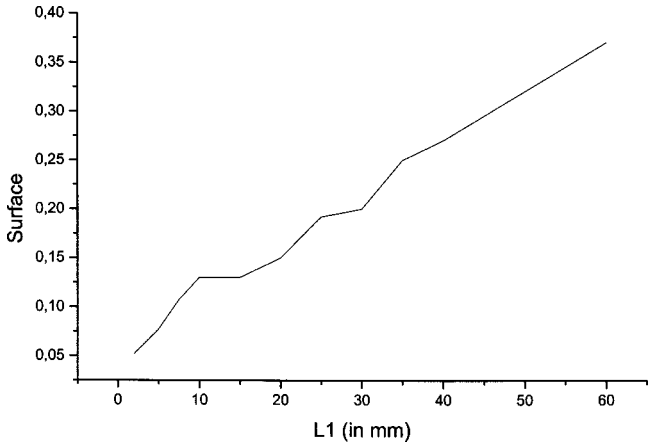


FIG. 12. Reflected energy (surface) as a function of the length of the crack.

we note very steep falloffs, which indicate the end of the crack very clearly. This can be explained by the disappearance of the crack at certain z values. As a result the resonance conditions are not fulfilled anymore and the acoustic energy is no longer blocked and no longer totally reflected by the crack. The propagation of the acoustic wave at $z \geqslant L_1$ results in multiple reflection and transmission which explains the several small peaks in amplitude [Fig. 11(a)] and the corresponding plane phase parts [Fig. 11(b)].

We calculated also the surface under the reflected amplitude field between $y''=0.04$ and $y''=0.10$. This is a measure for the amount of reflected energy. The result is sketched in Fig. 12, where the surface is plotted as a function of L_1.

E. Influence of the length of a crack on the transmitted field

We consider a fixed length of the crack, $L_1=5$ cm, and a critical angle, 11.84° of the L/S/V structure. We again move the transducer along the $y=y_0$ axis (parallel to the plate). We explore the transmitted field at the plane $y=6$ cm for different values of z^*. The transmitted field (Fig. 13) represents the total interference of the scattered field by the crack, the geometrical transmitted field (by means of the

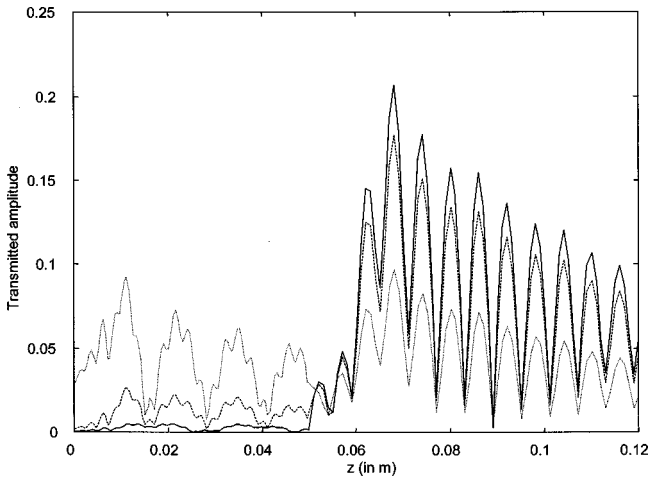


FIG. 13. Transmitted amplitude profile at $y=6$ cm. $L_1=5$ cm, the dotted line represents $z^*=0$ cm; the dashed line: $z^*=1.5$ cm and the full line: $z^*=2.5$ cm.

refraction laws), and the multiple transmitted field. Of course, the amplitude for $z \geqslant L_1$ is much larger than the amplitude in structure 2, since the incident wave, impinging on the plate with the crack, is blocked and reflected for a substantial amount. For $z^* = 0$ cm, there is a part of the incident beam which transmits through the plate at the left side of the crack. This is not the case anymore for larger values of z^* because the total beam is blocked. Analyzing Fig. 13, we see a clear difference between the curves left of $z = 0.05$ and those to the right of that point. At the right side of $z = 0.05$, there is no longer a crack and transmission is possible again. Thus, if z^* moves to the right, the profiles left of the point $z = 0.05$ disappear and the right ones increase. It is clear that the discrepancy between ''amplitudes for $z \leqslant L_1$'' and ''amplitudes for $z \geqslant L_1$'' increases for increasing values of z^*. When really the middle of the crack is insonified ($z^* = 2.5$ cm) (full line), a very steep increase of the amplitude is observed which enables us to detect the end or to measure the length of the crack very accurately.

VI. CONCLUSION

The radiation mode model gives an accurate description of the reflection and transmission of a bounded acoustic beam impinging on a plate with a crack parallel to the surface. We assumed oblique incidence and the crack to be sufficiently thin. Dimensions and geometrical situation of the crack inside the plate disturb the reflections and transmissions under well-understood rules. The effects on reflected and transmitted Gaussian beams, incident at critical angles of the L/S/L or L/S/V structure, were studied and physically clarified. The influence on focused beams was studied too. In general, is was demonstrated that bounded beam scattering by a crack is a very useful tool in NDT. The variant problem of Stoneley wave diffraction by subsurface cracks is being handled at the moment and will be published soon.

ACKNOWLEDGMENTS

The authors would like to thank the Research Council of the K. U. Leuven for the fellowship awarded to Professor G. N. Shkerdin, which enabled us to write this paper.

APPENDIX:

$$\sin(\varphi^+) = \frac{2k_{ly}\left[\cos(k_{ty}d_1)\sin(k_{ly}d_1) + \left(\frac{b}{a}\right)\cos(k_{ly}d_1)\sin(k_{ty}d_1)\right]}{k_{y,2}\left(1 - \frac{2k_{z,2}^2}{q_t^2}\right)|D_1|}, \tag{A1}$$

$$\cos(\varphi^+) = \frac{\frac{2\rho_1}{\rho}\left(1 - \frac{2k_{z,2}^2}{k_t^2}\right)\left[2 - 2\cos(k_{ty}d_1)\cos(k_{ly}d_1) + \left(\frac{b}{a} + \frac{a}{b}\right)\sin(k_{ty}d_1)\sin(k_{ty}d_1)\right]}{|D_1|} \tag{A2}$$

with $d_1 = d - c$. $\mathrm{Sin}(\varphi^-)$ and $\cos(\varphi^-)$ are the same with changes from d_1 into g, and D_1 into D_2.

$$\tilde{I}_{12,0}^{1\pm}(k_{y,1};k_{y,2}) = \sin(\varphi^{\pm}(k_{y,2}) - \varphi(k_{y,1}))P\left(\frac{1}{k_{y,2} - k_{y,1}}\right) + \frac{\sin(\varphi^{\pm}(k_{y,2}) + \varphi(k_{y,1}))}{k_{y,2} + k_{y,1}}, \tag{A3}$$

$$\tilde{I}_{12,0}^{2\pm}(k_{y,1};k_{y,2}) = \sin(\varphi^{\pm}(k_{y,2}) - \psi(k_{y,1}))P\left(\frac{1}{k_{y,2} - k_{y,1}}\right) + \frac{\sin(\varphi^{\pm}(k_{y,2}) + \psi(k_{y,1}))}{k_{y,2} + k_{y,1}}, \tag{A4}$$

where P denotes the principal value function.

[1] L. Guan and A. Norris, ''Elastic wave scattering by rectangular cracks,'' Int. J. Solids Struct. **29**, No. 12, 1549–1565 (1992).
[2] W. M. Visscher, ''Scattering of elastic waves from planar cracks in isotropic media,'' J. Acoust. Soc. Am. **69**(1), 50–53 (1981).
[3] S. Hirose, ''Higher harmonics in the far field due to dynamic crack-face contacting,'' J. Acoust. Soc. Am. **93**(1), 142–147 (1993).
[4] A. Boström and H. Wirdelius, ''Ultrasonic probe modeling and nondestructive crack detection,'' J. Acoust. Soc. Am. **97**(5), 2836–2848 (1995).
[5] K. I. Maslov, T. Kundu, and O. I. Lobkis, ''Acoustic microscopy for spherical inclusion characterization,'' J. Acoust. Soc. Am. **100**(1), 80–85 (1996).
[6] J. H. M. T. van der Hijden and F. L. Neerhoff, ''Diffraction of elastic waves by a sub-surface crack (in-plane motion),'' J. Acoust. Soc. Am. **75**(6), 1694–1704 (1984).
[7] J. D. Achenbach, W. Lin, and L. M. Keer, ''Mathematical modeling of ultrasonic wave scattering by sub-surface cracks,'' Ultrasonics **24**, 207–215 (1986).
[8] J. D. Achenbach, W. Lin, and L. M. Keer, ''Surface waves due to scattering by a near-surface parallel crack,'' IEEE Trans. Sonics Ultrason. **SU-30**, No. 4, 270–276 (1983).
[9] R. Briers, O. Leroy, G. N. Shkerdin, and Yu. V. Gulyaev, ''Mode theory as

a framework for the investigation of the generation of a Stoneley wave at a liquid/solid interface,'' J. Acoust. Soc. Am. **95**, 1953–1966 (1994).
[10] R. Briers, O. Leroy, and G. N. Shkerdin, ''Conversion of a Stoneley wave at the extremity of a fluid loaded plate,'' J. Acoust. Soc. Am. **101**, 1347–1357 (1997).
[11] J. Vandeputte, O. Leroy, R. Briers, and G. N. Shkerdin, ''Extension of the mode method for viscoelastic media and focused ultrasonic beams,'' J. Acoust. Soc. Am. **108**(4), 1614–1621 (2000).
[12] L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1960).
[13] R. Briers, O. Leroy, and G. N. Shkerdin, ''Bounded beam interaction with thin inclusions. Characterization by phase differences at Rayleigh angle incidence,'' J. Acoust. Soc. Am. **108**(4), 1622–1630 (2000).
[14] T. J. Plona, L. E. Pitts, and W. G. Mayer, ''Ultrasonic bounded beam reflection and transmission effects at a liquid/solid–plate/liquid interface,'' J. Acoust. Soc. Am. **59**(6), 1324–1328 (1976).
[15] T. D. K. Ngoc and W. G. Mayer, ''A general description of ultrasonic nonspecular reflection and transmission effects for layered media,'' IEEE Trans. Sonics Ultrason. **SU-27**, No. 5, 229–236 (1980).

Elastic constants of natural quartz

Paul Heyliger[a)]
Department of Civil Engineering, Colorado State University, Fort Collins, Colorado 80523

Hassel Ledbetter
Los Alamos National Laboratory (E536), Los Alamos, New Mexico 87545

Sudook Kim
Materials Science and Engineering Laboratory, National Institute of Standards and Technology, Boulder, Colorado 80305

(Received 21 December 2002; revised 18 May 2003; accepted 30 May 2003)

The elastic constants of a natural-quartz sphere using resonance-ultrasound spectroscopy (RUS) are measured. The measurements of the near-traction-free vibrational frequencies of the sphere are matched with the predicted frequencies from the dynamic theory of elasticity, with optimized estimates for the elastic constants driving the differences between these sets of frequencies to a minimal value. The present computational model, although based on earlier approaches, is the first application of RUS to trigonal-symmetry spheres. Quartz shows six independent elastic constants, and our estimates of these constants are close to those computed by other means. Except for C_{14}, after a 1% mass-density correction, natural quartz and cultured quartz show the same elastic constants. Natural quartz shows higher internal frictions. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1593063]

PACS numbers: 43.20.Hq, 43.20.Ks, 43.35.Cg [RR]

I. INTRODUCTION

Silicon dioxide (SiO_2) in the form of quartz is Earth's most ubiquitous surface mineral. Perhaps the first measurements of quartz's elastic constants originated with Groth,[1] whose 1895 book showed the representation surface of the Young (extension) modulus E_{ii} measured statically. This surface revealed quartz's moderate elastic anisotropy: $E_{33}/E_{11} = 1.31$.

Quartz's complete elastic constants were studied first by Woldemar Voigt, and his static-measurement results appeared in his epochal *Lehrbuch der Kristallphysik* in 1910.[2] Giebe and Scheibe[3] made, perhaps, the first dynamic measurements. Subsequently, quartz's elastic constants were reported in more than 20 studies.[4–6]

Interest in quartz's elastic constants continues for three reasons: First, quartz's elastic constants remain incompletely understood, mainly because quartz is piezoelectric and we lack a good theory for the elastic constants of piezoelectric crystals.[7] Second, quartz occupies a central place in crystal chemistry-physics because of its many polymorphs. We can study these polymorphs, both experimentally and theoretically, through the elastic constants. Third, quartz's device applications exceed one billion per year.[5] Most of these applications use quartz's macroscopic resonance frequencies, which depend mainly on crystal geometry and elastic constants. Elastic constants can be measured accurately (within a few parts in 10^4) and, because of their tensor nature, they provide essential information for preparing crystal cuts with various piezoelectric properties.[8]

Quartz's elastic constants received extensive review, notably by Cady,[4] Brice,[9] James,[10] and, recently, Ballato.[5]

The present study proceeded with two principal objectives: First, measure the complete elastic constants of natural quartz. Second, analyze the macroscopic vibration frequencies of a trigonal-crystal-symmetry sphere.

We used resonance-ultrasound spectroscopy (RUS) to determine these elastic constants. This method was described thoroughly by Migliori and Sarrao,[11] and, although our approach proceeds similarly, the material's trigonal symmetry requires several modifications to existing models.

Applications of RUS to trigonal crystals are rare. Ohno *et al.*[12] considered the case of rectangular parallelepipeds with the trigonal axis perpendicular to one of the faces in the first attempt to determine the elastic constants of trigonal crystals. Ledbetter *et al.*[6] studied the elastic stiffnesses and internal friction of monocrystal cultured quartz for cylindrical geometries. In both of these studies, the general eigenvalue problem that results from the application of the Ritz method to the basic elasticity problem of a solid in traction-free vibration was separated into four symmery groups based on material symmetry and geometrical symmetry. This reduces the size of the initial eigenvalue problem and results in much faster and more accurate frequency calculations. Our approach requires multiple solutions of this eigenvalue problem for a continuously changing elastic-stiffness tensor, and this separation is a critical step. Willis and colleagues[13] used RUS to study the hexagonal–trigonal phase transformation in $LiKSO_4$. Their study provides a good example of applying Landau theory to a second-order or near-second-order phase transition. Quartz's α–β transition at 573 °C provides another example of such a phase transition.

In the following we describe our theoretical model, measurement system, and results of their combination.

a)Electronic mail: prh@engr.colostate.edu

II. THEORETICAL FREQUENCIES

A. Geometry and equations of motion

Our specimen was a natural quartz sphere. The predicted resonance frequencies depend on the sphere radius because our variational method of approximation evaluates terms proportional to elastic stiffness over the volume of the solid. Density ρ appears on the right-hand side of the equations of motion:

$$\sigma_{ij,j}=\rho\ddot{u}_i. \tag{1}$$

Here, σ_{ij} denotes stress and u_i displacement. We do not explicitly solve these equations. Instead, we seek an alternative solution to their weak form.

B. Variational formulation

Hamilton's principle for an elastic medium is given by[14]

$$\delta\int_{t_0}^{t}dt\int_V\left[\frac{1}{2}\rho\dot{u}_j\dot{u}_j-\frac{1}{2}C_{ijkl}\epsilon_{ij}\epsilon_{kl}\right]dV+\int_{t_0}^{t}dt\int_S\overline{t}_k\delta u_k dS=0. \tag{2}$$

Here t denotes time, V and S the volume and surface occupied by and bounding the solid, $\overline{t}_k$ the components of the specified surface tractions, δ the variational operator, the overdot differentiation with respect to time, ϵ_{ij} the components of infinitesimal strain, and C_{ijkl} the components of the fourth-order elastic-stiffness tensor. The general constitutive relation for 32-point-group trigonal symmetry represented in rectangular Cartesian coordinates can be expressed in a contracted form of the stiffness tensor as a six-by-six matrix:

$$\begin{Bmatrix}\sigma_1\\ \sigma_2\\ \sigma_3\\ \sigma_4\\ \sigma_5\\ \sigma_6\end{Bmatrix}=\begin{bmatrix}C_{11}&C_{12}&C_{13}&C_{14}&0&0\\ C_{12}&C_{22}&C_{23}&C_{24}&0&0\\ C_{13}&C_{23}&C_{33}&0&0&0\\ C_{14}&C_{24}&0&C_{44}&0&0\\ 0&0&0&0&C_{55}&C_{56}\\ 0&0&0&0&C_{56}&C_{66}\end{bmatrix}\begin{Bmatrix}\epsilon_1\\ \epsilon_2\\ \epsilon_3\\ \epsilon_4\\ \epsilon_5\\ \epsilon_6\end{Bmatrix}. \tag{3}$$

Here we used the conventional contracted notation ($\sigma_{11}=\sigma_1$, $\sigma_{23}=\sigma_4$, $\epsilon_{11}=\epsilon_1$, $2\epsilon_{23}=\epsilon_4$, $C_{1111}=C_{11}$, $C_{1123}=C_{14}$, and so on), and it is understood that the 1, 2, 3 directions are $x_1=x$, $x_2=y$, and $x_3=z$.

Hamilton's principle in contracted notation becomes

$$\begin{aligned}0=&-\int_0^t\int_V\{\sigma_1\delta\epsilon_1+\sigma_2\delta\epsilon_2+\sigma_3\delta\epsilon_3+\sigma_4\delta\epsilon_4+\sigma_5\delta\epsilon_5\\&+\sigma_6\delta\epsilon_6\}dV\,dt+\frac{1}{2}\delta\int_0^t\int_V\rho(\dot{u}^2+\dot{v}^2+\dot{w}^2)dV\,dt\\&+\int_{t_0}^{t}dt\int_S\overline{t}_k\delta u_k dS.\end{aligned} \tag{4}$$

The surface integral in Eq. (4) is close to zero because the only forces engendered by our measurement system are small point forces at the locations where the pinducers contact the specimen. These forces cause little effect on our results; therefore, we neglect them.

The strain-displacement relations are

$$\epsilon_{ij}=\tfrac{1}{2}(u_{i,j}+u_{j,i}). \tag{5}$$

Our displacement components are expressed in terms of the coordinates $x_1=x$, $x_2=y$, $x_3=z$ with the corresponding displacement components $u_1=u(x,y,z)$, $u_2=v(x,y,z)$, and $u_3=w(x,y,z)$. Substitution of the constitutive and strain-displacement equations into Eq. (4) yields the final weak form of the equations of motion. If we integrate by parts and isolate the volume integral, the resulting Euler equations of this statement yield the equations of motion in Eq. (1). This, however, is not our goal. Instead, we seek the solution to the weak form to solve for the unknown frequencies ω and the mode shapes represented by the functions for u, v, w. This requires an assumed form for the displacements.

C. Ritz approximations

In the Ritz method, we seek approximations to the three displacements using finite linear combinations of the form

$$\begin{aligned}u_1(x_1,x_2,x_3)&=\sum_{i=1}^{n}a_j\Psi_j^1(x_1,x_2,x_3),\\ u_2(x_1,x_2,x_3)&=\sum_{i=1}^{n}b_j\Psi_j^2(x_1,x_2,x_3),\\ u_3(x_1,x_2,x_3)&=\sum_{i=1}^{n}d_j\Psi_j^3(x_1,x_2,x_3).\end{aligned} \tag{6}$$

Here the a, b, and d components are unknown constants, and the functions Ψ are known functions of the spatial coordinates used in the specific formulation. Substituting these equations into the weak form yields the matrix equation

$$\begin{aligned}&\begin{bmatrix}[M^{11}]&[0]&[0]\\ [0]&[M^{22}]&[0]\\ [0]&[0]&[M^{33}]\end{bmatrix}\begin{Bmatrix}\{a\}\\ \{b\}\\ \{c\}\end{Bmatrix}\omega^2\\&+\begin{bmatrix}[K^{11}]&[K^{12}]&[K^{13}]\\ [K^{21}]&[K^{22}]&[K^{23}]\\ [K^{31}]&[K^{32}]&[K^{33}]\end{bmatrix}\begin{Bmatrix}\{a\}\\ \{b\}\\ \{c\}\end{Bmatrix}=\begin{Bmatrix}\{0\}\\ \{0\}\\ \{0\}\end{Bmatrix}.\end{aligned} \tag{7}$$

The element matrices used for each of the formulations are given in the Appendix. The resulting eigenvalue problem is solved using the QR algorithm.[15]

Our choice of approximation function is the same as that of Visscher and colleagues.[16] For each of the three displacement components, the following class of basis functions was selected:

$$\Psi(x,y,z)=x^iy^jz^k. \tag{8}$$

Here x, y, z denote the rectangular-cylindrical-coordinate directions of the cylinder and i, j, and k are integers. The primary advantage of casting the problem in this form is that the integrals required by the weak form are simple to evaluate analytically. For the indices i, j, k, the volume integrals over the sphere given in the Appendix (without the elastic constants) have the form[16]

TABLE I. Classification and description of motion and approximation functions for vibrational modes (from Ohno—Ref. 17), with the modal groups for trigonal symmetry shown as four combinations of two subgroups.

Trigonal subset	Type	Component	Subset x	Subset y	Subset z	Classification
AG	OD	u	O	E	E	Longitudinal vibration
		v	E	O	E	
		w	E	E	O	
	OX	u	O	O	O	Torsion along x axis
		v	E	E	O	
		w	E	O	E	
AU	EX	u	E	E	E	Flexure along x axis
		v	O	O	E	
		w	O	E	O	
	EV	u	E	O	O	Axisymmetric flexure
		v	O	E	O	
		w	O	O	E	
BU	EY	u	O	O	E	Flexure along y axis
		v	E	E	E	
		w	E	O	O	
	EZ	u	O	E	O	Flexure along z axis
		v	E	O	O	
		w	E	E	E	
BG	OY	u	E	E	O	Torsion along y axis
		v	O	O	O	
		w	O	E	E	
	OZ	u	E	O	E	Torsion along z axis
		v	O	E	E	
		w	O	O	O	

$$F(i,j,k)=\left\{1,\frac{\pi}{2}\right\}\frac{R^{i+j+k+3}(i-1)!!(j-1)!!(k-1)!!}{(i+j+k+3)!!}. \tag{9}$$

Here R denotes the sphere radius. This is the $(+++)$ octant of the sphere and is one of eight that must be evaluated for each of the terms in these matrices. The braced term is one if two or three of the integers i, j, k are odd. Otherwise it is $\pi/2$. The remaining seven integrals are similar in form except for the sign, which depends on the evenness or oddness of (i,j,k).

Ohno[17] showed that for a trigonal symmetry parallelepiped it is possible to split the final eigenvalue problem into smaller eigenvalue problems based on the symmetries of the material, specimen geometry, and approximation functions. Heyliger and Johnson[18] extended this approach to a trigonal cylinder, and showed that the modal vibration patterns can be split into six subsets. For the sphere, we use the four subsets defined by Ohno and classify each of the functions in Eq. (8) according to the grouping shown in Table I.

III. MEASUREMENTS

We obtained a sphere of natural Brazilian quartz from a local rock shop. The specimen showed a mass density of 2.6466 g/cm^3, determined from the mass and the volume (diameter=4.4998 cm), 0.7% lower than the x-ray-diffraction mass density of a ''perfect'' quartz crystal: 2.6655 g/cm^3. The quartz showed several randomly oriented 2-μm-diam rutile (TiO_2) fibers varying in length from 0.5 to 3 cm with an average of 2 cm. Because the volume fraction of the fibers is small, we ignored them. Except for the rutile fibers, the specimen showed ''perfect'' optical properties, indicating freedom from macroscopic twins. Also, any significant twinning would appear as irregular apparent C_{ij} values.

Quartz possesses trigonal symmetry, the point group 32, and six independent elastic-stiffness coefficients $(C_{11},C_{12},C_{13},C_{14},C_{33},C_{44})$:

$$[C_{ij}]=\begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & 0 & 0 \\ & C_{11} & C_{13} & -C_{14} & 0 & 0 \\ & & C_{33} & 0 & 0 & 0 \\ & & & C_{44} & 0 & 0 \\ & & & & C_{44} & C_{14} \\ & & & & & \tfrac{1}{2}(C_{11}-C_{12}) \end{bmatrix}. \tag{10}$$

Quartz possesses a troublesome elastic coefficient $C_{14}=-C_{24}=C_{56}$, which complicates the mathematical analysis for relating macroscopic-vibration resonance frequencies. For quartz, $|C_{14}|$ exceeds C_{12} and C_{13}, and thus must be considered. When we invert Eq. (10) to obtain the elastic-compliance matrix $[S_{ij}]$, we see that the familiar relationship for the Young moduli, $E_{ii}=S_{ii}^{-1}$, fails to result. The E_{ii} depend also on C_{14}. Inversion leads to the further curiosity that $C_{44}\neq S_{44}^{-1}$. By inspection of Eq. (10), we see that $C_{14}=0$ leads to a symmetry increase to hexagonal (transverse-isotropic) and five independent C_{ij}, the symmetry of high-temperature (beta) quartz.

We determined the C_{ij} by using resonance-ultrasound spectroscopy (RUS), which we described previously.[19] Briefly, two 10 MHz PZT transducers hold the sphere specimen at two points. One transducer transmits continuous sinusoidal waves to the specimen, and the other transducer detects the specimen's displacement response. We made measurements at 22 °C in vacuum (1.33 Pa [10^{-3} Torr]). This method's principal advantage is that a single frequency sweep on a single specimen yields the complete elastic-stiffness tensor, both the real part C_{ijkl} and the imaginary part Q_{ijkl}^{-1}.

The resonance frequencies depend on specimen shape, size, mass (or mass density), and elastic-stiffnesses. (For nonspherical specimens, frequencies depend also on crystal orientation.) Each resonance consists of linear combinations of independent elastic-stiffness coefficients. Although not reported here, the complete internal-friction ''tensor'' can be calculated from the half-power width of associated resonance peaks.

We obtained the C_{ij} from the resonance frequencies by using an iterative process that minimizes the differences between measured and calculated resonance frequencies with the given C_{ij}, which take modified values for the next iteration. Because of small differences between the measured and calculated resonance frequencies, the C_{ij} calculations excluded the piezoelectric and dielectric effects, the ''piezoelectric stiffening.''

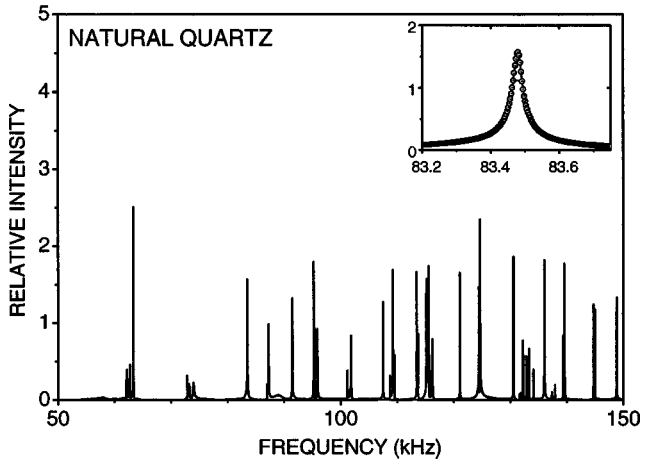


FIG. 1. Resonance spectrum of trigonal-crystal-symmetry, monocrystal quartz sphere. Resonance frequencies give the elastic-stiffness tensor C_{ijkl}. Resonance-peak widths give the imaginary parts of the C_{ijkl}, the internal friction Q^{-1}_{ijkl}.

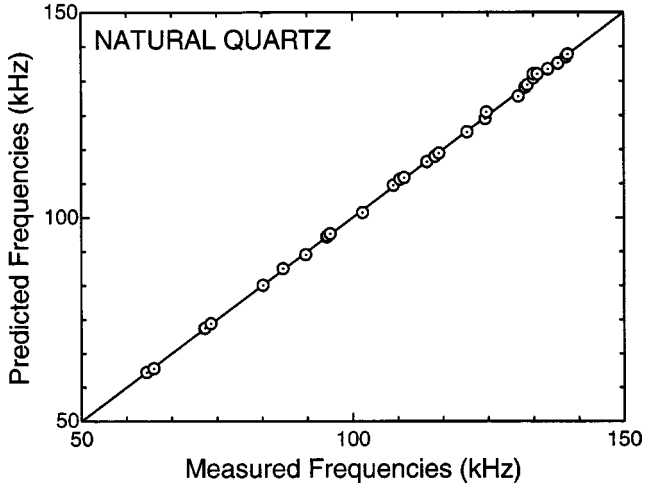


FIG. 2. Predicted versus measured resonance frequencies. Predicted values based on a purely elastic model, neglecting piezoelectric stiffening.

IV. RESULTS

Figure 1 shows the resonance spectrum. We measured more than 100 resonance frequencies, but selected only the first 45 (30 of which are distinct) for our analysis. Our Ritz model decreases in accuracy with increase in frequency; hence, higher modes are much less accurate than lower modes in our approximation.

The resulting elastic constants are shown in Table II, which also shows our results on cultured quartz, to be re-

TABLE II. Elastic constants and related properties of monocrystal α-quartz.

Property	Natural	Cultured (ordinary)	Cultured (premium)
ρ (g/cm^3)	2.6466±0.01	2.6497±0.0002	2.6497±0.0002
C_{11} (GPa)	87.26	87.16±0.14	87.17±0.05
C_{33} (GPa)	105.8	106.00±0.20	105.80±0.07
C_{44} (GPa)	57.15	58.14±0.08	58.27±0.03
C_{66} (GPa)	40.35	40.26±0.05	40.28±0.02
C_{12} (GPa)	6.57	6.64±0.10	6.610±0.035
C_{13} (GPa)	11.95	12.09±0.21	12.020±0.086
C_{14} (GPa)	−17.18	−18.15±0.08	−18.23±0.03
E_{11} (GPa)	79.41	78.62±0.32	78.61±0.15
E_{33} (GPa)	102.8	102.90±0.23	102.70±0.19
ν_{12}	0.1286	0.1364±0.0059	0.1367±0.0002
ν_{13}	0.0984	0.0985±0.0025	0.0981±0.0008
ν_{31}	0.1274	0.1289±0.0023	0.1282±0.0010
Θ_D (K)[a]	564.0	563.0	563.1
Quasi-isotropic polycrystalline elastic constants obtained by Voigt–Reuss–Hill averaging			
C_l (GPa)	100.2	97.14	100.7
B (GPa)	37.67	37.74	37.69
G (GPa)	46.91	44.55	47.23
E (GPa)	99.45	95.91	99.95
ν	0.0600	0.076 44	0.058 01

[a]Computed numerically from the C_{ij} by averaging the sound velocities from the Christoffel equations' 23 472 directions.

TABLE III. Measured and calculated resonance frequencies (in MHz) for quartz sphere.

Mode	ω_m	ω_c	Diff. (%)	Group
1	0.062 046	0.062 085	0.06	1
2	0.062 046	0.062 085	0.06	4
3	0.063 338	0.062 987	−0.55	2
4	0.063 338	0.062 987	−0.55	3
5	0.072 716	0.072 871	0.21	1
6	0.072 847	0.072 929	0.11	3
7	0.073 886	0.074 004	0.16	1
8	0.073 886	0.074 004	0.16	4
9	0.083 449	0.083 408	−0.05	4
10	0.087 210	0.087 468	0.30	1
11	0.087 210	0.087 468	0.30	4
12	0.091 391	0.090 919	−0.52	3
13	0.091 391	0.090 919	−0.52	2
14	0.095 250	0.095 285	0.04	3
15	0.095 361	0.095 513	0.16	2
16	0.095 908	0.096 041	0.14	4
17	0.101 855	0.101 206	−0.64	3
18	0.101 855	0.101 206	−0.64	2
19	0.107 530	0.107 857	0.30	3
20	0.108 727	0.109 350	0.57	1
21	0.109 526	0.109 801	0.25	2
22	0.109 526	0.109 801	0.25	3
23	0.113 717	0.113 633	−0.07	4
24	0.113 717	0.113 633	−0.07	1
25	0.115 216	0.114 980	−0.21	3
26	0.115 216	0.114 980	−0.21	2
27	0.115 955	0.115 728	−0.20	4
28	0.115 955	0.115 728	−0.20	1
29	0.121 118	0.120 943	−0.14	4
30	0.124 481	0.124 158	−0.26	3
31	0.124 684	0.125 763	0.87	2
32	0.130 591	0.129 580	−0.77	1
33	0.131 800	0.131 770	−0.02	1
34	0.131 800	0.131 770	−0.02	4
35	0.132 240	0.132 335	0.07	2
36	0.132 240	0.132 335	0.07	3
37	0.133 352	0.134 049	0.52	2
38	0.133 352	0.134 988	1.23	3
39	0.134 091	0.134 988	0.67	2
40	0.136 039	0.136 227	0.14	4
41	0.137 910	0.137 576	−0.24	1
42	0.137 910	0.137 576	−0.24	4
43	0.139 353	0.139 167	−0.13	3
44	0.139 687	0.139 777	0.06	2
45	0.139 687	0.139 777	0.06	3

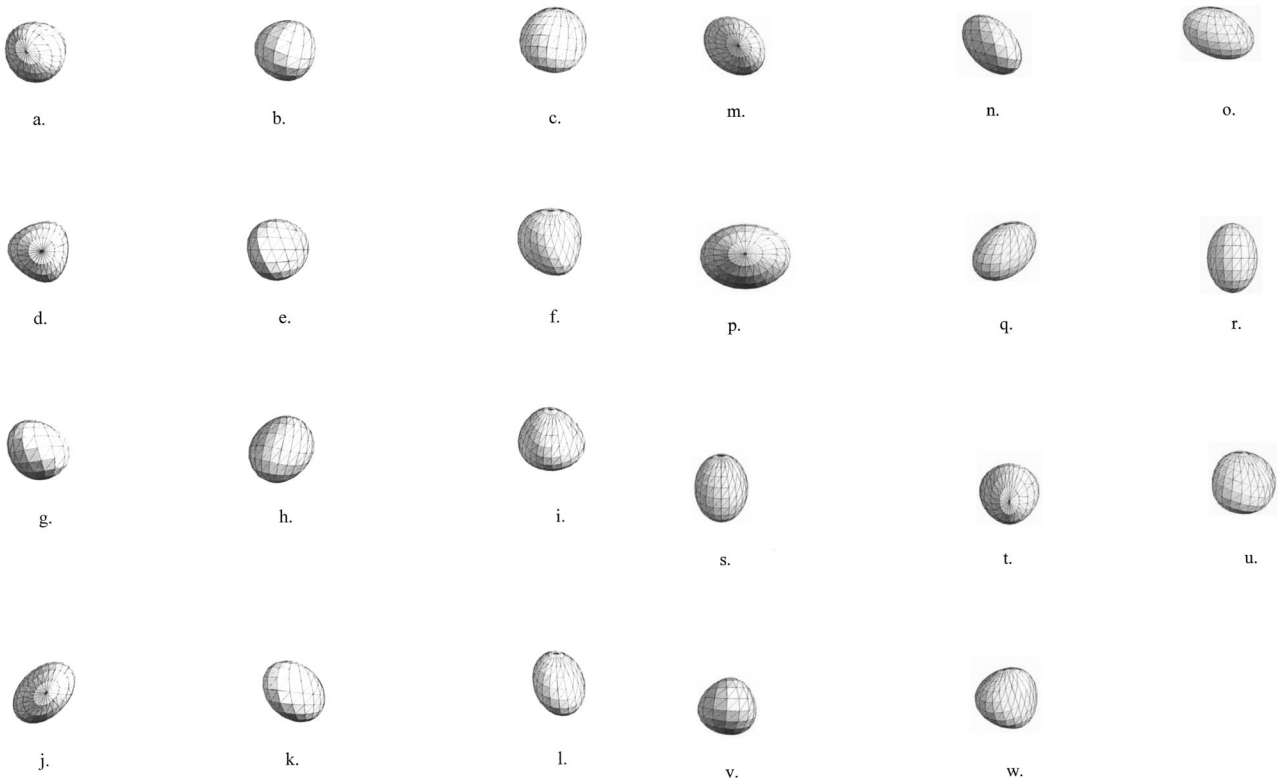


FIG. 3. Deformation diagrams for selected resonances. The three labels for each figure indicate group number, frequency within the group, and the orientation of the viewing perspective. For example, Fig. 1 shows the first mode from group 2 as seen looking toward the origin of the sphere (and coordinate system) along the direction [1,1,1]. (a) 1-1-001, (b) 1-1-010, (c) 1-1-111, (d) 1-2-001, (e) 1-2-010, (f) 1-2-111, (g) 1-3-010, (h) 1-3-100, (i) 1-3-111, (j) 2-1-001, (k) 2-1-010, (l) 2-1-111, (m) 2-2-001, (n) 2-2-010, (o) 2-2-111, (p) 3-1-001, (q) 3-1-111, (r) 3-2-100, (s) 3-2-111, (t) 4-1-001, (u) 4-1-111, (v) 4-2-100, (w) 4-2-111.

ported in detail elsewhere.[6,20] Measurement uncertainties for the natural crystal fall in the same range as those shown for the cultured crystals. Figure 2 shows the relationship between measured and predicted resonance frequencies and provides a basis for neglecting the piezoelectric stiffening. Thus, for quartz, piezoelectric stiffening is relatively small. For other materials, for example, $LiNbO_3$,[21] much larger differences arise between elastic and piezoelectric-dielectric predictions. In piezoelectric crystals, waves propagate with velocities determined by C_{ij}^{E}, E designating constant electric-field intensity, which would be obtained if the crystal were plated with a conductor. Nonplated crystals yield the C_{ij}^{D}, D designating constant electric displacement.

Our fitted frequencies yielded a rms error of 0.36%, slightly higher than typical values for higher-symmetry materials, but still in an acceptable range. Probably, the principal error arises from the specimen's slight nonsphericity. Our fitted frequencies, along with the percentage error and modal group, are shown in Table III. Most frequencies show excellent measurement–model agreement, with only one or two exceptions.

The first ten resonance peaks showed an average internal friction Q^{-1} of 4.7×10^{-4}.

Table II also shows the Debye characteristic temperature Θ_D calculated from the elastic constants.

Figure 3 shows deformation diagrams for selected resonances.

V. DISCUSSION

First, we consider the internal friction Q^{-1}, which arises from various lattice defects. The average Q_{ij}^{-1}, 4.7×10^{-4}, is approximately what we reported for a cultured-quartz crystal with a high dislocation content, $Q^{-1}=1.9\times10^{-4}$,[6] but much higher (by a factor of 34) than the average found for a high-quality (low-dislocation-content), cultured-quartz crystal, 1.4×10^{-5}.[20] Further study of this spherical crystal (strain dependence, frequency dependence, temperature dependence) should identify the source of its higher internal friction. Because of Q^{-1} values as small as 10^{-6} found for other quartz crystals with the same apparatus, we believe the higher Q^{-1} reflects the specimen, not mounting losses.

Second, we consider the natural-quartz versus cultured-quartz elastic constants (Table II). Within measurement errors, no remarkable differences appear. Neglecting C_{14}, the average difference is 0.8%, corresponding closely with the natural crystal's lower mass density and consistent with higher defect content and higher internal friction. Thus, except for a mass-density difference, natural quartz and cultured quartz possess the same elastic constants within a small fraction of 1%, except for C_{14}, where the natural-quartz value is lower by about 5%.

The large specimen-to-specimen variation in C_{14} may result from the lower symmetry of the 32 trigonal point group. When viewed along the c axis, [0001], the crystal structure shows ion shifts away from sixfold symmetry. (See

Fig. 2.32 in Kingery *et al.*[22]) These shifts give rise to the additional elastic stiffness C_{14}. This distorted, or derivative, structure possesses a degree of freedom (the exact ionic positions) that can easily vary among different quartz crystals. Ikeda[20] described C_{14} as the order parameter for the beta–alpha phase transformation. Order-parameter thermodynamics was treated in detail by Landau and Lifshitz.[23]

Our Debye characteristic temperature $\Theta_D = 564\,\mathrm{K}$, calculated as described elsewhere,[24] agrees favorably (but not exactly) with those given by Anderson.[25] $\Theta_D = 545$, 585, 585 K. The spread among the four values ($\pm 3.4\%$) is surprising because if the C_{ij} are measured within 1% (a simple feat), then Θ_D should vary only 0.5%. At zero temperature, the elastic and specific-heat Debye temperatures are theoretically identical.[26] For quartz, Barron and colleagues[27] gave a specific-heat value of $\Theta_D = 558\,\mathrm{K} \pm 6\,\mathrm{K}$. Because Θ_D varies as $G^{1/2}$ (G denotes effective shear modulus), we used the shear modulus to correct our Θ_D to zero temperature, and we obtained $\Theta_D^0\,(\mathrm{elastic}) = 567\,\mathrm{K}$, 1.5% higher than the Barron *et al.* value and 0.5% above their estimated upper bound, 564 K.

VI. CONCLUSIONS

(1) One can solve the problem of macroscopic vibration frequencies of a trigonal-crystal-symmetry sphere. And one can use this analysis to determine the complete elastic-stiffness tensor C_{ijkl} of a material such as quartz.

(2) Natural quartz shows a high internal friction, $Q^{-1} \approx 5 \times 10^{-4}$, but only slightly higher than shown by a poor grade of cultured quartz with moderate dislocation content.

(3) Natural and cultured quartz possess nearly identical C_{ijkl} when corrected about 1% for mass density, except for $C_{1123}\,(C_{14})$, which is about 5% lower in natural quartz. We hypothesize that this difference originates in the extra degree of freedom associated with C_{14}, the effective order parameter for the β–α phase transition.

(4) The elastic Debye temperature agrees with the specific-heat Debye temperature within about 1% when the elastic constants are adjusted to 0 K.

ACKNOWLEDGMENT

Part of this study was completed when P.R.H. was an Alexander von Humboldt research fellow at the University of Stuttgart. He gratefully acknowledges this support.

APPENDIX: COEFFICIENT MATRICES

Rectangular Cartesian Coordinates. The elements of the coefficient matrices are given by

$$K_{ij}^{11} = \int_V \left(C_{11} \frac{\partial \varphi_i^u}{\partial x} \frac{\partial \varphi_j^u}{\partial x} + C_{55} \frac{\partial \varphi_i^u}{\partial z} \frac{\partial \varphi_j^u}{\partial z} + C_{56} \frac{\partial \varphi_i^u}{\partial z} \frac{\partial \varphi_j^u}{\partial y} + C_{56} \frac{\partial \varphi_i^u}{\partial y} \frac{\partial \varphi_j^u}{\partial z} + C_{66} \frac{\partial \varphi_i^u}{\partial y} \frac{\partial \varphi_j^u}{\partial y} \right) dV, \tag{A1}$$

$$K_{ij}^{12} = K_{ji}^{21} = \int_V \left(C_{12} \frac{\partial \varphi_i^u}{\partial x} \frac{\partial \varphi_j^v}{\partial y} + C_{14} \frac{\partial \varphi_i^u}{\partial x} \frac{\partial \varphi_j^v}{\partial z} + C_{56} \frac{\partial \varphi_i^u}{\partial z} \frac{\partial \varphi_j^v}{\partial x} + C_{66} \frac{\partial \varphi_i^u}{\partial y} \frac{\partial \varphi_j^v}{\partial x} \right) dV, \tag{A2}$$

$$K_{ij}^{13} = K_{ji}^{31} = \int_V \left(C_{13} \frac{\partial \varphi_i^u}{\partial x} \frac{\partial \varphi_j^w}{\partial z} + C_{14} \frac{\partial \varphi_i^u}{\partial x} \frac{\partial \varphi_j^w}{\partial y} + C_{55} \frac{\partial \varphi_i^u}{\partial z} \frac{\partial \varphi_j^w}{\partial x} + C_{56} \frac{\partial \varphi_i^u}{\partial y} \frac{\partial \varphi_j^w}{\partial x} \right) dV, \tag{A3}$$

$$K_{ij}^{22} = \int_V \left(C_{22} \frac{\partial \varphi_i^v}{\partial y} \frac{\partial \varphi_j^v}{\partial y} + C_{24} \frac{\partial \varphi_i^v}{\partial y} \frac{\partial \varphi_j^v}{\partial z} + C_{24} \frac{\partial \varphi_i^v}{\partial z} \frac{\partial \varphi_j^v}{\partial y} + C_{44} \frac{\partial \varphi_i^v}{\partial z} \frac{\partial \varphi_j^v}{\partial z} + C_{66} \frac{\partial \varphi_i^v}{\partial x} \frac{\partial \varphi_j^v}{\partial x} \right) dV, \tag{A4}$$

$$K_{ij}^{23} = K_{ji}^{32} = \int_V \left(C_{23} \frac{\partial \varphi_i^v}{\partial y} \frac{\partial \varphi_j^w}{\partial z} + C_{24} \frac{\partial \varphi_i^v}{\partial y} \frac{\partial \varphi_j^w}{\partial y} + C_{44} \frac{\partial \varphi_i^v}{\partial z} \frac{\partial \varphi_j^w}{\partial y} + C_{56} \frac{\partial \varphi_i^v}{\partial x} \frac{\partial \varphi_j^w}{\partial x} \right) dV, \tag{A5}$$

$$K_{ij}^{33} = \int_V \left(C_{33} \frac{\partial \varphi_i^w}{\partial z} \frac{\partial \varphi_j^w}{\partial z} + C_{44} \frac{\partial \varphi_i^w}{\partial y} \frac{\partial \varphi_j^w}{\partial y} + C_{55} \frac{\partial \varphi_i^w}{\partial x} \frac{\partial \varphi_j^w}{\partial x} \right) dV. \tag{A6}$$

[1] P. Groth, *Physicalische Kristallographie und Einleitung in die Kristallographische Kenntniss der Wichtigsten Substanzen* (Oldenbourg, Berlin, 1895).
[2] W. Voigt, *Lehrbuch der Kristallphysik* (Teubner, Leipzig, 1910).
[3] E. Giebe and A. Scheibe, "On the series relationships of the natural elastic frequencies of quartz rods," Ann. Phys. (Leipzig) **9**, 93–175 (1931).
[4] W. Cady, *Piezoelectricity* (McGraw–Hill, New York, 1946).
[5] A. Ballato, "Elastic properties of crystalline quartz," in *Handbook of Elastic Properties of Solids, Liquids, Gases* (Academic, San Diego, 2001), Vol. II, pp. 257–275.
[6] H. Ledbetter, S. Kim, and M. Lei, "Elastic stiffnesses and internal friction of monocrystal quartz" (unpublished).
[7] M. Born and K. Huang, *Dynamic Theory of Crystal Lattices* (Oxford, University Press, London, 1954), especially Secs. 25 and 32.
[8] T. Ikeda, *Fundamentals of Piezoelectricity* (Oxford University Press, Oxford, 1990), 241 pp.
[9] J. Brice, "Crystals for quartz resonators," Rev. Mod. Phys. **57**, 105–146 (1985).
[10] B. James, "Determination of the elastic and dielectric properties of quartz," Ph.D. thesis, University of London, London, 1987.
[11] A. Migliori and J. Sarrao, *Resonant Ultasound Spectroscopy* (Wiley Interscience, New York, 1997).
[12] I. Ohno, S. Yamamoto, O. Anderson, and J. Noda, "Determination of elastic constants of trigonal crystals by the rectangular parallelepiped resonance method," J. Phys. Chem. Solids **12**, 1103–1108 (1986).
[13] F. Willis, R. Leisure, and T. Kanashiro, Phys. Rev. B **54**, 9077–9085 (1996).
[14] H. Tiersten, *Linear Piezoelectric Plate Vibrations* (Plenum, New York, 1969).
[15] L. Mierovitch, *Computational Methods in Structural Dynamics* (Sijthoff & Noordhoff, Alphen aan den Rijn, Netherlands, 1980).
[16] W. Visscher, A. Migliori, T. Bell, and R. Reinert, "On the normal modes of free vibration of inhomogeneous and anisotropic elastic objects," J. Acoust. Soc. Am. **90**, 2154–2162 (1991).
[17] I. Ohno, "Free vibration of a rectangular parallelepiped crystal and its

application to determination of elastic constants of orthorhombic crystals,'' J. Phys. Earth **24**, 355–379 (1976).
[18] P. Heyliger and W. Johnson, ''Traction-free vibrations of finite trigonal elastic cylinders,'' J. Acoust. Soc. Am. **113**, 1812–1825 (2003).
[19] H. Ledbetter, C. Fortunko, and P. Heyliger, ''Orthotropic elastic constants of a boron–aluminum fiber-reinforced composite: An acoustic-resonance-spectroscopy study,'' J. Appl. Phys. **78**, 1542–1546 (1995).
[20] H. Ledbetter and S. Kim, ''Elastic constants and internal frictions of premium cultured quartz'' (unpublished).
[21] H. Ogi, Y. Kawasaki, M. Hirao, and H. Ledbetter, ''Acoustic spectroscopy of lithium niobate: Elastic and piezoelectric coefficients,'' J. Appl. Phys. **92**, 2451–2456 (2002).
[22] W. Kingery, H. Bowan, and D. Uhlmann, *Introduction to Ceramics* (Wiley, New York, 1976), p. 74. Also in Ref. 4, p. 735, Fig. 163.
[23] L. Landau and E. Lifshitz, *Statistical Physics*, Course of Theoretical Physics, Volume 5 (Pergamon, Oxford, 1980), Secs. 142–146.
[24] H. Ledbetter and S. Kim, ''Monocrystal elastic constants and derived properties of the cubic and hexagonal elements,'' in *Handbook of Elastic Properties of Solids, Liquids, and Gases* (Academic, San Diego, 2001), Vol. II, pp. 97–106.
[25] O. Anderson, ''Determination and some uses of isotropic elastic constants of polycrystalline aggregates using single-crystal data,'' Phys. Acoust. **III-B**, 43–95 (1965).
[26] G. Leibfried and W. Ludwig, ''Theory of anharmonic effects in crystals,'' Solid State Phys. **12**, 275–444 (1961).
[27] T. Barron, J. Collins, T. Smith, and G. White, J. Phys. C **15**, 4311–4326 (1980).

The radiation Q factors obtained from the partial derivatives of the phase of the reflection coefficient of an elastic plate

O. Lenoir, J. M. Conoir, and J. L. Izbicki
Laboratoire d'Acoustique Ultrasonore et d'Electronique (LAUE), CNRS UMR 6068, Université du Havre, Place Robert Schuman, 76610 Le Havre, France

(Received 28 November 2001; accepted for publication 30 May 2003)

The phase gradient method is applied to study the partial derivatives of the phase of the reflection coefficient of a fluid-loaded elastic plate. We consider the derivatives with respect to the frequency f, the incidence angle θ, the phase velocities of the longitudinal and transverse waves propagating in the plate, c_L and c_T, respectively, and the phase velocity in the fluid c_F. The partial derivatives with respect to f, c_L, c_T, c_F are linked by a relation involving products of one of these variables with the corresponding partial derivative. At a resonance frequency, the product of frequency with the frequency phase derivative can be identified as a radiation quality factor. By analogy, the other products correspond to quality factors. It can be shown that the product assigned to the fluid phase velocity corresponds to an angular radiation quality factor. The products assigned to the longitudinal and transverse phase velocities are identified as longitudinal and transverse radiation quality factors. These quality factors are shown to be related to stored energies associated with either standing waves across the plate, guided waves, longitudinal waves or transverse waves. A reactive power balance between the plate and the fluid is also established.
[DOI: 10.1121/1.1593061]

PACS numbers: 43.20.Ks, 43.40.Dx, 43.40.Rj [ANN]

I. INTRODUCTION

The first aim of this paper is to show that the study of all partial derivatives of the phase of the reflection coefficient of an immersed elastic plate leads to the introduction of different radiation quality factors. We consider frequential, angular, longitudinal, and transverse quality factors. To our knowledge, only Auld[1] briefly talked about these types of factors for elastic media. They are different from those due to internal losses which are used for viscoelastic or porous media in acoustics[1,2] and geophysics.[3,4] The second aim of this paper deals with the physical meaning of each of the radiation quality factors. They are analyzed in terms of energies related to standing, propagative, longitudinal, and transverse waves. Fundamental relations linking these quality factors are also established.

In a previous paper,[5] the interest of studying the frequency and angular derivatives of the phase ϕ of the reflection coefficient $\underline{R}$ in order to characterize its frequency and angular resonances (location and half-width) was shown. These resonances respectively correspond to the frequency poles of the reflection coefficient, denoted as $\underline{f}_P = f_{\text{res}} - j(\Gamma/2)$ and to its angular poles denoted as $\underline{y}_P = y_{\text{res}} + j(\gamma/2)$ ($\underline{y}_P = \sin \underline{\theta}_P$ where $\underline{\theta}_P$ is the complex reemission angle). Indeed, the frequency phase derivative $(\partial\phi/\partial f)$ plotted versus f in the vicinity of f_{Res}, at a given incidence angle θ, exhibits a Breit-Wigner shaped curve. Its maximum is located at f_{Res} and its maximum amplitude is equal to $(2/\Gamma)$ ($\Gamma/2$: frequency resonance half-width). In the same way, at a given frequency f, the angular phase derivative $(\partial\phi/\partial y)$ ($y = \sin\theta$) plotted versus y also exhibits a Breit-Wigner shaped curve. Its minimum is located at y_{Res} and its minimum amplitude is equal to $(2/\gamma)$ ($\gamma/2$: angular resonance half-width). These phase derivative properties make it possible to avoid heavy calculations of complex frequency and angular roots. The results obtained are accurate except in the case where the pole imaginary parts are too large as compared to their real parts.

A frequency resonance may also be characterized by a frequency Q_x factor defined as (f_{Res}/Γ), which is thus connected to the inverse of the temporal damping (Γ) of the reflected wave. According to the property of the function $(\partial\phi/\partial f)$ recalled above and the definition of Q_x, it is shown in this paper that the study of the function $f(\partial\phi/\partial f)$ yields the determination of the Q_x value.

Similarly, an angular resonance may be characterized by a angular Q_y factor defined as (y_{Res}/γ). It is shown that the study of the function $-y(\partial\phi/\partial y)$ yields the determination of the Q_y value. In Ref. 5, it was shown that γ is related to the spatial damping of a leaky Lamb mode and the following relation was established:

$$\frac{\gamma}{\Gamma} = \frac{\partial y}{\partial f}(f_{\text{Res}}).$$

According to the definitions of Q_x and Q_y, the previous relation obviously proves that a relation between these quality factors exists. This relation will be developed in the following sections.

Even if this paper is only concerned with theoretical results, we can indicate that experimental methods based on the phase properties have been successfully carried out in order to obtain reliable values for frequency resonance[6] (Γ) and angular resonance[7] (γ) widths. Therefore, it follows that obtaining accurate values for frequency and angular quality

factors experimentally seems possible *a priori*. However, comparing our theoretical work to experimental data is not the subject of the present paper.

The phase of the reflection coefficient of an immersed elastic plate also depends on other parameters such as phase velocities and its density. Similar to the case of cylindrical shells first studied by Conoir,[8,9] it is possible in the case of plates, as developed in Sec. II, to establish a fundamental relation that links the frequency phase derivative with the partial phase derivatives with respect to the phase velocities:

$$f\frac{\partial\phi}{\partial f}+c_L\frac{\partial\phi}{\partial c_L}+c_T\frac{\partial\phi}{\partial c_T}+c_F\frac{\partial\phi}{\partial c_F}=0,$$

which is valid whatever the frequency or the incidence angle. This relation may appear to be rather formal at first sight. Indeed, the physical meaning of functions such as $c_{L,T,F}(\partial\phi/\partial c_{L,T,F})$ is not immediately apparent. It is the main goal of this paper to prove that this relation can be interpreted in terms of radiation quality factors and also leads to a new form of energy balance.

Because a quality factor is by nature related to a resonant phenomenon, we first have to verify the resonant feature of $c_{L,T,F}(\partial\phi/\partial c_{L,T,F})$. This resonant feature is easily shown because the frequency plots of these functions exhibit resonant Breit-Wigner shaped curves in the vicinity of a resonance frequency. As $f(\partial\phi/\partial f)$ is proportional to the frequency quality factor Q_x at the resonance frequency, then $c_{L,T,F}(\partial\phi/\partial c_{L,T,F})$ may also be related to quality factors. The terms involving the longitudinal phase velocity c_L and the transverse velocity c_T are naturally qualified as longitudinal or transverse quality factors, denoted as Q_L and Q_T, respectively.

In contrast, it is more difficult to relate $c_F(\partial\phi/\partial c_F)$ to a quality factor. Nevertheless, it is shown that the term depending on the phase velocity in the fluid, c_F, can be decomposed into the sum of two terms. The prevailing one of these two terms is associated with the angular phase derivative and the residual one is related to the ratio of the plate and fluid densities. In close vicinity of the resonance frequency, the following equality is established:

$$c_F\frac{\partial\phi}{\partial c_F}=-y\frac{\partial\phi}{\partial y}.$$

Thus, the term $c_F(\partial\phi/\partial c_F)$ can be related to the angular quality factor Q_y.

To conclude the first part of this paper, it is shown that the relation

$$f\frac{\partial\phi}{\partial f}+c_L\frac{\partial\phi}{\partial c_L}+c_T\frac{\partial\phi}{\partial c_T}+c_F\frac{\partial\phi}{\partial c_F}=0,$$

written at a resonance frequency, leads to the following balance between the different Q factors:

$$Q_x+Q_y=Q_L+Q_T.$$

The physical interpretation of such a balance is developed in the second part.

The second part of this paper is devoted to the obtaining of the energy expressions of the different radiation quality factors. The quality factors defined in the first part are of the type $(\Delta f/f_{\mathrm{Res}})$ (Δf: band-width), which is commonly used to describe frequency resonant phenomena in mechanics (spring mass system) or electronics (RLC circuit) for instance. We now use the energy definition of a quality factor that states that it is proportional to the ratio of a mean stored energy to a mean dissipated power. The energy definitions allow us to give the physical meaning of each of the radiation quality factors clearly.

It is thus shown that the longitudinal and transverse $Q_{L,T}$ factors introduced via the Phase Gradient Method are respectively related to mean energies carried by the longitudinal and transverse waves which propagate in the elastic plate. In addition, it is demonstrated that the frequency Q_x factor is related to the mean energy associated to standing waves through the thickness of the plate and that the angular Q_y factor is related to the mean energy associated to guided waves propagating in the plate. It is finally shown that the previous relation between the Q factors describes the process in which the mean energy absorbed by the plate, at a resonance frequency, is re-radiated towards the surrounding fluid.

In Sec. II A, the relations linking the different partial derivatives of the phase are established. Numerical results are presented and discussed in Sec. II B. In Sec. III, each partial derivative of the phase is assigned to an energy quality factor whose physical meaning is given.

II. RELATIONS BETWEEN THE PARTIAL DERIVATIVES OF THE PHASE

This section deals with the establishment of the fundamental relations linking the partial derivatives of the phase ϕ of the reflection coefficient $\underline{R}$ of an elastic plate immersed in water. As a preliminary remark, we would indicate that the first part of this section gives the proof of a series of analytical expressions based upon the mathematical properties of the phase ϕ. The physical meaning of the main equations will be developed in Sec. III.

A. Analytical relations linking the phase partial derivatives

According to Ref. 10, a factorized expression of the reflection coefficient of the immersed plate can be written as

$$\underline{R}=\frac{C_AC_S-\tau^2}{(C_A+j\tau)(C_S-j\tau)}. \tag{1}$$

The roots of $C_{A,S}$ correspond to the antisymmetric and symmetric normal modes of the plate in vacuum and τ is the ratio of the acoustic impedances in the fluid and in the plate. $C_{S,A}$ depend on the frequency f, the plate thickness d, $y=\sin\theta$ and on the phase velocities $c_{L,T,F}$. The acoustic impedance ratio is independent of f and c_T, but depends on the plate density ρ_S and on the fluid density ρ_F. This gives:[5,10]

$$C_A(f,y,c_L,c_T,c_F)$$
$$=\left(1-2\left(\frac{c_T}{c_F}y\right)^2\right)^2 \times\tan\left(\frac{\pi f d}{c_L}\sqrt{1-\left(\frac{c_L}{c_F}y\right)^2}\right), +4\frac{c_T^3y^2}{c_Lc_F^2}\sqrt{1-\left(\frac{c_L}{c_F}y\right)^2}\sqrt{1-\left(\frac{c_T}{c_F}y\right)^2} \times\tan\left(\frac{\pi f d}{c_T}\sqrt{1-\left(\frac{c_T}{c_F}y\right)^2}\right), \tag{2a}$$

$$C_S(f,y,c_L,c_T,c_F)$$
$$=\left(1-2\left(\frac{c_T}{c_F}y\right)^2\right)^2 \operatorname{cotan}\left(\frac{\pi f d}{c_L}\sqrt{1-\left(\frac{c_L}{c_F}y\right)^2}\right), +4\frac{c_T^3y^2}{c_Lc_F^2}\sqrt{1-\left(\frac{c_L}{c_F}y\right)^2}\sqrt{1-\left(\frac{c_T}{c_F}y\right)^2} \times\operatorname{cotan}\left(\frac{\pi f d}{c_T}\sqrt{1-\left(\frac{c_T}{c_F}y\right)^2}\right), \tag{2b}$$

$$\tau(y,c_L,c_F,\rho_F,\rho_S)=\frac{\rho_F c_F}{\rho_S c_L}\frac{\sqrt{1-\left(\frac{c_L}{c_F}y\right)^2}}{\sqrt{1-y^2}}. \tag{2c}$$

The phase ϕ of the reflection coefficient can be written as the sum of two phases, ϕ_A due to antisymmetric modes and ϕ_S due to symmetric modes

$$\phi=\phi_S+\phi_A=\arctan\left(\frac{\tau}{C_S}\right)-\arctan\left(\frac{\tau}{C_A}\right). \tag{3}$$

The phase differentials have the following general expressions

$$d\phi_{S,A}=\pm\frac{C_{S,A}\,d\tau-\tau\,dC_{S,A}}{C_{S,A}^2+\tau^2}. \tag{4}$$

For the set of parameters (f,c_L,c_T,c_F), $C_{S,A}$ are homogeneous functions of zero order: they satisfy the following property:

$$C_{S,A}(\lambda f,\lambda c_L,\lambda c_T,\lambda c_F)=C_{S,A}(f,c_L,c_T,c_F)\quad \forall\lambda\neq 0. \tag{5}$$

Such functions satisfy Euler's equation:

$$f\frac{\partial C_{S,A}}{\partial f}+c_L\frac{\partial C_{S,A}}{\partial c_L}+c_T\frac{\partial C_{S,A}}{\partial c_T}+c_F\frac{\partial C_{S,A}}{\partial c_F}=0. \tag{6a}$$

If we consider $C_{S,A}$ as functions of the set of parameters (c_F,y), it is verified that they are still homogeneous functions of zero order, and consequently, we can also write

$$c_F\frac{\partial C_{S,A}}{\partial c_F}+y\frac{\partial C_{S,A}}{\partial y}=0. \tag{6b}$$

In addition, the acoustic impedance ratio τ considered as a function of the set of parameters (c_L,c_F) is also an homogeneous function of zero order. So, we can write

$$c_L\frac{\partial\tau}{\partial c_L}+c_F\frac{\partial\tau}{\partial c_F}=0. \tag{7}$$

Due to the previous relations, the partial derivatives of the phase with respect to f, c_L, c_T, c_F and y can be written

$$f\frac{\partial\phi_{S,A}}{\partial f}=\mp\left(\frac{\tau}{C_{S,A}^2+\tau^2}\right)f\frac{\partial C_{S,A}}{\partial f}, \tag{8a}$$

$$c_L\frac{\partial\phi_{S,A}}{\partial c_L}=\mp\left(\frac{\tau}{C_{S,A}^2+\tau^2}\right)c_L\frac{\partial C_{S,A}}{\partial c_L} \pm\left(\frac{C_{S,A}}{C_{S,A}^2+\tau^2}\right)c_L\frac{\partial\tau}{\partial c_L}, \tag{8b}$$

$$c_T\frac{\partial\phi_{S,A}}{\partial c_T}=\mp\left(\frac{\tau}{C_{S,A}^2+\tau^2}\right)c_T\frac{\partial C_{S,A}}{\partial c_T}, \tag{8c}$$

$$c_F\frac{\partial\phi_{S,A}}{\partial c_F}=\mp\left(\frac{\tau}{C_{S,A}^2+\tau^2}\right)c_F\frac{\partial C_{S,A}}{\partial c_F} \pm\left(\frac{C_{S,A}}{C_{S,A}^2+\tau^2}\right)c_F\frac{\partial\tau}{\partial c_F}, \tag{8d}$$

$$y\frac{\partial\phi_{S,A}}{\partial y}=\mp\left(\frac{\tau}{C_{S,A}^2+\tau^2}\right)y\frac{\partial C_{S,A}}{\partial y}\pm\left(\frac{C_{S,A}}{C_{S,A}^2+\tau^2}\right)y\frac{\partial\tau}{\partial y}. \tag{8e}$$

According to Eqs. (6a) and (7), the sum of Eqs. (8a), (8b), (8c), and (8d) yields

$$f\frac{\partial\phi}{\partial f}+c_L\frac{\partial\phi}{\partial c_L}+c_T\frac{\partial\phi}{\partial c_T}+c_F\frac{\partial\phi}{\partial c_F}=0. \tag{9}$$

This equation is valid for the phases ϕ_A, ϕ_S and ϕ, whatever the values of f and y. This fundamental relation (a similar one was first established by J. M. Conoir for cylindrical shells[8]) is the basis of what has been called the Phase Gradient Method (PGM). Of course, we are particularly interested in this relation in the vicinity of a resonance.

If we denote as $f_{C_{S,A}}$ the roots of $C_{S,A}$, we can also write, according to Eqs. (6b), (8d), and (8e), the following relation:

$$c_F\frac{\partial\phi_{S,A}}{\partial c_F}+y\frac{\partial\phi_{S,A}}{\partial y}=0,\quad \text{at } f=f_{C_{S,A}}. \tag{10}$$

This relation holds true for the resonance frequencies which are found to be very close to $f_{C_{S,A}}$ for elastic plates.[5,10]

It is also useful to consider as variables the trace velocities of the reflected wave front on the axis parallel to the plate, $0x$, and on the axis normal to the plate, $0z$. They are, respectively, defined by

$$v_x=\frac{c_F}{\sin\theta}=\frac{c_F}{y}, \tag{11a}$$

and

$$v_z = \frac{c_F}{\cos\theta} = \frac{c_F}{\sqrt{1-y^2}}. \tag{11b}$$

According to the Cremer coincidence rule, at resonance frequencies the trace velocity v_x corresponds to the phase velocity of the Lamb wave propagating in the plate. $C_{S,A}$ and τ can be expressed as functions of the new set of parameters (f, c_L, c_T, v_x, v_z), replacing (c_F/y) by v_x and $(c_F/\sqrt{1-y^2})$ by v_z in Eqs. (2a), (2b), and (2c). It can be noticed that $C_{S,A}$ do not depend explicitly on v_z. With this set of variables, we can establish the relation

$$f\frac{\partial\phi}{\partial f} + c_L\frac{\partial\phi}{\partial c_L} + c_T\frac{\partial\phi}{\partial c_T} + v_x\frac{\partial\phi}{\partial v_x} + v_z\frac{\partial\phi}{\partial v_z} = 0. \tag{12}$$

Comparing this last equation to Eq. (9), we obtain the following relation

$$c_F\frac{\partial\phi}{\partial c_F} = v_x\frac{\partial\phi}{\partial v_x} + v_z\frac{\partial\phi}{\partial v_z}. \tag{13}$$

We can also use the ratio of the densities in the solid and in the fluid as a variable, denoted as $\rho = (\rho_S/\rho_F)$. Considering the impedance ratio τ as a function of the variables (ρ, v_z), it verifies the property of a homogeneous function of zero order, so we can write

$$v_z\frac{\partial\tau}{\partial v_z} + \rho\frac{\partial\tau}{\partial\rho} = 0 \tag{14}$$

Using Eq. (1) and Eq. (14), we can establish the following relation

$$v_z\frac{\partial\phi}{\partial v_z} = \mathrm{Im}(\underline{R}), \tag{15}$$

where Im stands for the imaginary part. Finally, making use of Eqs. (12) and (15) we can rewrite Eq. (9) in the following form

$$f\frac{\partial\phi}{\partial f} + c_L\frac{\partial\phi}{\partial c_L} + c_T\frac{\partial\phi}{\partial c_T} + v_x\frac{\partial\phi}{\partial v_x} + \mathrm{Im}(\underline{R}) = 0. \tag{16}$$

The interest of Eq. (16) will be developed in Sec. III.

The part which follows deals with numerical results; it will be shown that all the partial derivatives of the phase, except the one with respect to v_Z, present resonant forms of the Breit-Wigner type as well as the frequency and angular derivatives.[5] The resonant aspect of the scattering by a plate is thus proved for nearly all the variables involved in the problem.

B. Numerical results and discussion

In this sub-section, we present numerical results dealing with an aluminum plate immersed in water. The reference values of the parameters used for the aluminum plate are as follows: thickness $d = 5$ mm, density $\rho_{S0} = 2800\,\mathrm{kg/m^3}$, bulk longitudinal phase velocity $c_{L0} = 6380$ m/s, bulk transverse phase velocity $c_{T0} = 3100$ m/s. The parameter values used for water are: density $\rho_{F0} = 1000\,\mathrm{kg/m^3}$, phase velocity $c_{F0} = 1470$ m/s. The readers can find the expressions of the exact frequency and angular derivatives in Ref. 5. The other partial derivatives of the phase are given in Appendix A.

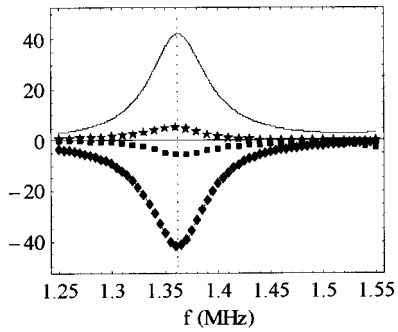


FIG. 1. Plots of $f(\partial\varphi/\partial f)$ (solid line), $c_L(\partial\varphi/\partial c_L)$ (triangle line), $c_T(\partial\varphi/\partial c_T)$ (box line), and $c_F(\partial\varphi/\partial c_F)$ (star line) in the vicinity of mode A_3 at $\theta = 5°$.

In Fig. 1, we have plotted the different functions which appear in Eq. (9). The computation has been performed in the vicinity of the resonance frequency $f_{\mathrm{Res}} = 1.3623$ MHz, corresponding to mode A_3 at the incidence angle $\theta_0 = 5°$. The vertical line indicates the position of the resonance frequency f_{Res}. The plot of the exact function $f(\partial\phi/\partial f)$ exhibits a so-called resonant Breit-Wigner form. Indeed, it has been shown[5] that an approximate expression of this function can be written

$$f\frac{\partial\phi_{\mathrm{app}}}{\partial f} = \frac{\dfrac{\Gamma}{f_{\mathrm{Res}}}}{(f-f_{\mathrm{Res}})^2 + \left(\dfrac{\Gamma}{2}\right)^2}. \tag{17}$$

At $f = f_{\mathrm{Res}}$, we obtain:

$$f\frac{\partial\phi_{\mathrm{app}}}{\partial f} = 2\frac{f_{\mathrm{Res}}}{\Gamma} = 2Q_x. \tag{18}$$

We can also observe, in Fig. 1, that the plots of all the other partial derivatives of the phase exhibit such a resonant form. All their extrema are located in the close vicinity of the resonance frequency f_{Res}. We can verify that for all Lamb modes whatever the incidence angle, the derivatives with respect to the bulk longitudinal and transverse phase velocities are always negative. The one depending on the phase velocity in the fluid is generally positive. In addition, it is important to note that the width of each resonant form is Γ. Therefore, each function of the type $u(\partial\phi/\partial u)$, where $u = c_{L,T,F}$, may be approximated using

$$u\frac{\partial\phi_{\mathrm{app}}}{\partial u} = \epsilon_u\frac{u_0\left(\dfrac{2}{\Gamma_u}\right)\left(\dfrac{\Gamma}{2}\right)^2}{(f-f_{\mathrm{Res}})^2 + \left(\dfrac{\Gamma}{2}\right)^2}. \tag{19}$$

The subscript 0 indicates that the variable u is fixed to its reference value. The function being dimensionless, Γ_u has the same dimension as the variable u. The extremum is given by $(2u_0/\Gamma_u)$. ϵ_u indicates the sign ($\epsilon_{c_{L,T}} = -1$, $\epsilon_{c_F} = 1$).

In Fig. 2, we compare the frequency plots of the exact function $c_L(\partial\phi/\partial c_L)$ (solid line) and its approximate expression $c_L(\partial\phi_{\mathrm{app}}/\partial c_L)$ (dotted line), for the same conditions as

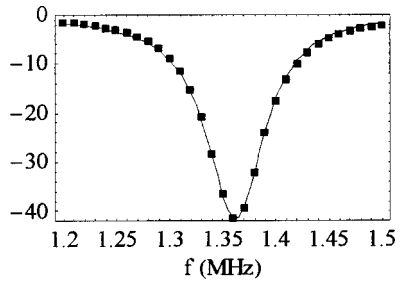


FIG. 2. Plots of $c_L(\partial\varphi/\partial c_L)$ (solid line) and $c_L(\partial\varphi_{\text{app}}/\partial c_L)$ (box line) in the vicinity of mode A_3 at $\theta=5°$.

in Fig. 1 in order to demonstrate the validity of the model. It is reliable for all functions $u(\partial\phi/\partial u)$, provided that the resonance is isolated as in Refs. 10–12.

In the same way as in Eq. (18), we assume that the values reached by the functions $u(\partial\phi/\partial u)(u=c_{L,T,F})$, at the resonance frequency f_{res}, are proportional to the quality factors denoted as $Q_{L,T,F}$. Therefore, we can write

$$\left(c_{L,T,F}\frac{\partial\phi}{\partial c_{L,T,F}}\right)_{f_{\text{Res}}}=2\epsilon_{L,T,F}Q_{L,T,F}, \tag{20}$$

where $Q_{L,T,F}\geqslant 0$. The terms $Q_{L,T}$ are called the longitudinal and transverse quality factors.

We now show that Q_F is equal to the angular Q_y factor. Equation (19) is also valid when the variable u is $y=\sin\theta$. In this latter case, $\epsilon_y=-1$ and $\Gamma_u=\gamma$. At $f=f_{\text{Res}}$, $y_0=\sin\theta_0$ is generally very close to an angular resonance y_{res}, so we can write

$$-y\frac{\partial\phi}{\partial y}=2\frac{y_{\text{res}}}{\gamma}=2Q_y. \tag{21}$$

In Fig. 3, we have plotted $c_F(\partial\phi/\partial c_F)$ (solid line) and $-y(\partial\phi/\partial y)$ (dotted line) as functions of the frequency. The solid vertical line indicates the value of the resonance frequency $f_{\text{Res}}=1.3623$ MHz (very close to $f_{C_A}=1.3619$ MHz for which C_A is null at θ_0). We observe that the two curves are similar but the maximum of $-y(\partial\phi/\partial y)$ is located at the resonance frequency f_{Res} while that of $c_F(\partial\phi/\partial c_F)$ is slightly shifted towards the low frequencies. As indicated by Eq. (10), the two curves intersect at f_{C_A}. At $f_{\text{res}}\cong f_{C_A}$, $c_F(\partial\phi/\partial c_F)=2Q_F$ and $-y(\partial\phi/\partial y)=2Q_y$. Therefore, we can identify the quality factor Q_F to Q_y. According to Eq. (20), Eq. (9) can be written at a resonance frequency as

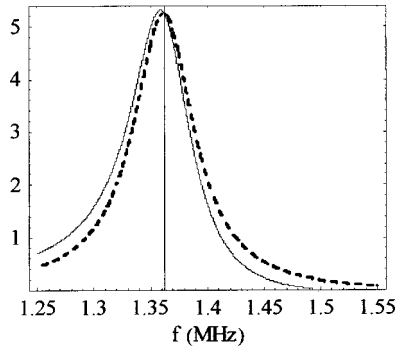


FIG. 3. Plots of $c_F(\partial\varphi/\partial c_F)$ (solid line) and $-y(\partial\varphi/\partial y)$ (dashed line) in the vicinity of mode A_3 at $\theta=5°$.

TABLE I. Comparison of the Q factors of the A_1 and A_3 modes, at $\theta=5°$.

	Q_x	Q_L	Q_T	Q_F	Q_y
A_3 mode	42.23	41.74	5.74	5.25	5.25
A_1 mode	255.6	6.3	282.6	33.3	33.3

$$Q_x+Q_y=Q_L+Q_T. \tag{22}$$

The physical meaning of this relation will appear more clearly in the following developments. However, we may make some comments about it right now. In Table I, we have collected the values of the different Q factors associated to the A_1 and A_3 modes at 5°. The difference between the two modes is that the A_3 mode at low incidence angles is longitudinal (its resonance frequency is close to a multiple of $(c_L/2d\cos\theta_L)$ where θ_L is the longitudinal refraction angle), whereas the A_1 mode is transverse (its resonance frequency is close to a multiple of $(c_T/2d\cos\theta_T)$ where θ_T is the transverse refraction angle). We notice, in Table I, that the Q_L factor value of mode A_3 is higher than that of mode A_1, and that the Q_T factor value of mode A_3 is considerably lower than that of mode A_1. So, the $Q_{L,T}$ factors seem to give the prevailing nature (longitudinal or transverse) of the associated Lamb mode quantitatively. We also notice that the Q_T value can be greater than that of the Q_x one, as for the A_1 mode. The comparison of the Q_x and Q_y values for the two modes indicates that the A_1 mode is both temporally and spatially less attenuated than the A_3 mode. At this point these conclusions may appear rather empirical. It is the reason why in the following section we develop the analytical relations existing between the Q factors introduced by means of the partial phase derivative properties and the mean energies absorbed and dissipated by the plate.

Let us now use the energy definition of a quality factor[1]

$$Q=\frac{\omega\langle\text{Stored Energy}\rangle}{\langle\text{Dissipated Power}\rangle}, \tag{23}$$

where the brackets $\langle\ \rangle$ indicate that we are dealing with the mean values of the considered quantities. In the following section, the factors $Q_{L,T}$ are shown to be related to mean energies of the longitudinal and transverse waves propagating in the plate. It will also be shown that the frequency quality factor is related to the mean energy of the standing waves across the plate and the angular quality factor is related to the mean energy of the guided waves propagating along the plate.

III. RELATIONS BETWEEN THE *Q* FACTORS AND MEAN STORED ENERGIES

We consider an elementary volume $dV=dx\,d\eta\,dz$ in the plate as shown in Fig. 4. It is excited by a plane harmonic wave incident at an incidence angle θ on its upper interface, at $x=0$ and $z=-(d/2)$. The incident wave may be represented in the form of the scalar potential

$$\xi_{\text{inc}}(x,z,t)=e^{jk_{F_z}[z+(d/2)]}e^{j(k_x x-\omega t)}. \tag{24}$$

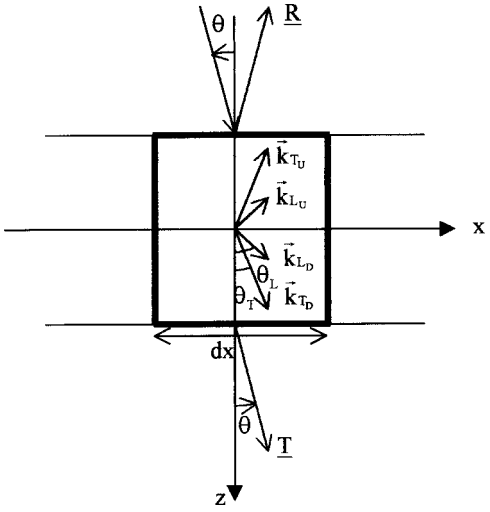


FIG. 4. Scheme of the problem.

ω is the angular frequency, k_{F_z} is the z-component of the incident wave vector, k_X is the x-component of the wave vectors of the incident, reflected and transmitted waves in the upper and lower fluids and of the wave vectors $\mathbf{k}_{L_{U,D}}$ and $\mathbf{k}_{T_{U,D}}$ of the longitudinal waves and transverse waves, which propagate up and down in the elastic plate. According to Snell's law, we can write

$$k_x = k_F \sin\theta = k_L \sin\theta_L = k_T \sin\theta_T, \tag{25}$$

where $k_{L,T,F}$ are the wave numbers related to the longitudinal and transverse waves in the plate and to the waves in the fluid. The potential in the upper fluid medium may be expressed as

$$\xi_{F_{\text{upp}}}(x,z,t) = (e^{jk_{F_z}[z+(d/2)]} + \underline{R}e^{-jk_{F_z}[z+(d/2)]})e^{j(k_x x-\omega t)}, \tag{26}$$

where $\underline{R}$ is the plate reflection coefficient. The scalar potential assigned to the longitudinal waves in the elastic plate is written

$$\varphi_L(x,z,t) = (A_L \cos k_{L_z} z + B_L \sin k_{L_z} z)e^{j(k_x x-\omega t)}, \tag{27}$$

where $k_{L_z} = k_L \cos\theta_L$. The vector potential $\boldsymbol{\psi}_T$ is assumed to be $(0,\psi_T,0)$. Its nonzero component is expressed as

$$\psi_T(x,z,t) = (A_T \cos k_{T_z} z + B_T \sin k_{T_z} z)e^{j(k_x x-\omega t)}, \tag{28}$$

where $k_{T_z} = k_T \cos\theta_T$. The potential in the lower fluid medium may be expressed as

$$\xi_{F_{\text{low}}}(x,z,t) = \underline{T}e^{jk_{F_z}[z-(d/2)]}e^{j(k_x x-\omega t)}, \tag{29}$$

where $\underline{T}$ is the plate transmission coefficient. The displacement velocity vector $\mathbf{v}$ can be obtained by

$$\mathbf{v} = \mathbf{v}_L + \mathbf{v}_T = (-j\omega)(\nabla\varphi_L + \nabla\times\boldsymbol{\psi}_T). \tag{30}$$

The global strain and stress tensor components are given by $\epsilon_{ij} = 1/2((\partial u_i/\partial x_j) + (\partial u_j/\partial x_i))$ and $\sigma_{ij} = \lambda\delta_{ij}\epsilon_{kk} + 2\mu\epsilon_{ij}$, where λ and μ are the Lamé coefficients of the plate. Using the classical boundary conditions on both plate surfaces, it is possible to show that the coefficients $A_{L,T}$ and $B_{L,T}$ appearing in the definitions of the longitudinal and transverse potentials can be expressed as

$$A_L = j\frac{c_L\cos\theta}{c_F\cos(\theta_L)}\frac{\cos(2\theta_T)}{\sin\left(k_{L_z}\frac{d}{2}\right)}\left(\frac{j\tau}{C_S - j\tau}\right), \tag{31a}$$

$$B_L = -j\frac{c_L\cos\theta}{c_F\cos(\theta_L)}\frac{\cos(2\theta_T)}{\cos\left(k_{L_z}\frac{d}{2}\right)}\left(\frac{j\tau}{C_A + j\tau}\right), \tag{31b}$$

$$A_T = -\frac{2c_T\sin(\theta_T)\cos\theta}{c_F\cos\left(k_{T_z}\frac{d}{2}\right)}\left(\frac{j\tau}{C_A + j\tau}\right), \tag{31c}$$

$$B_T = -\frac{2c_T\sin(\theta_T)\cos\theta}{c_F\sin\left(k_{T_z}\frac{d}{2}\right)}\left(\frac{j\tau}{C_S - j\tau}\right). \tag{31d}$$

We now establish the expressions of the mean values of the total and partial kinetic and strain energies stored in the volume dV. We have assumed the invariance of the problem according to the axis normal to the plane $x0z$, so $d\eta$ can be arbitrarily fixed at unity. The length dx is assumed to be very small compared to the wavelength in order to consider all the physical quantities constant between x and $x+dx$, in particular the kinetic and strain densities denoted as e_K and e_S. So, the expressions of the total kinetic and strain energies E_K and E_S are simply obtained by integrating the corresponding energy densities according to the variable z. The kinetic energy density mean value may be obtained by

$$\langle e_K\rangle = \tfrac{1}{4}\rho_S \operatorname{Re}(\mathbf{v}\cdot\mathbf{v}^*), \tag{32a}$$

where $\mathbf{v}$ is the total displacement velocity vector in the plate, the asterisk indicates the complex conjugate and Re denotes the real part. The strain energy density mean value is defined by

$$\langle e_S\rangle = \tfrac{1}{4}\operatorname{Re}(\sigma:\boldsymbol{\epsilon}^*), \tag{32b}$$

where σ and $\boldsymbol{\epsilon}$ are the stress and strain tensors respectively and: stands for the double dot product. The total mean values of the kinetic and strain energies are obtained by

$$\begin{aligned}\langle E_K\rangle &= dx\int_{z=-d/2}^{z=d/2}\langle e_K\rangle dz \\ &= \frac{1}{4}\rho_S\,dx\int_{z=-d/2}^{z=d/2}\operatorname{Re}(\mathbf{v}\cdot\mathbf{v}^*)dz \\ &= \frac{1}{4}\rho_S\,dx\int_{z=-d/2}^{z=d/2}\operatorname{Re}(v_x\cdot v_x^* + v_z\cdot v_z^*)dz\end{aligned} \tag{33a}$$

and

$$\langle E_S\rangle = dx\int_{z=-d/2}^{z=d/2}\langle e_S\rangle dz$$

$$= \frac{1}{4}dx\int_{z=-d/2}^{z=d/2}\mathrm{Re}(\sigma_{xx}\epsilon_{xx}^* + 2\sigma_{xz}\epsilon_{xz}^* + \sigma_{zz}\epsilon_{zz}^*)dz. \tag{33b}$$

In order to establish the relations between the Q factors introduced before and the mean stored energies, we also define partial kinetic and strain energies assigned to the longitudinal and transverse waves propagating in the plate.

The longitudinal and transverse kinetic energies may be obtained by

$$\langle E_{KL,T}\rangle = \frac{1}{4}\rho_S\, dx\int_{z=-d/2}^{z=d/2}(\mathbf{v}_{L,T}\cdot\mathbf{v}_{L,T}^*)dz$$

$$= \frac{1}{4}\rho_S\, dx\int_{z=-d/2}^{z=d/2}(v_{L,T_x}\cdot v_{L,T_x}^* + v_{L,T_z}\cdot v_{L,T_z}^*)dz. \tag{34a}$$

These partial energies may also be decomposed into two terms, the one corresponding to a displacement along the x-direction and the other one to a displacement along the z-direction. We choose the following notations for these terms

$$\langle E_{K_{L,T_{x,z}}}\rangle = \frac{1}{4}\rho_S\, dx\int_{z=-d/2}^{z=d/2}(v_{L,T_{x,z}}\cdot v_{L,T_{x,z}}^*)dz. \tag{34b}$$

Below, we only give the detail of the calculation of $\langle E_{K_{L_z}}\rangle$. We assume in the following that both the frequency and the components of the wave vectors are real. The term $\langle E_{K_{L_z}}\rangle$ may be obtained by

$$\langle E_{K_{L_z}}\rangle = \frac{1}{4}\rho_S\, dx\int_{z=-d/2}^{z=d/2}(v_{L_z}\cdot v_{L_z}^*)dz.$$

After integration, we obtain

$$\langle E_{K_{L_z}}\rangle = \frac{\rho_S\, dx\omega^4 d}{8}\left(\frac{\cos\theta_L}{c_L}\right)^2 \times\Bigg(\underbrace{(|A_L|^2+|B_L|^2)}_{\text{Term 1}} - \underbrace{(|A_L|^2-|B_L|^2)\frac{\sin(k_{L_z}d)}{k_{L_z}d}}_{\text{Term 2}}\Bigg). \tag{35}$$

It is possible to show numerically in Eq. (35) that Term 1 prevails when the incidence angle is below the first critical angle defined as $\theta_C^L = \mathrm{Arcsin}(c_L/c_F) = 12.32°$ and that Term 2 prevails when $\theta > \theta_C^L$. So, when $\theta \leqslant \theta_C^L$, we may write the following approximate expressions for the different partial longitudinal kinetic stored energies

$$\langle E_{K_L}\rangle = \frac{\rho_S\, dx\omega^4 d}{8c_L^2}(|A_L|^2+|B_L|^2), \tag{36a}$$

$$\langle E_{K_{L_x}}\rangle = \langle E_{K_L}\rangle\sin^2\theta_L, \tag{36b}$$

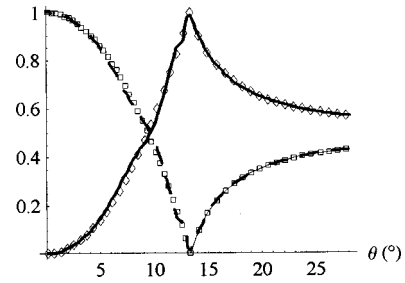


FIG. 5. Plots of the angular evolutions of the ratios ($\langle E_{K_{L_x}}\rangle/\langle E_{K_L}\rangle$) (exact: solid line, approximate: empty triangle line) and ($\langle E_{K_{L_z}}\rangle/\langle E_{K_L}\rangle$) (exact: dashed line, approximate: empty box line), following the dispersion curve of mode A_9.

$$\langle E_{K_{L_z}}\rangle = \langle E_{K_L}\rangle\cos^2\theta_L, \tag{36c}$$

where

$$|A_L|^2+|B_L|^2 = \tau^2\left(\frac{c_L\cos\theta}{c_F\cos\theta_L}\right)^2\cos^2(2\theta_T) \times\left(\frac{1}{\cos^2(k_{L_z}d/2)(C_A^2+\tau^2)} + \frac{1}{\sin^2(k_{L_z}d/2)(C_S^2+\tau^2)}\right). \tag{36d}$$

At a given depth in the plate, the longitudinal vibrations are due to the superposition of a longitudinal wave propagating towards the lower surface of the plate and of a longitudinal wave propagating towards its upper surface. They can be also described by the combination of a longitudinal wave which propagate in the x-direction, and of a longitudinal standing wave in the direction normal to the plate. Therefore, $\langle E_{K_{L_x}}\rangle$ may be interpreted as the mean kinetic energy of longitudinal guided waves propagating in the x-direction, parallel to the plate, and $\langle E_{K_{L_z}}\rangle$ as the mean kinetic energy of standing waves due to longitudinal waves propagating in the z and $-z$-directions, normal to the plate.

Numerical studies have proved that these approximate expressions are only valid at high frequencies, because Term 2 in Eq. (35) is then vanishing. In the case of a 5 mm thick aluminum plate, we can locate the beginning of the high frequency range at the cut-off frequency of mode S_3. In these conditions, we have chosen mode A_9 to show the numerical validity of the approximate expressions in Eqs. (36) at high frequencies.

In Fig. 5, following the dispersion curve of mode A_9, for an incidence angle ranging from 0° to the second critical angle $\theta_C^T = \mathrm{Arcsin}(c_T/c_F) = 28.3°$, we have plotted the exact (solid lines) and approximate (dotted lines) ratios $\langle E_{K_{L_x}}\rangle/\langle E_{K_L}\rangle$ and $\langle E_{K_{L_z}}\rangle/\langle E_{K_L}\rangle$ versus incidence angle. We can observe that the exact and approximate curves are nearly an exact match.

As for the partial longitudinal kinetic energies, it is possible to establish for the partial transverse kinetic stored energies the following approximate expressions

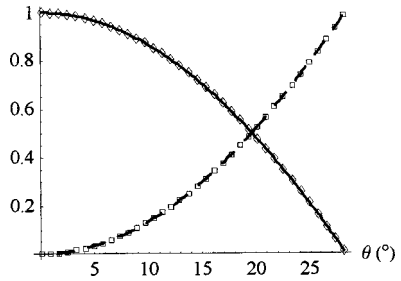


FIG. 6. Plots of the angular evolutions of the ratios ($\langle E_{K_{T_x}}\rangle/\langle E_{K_T}\rangle$) (exact: solid line, approximate: empty triangle line) and ($\langle E_{K_{T_z}}\rangle/\langle E_{K_T}\rangle$) (exact: dashed line, approximate: empty box line), following the dispersion curve of mode A_9.

$$\langle E_{K_T}\rangle = \frac{\rho_S\, dx\, \omega^4 d}{8c_T^2}(|A_T|^2+|B_T|^2), \tag{37a}$$

$$\langle E_{K_{T_x}}\rangle = \langle E_{K_T}\rangle \cos^2\theta_T, \tag{37b}$$

$$\langle E_{K_{T_z}}\rangle = \langle E_{K_T}\rangle \sin^2\theta_T, \tag{37c}$$

where

$$|A_T|^2+|B_T|^2 = \tau^2\left(\frac{2c_T\sin(\theta_T)\cos\theta}{c_F}\right)^2 \times\left(\frac{1}{\cos^2(k_{T_z}d/2)(C_A^2+\tau^2)} + \frac{1}{\sin^2(k_{T_z}d/2)(C_S^2+\tau^2)}\right). \tag{37d}$$

$\langle E_{K_{T_x}}\rangle$ can be interpreted as the mean kinetic energy of standing waves due to transverse waves propagating in the z and $-z$-directions, and $\langle E_{K_{T_z}}\rangle$ as the mean kinetic energy of transverse guided waves propagating in the x-direction. Equations (37a), (37b), and (37c) are valid for incidence angles lower than the second critical angle θ_C^T. Their numerical validity is shown in Fig. 6. We have plotted the exact (solid lines) and approximated (dotted lines) ratios $\langle E_{K_{T_x}}\rangle/\langle E_{K_T}\rangle$ and $\langle E_{K_{T_z}}\rangle/\langle E_{K_T}\rangle$ versus incidence angle. The comparison is as good as in the case of the longitudinal ratios.

Partial strain stored energies due to longitudinal and transverse waves may be defined in the same form as that of the total strain energy, but the strain and stress tensor components ϵ_{ij} and σ_{ij} are replaced by $\epsilon_{L,T_{ij}}$ and $\sigma_{L,T_{ij}}$. Decompositions of the partial strain energies are also possible. Their approximate expressions are given in Appendix B. Finally, by introducing mean total longitudinal and transverse stored energies denoted as $\langle E_{L,T}\rangle$, we can write

$$\langle E_{L,T}\rangle = \langle E_{K_{L,T}}\rangle + \langle E_{S_{L,T}}\rangle = 2\langle E_{K_{L,T}}\rangle. \tag{38}$$

It is worth noting that the sum $\langle E_L\rangle+\langle E_T\rangle$ is different from the mean value of the total stored energy, since the mean energy due to interactions between longitudinal and transverse waves is not taken into account, as in Ref. 13. The analytical expression of the total stored energy, which appears in the definition of the energy velocity,[1] can be found in Ref. 14.

The expressions of the stored energies of interest being established, we have yet to define the mean value of the power dissipated from the volume dV into the fluid denoted as $\langle P_d\rangle$. It may be obtained by

$$\langle P_d\rangle = \oint_{dS} \mathrm{Re}(\underline{\mathbf{P}})\cdot\mathbf{n}\, dS, \tag{39a}$$

where $\underline{\mathbf{P}}$ stands for the complex acoustic Poynting vector, $dS = dx\, d\eta$ is an elementary surface at the upper and lower interfaces between the plate and the fluids, and $\mathbf{n}$ is a unit normal on those surfaces. The complex Poynting vector is classically defined as

$$\underline{\mathbf{P}} = -\tfrac{1}{2}\boldsymbol{\sigma}\cdot\mathbf{v}^*. \tag{39b}$$

Interpretation of this vector and particularly of its imaginary part can be found in Refs. 1, 15. For the calculation of $\langle P_d\rangle$, only the z-component of $\underline{\mathbf{P}}$ is of interest. Its z-component is

$$\underline{P}_z = -\tfrac{1}{2}(\sigma_{xz}v_x^* + \sigma_{zz}v_z^*). \tag{40}$$

Studies of the real part of this component at given frequencies and given depths z can be found in Refs. 14, 16, 17. At the interfaces $z=\mp(d/2)$, the boundary conditions imply the continuity of the z-component of the complex Poynting vector. $\underline{P}_z$ can be simply written as a function of the pressure p_F in the fluid and of the z-component of the displacement velocity vector in the fluid, v_{F_z}. We can show that the total mean power dissipated in the surrounding medium is obtained by

$$\langle P_d\rangle = \rho_F\omega^4\frac{\cos\theta}{c_F}(1-|\underline{R}|^2)dx. \tag{41}$$

We set as a definition

$$P_{\mathrm{Im}} = \oint_{dS} \mathrm{Im}(\underline{\mathbf{P}})\cdot\mathbf{n}\, dS. \tag{42a}$$

We can show that P_{Im} is obtained by

$$P_{\mathrm{Im}} = \rho_F\omega^4\frac{\cos\theta}{c_F}\mathrm{Im}(\underline{R})dx. \tag{42b}$$

This imaginary part corresponds to the peak reactive power periodically exchange between the elastic plate and the upper and lower fluids.

We now have to identify the different Q factors with energy definitions of the type of Eq. (23). We are first interested in the frequency quality factor Q_x which can be obtained from the function $f(\partial\phi/\partial f)$. From Eq. (8a), it may be shown that the exact expression of this function below the first critical angle θ_C^L is

$$f\frac{\partial\varphi}{\partial f}=\tau\frac{\omega\, d\cos\theta_L}{2c_L}\left(\cos^2(2\theta_T)\left(\frac{1}{\cos^2\left(\frac{k_{L_z}d}{2}\right)(C_A^2+\tau^2)}+\frac{1}{\sin^2\left(\frac{k_{L_z}d}{2}\right)(C_S^2+\tau^2)}\right)+\sin^2(2\theta_T)\right.$$
$$\left.\times\left(\frac{1}{\cos^2\left(\frac{k_{T_z}d}{2}\right)(C_A^2+\tau^2)}+\frac{1}{\sin^2\left(\frac{k_{T_z}d}{2}\right)(C_S^2+\tau^2)}\right)\right). \tag{43}$$

The comparison of this equation with Eqs. (36c), (36d) and Eqs. (37b), (37d) allows us to write

$$f\frac{\partial\varphi}{\partial f}=\frac{2\omega(\langle E_{K_{L_z}}\rangle+\langle E_{K_{T_x}}\rangle)}{\frac{1}{2}\rho_F\omega^4\frac{\cos\theta}{c_F}}. \tag{44}$$

In Appendix B, it is shown that $\langle E_{K_{L_z}}\rangle$ is equal to a part of the longitudinal-strain stored energy denoted here for simplicity as $\langle E_{S_{L_z}}\rangle$ and that $\langle E_{K_{T_x}}\rangle$ is equal to a part of the transverse-strain stored energy $\langle E_{S_{T_x}}\rangle$. We may introduce the terms $\langle E_{L_z}\rangle$ and $\langle E_{T_x}\rangle$ equal to $2\langle E_{K_{L_z}}\rangle$ and $2\langle E_{K_{T_x}}\rangle$ respectively. They correspond to the mean total energies of longitudinal and transverse waves, each one propagating in the z and $-z$-directions normal to the plate. Finally, we can also define a mean total energy $\langle E_\perp\rangle$ associated to all the waves propagating normal to the plate

$$\langle E_\perp\rangle=\langle E_{L_z}\rangle+\langle E_{T_x}\rangle=2(\langle E_{K_{L_z}}\rangle+\langle E_{K_{T_x}}\rangle). \tag{45}$$

At the resonance frequency f_{res}, the left hand-side of Eq. (44) is $2Q_x$ and the reflection coefficient $\underline{R}$ is approximately null. So, the mean dissipated power according to Eq. (41) is obtained by

$$\langle P_d\rangle=\rho_F\omega^4\frac{\cos\theta}{c_F}dx. \tag{46}$$

Finally, we may write the frequency quality factor Q_x as

$$Q_x=\omega\frac{\langle E_\perp\rangle}{\langle P_d\rangle}. \tag{47}$$

Therefore, the frequency quality factor is related to a mean stored energy associated with both longitudinal and transverse standing waves in a direction normal to the propagation direction of the guided waves.

The numerical validity of Eq. (47) is shown in Fig. 7. We have plotted the inverse of the exact function $f(\partial\phi/\partial f)$ (solid line) and the inverse of $\omega(\langle E_\perp\rangle/\langle P_d\rangle)$ (dotted line) versus incidence angle. We observe that the two curves nearly coincide whatever the incidence angle.

In the same way as for $\langle E_\perp\rangle$, we can define a mean total energy associated with the waves propagating in the x-direction parallel to the plate: $\langle E_\parallel\rangle$. It is obtained by

$$\langle E_\parallel\rangle=\langle E_{L_x}\rangle+\langle E_{T_z}\rangle=2(\langle E_{K_{L_x}}\rangle+\langle E_{K_{T_z}}\rangle) \tag{48a}$$

Its expression below the first critical angle is

$$\langle E_\parallel\rangle=\frac{\rho_S\, dx\omega^4 d}{4}\left(\frac{\cos\theta}{c_F}\right)^2\tau^2\left(\tan^2\theta_L\cos^2 2\theta_T\left(\frac{1}{\sin^2\left(\frac{k_{L_z}d}{2}\right)(C_S^2+\tau^2)}+\frac{1}{\cos^2\left(\frac{k_{L_z}d}{2}\right)(C_A^2+\tau^2)}\right)\right.$$
$$\left.+\tan^2\theta_T\sin^2 2\theta_T\left(\frac{1}{\cos^2\left(\frac{k_{T_z}d}{2}\right)(C_A^2+\tau^2)}+\frac{1}{\sin^2\left(\frac{k_{T_z}d}{2}\right)(C_S^2+\tau^2)}\right)\right). \tag{48b}$$

If we compare Eq. (44b) to the expression of the prevailing term of $c_F(\partial\phi/\partial c_F)$ given in Appendix C, which corresponds to $2Q_F$ when calculated at f_{Res}, we may write

$$Q_F=\omega\frac{\langle E_\parallel\rangle}{\langle P_d\rangle}. \tag{49}$$

When the resonance frequency is close to a frequency for which $C_{S,A}$ is null, the quality factor Q_F corresponds to the angular quality factor Q_y. So the angular quality factor is proved to be related to a mean stored energy associated with waves propagating parallel to the direction of propagation of the Lamb waves generated in the plate.

In Fig. 8, we have plotted the inverse of Q_F (solid line) obtained by means of the exact expression of $c_F(\partial\phi/\partial c_F)$ and the inverse of $\omega(\langle E_\parallel\rangle/\langle P_d\rangle)$ (dotted line) versus incidence angle. We can observe that the two curves nearly coincide.

Finally, we can define mean total energies associated with the longitudinal and transverse waves in the plate as

$$\langle E_{L,T}\rangle=\langle E_{K_{L,T}}\rangle+\langle E_{S_{L,T}}\rangle=2\langle E_{K_{L,T}}\rangle. \tag{50}$$

The identity of the kinetic and strain energies is shown in Appendix B. If we compare the expressions of $\langle E_{L,T}\rangle$ to those of the prevailing terms of $c_{L,T}(\partial\phi/\partial c_{L,T})$ given in Ap-

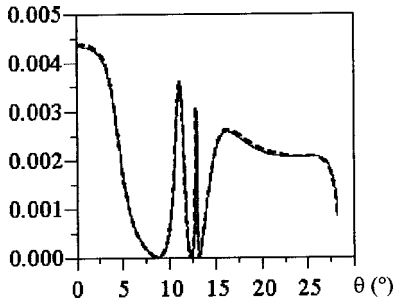


FIG. 7. Plots of the angular evolutions of the inverse of Q_x obtained from $f(\partial\varphi/\partial f)$ (solid line) and of the inverse of $\omega(\langle E_\perp\rangle/\langle P_d\rangle)$ (dotted line), following the dispersion curve of mode A_9.

pendix B, which correspond to $2Q_{L,T}$ at f_{Res}, we may write

$$Q_{L,T}=\omega\frac{\langle E_{L,T}\rangle}{\langle P_d\rangle}. \tag{51}$$

In Sec. II, it was outlined that the study of the maximums of $c_{L,T}(\partial\phi/\partial c_{L,T})$ could give the prevailing nature of a Lamb wave quantitatively, be it longitudinal or transverse. Equation (51) clearly establishes the relations between their maximums and the mean total stored energies associated to either longitudinal or transverse waves.

In Fig. 9(a), we have plotted θ for $\theta\leqslant\theta_C^L$, the inverse of Q_L (solid line) and the inverse of $\omega(\langle E_L\rangle/\langle P_d\rangle)$ (dotted line) versus incidence angle. In Fig. 9(b), we have plotted the evolutions of Q_L (solid line) and of $\omega(\langle E_L\rangle/\langle P_d\rangle)$ (dotted line) for $\theta>\theta_C^L$. In each figure, the two curves coincide well. In Fig. 10, we have compared the angular evolutions of the inverse of Q_T (solid line) and of the inverse of $\omega\langle E_T\rangle/\langle P_d\rangle$ (dotted line). Again, the coincidence is good.

We may now come back to the physical meaning of Eq. (22) stating that $Q_x+Q_y=Q_L+Q_T$. When a resonant phenomenon occurs in an elastic plate, it is due to constructive interferences of both longitudinal and transverse waves propagating up and down in the plate. This is the interpretation of the right-hand side of Eq. (22). According to the figures presented above, it seems that the interactions between the longitudinal and transverse waves can be neglected. The combination of these waves give rise on the one hand to standing waves whose resonant feature is described by Q_x, and on the other hand, to guided waves whose resonant feature is given by Q_y; this explains the left-hand side of Eq. (22).

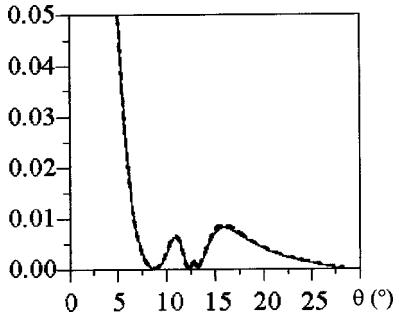


FIG. 8. Plots of the angular evolutions of the inverse of Q_F obtained from $c_F(\partial\varphi/\partial c_F)$ (solid line) and of the inverse of $\omega(\langle E_\parallel\rangle/\langle P_d\rangle)$ (dotted line), following the dispersion curve of mode A_9.

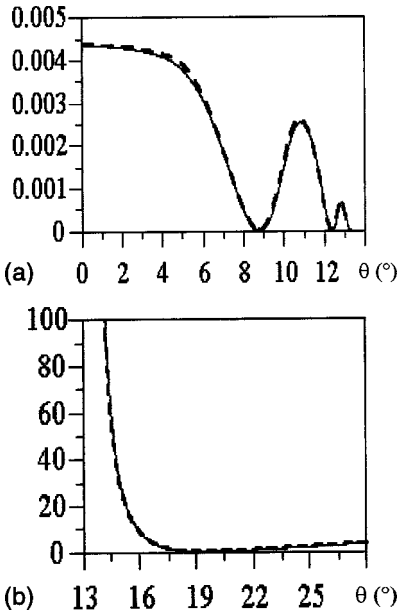


FIG. 9. (a) Plots of the angular evolutions of the inverse of Q_L obtained from $c_L(\partial\varphi/\partial c_L)$ (solid line) and of the inverse of $\omega(\langle E_L\rangle/\langle P_d\rangle)$ (dotted line), following the dispersion curve of mode A_9, for $\theta\leqslant\theta_C^L$. (b) Plots of the angular evolutions of Q_L obtained from $c_L(\partial\varphi/\partial c_L)$ (solid line) and of $\omega(\langle E_L\rangle/\langle P_d\rangle)$ (dotted line), following the dispersion curve of mode A_9, for $\theta\geqslant\theta_C^L$.

We may also come back to the physical meaning of Eq. (16), which may be rewritten, at the resonance frequency f_{Res}, as

$$\begin{aligned}\mathrm{Im}(\underline{R})&=\frac{P_{\mathrm{Im}}}{\langle P_d\rangle}=2(Q_L+Q_T-(Q_x+Q_y))\\&=\frac{2\omega(\langle E_L\rangle+\langle E_T\rangle-(\langle E_\parallel\rangle+\langle E_\perp\rangle))}{\langle P_d\rangle}.\end{aligned} \tag{52}$$

Remembering that the peak energy value is twice its mean value for harmonic waves, we can write

$$\begin{aligned}P_{\mathrm{Im}}&=\omega((E_L)_{\mathrm{peak}}+(E_T)_{\mathrm{peak}})-\omega((E_\parallel)_{\mathrm{peak}}+(E_\perp)_{\mathrm{peak}})\\&=P_{\mathrm{Im}}(\mathrm{plate})-P_{\mathrm{Im}}(\mathrm{upper\ \ fluid}).\end{aligned} \tag{53}$$

At f_{Res}, $\mathrm{Im}(\underline{R})$ is null, so P_{Im} is null as well. Therefore, the latter equation may be interpreted as a balance of reactive power. $(E_L)_{\mathrm{peak}}+(E_T)_{\mathrm{peak}}$ may be viewed as the peak energy which is periodically absorbed by the plate and $(E_\parallel)_{\mathrm{peak}}+(E_\perp)_{\mathrm{peak}}$ as the peak energy periodically reradiated into the

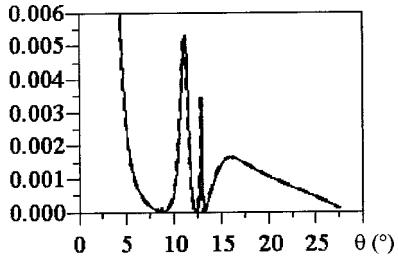


FIG. 10. Plots of the angular evolutions of the inverse of Q_T obtained from $c_T(\partial\phi/\partial c_T)$ (solid line) and from $\omega(\langle E_T\rangle/\langle P_d\rangle)$ (dotted line), following the dispersion curve of mode A_9.

fluid. The balance of active power transferred through the plate yields the energy conservation law. It is proved now that the balance of reactive power leads to Eq. (22) or conversely.

As a corollary to the expressions given by Eqs. (47), (49) and (51), we can also write

$$Q_x = Q_L \cos^2 \theta_L + Q_T \cos^2 \theta_T, \tag{54a}$$

$$Q_F = Q_y = Q_L \sin^2 \theta_L + Q_T \sin^2 \theta_T. \tag{54b}$$

So, if the frequency and angular quality factors can be measured experimentally, the longitudinal and transverse quality factors may be estimated thanks to the last equations.

IV. CONCLUSION

In this paper, we have shown the interest in studying the partial derivatives of the phase of the reflection coefficient of a fluid-loaded elastic plate, not only with respect to the frequency and angle variables, but also with respect to the bulk phase velocities involved in the problem. The use of the phase derivatives with respect to the longitudinal and transverse phase velocities as a means of determining the prevailing polarization state of a Lamb wave propagating along the plate was outlined. The fundamental relation linking the different phase derivatives naturally yielded the introduction of more or less common quality factors, in the field of elastic wave propagation, associated with resonant phenomena. The use of the energy definition of a quality factor allowed us to relate the frequency, angular, longitudinal, and transverse quality factors introduced to different types of mean stored energies which were detailed analytically in the second part. It was shown that the longitudinal and transverse quality factors depend on mean energies associated with longitudinal and transverse waves propagating in the plate. As well, the angular quality factor related to the spatial damping of a guided wave was proved to be connected to a mean energy associated with pure guided waves. The classical frequency quality factor, related to the temporal damping of the reflected wave, was proved to be connected to a mean energy associated with standing waves through the thickness of the plate. Finally, the base equation of the Phase Gradient Method was shown to correspond to a balance of reactive powers between the plate and the surrounding fluid.

ACKNOWLEDGMENTS

The authors wish to thank Dawn Hallidy for the rereading of this paper and the improvement of the English style.

APPENDIX A

Exact expressions of $c_L \partial \varphi_{S,A} / \partial c_L$

$$c_L \frac{\partial \varphi_{S,A}}{\partial c_L} = \mp \left(\frac{\tau}{C_{S,A}^2 + \tau^2} \right) c_L \frac{\partial C_{S,A}}{\partial c_L} \pm \left(\frac{C_{S,A}}{C_{S,A}^2 + \tau^2} \right) c_L \frac{\partial \tau}{\partial c_L},$$

$$c_L \frac{\partial C_S}{\partial c_L} = \frac{\dfrac{\pi f d}{c_L} \left(1 - 2 \left(\dfrac{c_T}{c_F} y \right)^2 \right)^2}{\sqrt{1 - \left(\dfrac{c_L}{c_F} y \right)^2} \sin^2 \left(\dfrac{\pi f d}{c_L} \sqrt{1 - \left(\dfrac{c_L}{c_F} y \right)^2} \right)} - \frac{4 c_T^3 y^2}{c_F^2 c_L} \frac{\sqrt{1 - \left(\dfrac{c_T}{c_F} y \right)^2}}{\sqrt{1 - \left(\dfrac{c_L}{c_F} y \right)^2}} \cot \left(\frac{\pi f d}{c_T} \sqrt{1 - \left(\frac{c_T}{c_F} y \right)^2} \right),$$

$$c_L \frac{\partial C_A}{\partial c_L} = - \frac{\dfrac{\pi f d}{c_L} \left(1 - 2 \left(\dfrac{c_T}{c_F} y \right)^2 \right)^2}{\sqrt{1 - \left(\dfrac{c_L}{c_F} y \right)^2} \cos^2 \left(\dfrac{\pi f d}{c_L} \sqrt{1 - \left(\dfrac{c_L}{c_F} y \right)^2} \right)} - \frac{4 c_T^3 y^2}{c_F^2 c_L} \frac{\sqrt{1 - \left(\dfrac{c_T}{c_F} y \right)^2}}{\sqrt{1 - \left(\dfrac{c_L}{c_F} y \right)^2}} \tan \left(\frac{\pi f d}{c_T} \sqrt{1 - \left(\frac{c_T}{c_F} y \right)^2} \right),$$

$$c_L \frac{\partial \tau}{\partial c_L} = - \frac{\rho_F c_F}{\rho_S c_L} \frac{1}{\sqrt{1 - y^2} \sqrt{1 - \left(\dfrac{c_L}{c_F} y \right)^2}}.$$

Exact expressions of $c_T (\partial \varphi_{S,A} / \partial c_T)$:

$$c_T \frac{\partial \varphi_{S,A}}{\partial c_T} = \mp \left(\frac{\tau}{C_{S,A}^2 + \tau^2} \right) c_T \frac{\partial C_{S,A}}{\partial c_T},$$

$$c_T\frac{\partial C_S}{\partial c_T}=4\frac{c_T^2y^2}{c_F^2c_L}\frac{\pi fd\sqrt{1-\left(\frac{c_L}{c_F}y\right)^2}}{\sin^2\left(\frac{\pi fd}{c_T}\sqrt{1-\left(\frac{c_T}{c_F}y\right)^2}\right)}-8\frac{c_T^2y^2}{c_F^2c_L}\left(1-2\left(\frac{c_T}{c_F}y\right)^2\right)^2\cot\left(\frac{\pi fd}{c_L}\sqrt{1-\left(\frac{c_L}{c_F}y\right)^2}\right)$$
$$+4\frac{c_T^3y^2}{c_F^2c_L}\frac{\sqrt{1-\left(\frac{c_L}{c_F}y\right)^2}}{\sqrt{1-\left(\frac{c_T}{c_F}y\right)^2}}\left(3-4\left(\frac{c_T}{c_F}y\right)^2\right)\cot\left(\frac{\pi fd}{c_T}\sqrt{1-\left(\frac{c_T}{c_F}y\right)^2}\right),$$

$$c_T\frac{\partial C_A}{\partial c_T}=-4\frac{c_T^2y^2}{c_F^2c_L}\frac{\pi fd\sqrt{1-\left(\frac{c_L}{c_F}y\right)^2}}{\cos^2\left(\frac{\pi fd}{c_T}\sqrt{1-\left(\frac{c_T}{c_F}y\right)^2}\right)}-8\frac{c_T^2y^2}{c_F^2c_L}\left(1-2\left(\frac{c_T}{c_F}y\right)^2\right)^2\tan\left(\frac{\pi fd}{c_L}\sqrt{1-\left(\frac{c_L}{c_F}y\right)^2}\right)$$
$$+4\frac{c_T^3y^2}{c_F^2c_L}\frac{\sqrt{1-\left(\frac{c_L}{c_F}y\right)^2}}{\sqrt{1-\left(\frac{c_T}{c_F}y\right)^2}}\left(3-4\left(\frac{c_T}{c_F}y\right)^2\right)\tan\left(\frac{\pi fd}{c_T}\sqrt{1-\left(\frac{c_T}{c_F}y\right)^2}\right).$$

Exact expressions of $c_F(\partial\varphi_{S,A}/\partial c_F)$

$$c_F\frac{\partial\varphi_{S,A}}{\partial c_F}=\mp\left(\frac{\tau}{C_{S,A}^2+\tau^2}\right)c_F\frac{\partial C_{S,A}}{\partial c_F}\pm\left(\frac{C_{S,A}}{C_{S,A}^2+\tau^2}\right)c_F\frac{\partial\tau}{\partial c_F},$$

$$c_F\frac{\partial C_S}{\partial c_F}=-\frac{c_L y^2}{c_F^2}\frac{\pi fd\left(1-2\left(\frac{c_T}{c_F}y\right)^2\right)^2}{\sqrt{1-\left(\frac{c_L}{c_F}y\right)^2}\sin^2\left(\frac{\pi fd}{c_L}\sqrt{1-\left(\frac{c_L}{c_F}y\right)^2}\right)}+8\frac{c_T^2y^2}{c_F^2}\left(1-2\left(\frac{c_T}{c_F}y\right)^2\right)$$
$$\times\cot\left(\frac{\pi fd}{c_L}\sqrt{1-\left(\frac{c_L}{c_F}y\right)^2}\right)-4\frac{c_T^4y^4}{c_F^4c_L}\frac{\pi fd\sqrt{1-\left(\frac{c_L}{c_F}y\right)^2}}{\sin^2\left(\frac{\pi fd}{c_T}\sqrt{1-\left(\frac{c_T}{c_F}y\right)^2}\right)}$$
$$-4\frac{c_T^2y^2}{c_F^2c_L}\frac{\left(-2-3\left(\left(\frac{c_L}{c_F}\right)^2+\left(\frac{c_T}{c_F}\right)^2\right)y^2+4\left(\frac{c_Lc_Ty^2}{c_F^2}\right)^2\right)}{\sqrt{1-\left(\frac{c_L}{c_F}y\right)^2}\sqrt{1-\left(\frac{c_T}{c_F}y\right)^2}}\cot\left(\frac{\pi fd}{c_T}\sqrt{1-\left(\frac{c_T}{c_F}y\right)^2}\right),$$

$$c_F\frac{\partial C_A}{\partial c_F}=\frac{c_L y^2}{c_F^2}\frac{\pi f d\left(1-2\left(\frac{c_T}{c_F}y\right)^2\right)^2}{\sqrt{1-\left(\frac{c_L}{c_F}y\right)^2}\cos^2\left(\frac{\pi f d}{c_L}\sqrt{1-\left(\frac{c_L}{c_F}y\right)^2}\right)}+8\frac{c_T^2y^2}{c_F^2}\left(1-2\left(\frac{c_T}{c_F}y\right)^2\right)$$

$$\times\cot\left(\frac{\pi f d}{c_L}\sqrt{1-\left(\frac{c_L}{c_F}y\right)^2}\right)+4\frac{c_T^4y^4}{c_F^4c_L}\frac{\pi f d\sqrt{1-\left(\frac{c_L}{c_F}y\right)^2}}{\cos^2\left(\frac{\pi f d}{c_T}\sqrt{1-\left(\frac{c_T}{c_F}y\right)^2}\right)}$$

$$-4\frac{c_T^2y^2}{c_F^2c_L}\frac{\left(-2-3\left(\left(\frac{c_L}{c_F}\right)^2+\left(\frac{c_T}{c_F}\right)^2\right)y^2+4\left(\frac{c_Lc_Ty^2}{c_F^2}\right)^2\right)}{\sqrt{1-\left(\frac{c_L}{c_F}y\right)^2}\sqrt{1-\left(\frac{c_T}{c_F}y\right)^2}}\cot\left(\frac{\pi f d}{c_T}\sqrt{1-\left(\frac{c_T}{c_F}y\right)^2}\right),$$

$$c_F\frac{\partial\tau}{\partial c_F}=-\frac{\rho_Fc_F^2}{\rho_Sc_L^2}\frac{1}{\sqrt{1-y^2}\sqrt{1-\left(\frac{c_L}{c_F}y\right)^2}}.$$

APPENDIX B: EXPRESSIONS OF THE MEAN STRAIN ENERGIES

Expressions of the stress and strain tensors components

$$\epsilon_{xx}(x,z,t)=\epsilon_{L_{xx}}(x,z,t)+\epsilon_{T_{xx}}(x,z,t),$$

$$\epsilon_{L_{xx}}(x,z,t)=-\omega^2\left(\frac{\sin\theta_L}{c_L}\right)^2(A_L\cos k_{L_z}z+B_L\sin k_{L_z}z)\times e^{j(k_xx-\omega t)},$$

$$\epsilon_{T_{xx}}(x,z,t)=-j\omega^2\frac{\sin 2\theta_T}{2c_T^2}(-A_T\sin k_{T_z}z+B_T\cos k_{T_z}z)\times e^{j(k_xx-\omega t)},$$

$$\epsilon_{xz}(x,z,t)=\epsilon_{L_{xz}}(x,z,t)+\epsilon_{T_{xz}}(x,z,t),$$

$$\epsilon_{L_{xz}}(x,z,t)=j\omega^2\frac{\sin\theta_T}{c_T}\frac{\cos\theta_L}{c_L}(-A_L\sin k_{L_z}z+B_L\cos k_{L_z}z)\times e^{j(k_xx-\omega t)},$$

$$\epsilon_{T_{xz}}(x,z,t)=\frac{1}{2}\omega^2\frac{\cos 2\theta_T}{c_T^2}(A_T\cos k_{T_z}z+B_T\sin k_{T_z}z)\times e^{j(k_xx-\omega t)},$$

$$\epsilon_{zz}(x,z,t)=\epsilon_{L_{zz}}(x,z,t)+\epsilon_{T_{zz}}(x,z,t),$$

$$\epsilon_{L_{zz}}(x,z,t)=-\omega^2\left(\frac{\cos\theta_L}{c_L}\right)^2(A_L\cos k_{L_z}z+B_L\sin k_{L_z}z)\times e^{j(k_xx-\omega t)},$$

$$\epsilon_{T_{zz}}(x,z,t)=-\epsilon_{T_{xx}}(x,z,t),$$

$$\sigma_{xx}(x,z,t)=\sigma_{L_{xx}}(x,z,t)+\sigma_{T_{xx}}(x,z,t),$$

$$\sigma_{L_{xx}}(x,z,t)=\rho_S\omega^2\left(-1+2c_T^2\left(\frac{\cos\theta_L}{c_L}\right)^2\right)\times(A_L\cos k_{L_z}z+B_L\sin k_{L_z}z)e^{j(k_xx-\omega t)},$$

$$\sigma_{T_{xx}}(x,z,t)=-j\rho_S\omega^2\sin 2\theta_T(-A_T\sin k_{T_z}z+B_T\cos k_{T_z}z)e^{j(k_xx-\omega t)},$$

$$\sigma_{xz}(x,z,t)=\sigma_{L_{xz}}(x,z,t)+\sigma_{T_{xz}}(x,z,t),$$

$$\sigma_{L_{xz}}(x,z,t)=2j\rho_S\omega^2c_T\sin\theta_T\frac{\cos\theta_L}{c_L}\times(-A_L\sin k_{L_z}z+B_L\cos k_{L_z}z)e^{j(k_xx-\omega t)},$$

$$\sigma_{T_{xz}}(x,z,t)=\rho_S\omega^2\cos 2\theta_T(A_T\cos k_{T_z}z+B_T\sin k_{T_z}z)e^{j(k_xx-\omega t)},$$

$$\sigma_{zz}(x,z,t)=\sigma_{L_{zz}}(x,z,t)+\sigma_{T_{zz}}(x,z,t),$$

$$\sigma_{L_{zz}}(x,z,t)=-\rho_S\omega^2\cos 2\theta_T(A_L\cos k_{L_z}z+B_L\sin k_{L_z}z)e^{j(k_xx-\omega t)},$$

$$\sigma_{T_{zz}}(x,z,t)=-\sigma_{T_{xx}}(x,z,t).$$

Expressions of the mean strain energies assigned to longitudinal waves

$$\langle E_{S_{L_{xx}}}\rangle = \frac{1}{4}\,dx\,\mathrm{Re}\left(\int_{z=-d/2}^{z=d/2}\sigma_{L_{xx}}\epsilon_{L_{xx}}^{*}\,dz\right)$$

$$= \frac{1}{8}\rho_S\,dx\omega^4 d\left(\left(\frac{\sin\theta_L}{c_L}\right)^2 - 2\sin^2\theta_T\left(\frac{\cos\theta_L}{c_L}\right)^2\right) \times(|A_L|^2+|B_L|^2),$$

$$\langle E_{S_{L_{xz}}}\rangle = \frac{1}{2}\,dx\,\mathrm{Re}\left(\int_{z=-d/2}^{z=d/2}\sigma_{L_{xz}}\epsilon_{L_{xz}}^{*}\,dz\right)$$

$$= \frac{1}{2}\rho_S\,dx\omega^4 d\sin^2\theta_T\left(\frac{\cos\theta_L}{c_L}\right)^2(|A_L|^2+|B_L|^2),$$

$$\langle E_{S_{L_{zz}}}\rangle = \frac{1}{4}\,dx\,\mathrm{Re}\left(\int_{z=-d/2}^{z=d/2}\sigma_{L_{zz}}\epsilon_{L_{zz}}^{*}\,dz\right)$$

$$= \frac{1}{8}\rho_S\,dx\omega^4 d\cos 2\theta_T\left(\frac{\cos\theta_L}{c_L}\right)^2(|A_L|^2+|B_L|^2),$$

$$\langle E_{S_L}\rangle = \langle E_{S_{L_{xx}}}\rangle + \langle E_{S_{L_{xz}}}\rangle + \langle E_{S_{L_{zz}}}\rangle$$

$$= \frac{1}{8}\rho_S\,dx\frac{\omega^4 d}{c_L^2}(|A_L|^2+|B_L|^2) = \langle E_{K_L}\rangle,$$

$$\langle E_{K_L}\rangle = \langle E_{S_L}\rangle$$

$$\langle E_{K_{L_x}}\rangle = \langle E_{S_{L_{xx}}}\rangle + \tfrac{1}{2}\langle E_{S_{L_{xz}}}\rangle$$

$$\langle E_{K_{L_z}}\rangle = \langle E_{S_{L_{zz}}}\rangle + \tfrac{1}{2}\langle E_{S_{L_{xz}}}\rangle$$

(these expressions are valid for $\theta < \theta_C^L$).

Expressions of the mean strain energies assigned to transverse waves

$$\langle E_{S_{T_{xx}}}\rangle = \frac{1}{4}\,dx\,\mathrm{Re}\left(\int_{z=-d/2}^{z=d/2}\sigma_{T_{xx}}\epsilon_{T_{xx}}^{*}\,dz\right)$$

$$= \frac{1}{16}\rho_S\,dx\omega^4 d\frac{\sin^2 2\theta_T}{c_T^2}(|A_T|^2+|B_T|^2)$$

$$= \langle E_{S_{T_{zz}}}\rangle,$$

$$\langle E_{S_{T_{xz}}}\rangle = \frac{1}{2}\,dx\,\mathrm{Re}\left(\int_{z=-d/2}^{z=d/2}\sigma_{T_{xz}}\epsilon_{T_{xz}}^{*}\,dz\right)$$

$$= \frac{1}{8}\rho_S\,dx\omega^4 d\frac{\cos^2 2\theta_T}{c_T^2}(|A_T|^2+|B_T|^2),$$

$$\langle E_{S_T}\rangle = \langle E_{S_{T_{xx}}}\rangle + \langle E_{S_{T_{xz}}}\rangle + \langle E_{S_{T_{zz}}}\rangle$$

$$= \frac{1}{8}\rho_S\,dx\frac{\omega^4 d}{c_T^2}(|A_T|^2+|B_T|^2) = \langle E_{K_T}\rangle,$$

$$\langle E_{K_T}\rangle = \langle E_{S_T}\rangle$$

$$\langle E_{K_{T_x}}\rangle = \langle E_{S_{T_{xz}}}\rangle$$

$$\langle E_{K_{T_z}}\rangle = \langle E_{S_{T_{xx}}}\rangle + \langle E_{S_{T_{zz}}}\rangle$$

(these expressions are valid for $\theta < \theta_C^T$).

APPENDIX C

$$c_F\frac{\partial\varphi_{\mathrm{app}}}{\partial c_F} = \frac{\omega d}{2c_L}\tau\cos\theta_L\left(\tan^2\theta_L\cos^2 2\theta_T\left(\frac{1}{\sin^2\left(\frac{k_{L_z}d}{2}\right)(C_S^2+\tau^2)} + \frac{1}{\cos^2\left(\frac{k_{L_z}d}{2}\right)(C_A^2+\tau^2)}\right)\right.$$

$$\left. + \tan^2\theta_T\sin^2 2\theta_T\left(\frac{1}{\cos^2\left(\frac{k_{T_z}d}{2}\right)(C_A^2+\tau^2)} + \frac{1}{\sin^2\left(\frac{k_{T_z}d}{2}\right)(C_S^2+\tau^2)}\right)\right),$$

$$c_L\frac{\partial\varphi_{\mathrm{app}}}{\partial c_L} = \frac{\omega d}{2c_L}\tau\cos\theta_L\frac{\cos^2 2\theta_T}{\cos^2\theta_L}\left(\frac{1}{\sin^2\left(\frac{k_{L_z}d}{2}\right)(C_S^2+\tau^2)} + \frac{1}{\cos^2\left(\frac{k_{L_z}d}{2}\right)(C_A^2+\tau^2)}\right),$$

$$c_T\frac{\partial\varphi_{\mathrm{app}}}{\partial c_T} = \frac{\omega d}{2c_L}\tau\cos\theta_L\frac{\sin^2 2\theta_T}{\cos^2\theta_T}\left(\frac{1}{\cos^2\left(\frac{k_{T_z}d}{2}\right)(C_A^2+\tau^2)} + \frac{1}{\sin^2\left(\frac{k_{T_z}d}{2}\right)(C_S^2+\tau^2)}\right).$$

[1] B. A. Auld, *Acoustic Fields and Waves in Solids*, Vols. I and II, 2nd ed. (Krieger Publishing, Malabar, Florida, 1990).

[2] J. M. Carcione and F. Cavallini, "Attenuation and quality factor surface in anisotropic-viscoelastic media," Mech. Mater. **19**, 311–327 (1995).

[3] A. Ben-Benahem and S. J. Singh, *Seismic Waves and Sources* (Springer, New York, 1981).

[4] X. P. Li, "Attenuation of Love waves in a two-layered half space," Wave Motion **22**, 349–370 (1995).

[5] O. Lenoir, J. Duclos, J. M. Conoir, and J. L. Izbicki, "Study of Lamb waves based upon the frequential and angular derivatives of the phase of the reflection coefficient," J. Acoust. Soc. Am. **94**, 330–343 (1993).

[6] S. Derible, P. Rembert, O. Lenoir, and J. L. Izbicki, "Elastic plate: Ex-

perimental measurements of resonances widths,'' Acoust. Lett. **16**, 208–213 (1993).
[7] G. Durinck, ''Experimental study of the transmission of pulsed ultrasound through elastic plates: Determination of the four resonance parameters. Interpretation in terms of mixed leaky-transient modes,'' Ph.D., Kortrijk, Belgium (1999).
[8] J. M. Conoir, ''Etude des résonances d'un tube élastique basées sur les propriétés de la dérivée de la phase de la matrice de diffusion'' (''Study of the resonances of an elastic tube based upon the properties of the derivatives of the phase of the scattering matrix,'') J. Acoust. **4**, 485–508 (1991).
[9] J. M. Conoir, J. L. Izbicki, and O. Lenoir, ''Phase gradient method applied to scattering by an elastic shell,'' Ultrasonics **35**, 157–169 (1997).
[10] R. Fiorito, W. Madigosky, and H. Überall, ''Resonance theory of acoustic waves interacting with an elastic plate,'' J. Acoust. Soc. Am. **66**, 1857–1866 (1979).
[11] A. Freedman, ''On the overlapping resonances: Concept of acoustic transmission through an elastic plate. I: An examination of properties,'' J. Sound Vib. **82**, 181–195 (1982).
[12] A. Freedman, ''On the overlapping resonances: Concept of acoustic transmission through an elastic plate. II: Numerical examples and physical properties,'' J. Sound Vib. **82**, 197–213 (1982).
[13] M. Deschamps and P. Chevée, ''Reflection and refraction of a heterogeneous wave by a solid layer,'' Wave Motion **15**, 61–75 (1992).
[14] A. Bernard, M. J. S. Lowe, and M. Deschamps, ''Guided waves energy velocity in absorbing and non-absorbing plates,'' J. Acoust. Soc. Am. **110**, 186–196 (2001).
[15] C. E. Dean and J. P. Braselton, ''A new geometric interpretation of the elastodynamic poynting vector,'' Acoust. Lett. **20**, 5–8 (1996).
[16] W. Maysenhölder, ''Rigourous computation of plate-wave intensity,'' Acustica **72**, 166–179 (1990).
[17] S. L. Means and R. R. Goodman, ''Structural intensity of acoustically excited waves in a fluid-loaded elastic plate,'' J. Acoust. Soc. Am. **102**, 335–341 (1997).

Some aspects of coupling-induced sound absorption in enclosures

K. S. Sum[a] and J. Pan
School of Mechanical Engineering, The University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009

(Received 20 August 2002; revised 9 May 2003; accepted 30 May 2003)

It is known that the coupling between a modally reactive boundary structure of an enclosure and the enclosed sound field induces absorption in the sound field. However, the effect of this absorption on the sound-field response can vary significantly, even when material properties of the structure and dimensions of the coupled system are not changed. Although there have been numerous investigations of coupling between a structure and an enclosed sound field, little work has been done in the area of sound absorption induced by the coupling. Therefore, characteristics of the absorption are not well understood and the extent of its influence on the behavior of the sound-field response is not clearly known. In this paper, the coupling of a boundary structure and an enclosed sound field in frequency bands above the low-frequency range is considered. Three aspects of the coupling-induced sound absorption are studied namely, the effects of exciting either the structure or the sound field directly, damping in the uncoupled sound field and damping in the uncoupled structure. The results provide an understanding of some features of the coupling-induced absorption and its significance to the sound-field response.
[DOI: 10.1121/1.1593062]

PACS numbers: 43.20.Ks, 43.20.Tb, 43.55.Br [MK]

LIST OF SYMBOLS

A_s, h, ρ_s	surface area, thickness and density of structure
c_0, ρ_0	speed of sound in air and density of air
F_p, Q_p	amplitudes of mechanical point force and volume velocity of monopole
M_s	modal overlap factor of uncoupled structure
p_{ext}	total complex sound pressure on external surface of structure
q	complex volume velocity of sound source per unit volume of enclosure
r, σ	location vectors of observation point in enclosure and on structure
r_0, σ_0	location vectors of monopole in enclosure and mechanical point force on structure
T_a, T_p	averaged 60-dB decay times of uncoupled acoustic modes and panel modes
t	time
V_0	volume of enclosure
η_p	averaged modal loss factor of uncoupled panel
θ, γ	initial phases of mechanical point force and noise emitted by monopole
ω	angular frequency
ω_L, ω_U	lower and upper angular frequency limits of excitation band
$\omega_0, \Delta\omega$	angular center frequency and bandwidth of excitation band

For the ith acoustic mode:

F_i, ξ_i	forcing and sound absorption functions
$\langle F_i \rangle$	space-averaged forcing function (over A_s for the location of mechanical point force and over V_0 for the location of monopole)
F_i^C, F_i^D	cross-modal and direct-modal forcing functions
P_i	complex modal amplitude of sound pressure
$\eta_{ai}^s, \overline{\eta}_{ai}^c$	loss factor due to acoustic–structural modal coupling only and total loss factor

For the ith acoustic mode and the jth structural mode in their uncoupled states:

$B_{j,i}, \Omega_{j,i}$	spatial- and frequency-matching factors for their coupling
M_{ai}, M_{sj}	modal masses
T_{ai}, T_{sj}	60-dB decay times
η_{ai}, η_{sj}	loss factors
Φ_{ai}, Φ_{sj}	mode-shape functions
ω_{ai}, ω_{sj}	angular resonance frequencies

I. INTRODUCTION

Traditionally, the sound absorption at a boundary of an enclosure and thus, the enclosed sound-field response, were described using the locally reactive concept.[1] This means that the motion of each surface element of the boundary depends only on the sound pressure in front of it and not on the pressure at other surface elements. However, the use of this concept is only a reasonable approximation in some cases. Experimental evidence showed that for a modally reactive boundary, the sound-field response actually depends on the

[a]Author to whom correspondence should be addressed. Electronic mail: ksum@mech.uwa.edu.au

coupling between acoustic modes of the sound field and structural modes of the boundary.[2] Acoustic–structural modal coupling is an area of continued research and the literature covering this area of investigation is very extensive. Reviews of some studies in this respect have been provided.[3,4] In general, a large amount of the previous work on acoustic–structural modal coupling can be summarized into three categories. The first focused on the effect of acoustical loading of a structure by a backing enclosure on the response of the structure. The studies were concerned with fatigue failure of structures, sound scattering or radiation from the structures (e.g., Refs. 5–9). The second category dealt with sound transmission into an enclosure through a boundary structure (e.g., Refs. 10–12). The work in the last category mainly focused on developments and validations, or applications of analytical or numerical methods for the prediction of the enclosed sound-field and/or boundary structural responses (e.g., Refs. 13–21). Therefore, characteristics of the sound absorption induced by acoustic–structural modal coupling have not been investigated in those studies. Some work on the effect of the coupling on the sound-field decay in an enclosure have been reported.[3,22] By comparing the behavior of the sound-field decay with and without the coupling, the influence of the coupling-induced absorption can be observed. However, the former[3] was only concerned with the sound-field response in the low-frequency range, where the decay time of individual acoustic modes was studied. The latter[22] was aimed at providing a physical interpretation of Sabine's absorption coefficient in terms of Statistical Energy Analysis in the high-frequency range. Knowledge of the significance of coupling-induced absorption to the response of an enclosed sound field at medium and high frequencies is required, as those frequencies cover a large part of the audible frequency range of human beings. Such knowledge may also be useful for architectural acoustics, since modally reactive structures are often found in rooms and auditoriums.

In this paper, the response behavior of an enclosed sound field that couples with a flexible vibrating boundary structure is studied analytically for frequency bands above the low-frequency range. An analytical model for a band-limited steady-state sound-field response is first presented briefly. This band-limited model allows the roles of acoustic–structural modal coupling in the determination of the sound-field response to be investigated separately. For the case where the structure is externally excited by an incident acoustic wave and/or a distributed mechanical force in which the sound field is driven only by the coupling, the model shows that the coupling not only acts as a forcing function of the sound field but also induces damping/absorption in the sound field. The former is responsible for sound radiation from the structure while the latter determines the amount of sound that is being absorbed back by the structure. However, the coupling only plays a role as the latter, in the case where the sound field is directly excited by an acoustic source and the structure is driven only by the coupling. The band-limited model is validated numerically for the direct structural excitation and it is then used to study the three aspects of the coupling-induced sound absorption mentioned in the abstract. Some results for the direct sound-field excitation are included for comparisons. In both cases, features of the coupling-induced absorption as well as its significance to the sound-field response are discussed.

II. BAND-LIMITED MODEL AND NUMERICAL VALIDATION

Consider an enclosed sound field that couples with a flat plate-type boundary structure. The sound-field pressure and structural velocity describe the response of the acoustic–structural coupled system. In a modal analysis, the sound pressure and the velocity are expanded in terms of orthogonal/normal modes of the sound field and the structure, respectively. By introducing these expansions into the equations of motion of the system, a set of modal equations is obtained where the response of each normal mode can be solved. The lengthy derivation of the modal equations is well known and widely available (e.g., Refs. 12, 13, 23, 24), but analytical solutions to the modal response are complicated.[25] However, a simplification of these solutions was justifiable for the prediction of band-limited response of the system above the low-frequency range and away from the first few low-order normal modes,[4] which is of concern here. Therefore, by approximating the response of each acoustic mode of the sound field in the same way as Ref. 4, the time- and space-averaged mean-square sound pressure can be derived as [see the Appendix for the derivation of Eqs. (1)–(11)]

$$\langle \overline{p^2} \rangle = \frac{\rho_0^3 c_0^4}{2V_0} \sum_{i=1}^{N} F_i(\omega)\xi_i(\omega). \tag{1}$$

In the above,

$$\xi_i(\omega) = \omega^2/[(\omega_{ai}^2-\omega^2)^2+(\overline{\eta}_{ai}^c\omega_{ai}^2)^2], \tag{2}$$

$$\overline{\eta}_{ai}^c = \eta_{ai}+\eta_{ai}^s, \tag{3}$$

$$\eta_{ai}^s = \frac{(A_s\rho_0 c_0)^2}{M_{ai}\omega_{ai}} \sum_{j=1}^{M} \frac{B_{j,i}^2}{M_{sj}} \Omega_{j,i}, \tag{4}$$

$$B_{j,i} = \frac{1}{A_s}\int_{A_s} \Phi_{sj}\Phi_{ai}\, dA, \tag{5}$$

$$\Omega_{j,i} = \begin{cases} 1/\eta_{sj}\omega_{sj}(\varepsilon_{j,i}^2+1), \\ \quad \text{for } T_{sj} \leqslant 1.0 \text{ s} \quad \text{or} \quad \omega_{sj}=\omega_{ai}; \\ \tan^{-1}\varepsilon_{j,i}/2(\omega_{sj}-\omega_{ai}), \\ \quad \text{for } T_{sj} > 1.0 \text{ s}, \quad T_{ai} \approx 1.0 \text{ s or below}; \\ [\tan^{-1}\varepsilon_{j,i} - \tan^{-1}(\varepsilon_{j,i}/2)](\omega_{sj}-\omega_{ai}), \\ \quad \text{for } T_{sj} > 1.0 \text{ s}, \quad T_{ai} \gg 1.0 \text{ s}, \end{cases} \tag{6}$$

$$\varepsilon_{j,i} = 2(\omega_{sj}-\omega_{ai})/\eta_{sj}\omega_{sj}. \tag{7}$$

For the case where the structure is externally excited by an incident acoustic wave and/or a distributed mechanical force

while the sound field is driven only through its coupling with the structure,

$$F_i(\omega)=F_i^D(\omega)+F_i^C(\omega), \tag{8}$$

$$F_i^D(\omega)=\frac{A_s^2\omega^2}{M_{ai}}\sum_{j=1}^{M}\frac{B_{j,i}^2[\int_{A_s}p_{\text{ext}}\Phi_{sj}\,dA][\int_{A_s}p_{\text{ext}}^*\Phi_{sj}\,dA]}{M_{sj}^2[(\omega_{sj}^2-\omega^2)^2+(\eta_{sj}\omega_{sj}^2)^2]}, \tag{9}$$

$$F_i^C(\omega)=\frac{2A_s^2\omega^2}{M_{ai}}\sum_{j=1}^{M-1}\sum_{k=j+1}^{M}\frac{B_{j,i}B_{k,i}[\int_{A_s}p_{\text{ext}}\Phi_{sj}\,dA][\int_{A_s}p_{\text{ext}}^*\Phi_{sk}\,dA]}{M_{sj}M_{sk}}\frac{[(\omega_{sj}^2-\omega^2)(\omega_{sk}^2-\omega^2)+(\eta_{sj}\eta_{sk}\omega_{sj}^2\omega_{sk}^2)]}{[(\omega_{sj}^2-\omega^2)^2+(\eta_{sj}\omega_{sj}^2)^2][(\omega_{sk}^2-\omega^2)^2+(\eta_{sk}\omega_{sk}^2)^2]}. \tag{10}$$

* denotes a complex conjugate. If the situation is reversed, where the sound field is directly excited by an acoustic source and the structure is driven only through its coupling with the sound field, then

$$F_i(\omega)=\frac{1}{M_{ai}}\left(\int_{V_0}q\Phi_{ai}\,dV\right)\left(\int_{V_0}q^*\Phi_{ai}\,dV\right). \tag{11}$$

Equation (1) is derived in such a way that it provides a means to define and to separately investigate the roles of acoustic–structural modal coupling in the determination of the sound-field response. The contribution of each acoustic mode to the sound pressure is split into a forcing function (F_i) and a sound absorption function (ξ_i). F_i describes characteristics of the excitation of the ith acoustic mode. On the other hand, ξ_i is in the form of the frequency response function of the mode, so it contains information of the resonance frequency and damping of the mode [see Eq. (2)].

As far as ξ_i is concerned, Eqs. (3) and (4) show that the modal coupling plays a role in providing damping to the ith acoustic mode, regardless of whether the structure or the sound field is directly excited. This added damping due to the coupling is denoted as η_{ai}^s. It depends on $B_{j,i}$ and $\Omega_{j,i}$, which, respectively, describe the extent of spatial and frequency matchings between the jth structural mode and the ith acoustic mode in their uncoupled states [see Eqs. (5) and (6)]. η_{ai} accounts only for the total sound absorption by air and by other boundaries of the enclosure. It is regarded as the background absorption in the sound field as far as the coupling between the sound field and the structure is concerned. Thus, $\overline{\eta}_{ai}^c$ given by Eq. (3) is the loss factor of the acoustic mode when the sound field is coupled to the structure. In other words, the coupling actually redistributes the loss factor of each acoustic mode.

As far as F_i is concerned, Eqs. (8)–(10) indicate that the modal coupling also plays a role as a forcing function of the ith acoustic mode, when only the structure is directly excited. In this case, the coupling acts in two ways: namely, as a direct-modal excitation (F_i^D) and as a cross-modal excitation (F_i^C) of the mode. In the former, the acoustic mode is excited directly by the jth uncoupled structural mode [see Eq. (9)]. In the latter, the acoustic mode is excited indirectly by the kth uncoupled structural mode through the jth uncoupled structural mode [see Eq. (10)]. On the other hand, the coupling does not have a role as a forcing function if only the sound field is directly excited, since F_i, given by Eq. (11), is independent of the coupling.

A band-limited sound pressure can be obtained by averaging Eq. (1) over a frequency band ($[\omega_L,\omega_U]$) with a bandwidth of $\Delta\omega$. For a direct sound-field excitation, the corresponding analytical results have been presented.[4] So, only the analytical results for a direct structural excitation are presented here [see the Appendix for the derivation of Eqs. (12) and (13)]:

$$\langle\overline{p^2}\rangle_{\Delta\omega}=\frac{\rho_0^3c_0^4}{2V_0\,\Delta\omega}\begin{cases}\displaystyle\sum_{i=1}^{N}\frac{K_iK_i^*}{M_{ai}}\int_{\omega_L}^{\omega_U}\xi_i(\omega)d\omega, & \text{for } M_s>1 \text{ or } T_p\ll 1.0 \text{ s};\\ \displaystyle\sum_{i=1}^{N}\frac{A_s^2}{M_{ai}}\sum_{j=1}^{M}\frac{B_{j,i}^2[\int_{A_s}p_{\text{ext}}\Phi_{sj}\,dA][\int_{A_s}p_{\text{ext}}^*\Phi_{sj}\,dA]}{M_{sj}^2}J_{j,i}(\omega_U,\omega_L,\omega_{sj},\omega_{ai},\eta_{sj},\overline{\eta}_{ai}^c), & \text{otherwise,}\end{cases} \tag{12}$$

$$K_i=A_s\sum_{j=1}^{M}\frac{B_{j,i}\omega_{ai}[\int_{A_s}p_{\text{ext}}\Phi_{sj}\,dA]}{M_{sj}[(\omega_{sj}^2-\omega_{ai}^2)+j\,\eta_{sj}\omega_{sj}^2]}. \tag{13}$$

Expressions for the frequency integral and $J_{j,i}$ in Eq. (12) are available from the Appendix.

The band-limited model derived previously for a direct sound-field excitation has been validated.[4] Thus, only a validation to the band-limited model derived for a direct structural excitation [i.e., Eq. (12)] is provided here. The sound-field response inside a rectangular parallelepiped enclosure predicted by the model is compared to that evaluated by the classical modal coupling method that employs a full mathematical solution.[13] The sound field is coupled to a simply supported panel located at an enclosure boundary, and various different dimensions of the system as well as dampings in the uncoupled panel and sound field were tested numeri-

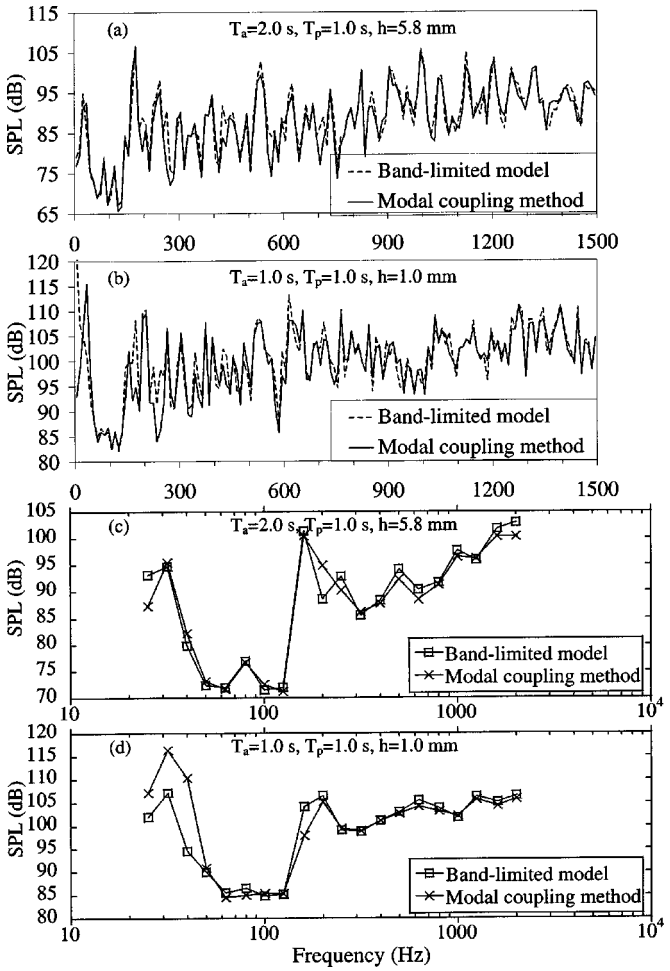


FIG. 1. Space-averaged sound-field response for a direct aluminum panel excitation at (0.152, 0.448) m in 9 Hz frequency bands [(a), (b)] and 1/3-octave bands [(c), (d)].

cally. Good agreements between both predictions were achieved and some examples of the results are shown in Fig. 1. In these examples, the enclosure has dimensions of (0.868, 1.150, 1.000) m. The panel is made of aluminum, located at $z=1.000$ m and has an area of 0.868×1.150 m^2. It is directly driven by a mechanical point force at a random position of (0.152, 0.448) m on the surface. Two different combinations of averaged uncoupled panel and acoustic modal decay times (i.e., T_p and T_a), and panel thicknesses ($h=1.0$ mm and $h=5.8$ mm) are used. The averaged decay times are then assigned to the uncoupled panel and acoustic modes (i.e., $T_{sj}=4.4/\eta_{sj}\omega_{sj}=T_p$ and $T_{ai}=4.4/\eta_{ai}\omega_{ai}=T_a$). The first two graphs present the sound-field response in a narrow frequency bandwidth of 9 Hz. Above the low-frequency range (e.g., >300 Hz), it can be seen that the results predicted by the band-limited model and the classical modal coupling method agree well. It is obvious that the former can still largely produce the envelope of the sound-field response as predicted by the latter, even though there are some discrepancies between the results. For example, the existence of maxima in various medium-frequency bands is well predicted by the band-limited model. For the case of 1/3-octave analysis, the trend of the sound-field response with frequency is also reasonably well predicted [see Figs. 1(c) and 1(d)]. The discrepancies are more pronounced in the low-frequency range because the band-limited model is formulated only for an analysis above this frequency range, where the coupling has been simplified mathematically (see the Appendix). Based on numerical tests, the simplification of the coupling also limits the scope of the model to $T_p<5$ s for discrepancies within 3 dB in medium- and high-frequency 1/3-octave bands. Nevertheless, this limit of T_p already represents a wide range of averaged uncoupled panel modal loss factor (i.e., $\eta_p=4.4\pi/T_p\omega_0$) in those frequency bands, starting from the order of 10^{-4}.

For an analysis above low frequencies, a large number of acoustic and/or structural modes is involved. So, the classical modal coupling method requires the manipulation of large complex matrices in order to achieve good convergence in the mathematical solution to the coupled sound-field response. However, this is avoided when the band-limited model is used. Therefore, the reduction in computational effort and time is very significant and similar to that shown previously,[4] where only a very small fraction of the calculation time using the classical method is required. The response in the required frequency band can also be obtained straightaway in one calculation without having to evaluate the response at each single frequency and then numerically average the results over the band, as in the classical method. In this case, the problem with frequency resolution is avoided, especially when the sound field and/or structure are lightly damped, where there is a risk of missing any peaks associated with maxima in the sound-field response above low frequencies when the selected resolution is too large. Since the model gives a band-limited description of the sound-field response, it provides a theoretical framework for studying the influence of modal properties of the uncoupled structure and sound field, and, thus, the overall band-averaged effect of modal coupling, on the response above low frequencies. As demonstrated in Fig. 1, the band-limited model is useful also for the case of narrow-band excitations, where there are insufficient resonance modes in the band for reliable estimates using statistical methods. It has a potential to be developed for the prediction and analysis of band-limited sound-field response in practical acoustic–structural coupled systems.

III. EFFECT OF COUPLING-INDUCED SOUND ABSORPTION

In this section, a case study is presented, where the averaged behavior of sound absorption induced by the coupling between the sound field and the panel is investigated for a direct panel excitation above the low-frequency range. A space-averaged location of a mechanical point force over the panel area is considered. The results are compared to those of a direct sound-field excitation for a space-averaged location of a monopole source over the enclosure volume. Although the averaged results are investigated, the features of the sound absorption obtained are generally similar to those for a given location of the point excitation because the coupling depends significantly on the modal properties of the uncoupled panel and sound field.[3] For a space average of the point excitation location, Eq. (1) becomes

$$\langle\langle\overline{p^2}\rangle\rangle = \frac{\rho_0^3 c_0^4}{2V_0}\sum_{i=1}^{N}\langle F_i(\omega)\rangle\xi_i(\omega), \tag{14}$$

where [the derivation of Eq. (15) from Eqs. (9), (10), and (11) is presented in the Appendix]

$$\langle F_i(\omega)\rangle = \begin{cases} \dfrac{A_s(F_p\omega)^2}{\rho_s h M_{ai}}\displaystyle\sum_{j=1}^{M}\frac{B_{j,i}^2}{M_{sj}[(\omega_{sj}^2-\omega^2)^2+(\eta_{sj}\omega_{sj}^2)^2]}, \\ \qquad \text{mechanical point force excitation;} \\ Q_p^2/\rho_0 V_0, \quad \text{monopole excitation,} \end{cases} \tag{15}$$

Then, by averaging Eq. (14) over $[\omega_L, \omega_U]$,

$$\begin{aligned}\langle\langle\overline{p^2}\rangle\rangle_{\Delta\omega} &= \frac{\rho_0^3 c_0^4}{2V_0\Delta\omega}\sum_{i=1}^{N}\int_{\omega_L}^{\omega_U}\langle F_i\rangle\xi_i d\omega, \\ &= \frac{\rho_0^3 c_0^4}{2V_0}\sum_{i=1}^{N}\langle\langle F_i\rangle\xi_i\rangle_{\Delta\omega}, \\ &= \begin{cases} \dfrac{A_s F_p^2\rho_0^3 c_0^4}{2\rho_s h V_0\Delta\omega}\displaystyle\sum_{i=1}^{N}\frac{1}{M_{ai}}\sum_{j=1}^{M}\frac{B_{j,i}^2}{M_{sj}}J_{j,i}(\omega_U,\omega_L,\omega_{sj},\omega_{ai},\eta_{sj},\overline{\eta}_{ai}^c), & \text{mechanical point force excitation;} \\ \dfrac{Q_p^2\rho_0^2 c_0^4}{2V_0^2\Delta\omega}\displaystyle\sum_{i=1}^{N}\int_{\omega_L}^{\omega_U}\xi_i(\omega)d\omega, & \text{monopole excitation.} \end{cases}\end{aligned} \tag{16}$$

The derivation of Eq. (16) is also explained in the Appendix. In Eq. (16), $F_p^2/\Delta\omega$ and $Q_p^2/\Delta\omega$ are, respectively, the power spectral densities of the band-limited mechanical force and noise emitted by the monopole.

The acoustic energy of the sound field can be deduced from the sound pressure evaluated by Eq. (16) and it is presented in 1/3-octave bands for the coupling with a chipboard panel of $h=15.9$ mm (see Fig. 2). As the 1/3-octave bandwidth and thus, the band energy, increase with frequency, the energy spectral density (ESD) is considered where the band energy is divided by the bandwidth. For each combination of T_p and T_a in the figure, it is assumed that $T_{sj}=T_p$ and $T_{ai}=T_a$ so that the extent of the coupling-induced absorption in one frequency band and another can be clearly observed and compared. Also included are the results where such absorption is not considered (i.e., negligence of the coupling-induced loss factor, η_{ai}^s). Three aspects of the absorption are studied for the same panel namely, the effects of the type of excitation (i.e., either a direct sound-field or a direct panel excitation), background absorption (i.e., T_a), and uncoupled panel damping (i.e., T_p). Frequency bands above the 200 Hz band are considered, where each of those bands have two or more uncoupled acoustic modes for the enclosure system used here. The product of the forcing and absorption functions (i.e., $\langle\langle F_i\rangle\xi_i\rangle_{\Delta\omega}$) is used to explain features of the three aspects mentioned. Physically, $\langle\langle F_i\rangle\xi_i\rangle_{\Delta\omega}$ represents the acoustic potential energy stored in the ith mode because it is directly related to $\langle\langle\overline{p^2}\rangle\rangle_{\Delta\omega}$ as in Eq. (16), and, thus, to the ESD of the sound field. It also describes the net contribution of the forcing and absorption terms, since both terms are combined together.

For each of the two background absorptions (i.e., $T_a=1$ s and $T_a=5$ s) shown in Figs. 2(a) and (b), it is obvious that the effect of coupling-induced absorption on the ESD for the direct panel excitation is much larger than that for the direct sound-field excitation, although the same panel is used. This observation can be explained as follows. For example, consider the 800 Hz band, which has 45 uncoupled acoustic modes (i.e., those with resonance frequencies that lie within the band: mode number 61–105 if the fundamental acoustic mode is numbered 1). As far as the energies of these modes due to the 800 Hz band-limited excitation is concerned, $\langle\langle F_i\rangle\xi_i\rangle_{\Delta\omega}\propto\langle F_i(\omega_{ai})\rangle/(\eta_{ai}+\eta_{ai}^s)$ at or around the resonance frequency of the ith mode (i.e., $\omega=\omega_{ai}$) since ξ_i is inversely proportional to $\overline{\eta}_{ai}^c$ [see Eq. (2)]. Figure 3 presents $\langle\langle F_i\rangle\xi_i\rangle_{\Delta\omega}$ of the modes for the direct sound-field excitation. Without the coupling-induced absorption, $\langle\langle F_i\rangle\xi_i\rangle_{\Delta\omega}\propto\langle F_i(\omega_{ai})\rangle/\eta_{ai}$ and the energies of all the modes are comparable, as $\langle F_i(\omega_{ai})\rangle$ is independent of the coupling [see Eq. (15) for the monopole excitation]. Therefore, all the modes initially dominate the band response. Since there is a large energy flow between an uncoupled acoustic and panel mode (i.e., large η_{ai}^s) only when the modes are well coupled in terms of proximity of their resonance frequencies and/or mode-shape matching,[3] the coupling is very selective. Hence, the inclusion of η_{ai}^s only significantly decreases $\langle\langle F_i\rangle\xi_i\rangle_{\Delta\omega}$ of a limited number of the dominant acoustic modes in the band (e.g., modes 63, 64, 65, 84, and 85, as shown in Fig. 3). As examples, the proximity of the first two modal indices and resonance frequencies of acoustic modes 63 and 65 to those of panel modes (5,3) and (4,5), respectively, can be seen in Table I. As a result of the small number of the dominant acoustic modes whose energies are significantly affected when the coupling-induced absorption is included, the overall ESD of the band does not change much. For the direct panel excitation, $\langle F_i(\omega_{ai})\rangle$ depends on the coupling. As $\langle F_i(\omega_{ai})\rangle$ relates to $B_{j,i}$ [see Eq. (15)] in the same way as η_{ai}^s [see Eq. (4)], the forcing function has the

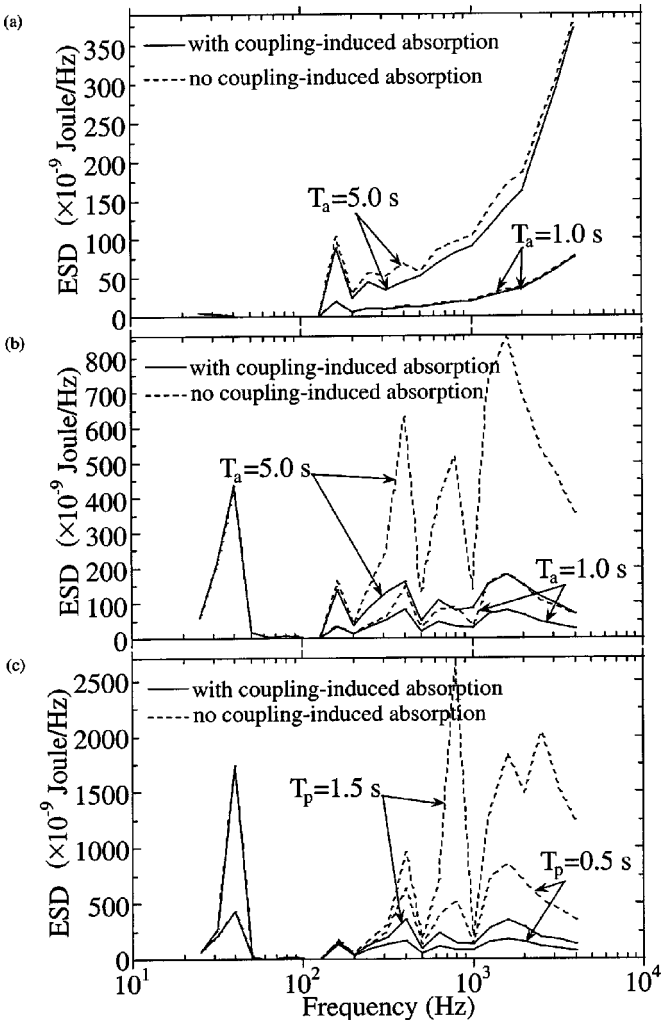


FIG. 2. Sound-field energy spectral density in 1/3-octave bands for the coupling with a chipboard panel of $h=15.9$ mm and a space average of excitation location. (a) Direct sound-field excitation by a monopole for $T_p=0.5$ s; Direct panel excitation for (b) $T_p=0.5$ s and (c) $T_a=5.0$ s.

same dependency of mode-shape matching as the coupling-induced absorption. In addition, Eq. (15) indicates that for $\omega=\omega_{ai}$, where the energy of the ith acoustic mode is defined for the band-limited excitation, $\langle F_i(\omega_{ai})\rangle$ increases with the closeness of ω_{sj} to ω_{ai}. These mean that similar to η_{ai}^s,

TABLE I. The dominant uncoupled acoustic modes in the 630 and 800 Hz bands and their main contributing uncoupled chipboard panel modes.

Acoustic modes			Panel modes	
Mode number	Modal indices	$\omega_{ai}/2\pi$ (Hz)	Modal indices	$\omega_{sj}/2\pi$ (Hz)
40	(1,3,2)	599.108	(4,4)	596.274
41	(2,2,2)	604.052	(5,1)	607.066
56	(0,0,4)	688.000	(1,7)	686.492
57	(0,4,2)	690.110		
63	(2,2,3)	716.100	(5,3)	715.270
65	(1,4,2)	717.996	(4,5)	718.004

$\langle F_i(\omega_{ai})\rangle$ is large only when the uncoupled acoustic and panel modes are well coupled. So, without including η_{ai}^s, only the five same well-coupled acoustic modes as before are driven significantly, rather than all the modes in the band. Thus, these modes initially dominate the band response [see Fig. 4(a)]. Since these modes are well coupled, they dissipate also a large amount of acoustic energy through the coupling when the panel absorbs back some energy from the sound field (i.e., the inclusion of η_{ai}^s). Hence, in contrast to the direct sound-field excitation, all the dominant modes in the band are attenuated significantly by the coupling-induced absorption [see Fig. 4(a)]. As a result, the contribution of such absorption to the ESD for the direct panel excitation is more than that for the direct sound-field excitation, as seen in Figs. 2(a) and (b), although the same panel is used in both cases. The above explanation for the 800 Hz band is also applicable for other frequency bands under consideration.

The extent of the sound absorption induced by the coupling also depends on the background absorption. The results in Figs. 2(a) and (b) show that the coupling-induced absorption gives only a little contribution to the sound-field response when the background absorption is high, and *vice versa* (for example, compare $T_a=1$ s to $T_a=5$ s). This contribution can be insignificant in the former case even though the panel is directly excited [e.g., Fig. 2(b) indicates that the coupling-induced absorption only attenuates the response in most bands by about half or less ($\leqslant 3$ dB) when $T_a=1$ s]. Figure 5 shows the uncoupled part (η_{ai}) and the coupled part (η_{ai}^s) of the total loss factor of each acoustic mode in the 800 Hz band. When $T_a=1$ s, η_{ai} is about twice the order

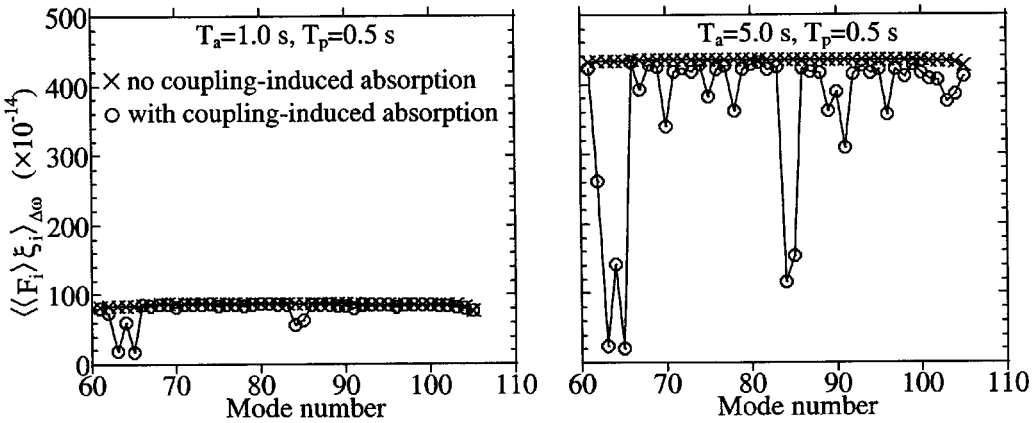


FIG. 3. The product of forcing and absorption functions of each acoustic mode in the 800 Hz band for the monopole excitation and coupling with the chipboard panel.

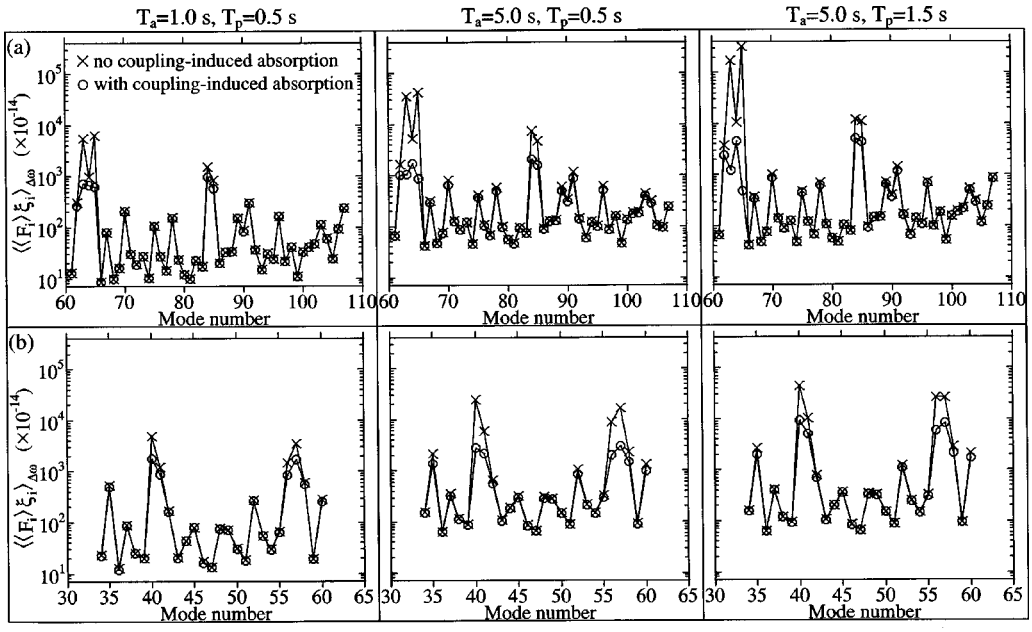


FIG. 4. The product of forcing and absorption functions of each acoustic mode in the (a) 800 Hz band and (b) 630 Hz band, for the direct chipboard panel excitation.

of magnitude larger than η_{ai}^s for most modes. It can also be seen that η_{ai}^s is larger than η_{ai} only for two well-coupled modes (i.e., modes 63 and 65). So, for the majority of the modes, the coupling-induced absorption is submerged in a high background absorption and contributes only a small fraction of the total sound absorption and hence the modal energy since $\langle\langle F_i\rangle\xi_i\rangle_{\Delta\omega} \propto \langle F_i(\omega_{ai})\rangle/\overline{\eta}_{ai}^c = \langle F_i(\omega_{ai})\rangle/(\eta_{ai} + \eta_{ai}^s)$. The consequence of the small contribution is shown in the first graphs in Figs. 3 and 4(a). It is obvious that all the acoustic modes have already been attenuated significantly before the coupling-induced absorption is included. The presence of this absorption further attenuates significantly modes 63 and 65 only, while all the other modes are affected by negligible amounts because of their very small η_{ai}^s compared to η_{ai}. When $T_a=5$ s, η_{ai} reduces by five times from that for $T_a=1$ s, and, thus, becomes about an order of magnitude larger than η_{ai}^s for most modes rather than twice the order as before (see Fig. 5). So, η_{ai}^s is now larger than η_{ai} for five well-coupled modes instead of two. Therefore, in this case, the contribution of the coupling-induced absorption to the overall sound absorption is larger than before. This increased influence as well as the decreased influence of the background absorption are reflected in the energy distribution of the acoustic modes in the band. When η_{ai}^s is not included, the second graph in Fig. 3 and the last two graphs in Fig. 4(a) for $T_a=5$ s show that all the modes are not attenuated by much initially as compared to those for $T_a=1$ s because of the low background absorption. Thus, more of the modes, especially those well-coupled ones, can be attenuated further by substantial amounts when η_{ai}^s is included. The same explanation can also be applied for other frequency bands [e.g., Fig. 4(b) for the 630 Hz band also illustrates the same phenomenon as in the 800 Hz band mentioned above]. It is why the coupling-induced absorption contributes more to the sound-field response when the background absorption is low, as seen in Figs. 2(a) and (b), even though material properties of the panel and the dimensions of the coupled system are not changed.

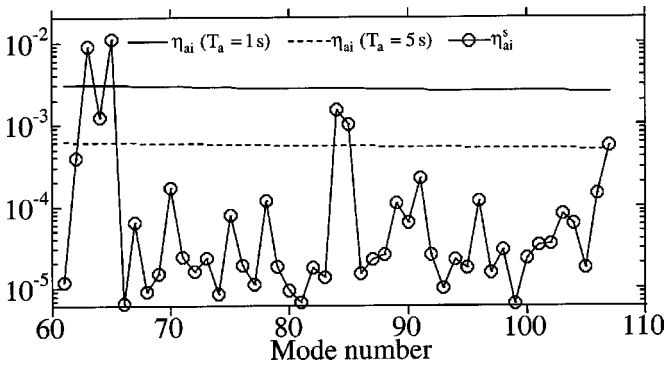


FIG. 5. The uncoupled and coupled parts of the total loss factor of each acoustic mode in the 800 Hz band for the coupling with the chipboard panel ($T_p=0.5$ s).

For the direct panel excitation, the behavior of $\langle\langle F_i\rangle\xi_i\rangle_{\Delta\omega}$ depends only on $\langle F_i(\omega_{ai})\rangle$ if η_{ai}^s in ξ_i is either not included or negligibly smaller than η_{ai}. For example, Fig. 2(b) for $T_a=1$ s shows that the shapes of the sound-field response with and without the coupling-induced absorption are about the same. Hence, the coupling plays a more important role in the forcing function than in the sound absorption, since the former solely determines the frequency trend of the response. When $T_a=5$ s, the sound-field response is also controlled by the coupling-induced absorption apart from the forcing function, and this is indicated by large differences in the response with and without the absorption [see Fig. 2(b)]. Such an influence of the absorption can be increased further by reducing the uncoupled panel damping (for example, increase T_p to 1.5 s). For the same panel, this damping varies with the panel mounting conditions and/or the characteristics of the sound field on the panel external surface. When ω

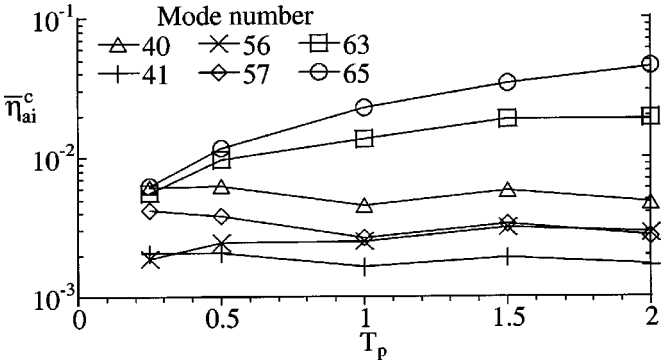


FIG. 6. The total loss factors of the dominant acoustic modes in the 630 and 800 Hz bands versus the averaged uncoupled panel modal decay time for the coupling with the chipboard panel ($T_a = 5.0$ s).

$=\omega_{ai}$ and ω_{sj} is close to ω_{ai}, the uncoupled acoustic and panel modes are well coupled in terms of resonance frequencies, and Eq. (15) indicates that $\langle F_i(\omega_{ai})\rangle$ is inversely proportional to η_{sj} (i.e., directly proportional to T_p). So, the sound-field response without the coupling-induced absorption increases significantly as T_p is changed from 0.5 to 1.5 s [see Fig. 2(c)]. It is obvious that the shapes of the response with and without the absorption are also partly different now. For instance, when $T_p = 1.5$ s, the response determined only by the forcing function has maxima in the 400, 800, 1600, and 2500 Hz bands. When the coupling-induced absorption is included, the resulting response in the 630 Hz band becomes a maximum instead of the 800 Hz band, and the maximum in the 2500 Hz band disappears. This observation is explained as follows. For example, consider the 630 and 800 Hz bands. The dominant acoustic modes before the inclusion of η^s_{ai} and their sole main contributing uncoupled panel modes are shown in Table I for both bands. Figure 6 shows $\overline{\eta}^c_{ai}$ of the acoustic modes for five values of T_p varying from 0.25 to 2 s. The corresponding half-power bandwidths $(=2.2/T_p)$ of the panel modes are thus, 8.8, 4.4, 2.2, 1.47, and 1.1 Hz. Guided by Table I, acoustic modes 63 and 65 in the 800 Hz band, respectively, always lie within the half-power bandwidths of panel modes (5,3) and (4,5) for most of the values of T_p considered (i.e., always well coupled in terms of resonance frequencies). So, $\overline{\eta}^c_{ai}$ of both acoustic modes increases significantly with T_p (see Fig. 6), which indicates that the resistance of the two panel modes to acquiring energy decreases. In the 630 Hz band, acoustic modes 40, 41, 56 and 57 lie within the half-power bandwidths of their main contributing panel modes only when $T_p = 0.25$ s (see Table I). As a result, $\overline{\eta}^c_{ai}$ of these acoustic modes does not vary much with the values of T_p under consideration. The change in $\overline{\eta}^c_{ai}$ of each dominant mode in both bands is reflected in the modal energy distribution shown in the last two graphs in Figs. 4(a) and (b), respectively. When T_p increases from 0.5 to 1.5 s in the 800 Hz band, the attenuation of modes 63 and 65 by the coupling-induced absorption increases from about an order of magnitude to more than twice the order. However, in the 630 Hz band, the attenuation of modes 40, 41, 56, and 57 does not change much when T_p is increased. In other words, the panel absorbs back a significantly larger amount of acoustic energy in the 800 Hz band than that in the 630 Hz band. Hence, in the resulting response shown in Fig. 2(c), the maximum in the former disappears but appears in the latter if the coupling-induced absorption is included. The above explanation can be used for the 2500 Hz band as well.

IV. CONCLUSIONS

An analytical model for a band-limited response of an enclosed sound field that couples with a flexible vibrating boundary structure has been presented. It was validated numerically for the case of a direct mechanical excitation of the structure where the sound field is driven only through acoustic–structural modal coupling. Some advantages of the band-limited model over the classical modal coupling method for an analysis above low frequencies have been reviewed and summarized. The model also provides a theoretical framework for studying the influence of modal properties of the uncoupled structure and sound field, and therefore the overall band-averaged effect of the modal coupling since it gives a band-limited description of the sound-field response. For the case of a direct structural excitation, this description allows the effects of the coupling in the forcing function and in the sound absorption to be investigated separately. The former is responsible for sound radiation from the structure while the latter describes the acoustic energy that is absorbed back by the structure. By using the band-limited model, three aspects of the coupling-induced sound absorption above the low-frequency range have been studied. First, when the structure is directly excited, such absorption is much more significant than that when the sound field is directly excited. In the latter, only a limited number of dominant acoustic modes are attenuated by the coupling-induced absorption, while in the former, all the dominant modes are attenuated since they are the same modes that dominate such absorption. Second, the effect of the coupling-induced absorption on the sound-field response increases with the decrease in the background absorption. This is because of an increased fractional contribution of the coupling-induced absorption to the total absorption in the sound field. Third, when the structure is directly excited and the background absorption is high and/or uncoupled panel is not lightly damped, only the forcing function (i.e., radiation from the structure) is in control of the sound-field behavior because the influence of the coupling-induced absorption is negligibly small. When the background absorption is low and/or uncoupled panel is lightly damped, the panel not only absorbs back a significant amount of acoustic energy from the sound field, but the coupling-induced absorption in certain bands is also much higher than that in others. Thus, the level and shape of the sound-field response determined only by the forcing function are changed considerably by the absorption. Hence, both the forcing function and the coupling-induced absorption (i.e., both radiation and absorption of the structure) are in control of the resulting sound-field response. This study has provided an understanding of some features of the coupling-induced absorption and its significance to the sound-field response. The band-limited model presented here

can be further developed into a more general format for the prediction of sound-field response in irregular enclosures and the work is underway.

APPENDIX: SUMMARY OF THE DERIVATION OF EQS. (1)–(16)

From a modal coupling analysis, an analytical solution to the response of the ith acoustic mode of an acoustic-structural coupled system has been obtained.[25] By neglecting acoustic modal cross-coupling terms of the mode and the shift of its resonance frequency due to acoustic–structural modal coupling, the solution can be written as

$$P_i \approx \frac{j\omega(\rho_0 c_0)^2}{M_{ai}(\omega_{ai}^2-\omega^2+j\eta_{ai}^c\omega_{ai}^2)} \times \begin{cases} j\omega A_s \sum_{j=1}^{M} \dfrac{B_{j,i}\int_{A_s} p_{\text{ext}}\Phi_{sj}\,dA}{M_{sj}(\omega_{sj}^2-\omega^2+j\eta_{sj}\omega_{sj}^2)}, \\ \quad \text{direct structural excitation;} \\ \int_{V_0} q\Phi_{ai}\,dV, \quad \text{direct sound field excitation,} \end{cases} \tag{A1}$$

where

$$\eta_{ai}^c = \eta_{ai} + \frac{(A_s\rho_0 c_0)^2\omega}{M_{ai}\omega_{ai}^2}\sum_{j=1}^{M}\frac{B_{j,i}^2}{\eta_{sj}\omega_{sj}M_{sj}}\left(\frac{1}{\varepsilon_{sj}^2+1}\right), \tag{A2}$$

$$\varepsilon_{sj} = 2(\omega_{sj}-\omega)/\eta_{sj}\omega_{sj}. \tag{A3}$$

The negligence of the cross-coupling terms has been justified for an analysis above low frequencies and away from the first few low-order normal modes of the system,[4,25] which is of concern here. It was also shown to be well-adapted for light fluid coupling.[19] On the other hand, the effect of acoustic–structural modal coupling on the shift of resonance frequency was shown to be significant only for the first few acoustic modes.[3] The same conclusion was obtained for the shift of resonance frequencies of structural modes.[6]

In Eq. (A1), the $\eta_{ai}^c\omega_{ai}^2$ term is dominant over the $\omega_{ai}^2-\omega^2$ term for an excitation frequency within the half-power bandwidth of the coupled acoustic mode, but $|\omega_{ai}^2-\omega^2| \gg \eta_{ai}^c\omega_{ai}^2$ at other excitation frequencies. Therefore, an averaged term of η_{ai}^c defined over the half-power bandwidth can be used to replace η_{ai}^c in Eq. (A1). This averaged term is the loss factor of the mode ($\overline{\eta}_{ai}^c$) that has a physical meaning only within the half-power bandwidth. In order to obtain $\overline{\eta}_{ai}^c$, Eq. (A2) is used where three conditions are considered. The first condition is $T_{sj} \leqslant 1.0$ s, where the half-power bandwidth of the jth uncoupled structural mode is quite broad. In this case, $1/(\varepsilon_{sj}^2+1)$ in Eq. (A2) varies slowly with ω in the half-power bandwidth of the acoustic mode, so it can be evaluated at ω_{ai}. Equation (A2) then yields

$$\overline{\eta}_{ai}^c = \eta_{ai} + \frac{(A_s\rho_0 c_0)^2}{M_{ai}\omega_{ai}}\sum_{j=1}^{M}\frac{B_{j,i}^2}{\eta_{sj}\omega_{sj}M_{sj}}\left(\frac{1}{\varepsilon_{j,i}^2+1}\right), \tag{A4}$$

$$\varepsilon_{j,i} = 2(\omega_{sj}-\omega_{ai})/\eta_{sj}\omega_{sj}. \tag{A5}$$

When $T_{sj} > 1.0$ s, the half-power bandwidth of the uncoupled structural mode is very narrow, and it can be shown that $1/(\varepsilon_{sj}^2+1)$ varies by an order of magnitude or more with ω around ω_{sj}. In this case, the evaluation of $1/(\varepsilon_{sj}^2+1)$ at ω_{ai} may not be sufficient to produce an accurate expression for $\overline{\eta}_{ai}^c$. Thus, it is necessary to establish accurate expressions from Eq. (A2), where the remaining two conditions are considered, namely, $T_{ai} \approx 1.0$ s or below, and $T_{ai} \gg 1.0$ s. When $T_{ai} \approx 1.0$ s or below, the half-power bandwidth of the acoustic mode is broad so a large portion of the response curve of the structural mode between ω_{ai} and ω_{sj} lies in the bandwidth if ω_{sj} is in the vicinity of ω_{ai}. So, an averaged value of η_{ai}^c over the full $[\omega_{ai},\omega_{sj}]$ interval can be used as an approximation to $\overline{\eta}_{ai}^c$, i.e.,

$$\frac{1}{\omega_{sj}-\omega_{ai}}\int_{\omega_{ai}}^{\omega_{sj}} 1/(\varepsilon_{sj}^2+1)\,d\omega$$

that holds for both $\omega_{sj} > \omega_{ai}$ and $\omega_{sj} < \omega_{ai}$. After the integration, Eq. (A2) yields

$$\overline{\eta}_{ai}^c = \eta_{ai} + \frac{(A_s\rho_0 c_0)^2}{2M_{ai}\omega_{ai}}\sum_{j=1}^{M}\frac{B_{j,i}^2}{M_{sj}(\omega_{sj}-\omega_{ai})}\tan^{-1}\varepsilon_{j,i}. \tag{A6}$$

In the case where ω_{sj} is far away from ω_{ai}, only a small portion of the response curve of the structural mode between ω_{ai} and ω_{sj} lies in the half-power bandwidth of the acoustic mode. Although the full $[\omega_{ai},\omega_{sj}]$ interval is used, the above integral is negligibly small and its effect on $\overline{\eta}_{ai}^c$ is insignificant. When $T_{ai} \gg 1.0$ s, the portion of the response curve lying within the half-power bandwidth is also small, even though ω_{sj} is very close to ω_{ai}, because the bandwidth is now very narrow. However, this can be accounted for by reducing the integration interval where half of the $[\omega_{ai},\omega_{sj}]$ interval on the ω_{ai} side is used as an approximation. The above integral then becomes

$$\frac{2}{\omega_{sj}-\omega_{ai}}\int_{\omega_{ai}}^{(\omega_{sj}+\omega_{ai})/2} 1/(\varepsilon_{sj}^2+1)\,d\omega.$$

Hence, Eq. (A2) yields

$$\overline{\eta}_{ai}^c = \eta_{ai} + \frac{(A_s\rho_0 c_0)^2}{M_{ai}\omega_{ai}}\sum_{j=1}^{M}\frac{B_{j,i}^2}{M_{sj}(\omega_{sj}-\omega_{ai})} \times\left(\tan^{-1}\varepsilon_{j,i} - \tan^{-1}\frac{\varepsilon_{j,i}}{2}\right). \tag{A7}$$

$\overline{\eta}_{ai}^c$ given by Eqs. (A4), (A6), and (A7) can be written in the form of Eq. (3) and then expressed in terms of a frequency-matching factor ($\Omega_{j,i}$), where Eq. (6) is obtained.

The time- and space-averaged mean-square sound-field pressure is given by

$$\langle \overline{p^2} \rangle = \frac{1}{2V_0} \int_{V_0} \left(\sum_{i=1}^{N} P_i \Phi_{ai} \right) \left(\sum_{i=1}^{N} P_i \Phi_{ai} \right)^* dV,$$

$$= \frac{1}{2\rho_0 V_0} \sum_{i=1}^{N} P_i P_i^* M_{ai}, \quad \text{for orthogonal conditions.} \tag{A8}$$

The second term in Eq. (A8) is derived from the first term by using the properties of orthogonality for all the acoustic modes,[13,23,24] i.e.,

$$\int_{V_0} \Phi_{ai} \Phi_{aj} \, dV = \begin{cases} 0, & \text{if } j \neq i; \\ M_{ai}/\rho_0, & \text{for } j = i. \end{cases} \tag{A9}$$

Upon substituting P_i from Eq. (A1) (with η_{ai}^c replaced by $\overline{\eta}_{ai}^c$) into the second term of Eq. (A8) and arranging mathematical terms, Eq. (1) is obtained. By an analogy to the case of a direct sound-field excitation.[4]

F_i^D and F_i^C, given by Eqs. (9) and (10) for the case of a direct structural excitation, are smooth functions of ω and their values are comparable if $M_s > 1$ or $T_p \ll 1.0$ s. Therefore, these forcing functions can be approximately evaluated at ω_{ai}, where the response of the acoustic mode is large and most important. Then, it can be shown that $F_i^D(\omega_{ai}) + F_i^C(\omega_{ai}) = K_i K_i^*/M_{ai}$, where K_i is given by Eq. (13). Subsequently, a band-limited sound pressure is obtained by integrating Eq. (1) over $[\omega_L, \omega_U]$ and dividing it by $\Delta\omega$:

$$\langle \overline{p^2} \rangle_{\Delta\omega} = \frac{\rho_0^3 c_0^4}{2V_0 \Delta\omega} \sum_{i=1}^{N} \frac{K_i K_i^*}{M_{ai}} \int_{\omega_L}^{\omega_U} \xi_i(\omega) \, d\omega. \tag{A10}$$

If the condition of $M_s > 1$ or $T_p \ll 1.0$ s is not satisfied, then by an analogy to the case of a direct sound-field excitation again,[4] F_i^C is negligibly small compared to F_i^D. Thus, $F_i \approx F_i^D$ and by using Eq. (9), the band-averaged sound pressure derived from Eq. (1) can be written as

$$\langle \overline{p^2} \rangle_{\Delta\omega} = \frac{\rho_0^3 c_0^4}{2V_0 \Delta\omega} \sum_{i=1}^{N} \frac{A_s^2}{M_{ai}} \sum_{j=1}^{M} \frac{B_{j,i}^2 [\int_{A_s} p_{\text{ext}} \Phi_{sj} \, dA][\int_{A_s} p_{\text{ext}}^* \Phi_{sj} \, dA]}{M_{sj}^2}$$

$$\times \int_{\omega_L}^{\omega_U} \frac{\omega^4}{[(\omega_{sj}^2 - \omega^2)^2 + (\eta_{sj}\omega_{sj}^2)^2][(\omega_{ai}^2 - \omega^2)^2 + (\overline{\eta}_{ai}^c \omega_{ai}^2)^2]} \, d\omega. \tag{A11}$$

Equations (A10) and (A11) are shown in Eq. (12), where $J_{j,i}$ in the latter denotes the frequency integral in Eq. (A11). The analytical solution to the integral in Eq. (A10) is obtained as $I(\omega_U) - I(\omega_L)$, where

$$I(\omega) = \frac{(c_{ai} A_{ai} + d_{ai} B_{ai})}{2(A_{ai}^2 + B_{ai}^2)} \ln \left[\frac{(\omega + A_{ai})^2 + B_{ai}^2}{(\omega - A_{ai})^2 + B_{ai}^2} \right] + \frac{(c_{ai} B_{ai} - d_{ai} A_{ai})}{2(A_{ai}^2 + B_{ai}^2)} \times \tan^{-1} \left[\frac{4 B_{ai} \omega (\omega^2 - A_{ai}^2 - B_{ai}^2)}{(\omega^2 - A_{ai}^2 - B_{ai}^2)^2 - (2 B_{ai} \omega)^2} \right], \tag{A12}$$

$$A_{ai} = \omega_{ai} \sqrt{(1 + \sqrt{1 + \overline{\eta}_{ai}^{c2}})/2}, \tag{A13}$$

$$B_{ai} = \omega_{ai} \sqrt{(\sqrt{1 + \overline{\eta}_{ai}^{c2}} - 1)/2}, \tag{A14}$$

$$c_{ai} = -0.5, \tag{A15}$$

$$d_{ai} = 0.5/\overline{\eta}_{ai}^c. \tag{A16}$$

On the other hand, the analytical solution to the frequency integral in Eq. (A11) is obtained as $I_1(\omega_U) - I_1(\omega_L) + I_2(\omega_U) - I_2(\omega_L)$, where

$$I_1(\omega) = \frac{(c_{j,i} A_{sj} + d_{j,i} B_{sj})}{2(A_{sj}^2 + B_{sj}^2)} \ln \left[\frac{(\omega + A_{sj})^2 + B_{sj}^2}{(\omega - A_{sj})^2 + B_{sj}^2} \right] + \frac{(c_{j,i} B_{sj} - d_{j,i} A_{sj})}{2(A_{sj}^2 + B_{sj}^2)} \times \tan^{-1} \left[\frac{4 B_{sj} \omega (\omega^2 - A_{sj}^2 - B_{sj}^2)}{(\omega^2 - A_{sj}^2 - B_{sj}^2)^2 - (2 B_{sj} \omega)^2} \right], \tag{A17}$$

$$I_2(\omega) = \frac{(e_{j,i} C_{ai} + f_{j,i} D_{ai})}{2(C_{ai}^2 + D_{ai}^2)} \ln \left[\frac{(\omega + C_{ai})^2 + D_{ai}^2}{(\omega - C_{ai})^2 + D_{ai}^2} \right] + \frac{(e_{j,i} D_{ai} - f_{j,i} C_{ai})}{2(C_{ai}^2 + D_{ai}^2)} \times \tan^{-1} \left[\frac{4 D_{ai} \omega (\omega^2 - C_{ai}^2 - D_{ai}^2)}{(\omega^2 - C_{ai}^2 - D_{ai}^2)^2 - (2 D_{ai} \omega)^2} \right], \tag{A18}$$

$$A_{sj} = \omega_{sj} \sqrt{(1 + \sqrt{1 + \eta_{sj}^2})/2}, \tag{A19}$$

$$B_{sj} = \omega_{sj} \sqrt{(\sqrt{1 + \eta_{sj}^2} - 1)/2}, \tag{A20}$$

$$C_{ai} = \omega_{ai} \sqrt{(1 + \sqrt{1 + \overline{\eta}_{ai}^{c2}})/2}, \tag{A21}$$

$$D_{ai} = \omega_{ai} \sqrt{(\sqrt{1 + \overline{\eta}_{ai}^{c2}} - 1)/2}, \tag{A22}$$

$$c_{j,i}=2\,\overline{\eta}^{c}_{ai}\,\eta_{sj}\omega^2_{ai}\omega^2_{sj}[\,\omega^2_{sj}(1+\eta^2_{sj})-\omega^2_{ai}(1+\overline{\eta}^{c2}_{ai})]/g_{j,i}\,, \tag{A23}$$

$$d_{j,i}=\overline{\eta}^{c}_{ai}\omega^2_{sj}\{\omega^4_{sj}(1+\eta^2_{sj})^2+\omega^2_{ai}[\,\omega^2_{ai}(1+\overline{\eta}^{c2}_{ai})(1-\eta^2_{sj}) -2\omega^2_{sj}(1+\eta^2_{sj})]\}/g_{j,i}\,, \tag{A24}$$

$$e_{j,i}=-c_{j,i}\,, \tag{A25}$$

$$f_{j,i}=\eta_{sj}\omega^2_{ai}\{\omega^4_{ai}(1+\overline{\eta}^{c2}_{ai})^2+\omega^2_{sj}[\,\omega^2_{sj}(1-\overline{\eta}^{c2}_{ai})(1+\eta^2_{sj}) -2\omega^2_{ai}(1+\overline{\eta}^{c2}_{ai})]\}/g_{j,i}\,, \tag{A26}$$

$$\begin{aligned} g_{j,i}=&\,2\,\overline{\eta}^{c}_{ai}\,\eta_{sj}\{\omega^8_{ai}(1+\overline{\eta}^{c2}_{ai})^2+\omega^8_{sj}(1+\eta^2_{sj})^2 \\ &-2\omega^2_{ai}\omega^2_{sj}[2(\omega^2_{ai}-\omega^2_{sj})^2+2(\overline{\eta}^{c2}_{ai}\omega^4_{ai}+\eta^2_{sj}\omega^4_{sj}) \\ &+\omega^2_{ai}\omega^2_{sj}(1-\overline{\eta}^{c2}_{ai})(1-\eta^2_{sj})]\}, \\ \approx&\,2\,\overline{\eta}^{c}_{ai}\,\eta_{sj}\omega^8_{ai}(\overline{\eta}^{c2}_{ai}-\eta^2_{sj})^2, \quad \text{if } \omega_{sj}\approx\omega_{ai}\,. \end{aligned} \tag{A27}$$

The values of the arctangent terms in Eqs. (A12), (A17), and (A18) vary from 0 to 2π.

If the structure is directly excited by a steady-state harmonic mechanical point force, then $p_{\text{ext}}=F_p\delta(\sigma-\sigma_0)e^{j(\omega t+\theta)}$, where δ is a Dirac delta function and has a unit of m^{-2}. If the sound field is directly excited by a steady-state harmonic monopole, then $q=Q_p\delta(r-r_0)e^{j(\omega t+\gamma)}$ where δ has a unit of m^{-3} in this case. By using these expressions, the product of integrals in Eqs. (9)–(11), respectively, becomes $F_p^2\Phi^2_{sj}(\sigma_0)$, $F_p^2\Phi_{sj}(\sigma_0)\Phi_{sk}(\sigma_0)$, and $Q_p^2\Phi^2_{ai}(r_0)$. By applying orthogonal conditions for all the structural modes,[13,23,24] i.e.,

$$\int_{A_s}\Phi_{si}\,\Phi_{sj}dA=\begin{cases}0, & \text{if } i\neq j;\\ M_{sj}/\rho_s h, & \text{if } i=j,\end{cases}$$

the first term subsequently becomes $F_p^2M_{sj}/\rho_s hA_s$ when integrated with respect to the position of the point force and divided by A_s, while the second term vanishes. On the other hand, by applying Eq. (A9), the third term becomes $Q_p^2M_{ai}/\rho_0V_0$ when integrated with respect to the position of the monopole and divided by V_0. If these results are used for the product of integrals in Eqs. (9)–(11), then a space-averaged forcing function ($\langle F_i\rangle$) shown in Eq. (15) is obtained. Hence, Eq. (16) can be obtained from Eq. (14) when one multiplies the expression for ξ_i given by Eq. (2) with those of $\langle F_i\rangle$, integrating over $[\omega_L,\omega_U]$ and then, dividing by $\Delta\omega$.

[1] P. M. Morse, "Some aspects of the theory of room acoustics," J. Acoust. Soc. Am. **11**, 56–66 (1939).

[2] J. Pan and D. A. Bies, "An experimental investigation into the interaction between a sound field and its boundaries," J. Acoust. Soc. Am. **83**, 1436–1444 (1988).

[3] J. Pan and D. A. Bies, "The effect of fluid-structural coupling on sound waves in an enclosure—Theoretical part," J. Acoust. Soc. Am. **87**, 691–707 (1990).

[4] K. S. Sum and J. Pan, "An analytical model for band-limited response of acoustic–structural coupled systems. I. Direct sound field excitation," J. Acoust. Soc. Am. **103**, 911–923 (1998).

[5] E. H. Dowell and H. M. Voss, "The effect of a cavity on panel vibration," AIAA J. **1**, 476–477 (1963).

[6] A. J. Pretlove, "Free vibrations of a rectangular panel backed by a closed rectangular cavity," J. Sound Vib. **2**, 197–209 (1965).

[7] A. J. Pretlove, "Forced vibrations of a rectangular panel backed by a closed rectangular cavity," J. Sound Vib. **3**, 252–261 (1966).

[8] F. J. Fahy, "Vibration of containing structures by sound in the contained fluid," J. Sound Vib. **10**, 490–512 (1969).

[9] R. W. Guy, "The response of a cavity backed panel to external airborne excitation: a general analysis," J. Acoust. Soc. Am. **65**, 719–731 (1979).

[10] T. Kihlman, "Sound radiation into a rectangular room. Applications to airborne sound transmission in buildings," Acustica **18**, 11–20 (1967).

[11] R. W. Guy and M. C. Bhattacharya, "The transmission of sound through a cavity-backed finite plate," J. Sound Vib. **27**, 207–223 (1973).

[12] W. B. McDonald, R. Vaicaitis, and M. K. Myers, "Noise transmission through plates into an enclosure," NASA Technical Paper No. 1173, NASA, 1978.

[13] E. H. Dowell, G. F. Gorman, and D. A. Smith, "Acoustoelasticity: general theory, acoustic natural modes and forced response to sinusoidal excitation, including comparisons with experiment," J. Sound Vib. **52**, 519–542 (1977).

[14] L. D. Pope, D. C. Rennison, C. M. Willis, and W. R. Mayes, "Development and validation of preliminary analytical models for aircraft interior noise prediction," J. Sound Vib. **82**, 541–575 (1982).

[15] A. Sestieri, D. D. Vescovo, and P. Lucibello, "Structural–acoustic couplings in complex shaped cavities," J. Sound Vib. **96**, 219–233 (1984).

[16] S. H. Sung and D. J. Nefske, "A coupled structural-acoustic finite element model for vehicle interior noise analysis," J. Vib. Acoust. Stress Rel. Design **106**, 314–318 (1984).

[17] L. F. Peretti and E. H. Dowell, "Asymptotic modal analysis of a rectangular acoustic cavity excited by wall vibration," AIAA J. **30**, 1191–1198 (1992).

[18] C. Soize, "A model and numerical method in the medium frequency range for vibroacoustic predictions using the theory of structural fuzzy," J. Acoust. Soc. Am. **94**, 849–865 (1993).

[19] J. L. Guyader and B. Laulagnet, "Structural acoustic radiation prediction: expanding the vibratory response on a functional basis," Appl. Acoust. **43**, 247–269 (1994).

[20] M. W. Bonilha and F. J. Fahy, "A hybrid probabilistic-deterministic approach for vibroacoustic studies," *Proceedings of the 15th International Congress on Acoustics*, Trondheim, Norway, 1995, pp. 483–486.

[21] W. Desmet, P. Sas, D. Vandepitte, and P. De Fonseca, "A new numerical technique for coupled vibro-acoustic predictions based on wave representation," *Proceedings of the 4th International Congress on Sound and Vibration*, St. Petersburg, Russia, 1996, pp. 857–868.

[22] J. Pan and D. A. Bies, "The effect of fluid-structural coupling on acoustical decays in a reverberation room in the high frequency range," J. Acoust. Soc. Am. **87**, 718–727 (1990).

[23] F. J. Fahy, "Acoustic coupling between structures and enclosed volumes of fluid," Chap. 6 in *Sound and Structural Vibration: Radiation, Transmission and Response* (Academic, London, 1987).

[24] J. Pan, S. J. Elliott, and K.-H. Baek, "Analysis of low frequency acoustic response in a damped rectangular enclosure," J. Sound Vib. **223**, 543–566 (1999).

[25] K. S. Sum and J. Pan, "On acoustic and structural modal cross-couplings in plate-cavity systems," J. Acoust. Soc. Am. **107**, 2021–2038 (2000).

Localization of notches with Lamb waves

Rüdiger Benz
School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0355

Marc Niethammer
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0250

Stefan Hurlebaus
Institute A of Mechanics, University of Stuttgart, Allmandring 5 b, D-70550 Stuttgart, Germany

Laurence J. Jacobs[a)]
School of Civil and Environmental Engineering and G. W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0355

(Received 5 February 2003; revised 20 May 2003; accepted 30 May 2003)

A time–frequency representation (TFR) is used to analyze the interaction of a multimode and dispersive Lamb wave with a notch, and then serves as the basis for a correlation technique to locate the notch. The experimental procedure uses a laser source and a dual-probe laser interferometer to generate and detect Lamb waves in a notched plate. The high fidelity, broad-bandwidth, point-like and noncontact nature of laser ultrasonics are critical to the success of this study, making it possible to experimentally measure transient Lamb waves without any frequency biases. A specific TFR, the reassigned spectrogram, is used to resolve the dispersion curves of the individual modes of the plate, and then the slowness-frequency representation (SFR) of the plate is calculated from this reassigned spectrogram. By considering the notch to be an additional (second) source, the reflected and transmitted contributions of each Lamb mode are automatically identified using the SFRs. These results are then used to develop a quantitative understanding of the interaction of an incident Lamb wave with a notch, helping to identify mode conversion. Finally, two complementary, automated localization techniques are developed based on this understanding of scattering of Lamb waves.
 [DOI: 10.1121/1.1593058]

PACS numbers: 43.20.Mv, 43.35.Cg [YHB]

I. INTRODUCTION

Our objective in this research is to develop an automated methodology that uses multimode and dispersive Lamb waves to locate a notch. The experimental procedure uses a laser source and a dual-probe laser interferometer to generate and detect Lamb waves in both notched and perfect plates. The high fidelity, broad-bandwidth, point-like, and noncontact nature of laser ultrasonics are critical to the success of this study, making it possible to measure transient Lamb waves without any frequency biases. Experimentally measured time–domain Lamb waves are first transformed into the time–frequency domain by calculating the reassigned spectrogram, a time–frequency representation (TFR); this TFR resolves the individual modes of the plate and generates its dispersion curves. The reassigned spectrogram is then normalized with respect to propagation distance by converting it to a slowness-frequency representation (SFR)—this is a quick and simple calculation.

A procedure is then developed to identify the reflected and transmitted contributions of each Lamb mode by considering the notch to be an additional (second) source. The SFRs from each of the interferometric probes (and in both the perfect and notched plates) are compared, and modes are automatically classified into different cases, depending on their interaction with the notch. These results are used to develop a quantitative understanding of the interaction of an incident Lamb wave with a notch, helping to identify mode conversion.

Finally, two complementary, automated localization techniques are developed based on this understanding of the scattering of Lamb waves. One technique isolates the contributions of the signal reflected from the notch by performing a correlation of a series of SFR spectra, each calculated with different, assumed propagation distances. This correlation technique uses an understanding of Lamb wave scattering—for example, the knowledge such as which modes are reflected, but not mode converted (and through which frequencies) helps refine the proposed correlation localization procedure. The second technique uses a goodness-of-fit metric when allocating (identifying) the transmitted, and mode converted wave field to determine the notch location. These two complementary techniques can be combined to provide two independent predictions (and thus increase accuracy), or can be used separately. Considering the arbitrary nature of the geometric relationship between the notch, the source, and each of the receivers in a real application, this robustness is especially important; the effectiveness of the proposed local-

[a)]Author to whom correspondence should be addressed. Electronic mail: laurence.jacobs@ce.gatech.edu

ization procedure is not dependent on the notch being located between the laser source and the receivers.

Previous researchers have used Lamb waves for material characterization (see Chimenti[1] for details), but a Lamb wave's multimode and dispersive nature makes an interpretation of time–domain signals difficult. In contrast, TFRs operate on time–domain signals, are capable of resolving the individual modes of a plate, and naturally lead to the slowness-frequency representation. TFRs are well known in the signal processing community (see Cohen[2] for a review of TFRs). Previous research has shown that TFRs based on the short-time Fourier transform (STFT)—spectrogram, reassigned spectrogram[3]—and the (pseudo) Wigner–Ville distribution[4] are particularly well suited for representing Lamb waves. The spectrogram is effective in this application because of its constant time–frequency resolution over all times and frequencies.[3] Lemistre *et al.*[5] and Wilcox *et al.*[6] used the TFRs of Lamb waves for the damage detection in composite plates, while Valle *et al.*[7] used time-of-flight information calculated from a reassigned spectrogram to locate notches in cylinders. Hurlebaus *et al.*[8] used the TFRs of Lamb waves and a correlation of the reflected contribution to localize notches in plates, but their research does not require (nor was it developed with) an understanding of the scattering of a Lamb wave by a notch.

Previous researchers have studied the interaction of guided Lamb waves with a crack. Liu *et al.*[9] investigated transient scattering of Lamb waves by a surface-breaking crack, while Cho and Rose[10] applied the boundary element method to determine the transmission and reflection coefficients from a surface breaking defect. Lowe and Diligent,[11] and Diligent *et al.*[12] examined the scattering of the lowest symmetric (s_0) mode from both a rectangular notch, and from a through thickness hole, respectively. Castaings *et al.*[13] used modal decomposition and an air-coupled ultrasonic receiver to model the interaction of Lamb waves with a crack; this study used the noncontact and high fidelity nature of air-coupled ultrasound, and generated selected modes (through a limited frequency bandwidth) with an interdigital-like (IDT) transmitter. This is in contrast to the current research that generates a broadband, multimode Lamb wave, and then relies on signal processing (the reassigned spectrogram) to resolve the individual modes of the plate. Note that an alternative approach is to use the two-dimensional Fourier transform (2D-FT)[14,15] to develop the dispersion curves. The 2D-FT is robust, but has the disadvantage that it requires multiple, equally spaced waveforms. In contrast, the current approach only requires a single waveform, while the processing involved in transforming this single time–domain Lamb wave to a SFR is computationally straightforward, and can be easily automated.

II. EXPERIMENTAL PROCEDURE, TIME–FREQUENCY AND SLOWNESS-FREQUENCY REPRESENTATIONS

The experimental procedure makes high fidelity, resonance-free, point-like noncontact measurements of Lamb waves over a wide frequency range (100 kHz to 10 MHz). Broadband Lamb waves are generated with an Nd:YAG laser source (see Scruby and Drain[16] for details on laser ultrasonics). Laser detection of these waves is accomplished with a dual-probe, heterodyne laser interferometer originally developed by Bruttomesso *et al.*,[17] and extended by Hurlebaus[18] to be a dual-probe receiver. This instrument uses the Doppler shift to simultaneously measure out-of-plane surface velocity (particle velocity) at two points on the surface of the plate. The Nd:YAG laser fires (at $t=0$) and generates a Lamb wave at the source location (ablation source). This Lamb wave is independently measured by each of the interferometric probes (as a transient, time–domain signal), discretized using a digital oscilloscope with a sampling frequency of 100 MHz, low-passed filtered (a linear phase analog Bessel filter) at 10 MHz, and averaged over 60 Nd:YAG shots (to increase the signal-to-noise ratio).

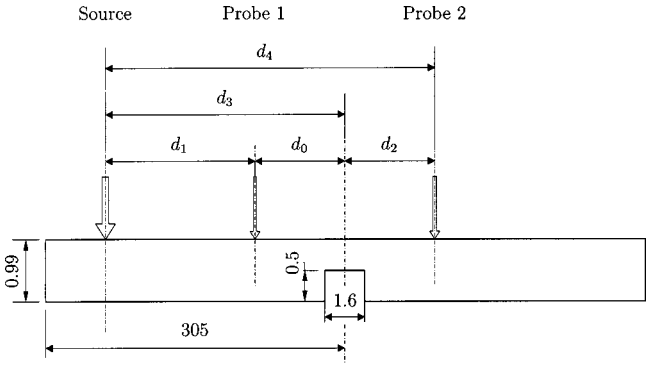


FIG. 1. Experimental setup (dimensions in mm).

Two different plate specimens are considered in this research. One plate specimen has no defects and is identified as the ''perfect plate,'' while the second has a single milled notch and is referred to as the ''notched plate''—the plates are identical otherwise, made of 3003 aluminum, with dimensions of 305 mm$\times$610 mm$\times$0.99 mm. The notch has a depth of 0.5 ± 0.1 mm, a width, w, of 1.6 ± 0.1 mm, a length of 305 mm (full plate width) and is located on the centerline of the 610 mm dimension of the plate. The relatively large plate size (305 mm$\times$610 mm) is needed to minimize the interference of reflections from the plate edges and all measurements are made in the center of the plate. Figure 1 shows the notched plate specimen, together with the location of the laser source and each of the two interferometric probes—the source and probe 1 are always on one side of the notch, while probe 2 is on the opposite side of the notch. Note that the source and both receiving probes are on the same face of the plate, while the notch is on the opposite (inaccessible) face. d_1 is the distance from source to probe 1, d_0 is the distance from probe 1 to the centerline of the notch, and d_2 is the distance from the centerline of the notch to probe 2. Define $d_3=d_1+d_0$, $d_4=d_3+d_2$, and $d_5=d_1+2d_0$. Lamb waves are measured at a variety of propagation distances (for comparison and verification purposes) by holding d_0 and d_2 fixed (30 mm each, which is approximately 30 plate thicknesses) for all experiments, and varying d_1 from 50 to 85 mm (in 5 mm increments).

A time–domain Lamb wave signal is transformed into the time–frequency domain using the STFT, essentially chopping the signal into a series of small overlapping pieces. Each of these pieces is windowed and then individually Fou-

rier transformed.[2] The STFT of a time–domain signal, $s(t)$, is defined as

$$S(\omega,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega\tau}s(\tau)h(\tau-t)\mathrm{d}\tau, \tag{1}$$

where $h(t)$ is the window function and ω is angular frequency. The energy density spectrum of a STFT is defined as

$$E_d(\omega,t)=|S(\omega,t)|^2, \tag{2}$$

and is called a spectrogram.

TFRs such as the spectrogram suffer from the Heisenberg uncertainty principle, making it impossible to simultaneously have perfect resolution in both time and frequency.[2] The resolutions in time and frequency are related to each other, and limited by the inequality, $\sigma_t^2\sigma_w^2\geqslant 0.25$, where σ_t and σ_w are the standard deviations for time and frequency, respectively. The equality holds for a Gaussian window [$h(t)$ in Eq. (1)]; the current research uses a Hanning window to compute the STFT because it allows for relatively small signal distortion, while ensuring smoothness of the windowed signal.[19] The time–frequency resolution of a spectrogram is solely controlled by the window size and type, and is independent of frequency. Choosing a narrow window provides for good resolution in time, but poor resolution in frequency, whereas a wide window leads to better frequency resolution and worse time resolution. This study uses a 384-point window for the time–domain signals measured at a 100 MHz sampling frequency (see Niethammer *et al.*[3] for details).

An improvement in the time–frequency resolution of a spectrogram is achieved by applying the reassignment method, developed by Auger and Flandrin.[20] The reassignment method reduces the time–frequency spread of a spectrogram by relocating ''energy'' from its old position, coordinates (t, ω), to new, reassigned coordinates, $(\hat{t},\hat{\omega})$.

The reassigned spectrogram transforms an experimentally measured time–domain Lamb wave signal into the time–frequency domain. This transformation enables a more quantitative interpretation of the Lamb wave,[3] but it is still difficult to compare reassigned spectrograms of measurements made with different propagation distances. This difficulty is due to the fact that arrival time is a function of propagation distance, which causes different time shifts for different propagation distances. It is possible to normalize these Lamb waves with respect to propagation distance by considering either the group velocity–frequency representation, or the slowness-frequency representation (SFR). Define group velocity, c_g, for any propagation distance, d, and time, t (for each frequency) as $c_g=d/t$, while (energy) slowness, sl_e, is defined as $sl_e=t/d$.

The calculation of group velocity from a TFR (like the reassigned spectrogram) with a fixed propagation distance d is a nonlinear operation in t that transforms an equally spaced time grid to an unequally spaced grid in group velocity, c_g. This is in contrast to the slowness transformation, which is a linear operation in t—the equally spaced time grid is transformed to an equally spaced grid in (energy) slowness, sl_e. Transformation nonlinearities are avoided by selecting the SFR to represent the experimentally measured time–domain Lamb wave signals.

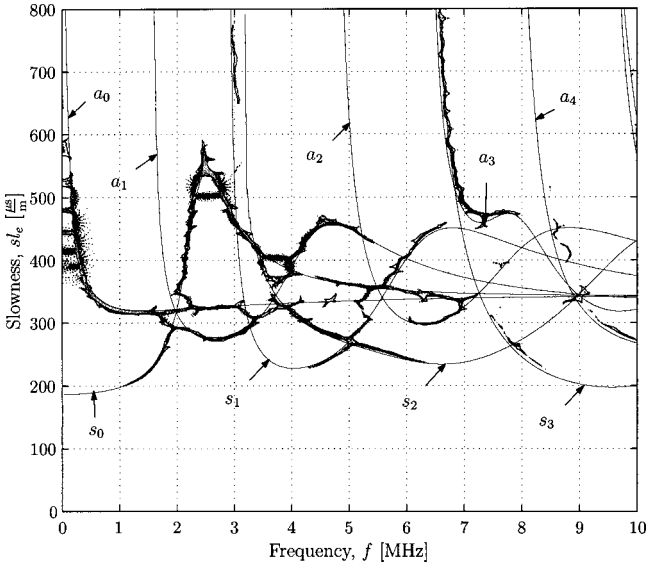


FIG. 2. SFR of perfect plate.

III. EXPERIMENTAL INTERACTION OF A LAMB WAVE WITH A NOTCH

Figure 2 shows the square root (for better visibility) of the energy density spectrum of the SFR (obtained from the reassigned spectrogram) measured in the perfect plate for a propagation distance of 70 mm, together with the theoretical Rayleigh–Lamb solution,[21] presented as solid lines and identified as either symmetric, s_i, or antisymmetric, a_i, modes. There is excellent definition of seven experimentally measured modes (s_0-s_2 and a_0-a_3) and all these experimental modes match the theoretical solution (no experimental modes are unaccounted for). Interference during the reassignment procedure causes some ''fuzziness'' in the vicinity of the intersection of modes.[3] Note that Fig. 2 is also a representation of the propensity of a particular mode to be excited by the laser source (and noting that the laser interferometer used in this study only measures out-of-plane motion), meaning that not all portions of a mode (e.g., a_1 below 2 MHz) can be considered in this study. However, Fig. 2 can be thought of as the SFR of the Lamb wave that is incident on the notch (in the notched plate specimen).

Now consider SFRs for the notched plate. Note that the notch location is now assumed to be *known* in order to study the scattering phenomena; this knowledge will then be used to determine an unknown notch location in Sec. IV. Figures 3(a)–(b) show the SFRs for propagation distances of 70 mm (d_1) and 130 mm (d_4)—simultaneously measured by probes 1 and 2, respectively—together with the theoretical solution for the *perfect plate*, presented as solid lines. Modes above a slowness of about 650 μs/m in Fig. 3(b) that do not match any theoretical mode lines are due to reflections from the plate edges, are spurious, and are not included in the following discussions. A comparison of SFRs of the perfect plate and the notched plate shows that a significant portion of the

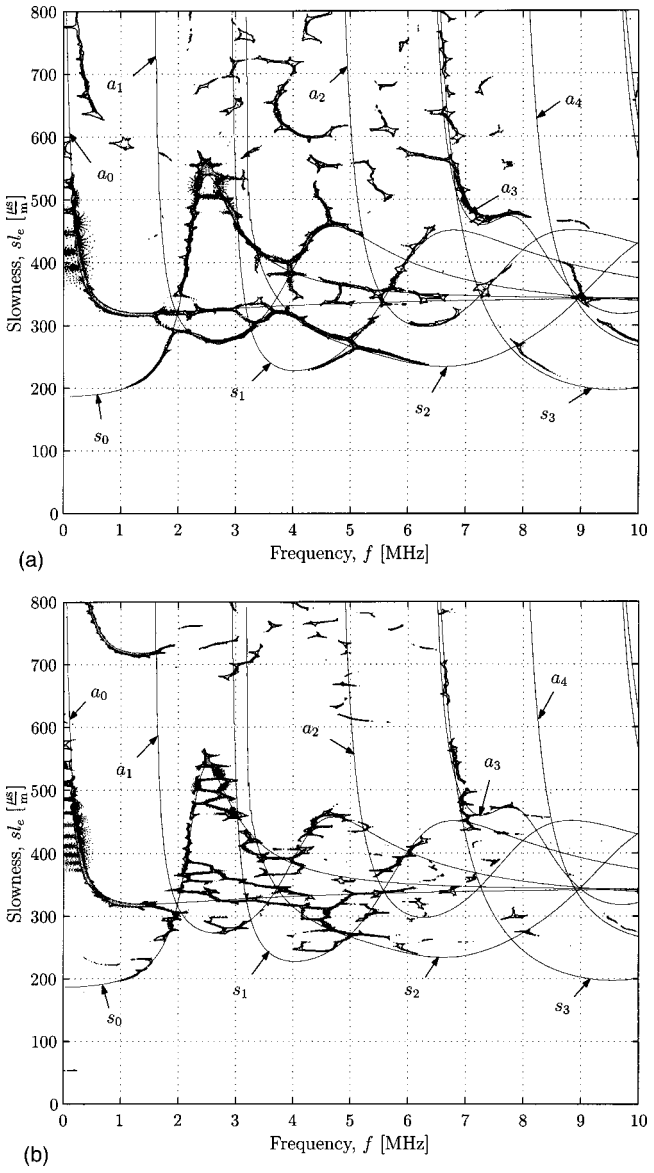


FIG. 3. (a). SFR of notched plate, Probe 1, source to receiver distance (d_1) of 70 mm. (b) SFR of notched plate, Probe 2, source to receiver distance (d_4) of 130 mm.

energy in the notched plate still corresponds to the theoretical solution of the perfect plate. However, the notched plate SFRs show additional modes that are not present in the perfect plate SFR, and these additional modes do not fit any theoretical mode lines.

First, compare the probe 1 signal of the notched plate [Fig. 3(a)] with the perfect plate (Fig. 2). The SFR of the notched plate has additional modes for slownesses larger than 500 μs/m that are the contribution of the Lamb wave that is reflected from the notch. The perfect plate SFR contains only the incident signal, while the probe 1 notched plate SFR contains both the incident and reflected (from the notch) signals. The reflected signal occurs at a later time (higher slowness) because it has a longer propagation distance. If the probe 1 TFR was transformed with a propagation distance, d, that corresponds to the source to notch to receiver distance, d_5, instead of the source to receiver distance [d_1 of 70 mm used to calculate the SFR in Fig. 3(a)], the resulting SFR would highlight the reflected modes.[19] The reflected (and nonconverted) modes in this SFR would coincide with the theoretical modes. This ''reflected'' SFR would also contain modes that do not fit the theoretical solution—referred to as *extraneous modes*—which will be discussed in detail later.

Now, compare the probe 2 signal for the notched plate [Fig. 3(b)] with the perfect plate (Fig. 2). The probe 2 SFR only contains the transmitted portion of the signal, and Fig. 3(b) shows modes that do not fit the theoretical solution. These modes (e.g., between 300–500 μs/m, 2–4 MHz) are also identified as extraneous modes, and are due to the interaction of the incident Lamb wave with the notch.

An explanation of extraneous modes is based on a hypothesis that treats the notch as an additional (second) source that ''creates'' the reflected and transmitted modes. In addition, this development recognizes that these reflected and transmitted modes all propagate within the same plate, so they should all be able to be represented as the theoretical modes of the perfect plate—all modes must satisfy the Rayleigh–Lamb equations of the perfect plate. These extraneous modes violate this requirement (they do not fall on a theoretical mode line) because the SFRs are calculated under the (possibly incorrect) assumption that the same mode propagates from the laser source to the notch, and then from the notch to the receiver (either probe 1 or probe 2). If this were truly the case, then there would be no extraneous modes present.

The scattering of Lamb waves by a notch is a complicated process, with mode conversion, and a number of other phenomena possible. Mode conversion in this context means that the energy propagates with mode i during the propagation from the laser source to the notch, and with a different mode, j, during the propagation from notch to receiver. Internal reflections (within the notch),[11] nonpropagating modes,[13] and local frequency (or phase) shifts are additional phenomena that can also be part of this scattering process. Instead of approaching this problem from an energy conservation approach, this research develops a systematic procedure to identify which modes (and through which frequencies) are transmitted and reflected, with or without mode conversion. Note that the proposed procedure does not consider internal reflections or local frequency shifts, but the success of the proposed localization technique (presented in Sec. IV) shows that these contributions are probably relatively small.

Consider a nomenclature where signals measured in the notched plate with probe 1 are classified as case P1a, incident; case P1b, reflected and non-mode-converted; and case P1c, reflected and mode converted. Signals measured in the notched plate with probe 2 are classified as case P2b, transmitted and non-mode-converted; and case P2c, transmitted and mode converted.

Since both the incident and the reflected signals are present in probe 1 measurements, case P1 is inherently more complicated, so examine case P2 (transmitted only) first. Consider a procedure that is used to accentuate the modes that coincide with the theoretical solution. This procedure entails multiplying a SFR with a *theoretical mode matrix*, $T(sl_e, f)$—this theoretical mode matrix is a three-

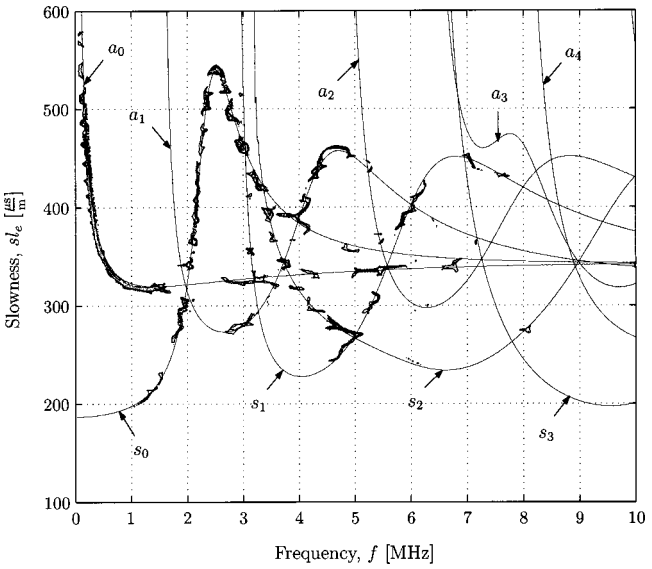


FIG. 4. SFR of modes belonging to case P2b, transmitted and non-mode-converted.

dimensional representation of the Rayleigh–Lamb equations that accounts for possible measurement uncertainties. $T(sl_e, f)$ equals unity on the set of slowness-frequency pairs that correspond to the solution of the Rayleigh–Lamb equations, and is smoothed out by Gaussian filtering in the slowness and frequency directions. Each mode is filtered individually, the respective maxima (the ridge values) are set equal to unity, and the individual modes are combined by taking the maximum over all modes (this guarantees a value of unity at mode intersections). The *inverse theoretical mode matrix*—a value of zero at the theoretical mode values (Gaussian curve), unity everywhere else—is obtained by subtracting the theoretical mode matrix from a matrix completely populated with ones.

Modes that belong to case P2b are comparatively easy to identify; modes that lie close to the theoretical mode lines in the SFR of the probe 2 signal, calculated with propagation distance d_4, are not mode converted at the notch. Thus, the SFR of Fig. 3(b) is multiplied by the theoretical mode matrix, and Fig. 4 shows the result—modes that are transmitted and not mode converted.

Modes not assigned to case P2b are considered to be extraneous and portions of these modes are systematically identified as belonging to case P2c. First, remove case P2b modes from the signal measured by probe 2 by multiplying the SFR in Fig. 3(b) with the inverse theoretical mode matrix. Define the propagation time from source to probe 2 as time t_4, and express this time as the sum of the time from source to notch, t_3, plus the time from notch to probe 2, t_2, or $t_4 = t_3 + t_2$. The slowness associated with the propagation over distance d_3, from source to notch, is defined as $sl_{e3} = t_3/d_3$, and the slowness associated with the propagation from notch to probe 2 is $sl_{e2} = t_2/d_2$. Finally, the slowness from source to probe 2 is $sl_{e4} = t_4/d_4$—this is the slowness shown in the SFR of Fig. 3(b). These slownesses are combined, and, after some simple algebra,

$$sl_{e3} = \frac{d_4 sl_{e4} - (d_4 - d_3) sl_{e2}}{d_3}. \tag{3}$$

Equation (3) relates the *unknown* slowness from source to notch (sl_{e3}), to the *known* slowness from source to receiver (sl_{e4}), and an *assumed* slowness from notch to receiver (sl_{e2}). The unknown slowness sl_{e3} can be calculated for every possible slowness-frequency combination through the relevant slowness range and frequency bandwidth. Consider only six modes—a_0, a_1, a_2, s_0, s_1, and s_2—these modes carry most of the experimentally measured energy. Six new SFRs are calculated (one for each of these six modes), each assuming all of the energy propagates from the notch to probe 2 with that particular mode. The slowness-frequency values for these modes are obtained from the theoretical solution[21] and go into Eq. (3) as the assumed sl_{e2}. All distances and sl_{e4} are *known* and thus the slowness sl_{e3} can be calculated for each frequency.

An algorithm that follows this procedure is programmed into MATLAB to develop an automated and objective way to (possibly) allocate the extraneous modes into one of these six possible transmitted modes. First, separate the modes that belong to case P2c, and then pick any point on the P2c SFR and do the following.

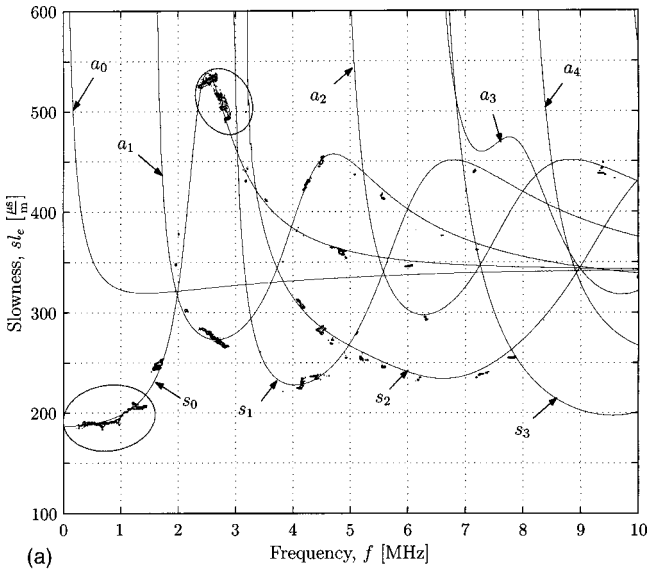
(1) Compute sl_{e3} for this point for each of the six possible modes: $a_0 - a_2$ and $s_0 - s_2$.

(2) Next, compare the magnitude of the theoretical mode matrix, $T(sl_{e3}, f)$, at each of the six slowness-frequency pairs calculated in step 1. Select the slowness-frequency pair that has the largest magnitude. This defines the mode type propagating from laser source to notch. This is a modified probabilistic approach that searches for the mode that is most likely to be the match, since $T(sl_{e3}, f)$ is similar to a two-dimensional probability distribution function, although not in the strict mathematical sense.

(3) The entire propagation path is now defined—from laser source to notch with the slowness corresponding to the mode selected in step (2), and then from the notch to the receiver (probe 2), by looking at the mode that it is mapped from.

For example, Fig. 5(a) shows the subset of the original extraneous modes, at their new slowness-frequency coordinates, which fit the theoretical solution, assuming that mode a_0 propagates from the notch to probe 2. The two circled regions show where significant energy matches the theoretical mode matrix—these regions propagate as mode s_0 from laser source to notch, and are converted at the notch to a_0 (and propagate as a_0 from notch to probe 2). Figure 5(b) shows the same modes as Fig. 5(a), but at their original slowness-frequency coordinates. Note that the energy of the incident s_0 mode below 1 MHz is much lower than the incident energy in the a_0 mode in the same frequency range, so this incident s_0 contribution is not obvious in Fig. 2. Figure 6 is a summary of the contributions of all six possible modes, and represents case P2c, with the circles specifying how the original incident Lamb wave is mode converted at the notch.

Now consider the probe 1 measurements that contain the incident and reflected signals. The incident modes (case P1a) are the SFR of probe 1, normalized with propagation dis-



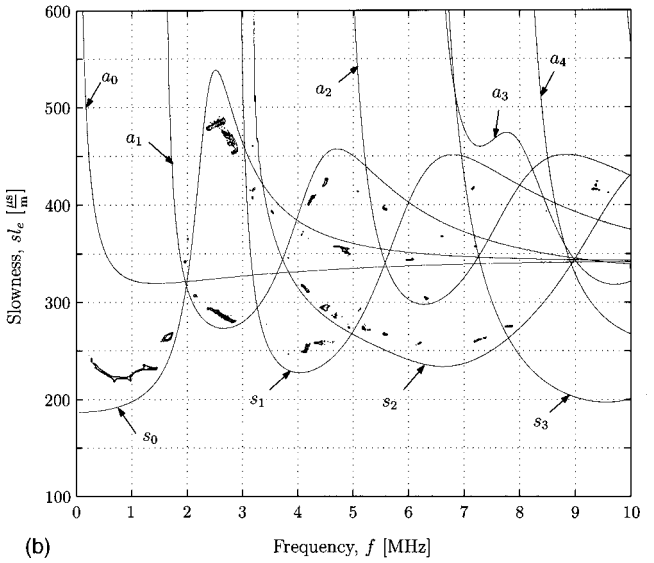


FIG. 5. (a). Subset of extraneous modes (transmitted) assuming mode conversion to mode a_0, new coordinates. (b) The subset of extraneous modes (transmitted) assuming mode conversion to mode a_0, original coordinates.

tance d_1 and shown in Fig. 3(a). The modes in Fig. 3(a) that are close to the theoretical mode lines can be accentuated by multiplying this SFR with the theoretical mode matrix and the resulting SFR is almost identical to Fig. 2, the SFR for the perfect plate.[19] The reflected and non-mode-converted portion (case P1b) of the probe 1 signal is identified by calculating a SFR with the propagation distance d_5, and multiplying with the theoretical mode matrix (to accentuate the modes coinciding with the theoretical solution). These case P1b modes are shown in Fig. 7.

In a procedure similar to case P2c, extraneous modes that cannot be allocated to case P1a (incident modes) or case P1b (reflected, non-mode-converted) are extracted by multiplying the SFR of the reflected signal with the inverse theoretical mode matrix. An automated algorithm similar to the one used for case P2c is developed for case P1c, using a modified version of Eq. (3). The SFR that shows the modes belonging to case P1c, including their converted mode assignment (notch to probe 1) is shown in Fig. 8.

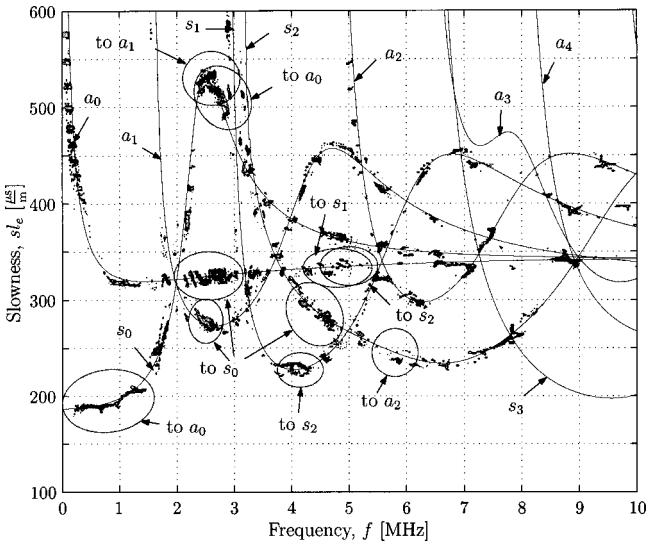


FIG. 6. SFR of modes belonging to case P2c, transmitted and converted.

The non-mode-converted, transmitted, and reflected SFRs (Figs. 4 and 7) are very similar to each other and do not appear to show any definite pattern that can be associated with notch width or depth. There are subtle differences—such as mode a_1 from 2–3 MHz and mode s_1 below 4 MHz are reflected but not transmitted—but it is difficult to make a quantitative interpretation as to how the incident Lamb wave is scattered into reflected and transmitted portions without a predictive theoretical model. However, this comparison is not an objective of this research. More importantly, note that the mode conversion at the notch for both the transmitted and reflected cases are very similar to each other—compare Figs. 6 and 8. This behavior validates the hypothesis that the notch acts as an additional source. Finally, note that a vast majority of the extraneous mode energy is identified with mode conversion,[19] so the amount of incident signal energy that is scattered at the notch as a local frequency shift or as an internal reflection is probably small in comparison.

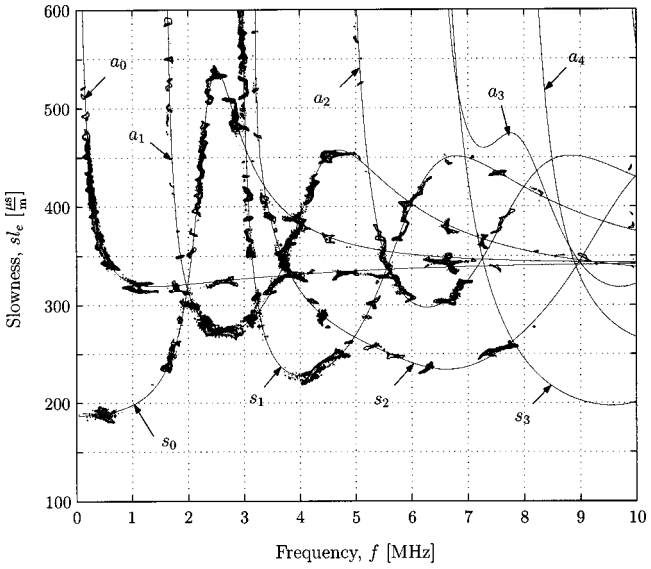


FIG. 7. SFR of modes belonging to case P1b, reflected and non-mode-converted.

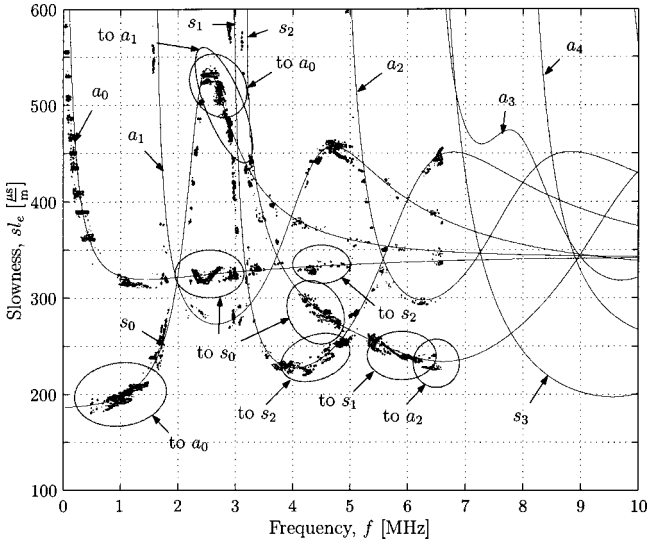


FIG. 8. SFR of modes belonging to case P1c, reflected and converted.

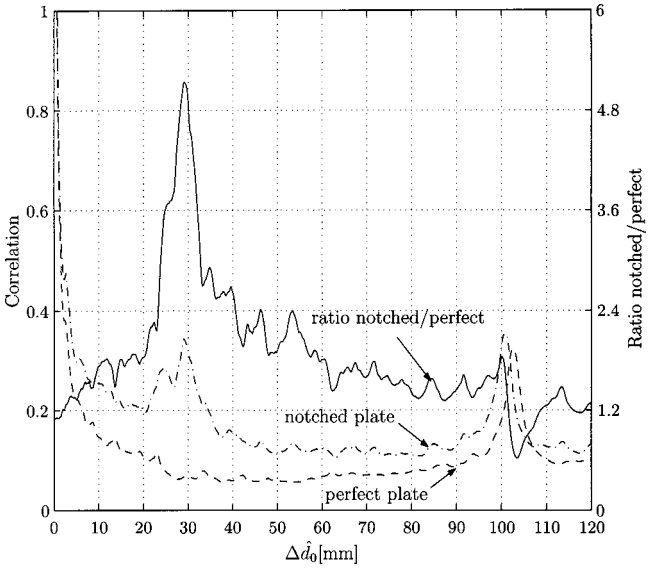


FIG. 9. Correlation curves for the perfect plate, notched plate, and a division of both curves, 0–10 MHz frequency bandwidth, reflected contribution.

IV. LOCALIZATION OF THE NOTCH

An automated localization methodology is now developed based on the understanding of the interaction of a Lamb wave with a notch—note that the notch location is now assumed to be unknown. Previous research[8] formulated a correlation technique that uses the non-mode-converted parts of the reflected modes to locate a notch in a plate. This correlation procedure does not require (nor was it developed with) an understanding of the scattering of a Lamb wave by a notch, and this procedure identifies the position where the non-mode-converted portions (and only these portions) of the reflected modes coincide with the incident modes. The current research uses an understanding of scattering to extend this technique, using, for example, a targeted frequency bandwidth to help refine the localization procedure.

Up until this point, the current research has neglected the width of the notch (the notch is treated as being infinitely thin), and has defined the location of the notch by its centerline (see Fig. 1). The proposed localization technique will be more accurate if the finite width of the notch is taken into account. Introduce three new distances ($\hat{d}_0, \hat{d}_3, \hat{d}_5$) that are based on the near edge (in relation to the laser source) of the notch: $\hat{d}_0 = d_0 - w/2$, $\hat{d}_3 = d_3 - w/2$, and $\hat{d}_5 = d_5 - w$, where w is the notch width. Note that there is no difference in the *visual appearance* of a SFR that is calculated with $\hat{d}_i$ instead of d_i—this small change in propagation distance has no visible effect on the SFRs presented in Figs. 2–8.

First, localize the notch with the reflected wave field by considering the probe 1 signal. Assume that $\Delta\hat{d}_0$ is the unknown distance from probe 1 to the notch. By systematically varying $\Delta\hat{d}_0$ (and the associated propagation distance $d = d_1 + 2\Delta\hat{d}_0$), SFRs are calculated for each distance. Next, correlate each of these SFRs with the incident modes (the SFR calculated with distance d_1), or

$$\mathrm{Corr}(d) = \sum_{sl_e} \sum_{f} \mathrm{SFR}(sl_e, f, d_1) \times \mathrm{SFR}(sl_e, f, d), \quad \forall d. \tag{4}$$

Obviously, this correlation reaches its maximum for a representation calculated with $\Delta\hat{d}_0 = 0$ (the correlation of two identical signals), but reflections (from both the notch and any plate edges) will introduce local maxima in the correlation curve at certain $\Delta\hat{d}_0$'s. These local maxima occur when the reflected modes within the SFR coincide with the incident modes—which, in turn, provides a measure of the receiver to notch distance. At $\Delta\hat{d}_0$ equal to the exact receiver to notch distance $\Delta\hat{d}_0 = \hat{d}_0$, the reflected (but not mode converted, case P1b) contribution matches the incident wave field.

Figure 9 shows the correlation of the SFRs of the notched plate (the dot–dashed line) for $\Delta\hat{d}_0$ varying between 0 and 120 mm with an increment step of 0.2 mm. Each of these SFRs (calculated with a propagation distance $d = d_1 + 2\Delta\hat{d}_0$) are correlated with the same SFR, calculated with $d = d_1$, as defined by Eq. (4). This procedure is repeated for the perfect plate (correlated with the incident modes of the perfect plate) and shown as the dashed line in Fig. 9. The solid line in Fig. 9 represents the ratio of the notched plate correlation curve divided by the perfect plate correlation curve. This ratio curve emphasizes the local maxima caused by features that exist in the notched plate, but not in the perfect plate; maxima in the perfect plate can only be caused by the edges, while maxima in the notched plate can be caused by the notch or the edges. As a result, the maximum of the ratio curve represents the reflection from the notch—the ratio curve has a single dominant peak at $\Delta\hat{d}_0 = 29$ mm (the actual distance is $\hat{d}_0 = 29.2$ mm). This correlation procedure calculates the receiver to notch distance with outstanding accuracy, and the error (0.2 mm) is on the order of the tolerance of the width of the notch.

Further improvement in localization accuracy is achieved using a frequency-limited bandwidth (such as from 0–2 MHz) of the SFRs for the correlations—the results are shown in Fig. 10. Note that the correlations in Fig. 9 use the

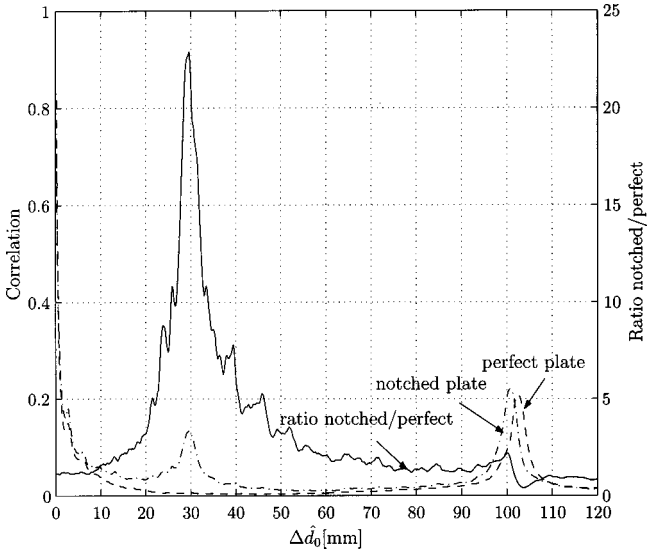


FIG. 10. Correlation curves for the perfect plate, notched plate, and a division of both curves, 0–2 MHz frequency bandwidth, reflected contribution.

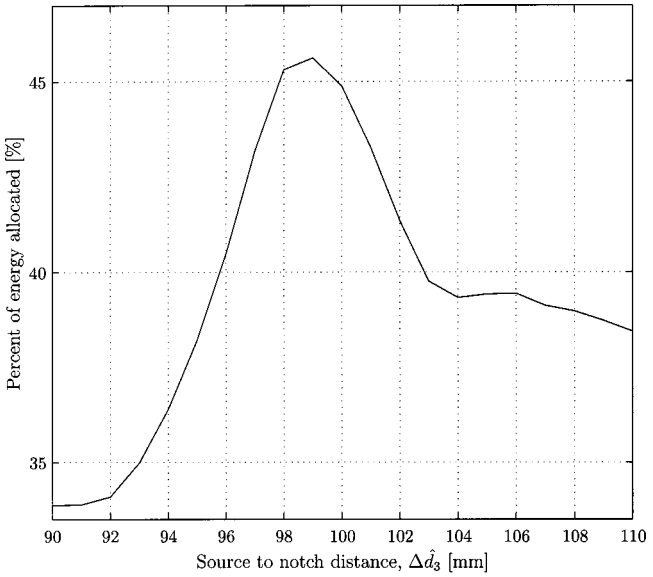


FIG. 11. Goodness-of-fit (as a function of percent allocated) of the allocation of extraneous modes to case P2c, used to locate the notch with a transmitted contribution.

entire frequency bandwidth, 0–10 MHz. This particular (limited-) frequency bandwidth is selected because the case P1b (reflected and non-mode-converted) SFR in Fig. 7 shows significant energy in the frequency bandwidth from 0–2 MHz—the same is also true for the SFR of the incident modes. The peak of the curve representing the ratio of notched/perfect plate in Fig. 10 is still at the same location, but it is now even sharper and its value is almost five times higher than in Fig. 9—the sharpness and value of the maximum are measures of the preciseness of the results. The significance of this frequency bandwidth (0–2 MHz) becomes more evident when correlation curves for the remaining bandwidth (2–10 MHz) are considered, which results in a much wider peak, on the order of 8 mm.[19] Note that the correlation curves in Figs. 9 and 10 are more accurate and definitive than the results presented in Hurlebaus *et al.*[8] These improvements are due to the use of the reassigned spectrogram in this study (as opposed to the unreassigned spectrogram in Hurlebaus *et al.*[8]) as the basis for the SFRs, and the ability to identify a targeted frequency range that contains significant P1b energy.[19]

Now, consider a second localization technique that uses the transmitted wave field of probe 2. Unfortunately, a correlation technique in terms of the case P2b transmitted, non-mode-converted modes is not possible, because the case P2b SFR is calculated with distance d_4, which is independent of notch location. Instead, consider a technique based on the goodness-of-fit when allocating the extraneous modes to case P2c, transmitted and mode converted. Assume that $\Delta\hat{d}_3$ is the unknown source to notch distance; varying $\Delta\hat{d}_3$ can be used to define a measure (as a function of distance) of how well portions of the extraneous modes can be allocated to case P2c. For each $\Delta\hat{d}_3$, a modified version of the allocation algorithm for case P2c [that uses Eq. (3) and is described in the previous section] is performed on the transmitted signal. Specifically, the goodness-of-fit for any $\Delta\hat{d}_3$ is given by summing the maximum possible magnitudes of the theoretical mode matrix, $T(sl_e, f)$, for the mapping of all the slowness-frequency pairs specified by the extraneous P2c modes being classified. Dividing this value by the number of slowness-frequency pairs to be mapped is the goodness-of-fit measure. A value of 100% means perfect mapping, with all the extraneous modes mapped perfectly to the theoretical solution. In reality, lower percentages will be obtained because of experimental uncertainties and the possibility that an extraneous mode is mapped to, but not exactly on top of the theoretical Rayleigh–Lamb solution.

Figure 11 shows the curve representing the percentage of energy that is allocated to the extraneous modes, as a function of source to notch distance, $\Delta\hat{d}_3$—there is a clear peak near the correct distance, $\Delta\hat{d}_3 = d_3 = 99.2$ mm. Note that this technique does not require an experimental signal from the perfect plate, but it does need the theoretical solution of Lamb modes in a perfect plate. Additionally, the success of this technique validates the accuracy of the assignment of the extraneous modes into case P2c in Sec. III.

V. CONCLUSION

This research establishes the effectiveness of combining laser ultrasonic techniques with a SFR to locate a notch in a plate. The high fidelity, broad-bandwidth, point-like, and noncontact nature of laser ultrasonics are critical to the success of this study. The dual-probe laser interferometer allows for the simultaneous interrogation of both the reflected and transmitted wave fields. The distance invariance of a SFR enables a quantitative interpretation of the scattering of a broadband, incident Lamb wave by a notch. The systematic (and unbiased) mode allocation procedure developed in this research establishes which modes are transmitted and reflected by the notch (and through which frequencies), and identifies which modes are mode converted.

Two automated localization methodologies are developed using this understanding of the scattering of Lamb waves. One technique isolates the contributions of the signal

reflected from the notch by performing a correlation of a series of SFR spectra, each calculated with different, assumed propagation distances. This correlation technique is refined with an understanding of which modes are reflected, but not mode converted (and through which frequencies). A second localization technique uses a goodness-of-fit metric when determining how the transmitted extraneous modes are mode converted. A combination of these two localization techniques is very robust, because it can use either the reflected or transmitted wave fields, or both—this allows for two independent measures of notch location, and can interrogate a variety of source, receiver, and notch, geometric relationships.

ACKNOWLEDGMENT

The Deutscher Akademischer Austausch Dienst (DAAD) provided partial support to Rüdiger Benz.

[1] D. E. Chimenti, ''Guided waves in plates and their use in materials characterization,'' Appl. Mech. Rev. **50**, 247–284 (1997).
[2] L. Cohen, *Time–Frequency Analysis* (Prentice–Hall, Englewood Cliffs, NJ, 1995).
[3] M. Niethammer, L. J. Jacobs, J. Qu, and J. Jarzynski, ''Time–frequency representations of Lamb waves,'' J. Acoust. Soc. Am. **109**, 1841–1847 (2001).
[4] W. H. Prosser, M. D. Seale, and B. T. Smith, ''Time–frequency analysis of the dispersion of Lamb modes,'' J. Acoust. Soc. Am. **105**, 2669–2676 (1999).
[5] M. Lemistre and D. Balageas, ''Structural health monitoring based on diffracted Lamb waves analysis by discrete wavelet transform,'' in *System Identification and Structural Health Monitoring*, edited by A. Guemes (Graficas, Madrid, 2000), pp. 561–569.
[6] P. D. Wilcox, R. P. Dalton, M. J. S. Lowe, and P. Cawley, ''Mode selection and transduction for structural monitoring using Lamb waves'' in *Structural Health Monitoring 2000*, edited by F.-K. Chang (Technomics, New York, 1999), pp. 703–712.
[7] C. Valle, M. Niethammer, J. Qu, and L. J. Jacobs, ''Crack characterization using guided circumferential waves,'' J. Acoust. Soc. Am. **110**, 1282–1290 (2001).
[8] S. Hurlebaus, M. Niethammer, L. J. Jacobs, and C. Valle, ''Automated methodology to locate notches with Lamb waves,'' Acoust. Res. Lett. Online **2**, 97–102 (2001).
[9] S. W. Liu, S. K. Datta, and T. H. Ju, ''Transient scattering of Rayleigh–Lamb waves by a surface-breaking crack: comparison of numerical simulation and experiment,'' J. Nondestruct. Eval. **10**, 111–126 (1991).
[10] Y. Cho and J. L. Rose, ''An elastodynamic hybrid boundary element study for guided wave interactions with a surface breaking defect,'' Int. J. Solids Struct. **37**, 4103–4124 (2000).
[11] M. J. S. Lowe and O. Diligent, ''Low-frequency reflection characteristics of the s_0 Lamb wave from a rectangular notch in a plate,'' J. Acoust. Soc. Am. **111**, 64–74 (2002).
[12] O. Diligent, T. Grahn, A. Boström, P. Cawley, and M. J. S. Lowe, ''The low-frequency reflection and scattering of the S_0 Lamb mode from a circular through-thickness hole in a plate: Finite element, analytical and experimental studies,'' J. Acoust. Soc. Am. **112**, 2589–2601 (2002).
[13] M. Castaings, E. Le Clezio, and B. Hosten, ''Modal decomposition method for modeling the interaction of Lamb waves with cracks,'' J. Acoust. Soc. Am. **112**, 2567–2582 (2002).
[14] D. Alleyne and P. Cawley, ''A two-dimensional Fourier transform method for measurement of propagating multimode signals,'' J. Acoust. Soc. Am. **89**, 1159–1168 (1991).
[15] C. Eisenhardt, L. J. Jacobs, and J. Qu, ''Application of laser ultrasonics to develop dispersion curves for elastic plates,'' J. Appl. Mech. **66**, 1043–1045 (1999).
[16] C. B. Scruby and L. E. Drain, *Laser Ultrasonics: Techniques and Applications* (Adam Hilger, Bristol, 1990).
[17] D. A. Bruttomesso, L. J. Jacobs, and R. D. Costley, ''Development of an interferometer for acoustic emission testing,'' J. Eng. Mech. **119**, 2303–2316 (1993).
[18] S. Hurlebaus, ''A contribution to structural health monitoring using elastic waves,'' Doctoral thesis, University of Stuttgart, Stuttgart, Germany, 2002.
[19] R. Benz, ''Localization of notches with Lamb waves,'' Masters thesis, Georgia Institute of Technology, Atlanta, Georgia, 2002.
[20] F. Auger and P. Flandrin, ''Improving the readability of time–frequency and time-scale representations by the reassignment method,'' IEEE Trans. Signal Process. **43**, 1068–1089 (1995).
[21] R. D. Mindlin, ''Waves and vibrations in isotropic elastic plates,'' in *Structural Mechanics*, edited by J. N. Goodier and N. J. Hoff (Pergamon, New York, 1960).

Determination of unsteady heat release distribution from acoustic pressure measurements: A reformulation of the inverse problem

P. Bala Subrahmanyam,[a] R. I. Sujith,[b] and M. Ramakrishna[c]
Department of Aerospace Engineering, Indian Institute of Technology Madras, Chennai 600036, India

(Received 18 December 2002; revised 23 May 2003; accepted 27 May 2003)

An integral method is developed to solve the inverse problem of determining the oscillatory heat release distribution from the knowledge of the acoustic pressure field within a combustor. Unlike earlier approaches, in which the problem is formulated in terms of Fredholm integral equation, the inverse problem is reformulated in terms of Volterra integral equation. This reformulation, valid for low Mach numbers ($M^2 \ll 1$), facilitates the recovery of heat release at all frequencies. The resulting Volterra integral equation is solved using both direct numerical method and implicit least-squares method. The results show that the implicit least-squares method is superior to the direct numerical method and yields accurate determination of heat release at all frequencies. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1592524]

PACS numbers: 43.28.Kt [MSH]

I. INTRODUCTION

The occurrence of combustion instabilities has been a plaguing problem in the development of combustors for rockets, jet engines, and power generating gas turbines.[1] Predicting and controlling combustion instability requires an understanding of the interactions between the combustion process and the acoustic waves. In this context, the knowledge of oscillatory heat release distribution and its relation with the oscillatory pressure is very essential for the understanding of any combustor.

To develop a thorough understanding of these combustion–acoustic interactions, the quantification of various physical processes occurring in the combustion zone is needed. This would require measuring unsteady pressures, velocities, temperatures, and species concentrations along the length of the combustor. In practice, it is very difficult to measure these unsteady quantities with the currently available techniques. However, with the existing techniques, it is fairly easy to measure the oscillatory pressure. Therefore, the possibility of recovering oscillatory heat release distribution using acoustic pressure measurements alone needs to be considered. This is referred to as the inverse problem of determining the oscillatory heat release distribution and it belongs to a general class of inverse problems.

In such inverse problems, we try to recover the sources by measuring the effect produced by the sources. Some of the other applications include digital tomography, detection of tumors using ultrasound techniques, geological/oceanographic analysis using seismography/sonar techniques, inverse heat conduction problem, etc. The characteristic of most of the inverse problems is that they are likely to be stiff or even ill posed. The solution of the problem is very sensitive to the actual data values, i.e., small errors in the measured data values can produce very large errors in the solution. The inverse problems have been a subject of extensive study in recent times due to their application in various engineering disciplines.[2]

The inverse problem of acoustic radiation was first addressed by Ramachandra and Strahle.[3] They attempted to recover the heat release distribution in open premixed turbulent flames from acoustic pressure measurements. This was done by neglecting the mean flow effects and using linear acoustic theory. The inverse problem was then reduced to solving a Fredholm integral equation of the first kind. The resulting equation was solved using the augmented Galerkin method, where two constraints had to be artificially adjusted. The accuracy of the results was assessed by comparison with C_2 radical emission studies. They concluded that in principle the oscillatory heat release can be successfully recovered from acoustic pressure measurements.

Chao[4] attempted to extend the inverse formulation to recover the heat release distribution in a gas turbine combustor. Unlike the Galerkin method, where two artificial constraints have to be specified, he used a parameter free implicit least-squares method. In this method, the solution is approximated by a linear combination of a set of known functions referred to as basis functions. The coefficients of these basis functions are evaluated by minimizing the error function. He showed that this is one of the best linear estimations of the ill-posed problem and is equivalent to the pseudo-inverse method.[5] He found that the condition number (which gives an indication of stiffness of the problem) of the resulting matrix is less than that obtained by the direct inversion process. Therefore, this method gives more accurate predictions than the Galerkin method for open premixed flames. However, in the case of confined systems such as a gas turbine combustor, he was able to recover heat release successfully only in certain frequency bands. In other frequency bands the recovery was of poor accuracy. This, he

[a]Graduate student; electronic mail: bala_aerospace@yahoo.co.in
[b]Author to whom correspondence should be addressed; Assistant Professor; electronic mail: sujith@aero.iitm.ernet.in
[c]Associate Professor; electronic mail: krishna@aero.iitm.ernet.in

concluded was due to high sensitivity of the solution to the noise in the boundary condition.

Subsequently, Lieuwen *et al.*[6] addressed this problem and gave a plausible explanation for the poor accuracy in certain frequency bands. Using an example, they illustrated that the frequency bands that rendered poor results corresponded to the natural acoustic modes of the combustor. When formulated as a Fredholm integral equation of the first kind, they argued that the minimum and maximum sensitivity of the solution occurs at off-resonant and resonant frequencies, respectively. Since in ducted systems, combustion instabilities generally occur when the combustion process excites the natural modes, they concluded that it is not feasible to recover the oscillatory heat release from acoustic pressure measurements in these systems using the investigated technique.

Clearly, to recover the heat release distribution for all frequency bands, it is necessary to adopt a different approach for the problem. This paper explores the possibility of reformulating the inverse problem in terms of Volterra integral equations. In Sec. II, the governing differential equation for an inviscid, nonheat conducting, and perfect gas is derived in the presence of oscillatory heat release. In Sec. III, two approaches are presented to formulate the inverse problem in terms of an integral equation for unknown heat release distribution. First, the differential equation is transformed to a Fredholm integral equation of the first kind using Green's function approach and the nonfeasibility of the method at resonance frequencies is discussed. As an alternative approach, the inverse problem is reformulated in terms of the Volterra integral equation. Next, a numerical method is described to solve this integral equation and the stability and sensitivity issues are discussed in detail. In Sec. III B 2, an implicit least-squares method is described to solve the Volterra integral equation. In Sec. III B 3, the implicit least-squares formulation is generalized to include the effects of variable Mach number. In Sec. IV, the application of the above-mentioned methods to recover heat release distribution in a combustor is illustrated using examples, and the superiority of the implicit least-squares method in yielding accurate determination of the heat release distribution is established.

II. THEORY

This section describes the theory and the derivation of the required governing equations. For the sake of simplicity, the following assumptions are made:

(a) All the physical variables are assumed to be functions of the axial distance alone.
(b) The variation of the cross-sectional area along the length of the combustor is neglected.
(c) Inviscid, nonheat conducting, and perfect gas is assumed in the combustor.

With the aid of the above assumptions, the one-dimensional conservation equations and the equation of state can be written in the following form.[7]

Momentum:

$$\rho \frac{\partial u}{\partial t} + \rho u \frac{\partial u}{\partial x} = -\frac{\partial p}{\partial x}. \tag{1}$$

Energy (in modified form):

$$\frac{\partial p}{\partial t} + u \frac{\partial p}{\partial x} + \gamma p \frac{\partial u}{\partial x} = (\gamma - 1) q. \tag{2}$$

State:

$$p = \rho R T, \tag{3}$$

where ρ, u, p, T, q are the density, axial velocity, pressure, temperature, and rate of heat addition per unit volume, respectively. Expressing any physical variable $v(x,t)$ as the sum of a mean quantity, $\overline{v}(x)$, and a fluctuating quantity, $v'(x,t)$, [where $v'(x,t)/\overline{v}(x) \ll 1$], one can write

$$\rho(x,t) = \overline{\rho}(x) + \rho'(x,t), \tag{4}$$

$$p(x,t) = \overline{p}(x) + p'(x,t), \tag{5}$$

$$u(x,t) = \overline{u}(x) + u'(x,t). \tag{6}$$

Substitution of the above-noted expressions in Eqs. (1)–(3) yields the steady conservation equations and equation of state as

$$\overline{p}(1 + \gamma M^2) = \text{constant}, \tag{7}$$

$$\overline{u}\frac{d\overline{p}}{dx} + \gamma \overline{p}\frac{d\overline{u}}{dx} = (\gamma - 1)\overline{q}(x), \tag{8}$$

$$\overline{p} = \overline{\rho} R \overline{T}, \tag{9}$$

where $M = u/c$ is the Mach number. For very low Mach numbers, $M^2 \ll 1$, the steady momentum equation implies a negligible variation of mean pressure along the length of the combustor. Therefore, from Eq. (9) we can write $(1/\overline{\rho})d\overline{\rho}/dx = -(1/\overline{T})d\overline{T}/dx$. Instead of solving the steady state energy equation with chemical reactions and wall heat transfer taken into account, a corresponding mean temperature distribution can be prescribed along the length of the combustor.

The first-order acoustic momentum and energy equations can then be reduced to

$$\overline{\rho}\frac{\partial u'}{\partial t} + \overline{\rho u}\frac{\partial u'}{\partial x} + \overline{\rho} u' \frac{d\overline{u}}{dx} + \rho' \overline{u}\frac{d\overline{u}}{dx} = -\frac{\partial p'}{\partial x}, \tag{10}$$

$$\frac{\partial p'}{\partial t} + \overline{u}\frac{\partial p'}{\partial x} + \gamma \overline{p}\frac{\partial u'}{\partial x} + \gamma p' \frac{d\overline{u}}{dx} = (\gamma - 1) q'. \tag{11}$$

It has been shown by Karthik *et al.*[7] that the ratio between $\rho' \overline{u}\, d\overline{u}/dx$ and $\partial p'/\partial x$ is of the order of square of the mean Mach number. Therefore, the term $\rho' \overline{u}\, d\overline{u}/dx$ in Eq. (10) can be neglected for low Mach numbers. Assuming a harmonic dependence of each unsteady quantity, we can write

$$p'(x,t) = P(x) e^{i\omega t}, \tag{12}$$

$$u'(x,t) = U(x) e^{i\omega t}, \tag{13}$$

$$q'(x,t) = Q(x) e^{i\omega t}. \tag{14}$$

With the aid of Eqs. (12)–(14), the acoustic equations can be simplified to

$$\left(i\omega+\frac{d\bar{u}}{dx}\right)U+\bar{u}\frac{dU}{dx}+\frac{1}{\bar{\rho}}\frac{dP}{dx}=0, \tag{15}$$

$$\left(i\omega+\gamma\frac{d\bar{u}}{dx}\right)P+\bar{u}\frac{dP}{dx}+\gamma\bar{p}\frac{dU}{dx}=(\gamma-1)Q. \tag{16}$$

Using Eqs. (15) and (16), U and its derivative dU/dx can be eliminated. This leads to the following second-order differential equation for the acoustic pressure amplitude

$$\begin{aligned}[1-M^2]\frac{d^2P}{dx^2}+&\left[(1-(3+\gamma)M^2)\frac{2}{M}\frac{dM}{dx}\right.\\&\left.-g(M)(1-M^2)-2ikM\right]\frac{dP}{dx}-\left[8\gamma\left(\frac{dM}{dx}\right)^2\right.\\&\left.+ik(\gamma-1)Mg(M)+2ik(2+\gamma)\frac{dM}{dx}-k^2\right]P\\&=-\frac{\gamma-1}{\bar{c}}\left[M\frac{dQ}{dx}+\left(ik+2\frac{dM}{dx}+Mg(M)\right)Q\right],\end{aligned} \tag{17}$$

where

$$g(M)=2\frac{\dfrac{d^2M}{dx^2}+\dfrac{1}{M}\left(\dfrac{dM}{dx}\right)^2}{ik+2\dfrac{dM}{dx}},$$

$M=\bar{u}/\bar{c}$ is the mean Mach number, and $k=\omega/\bar{c}$ is the wave number. In order to obtain the oscillatory heat release from Eq. (17), numerical derivatives of the acoustic pressure and the mean velocity need to be computed. This differentiation process always amplifies the errors in the experimental data. Hence, the differential equation is transformed into an integral equation.

III. ANALYSIS

In this section the governing differential equation is transformed to an integral equation in the heat release distribution. To reduce the complexity, a constant mean flow velocity in the combustor is initially assumed. In Sec. III B 3, the formulation will be generalized to account for an arbitrary mean flow variation in the domain. Under the constant mean flow velocity assumption, the governing equation simplifies to

$$\begin{aligned}(1-M^2)\frac{d^2P}{dx^2}-2iKM\frac{dP}{dx}+k^2P\\=-\frac{\gamma-1}{\bar{c}}\left(ikQ+M\frac{dQ}{dx}\right).\end{aligned} \tag{18}$$

It should be noted that though the validity of Eq. (17) is limited to low Mach number ($M^2\ll1$) flows, Eq. (18) is exact and valid for all Mach numbers. Equation (18) can be solved exactly for $Q(x)$ by multiplying both sides by $e^{ikx/M}$ and then integrating Eq. (18) from 0 to x. This results in an integral representation for $Q(x)$ in terms of the pressure $P(x)$, and its first derivative $P'(x)$. In order to compute the heat release from the above-mentioned representation, the values of acoustic pressure and its derivative need to be prescribed at various locations along the length of the combustor. In reality, the acoustic pressure measurements can only be performed at a finite number of locations in the combustor. Therefore, computation of the derivative of acoustic pressure at these locations involves differentiation of the experimental data that always amplifies the errors in the data. The highly sensitive nature of the solution to the exact acoustic pressure values makes this method impractical in this case. Moreover, the above-mentioned method does not provide any information about the sensitivity of the solution to the exact pressure values. To overcome these difficulties, the inverse problem is formulated in terms of integral equations. The differential equation (18) can be transformed into an integral equation for $Q(x)$ in two ways:

(1) The governing differential equation can be transformed to a Fredholm integral equation of the first kind, using the Green's function approach.
(2) The equation can be transformed to a Volterra integral equation of the first kind by repeated integration.[8]

A. Method (1): The Green's function approach

In this section the governing differential equation is transformed to a Fredholm integral equation. This method will be referred to as the Green's function approach. Previously, Ramachandra and Strahle[3] used this method to recover the heat release distribution for open premixed turbulent flames. Chao[4] attempted to use this technique to recover the heat release distribution in a gas turbine combustor. However, with this method he was able to recover the heat release successfully only at certain frequencies. He remarked that the reason for the poor recovery at other frequencies is not fully understood. The reason for this problem is investigated in this section.

The governing differential equation can be rewritten as

$$\frac{d^2P}{dx^2}+Z_1\frac{dP}{dx}+Z_2P=Z_3\left(ikQ+M\frac{dQ}{dx}\right), \tag{19}$$

where

$$Z_1=-\frac{2ikM}{1-M^2},\quad Z_2=\frac{k^2}{1-M^2},\quad Z_3=-\frac{\gamma-1}{\bar{c}(1-M^2)}.$$

To construct the Green's function for this problem, the above-mentioned differential equation is transformed to its self-adjoint form. Multiplying Eq. (19) with the factor e^{Z_1x} leads to the standard self-adjoint form:[9]

$$\frac{d}{dx}\left(r(x)\frac{dP}{dx}\right)+q(x)P=f(x), \tag{20}$$

where $r(x)=e^{Z_1x}$, $q(x)=Z_2e^{Z_1x}$, and $f(x)=Z_3e^{Z_1x}(ikQ+M(dQ/dx))$.
The self-adjoint operator L can be defined such that

$$LP(x)=f(x).$$

Therefore, the Green's function satisfies the differential equation

$$LG(x,\xi)=\delta(x-\xi). \tag{21}$$

It is well known[9] that for any two arbitrary functions u and v, the following relation is satisfied:

$$\int_a^b (vLu-uLv)d\xi=[r(\xi)(vu'-uv')]_a^b. \tag{22}$$

If u, v are chosen such that $u=G(x,\xi)$ and $v=P(\xi)$, then Eq. (22) simplifies to the Fredholm integral equation of the first kind,

$$P(x)=[r(\xi)(vu'-uv')]_a^b+\int_a^b G(x,\xi)f(\xi)d\xi. \tag{23}$$

Equation (23) is usually written as

$$P(x)=P_h(x)+\int_a^b G(x,\xi)f(\xi)d\xi, \tag{24}$$

where the homogeneous part $P_h(x)$ is the contribution due to the boundary conditions.

It should be noted that the boundary conditions for the Green's function are chosen according to the boundary conditions for the corresponding homogeneous problem. Once the boundary conditions are specified, the Green's function can be constructed[10] using the method of variation of parameters or the method of eigenfunction expansion.

As an illustration of eigenfunction expansion of Green's function, a sample problem is considered. For simplicity, the Mach number is assumed to be zero. Therefore, the differential equation (20) reduces to

$$\frac{d^2P}{dx^2}+k^2P=f(x). \tag{25}$$

If the homogeneous boundary conditions $[P(a)=P(b)=0]$ are assumed, the solution of the problem from Eq. (24) can be written as

$$P(x)=\int_a^b G(\xi,x)f(\xi)d\xi. \tag{26}$$

To find the Green's function for the problem, both $P(x)$ and $f(x)$ are expanded in terms of eigenmodes of the problem as

$$P(x)=\sum_{l=1}^{\infty} a_l u_l(x), \quad f(x)=\sum_{l=1}^{\infty} b_l u_l(x), \tag{27}$$

where $u_l(x)$ is the solution of the eigenvalue problem

$$\frac{d^2u_l}{dx^2}+k_l^2u_l(x)=0. \tag{28}$$

Using Eqs. (27) and (28) it can be shown[9] that the solution $P(x)$ can be obtained as

$$P(x)=\int_a^b\left(\sum_{l=1}^{\infty}\frac{u_l(x)u_l(\xi)}{k^2-k_l^2}\right)f(\xi)d\xi. \tag{29}$$

Comparing Eqs. (26) and (29), the Green's function of the problem can be written as

$$G(x,\xi)=\sum_{l=1}^{\infty}\frac{u_l(x)u_l(\xi)}{k^2-k_l^2}. \tag{30}$$

It can be seen that the Green's function becomes indefinite at the eigenmodes of the combustor. Therefore, any attempt to solve the integral equation at the resonant frequencies will induce large errors in the process. Hence, as Lieuwen *et al.*[6] remarked, this method in general is not feasible to recover heat release in ducted systems. To avoid this difficulty, the problem is reformulated using a different approach in Sec. III B.

B. Method (2): Transformation to a Volterra integral equation

In this section, the governing differential equation is transformed to a Volterra integral equation and the methods of solution of the resulting equation are discussed. The governing differential equation can be written as

$$\frac{d^2P}{dx^2}+Z_1\frac{dP}{dx}+Z_2P=Z_3\left(ikQ+M\frac{dQ}{dx}\right). \tag{31}$$

Integrating Eq. (31) from 0 to ξ gives

$$P'(\xi)-P'(0)+Z_1(P(\xi)-P(0))+Z_2\int_0^{\xi}P(x')dx'$$
$$=Z_3\left[M(Q(\xi)-Q(0))+ik\int_0^{\xi}Q(x')dx'\right]. \tag{32}$$

Integrating Eq. (32) once again from 0 to x gives

$$P(x)+Z_1\int_0^x P(\xi)d\xi+Z_2\int_0^x\int_0^{\xi}P(x')dx'\,d\xi$$
$$-(1+Z_1x)P(0)-xP'(0)$$
$$=Z_3\left[M\int_0^x Q(\xi)d\xi+ik\int_0^x\int_0^{\xi}Q(x')dx'd\xi-MxQ(0)\right]. \tag{33}$$

Using the identity,

$$\int_0^x\int_0^{\xi}f(x')dx'\,d\xi=\int_0^x(x-\xi)f(\xi)d\xi,$$

Eq. (33) can be simplified as

$$P(x)+Z_1\int_0^x P(\xi)d\xi+Z_2\int_0^x(x-\xi)P(\xi)d\xi$$
$$-(1+Z_1x)P(0)-xP'(0)$$
$$=Z_3\left[M\int_0^x Q(\xi)d\xi+ik\int_0^x(x-\xi)Q(\xi)d\xi-MxQ(0)\right]. \tag{34}$$

Equation (34) is a Volterra integral equation of the first kind for unknown heat release distribution. It can be recast into the standard form[9] as

$$f[P(x)]+Z_3MxQ(0)=\int_0^x k(x,\xi)Q(\xi)d\xi, \tag{35}$$

where the kernel $k(x,\xi)=Z_3(M+ik(x-\xi))$.

Equation (34) can be solved sequentially for heat release along the length of the combustor starting from the initial

point $x=0$. However, it has to be supplemented with a boundary condition in terms of heat release. The mixing of fuel and air in a gas turbine combustor takes a finite amount of distance. Therefore, most of the combustion region is distributed about the center of the combustor and usually the pressure measurements are done at all locations along the length of the combustor. Therefore, the appropriate boundary condition would be a Neumann boundary condition, i.e., $Q'(0)=0$.

1. Numerical method

In this section a numerical scheme is discussed to solve the governing integral equation at discrete points located at equally spaced intervals along the length of the combustor. This will be referred as the direct numerical method (DNM). The distance between any two grid points, referred as step size, is $h=L/n$, where L is the length of the domain. Each step is denoted by the index j, where j ranges from 1 to N. The starting and the ending coordinates of the jth step are designated by x_{j-1} and x_j, respectively, where $x_j=j*h$. Using the above-given notation, Eq. (34) can be written in discretized form as

$$\begin{aligned} &P_j+\sum_{l=1}^{j}\int_{x_{l-1}}^{x_l}[Z_1+Z_2(x_j-\xi)]P(\xi)d\xi \\ &-(1+Z_1x_j)P(0)-x_jP'(0) \\ &-Z_3\left[\sum_{l=1}^{j}\int_{x_{l-1}}^{x_l}[M+ik(x_j-\xi)]Q(\xi)d\xi-Mx_jQ(0)\right] \\ &=0 \quad \text{for } j=1 \text{ to } N. \end{aligned} \tag{36}$$

Here P_j refers to the value of pressure amplitude at x_j. To evaluate the integrals, a linear shape function for the pressure and heat release in the step is assumed, i.e.,

$$P(x)=P_{j-1}+\left(\frac{P_j-P_{j-1}}{h}\right)(x-x_{j-1}) \quad \text{for } x_{j-1}\leqslant x\leqslant x_j$$

and (37)

$$Q(x)=Q_{j-1}+\left(\frac{Q_j-Q_{j-1}}{h}\right)(x-x_{j-1}) \quad \text{for } x_{j-1}\leqslant x\leqslant x_j.$$

Using the above-given relations, Eq. (36) can be simplified as

$$\begin{aligned} &P_j+Z_1\sum_{l=1}^{j}0.5h(P_{l-1}+P_l)+Z_2\sum_{l=1}^{j}h^2\left[P_{l-1}\left(\frac{j-l}{2}+\frac{1}{3}\right)\right. \\ &\left.+P_l\left(\frac{j-l}{2}+\frac{1}{6}\right)\right]+(1+Z_1jh)P_0-jhP_0' \\ &=Z_3\left[\sum_{l=1}^{j}0.5Mh(Q_{l-1}+Q_1)+ikh^2\left[Q_{l-1}\left(\frac{j-l}{2}+\frac{1}{3}\right)\right.\right. \\ &\left.\left.+Q_l\left(\frac{j-l}{2}+\frac{1}{6}\right)\right]-jhQ_0\right] \quad \text{for } j=1 \text{ to } N. \end{aligned} \tag{38}$$

Equation (38) is equivalent to N linear equations in terms of N unknown values of the heat release along the length of the combustor. It can be written in the matrix form as $A\mathbf{Q}=B$, where A is an $N\times N$ matrix, $\mathbf{Q}$ is an N dimensional column vector containing the unknown heat release values Q_j $(j=1-N)$, and B is a function of experimental pressure values. It can be seen that the matrix A is a lower triangular matrix. Therefore, the system of N equations can be solved sequentially for Q_j values starting from the index $j=1$.

a. Stability analysis. The stability of the method can be analyzed by computing the amplification of the error in the heat release from $(j-1)$th step to jth step. If Q is the exact solution to Eq. (36) and Q_d is the solution obtained by discretization, then the error δQ is defined as $\delta Q=Q-Q_d$. As the equation is linear, the errors δQ in the heat release obey the same governing equation as the exact quantity Q_d so that

$$\begin{aligned} &\sum_{l=1}^{j}0.5Mh(\delta Q_{l-1}+\delta Q_l)+ikh^2\left[\delta Q_{l-1}\left(\frac{j-l}{2}+\frac{1}{3}\right)\right. \\ &\left.+\delta Q_l\left(\frac{j-l}{2}+\frac{1}{6}\right)\right]=\text{known terms}. \end{aligned} \tag{39}$$

Collecting the coefficients of δQ_j and δQ_{j-1} in Eq. (39) leads to

$$\begin{aligned} &(0.5Mh+ikh^2/6)\,\delta Q_j+(Mh+ikh^2)\,\delta Q_{j-1} \\ &+\sum_{l=1}^{j-2}f(\delta Q_j)=\text{known terms}. \end{aligned} \tag{40}$$

The amplification factor "g" in the jth step is defined as

$$g_j=\frac{\text{Coefficient of } (\delta Q_{j-1})}{\text{Coefficient of } (\delta Q_j)}.$$

From Eq. (40), the amplification factor can be evaluated as

$$g_j=\frac{M+ikh}{M/2+ikh/6}. \tag{41}$$

It can be readily seen from Eq. (41) that the amplification factor is independent of j. Therefore, the stability of the method requires $g_j\overline{g_j}<1$, where $\overline{g_j}$ is the complex conjugate of g_j.

From Eq. (41),

$$g_j\overline{g_j}=\frac{M^2+k^2h^2}{M^2/4+k^2h^2/36}. \tag{42}$$

For all values of M, k, and h,

$$g_j\overline{g_j}>1.$$

This shows that the method is unconditionally unstable. This also means that any deviation in the value of Q from the exact value would be subsequently amplified. This is unfortunate in the present scenario because the experimental values of the pressure will always be contaminated by noise. Therefore, the heat release computed from the contaminated data could be far from the exact value. In order to make the method stable, the following transformation is applied:

$$S(x)=\int_0^x Q(\xi)d\xi. \tag{43}$$

The governing differential equation (18) reduces to

$$(1-M^2)\frac{d^2P}{dx^2}-2ikM\frac{dP}{dx}+k^2P$$
$$=-\frac{\gamma-1}{\bar{c}}\left(ik\frac{dS}{dx}+M\frac{d^2S}{dx^2}\right). \tag{44}$$

Integrating Eq. (44) from 0 to ξ and then from 0 to x yields the following integral equation for $S(x)$:

$$P(x)+Z_1\int_0^x P(\xi)d\xi+Z_2\int_0^x\int_0^\xi P(x')dx'\,d\xi$$
$$-(1+Z_1x)P(0)-xP'(0)$$
$$=Z_3\left[MS(x)+ik\int_0^x S(\xi)d\xi-MxS'(0)\right]. \tag{45}$$

Equation (45) is a Volterra equation of the second kind for unknown $S(x)$. This can be written in the standard form as

$$f[P(x)]+xS'(0)=S(x)+\int_0^x k(x,\xi)S(\xi)d\xi, \tag{46}$$

where the kernel $k(x,\xi)=ik/M$.

Adopting the same numerical method as discussed earlier, Eq. (45) in discretized form can be written as

$$P_j+\sum_{l=1}^{j}\int_{x_{l-1}}^{x_l}[Z_1+Z_2(x_j-\xi)]P(\xi)d\xi-(1+Z_1x_j)P(0)$$
$$-x_jP'(0)-Z_3\left[MS_j+ik\sum_{l=1}^{j}S(\xi)d\xi-Mx_jS_0'\right]=0$$
$$\text{for } j=1 \text{ to } N. \tag{47}$$

Using the linear shape functions [Eq. (37)] Eq. (47) can be simplified as

$$P_j+Z_1\sum_{l=1}^{j}0.5h(P_{l-1}+P_l)+Z_2\sum_{l=1}^{j}h^2\left[P_{l-1}\left(\frac{j-l}{2}+\frac{1}{3}\right)\right.$$
$$\left.+P_l\left(\frac{j-l}{2}+\frac{1}{6}\right)\right]+(1+Z_1jh)P_0-jhP_0'$$
$$=Z_3\left[MS_j+\sum_{l=1}^{j}0.5ikh(S_{l-1}+S_l)-jhS_0'\right]$$
$$\text{for } j=1 \text{ to } N. \tag{48}$$

Therefore, the equation governing the amplification of errors can be deduced as

$$(M+Ikh/2)\delta S_j+(Ikh)\delta S_{j-1}+\sum_{l=1}^{j-2}f(\delta S_j)=\text{known terms}.$$

The amplification factor "g" for this modified method is

$$g=\frac{Ikh}{M+Ikh/2}. \tag{49}$$

From Eq. (49), the stability condition for the method is

$$k<\sqrt{\frac{4}{3}}\frac{M}{h}. \tag{50}$$

It can be readily seen that the stability of the method increases as the Mach number increases. This obviously follows from the fact that the coefficient of $S''(x)$ in Eq. (44) facilitated the transformation of the Volterra integral equation from the first kind to the second kind. Another important observation from Eq. (50) is that the stability of the method decreases gradually as the wave number increases. This type of stability is different from the stability of the Fredholm integral equation discussed in Sec. III A, where the method is highly unstable at the eigenvalues of the wave number. Clearly, the reformulation in terms of Volterra integral equations helped in improving the nature of the stability of the inversion process.

b. Sensitivity analysis of the problem. The measured pressure data always contain a certain amount of errors. These errors might be caused due to errors in data acquisition system, structural noise, flow turbulence leading to random oscillations of pressure, etc. The objective of this section is to estimate the error in the computed acoustic heat release (ΔQ) due to a unit error in the acoustic pressure (ΔP). For this purpose, we can define the sensitivity factor (SF) mathematically as $\text{SF}=|\Delta Q/\Delta P|$. It can be seen that the sensitivity factor gives an indication of the amplification of the errors in the process. The higher the SF, the more the amplification of noise would be. Therefore, any small change in the data will produce a large change in the solution. Higher SFs are typical of stiff inverse problems. From Eq. (48)

$$\delta P_j+Z_1\sum_{l=1}^{j}0.5h(\delta P_{l-1}+\delta P_l)$$
$$+Z_2\sum_{l=1}^{j}h^2\left[\delta P_{l-1}\left(\frac{j-l}{2}+\frac{1}{3}\right)+\delta P_l\left(\frac{j-l}{2}+\frac{1}{6}\right)\right]$$
$$+(1+Z_1jh)\delta P_0-jh\delta P_0'$$
$$=Z_3\left[M\delta S_j+\sum_{l=1}^{j}0.5ikh(\delta S_{l-1}+\delta S_l)-jh\delta S_0'\right]$$
$$\text{for } j=1 \text{ to } N. \tag{51}$$

To get a rough estimate of SF, the following simplifications are made. The noise in the measured pressure is assumed to have zero mean. So, the contribution of the second and third terms on the left-hand side of Eq. (51) involving summation of the noise in pressure data would be negligible. Moreover, it should be noted that these terms involve multiplication by h and h^2, respectively, which would further reduce their contribution. The coefficient of δS_j can be evaluated as $Z_3(M+Ikh/2)$. From Eq. (43), $\delta Q_j\approx\delta S_j/h$. Therefore, the sensitivity factor becomes

$$\text{SF}=\left|\frac{\Delta Q}{\Delta P}\right|\approx\left|\frac{\bar{c}(1-M^2)}{h(\gamma-1)(M+ikh/2)}\right|. \tag{52}$$

For typical values of $\bar{c}=1000$ K, $M=0.1$, $k=0.5$, and $h=0.1$ m the sensitivity factor evaluates to $\text{SF}\approx5\times10^5$. This shows that the noise in the experimental data gets amplified

by a huge factor even though the method is stable. To avoid these problems a parameter free implicit least-squares method discussed by Chao[4] is used.

2. Implicit least-squares method

This method will be referred to as ILSM. In this method, the oscillatory heat release distribution is approximated by a set of known basis functions. Usually these basis functions are chosen to be polynomials. The heat release distribution in the domain is approximated by

$$\tilde{Q}(\xi)=\sum_{i=1}^{m} a_l \Phi_l(\xi). \tag{53}$$

The governing integral equation (35) is

$$f[P(x)]=\int_0^x k(x,\xi)Q(\xi)d\xi - Z_3 Mx Q(0),$$

where the kernel $k(x,\xi)=Z_3(M+ik(x-\xi))$. In reality, the function $f(x)$ is contaminated with noise. Let the function $f(x)$ obtained from experimental data be $f(x)_{\text{expt}}$. The mean square error (MSE) for the problem can be defined as

$$E=\frac{1}{L}\int_0^L [f(x)_{\text{expt}}-f(x)_{\text{calc}}]^2 dx, \tag{54}$$

where the function $f(x)_{\text{calc}}$ is obtained by substituting for $\tilde{Q}(\xi)$ in Eq. (53) into Eq. (35).

The unknown coefficients a_l in Eq. (53) can be determined by minimizing the MSE with respect to each of the coefficients. This minimum can be obtained from

$$\nabla E=\frac{1}{L}\int_0^L 2[f(x)]_{\text{expt}}-f(x)_{\text{calc}}]\nabla f\, dx=0, \tag{55}$$

where the operator $\nabla=\partial/\partial a_l$ and $l=1$ *to* m.

Using Eq. (35), ∇f can be evaluated as

$$\nabla f=\int_0^x k(x,\xi)\Phi_l(\xi)d\xi - Z_3 Mx\Phi_l(0). \tag{56}$$

Substituting for f and ∇f in Eq. (55) yields

$$\begin{aligned}\nabla E=\int_0^L \Bigg[& f(x)_{\text{expt}}-\sum_{j=1}^{m} a_j \int_0^x k(x,\xi)\Phi_j(\xi)d\xi \\ & - Z_3 Mx\Phi_j(0)\Bigg]\Bigg(\int_0^x k(x,\xi)\Phi_l(\xi)d\xi \\ & - Z_3 Mx\Phi_l(0)\Bigg)dx=0.\end{aligned} \tag{57}$$

Equation (57) can be written in a compact form as

$$\int_0^L f(x)_{\text{expt}} R_l(x)dx=\sum_{j=1}^{m} a_j \int_0^L R_j(x)R_l(x)dx, \quad 1\leqslant l\leqslant m, \tag{58}$$

where $R_l(x)=\int_0^x k(x,\xi)\Phi_l(\xi)d\xi - Z_3 Mx\Phi_l(0)$.

The above-given expression is a set of m linear equations in terms of m unknown coefficients a_l. It can be easily solved using Gaussian elimination or standard LU decomposition techniques. In the present method, $\tilde{Q}(\xi)$ is assumed to be a third-order polynomial. In general, any analytic form for basis functions can be assumed. Usually, the sinusoidal functions with highly oscillatory behavior are not appropriate for the method because they fit more into noise than the signal. For the same reason, higher powers of x are also not recommended.

3. Integral formulation including the effects of variable Mach number

In a real combustor the occurrence of chemical reactions would induce nonzero temperature gradients along the length of the combustor. This would result in the corresponding axial variation in the Mach number. In this section, the implicit least-squares method will be extended to the general case where the Mach number varies along the length of the combustor. The steady state conservation of mass is represented as

$$\dot{m}=\overline{\rho u}(\text{const}). \tag{59}$$

Using Eq. (59), Eq. (9), and the definition of the mean Mach number, i.e., $M=\overline{u}/\overline{c}$, the governing differential equation (17) can be rewritten in terms of mean velocity as

$$\begin{aligned}&\left[\overline{u}^2-\frac{\gamma\overline{pu}}{\dot{m}}\right]\frac{d^2P}{dx^2}+\left[\left((3+\gamma)\overline{u}-\frac{\gamma\overline{p}}{\dot{m}}\right)\frac{d\overline{u}}{dx}\right. \\ &\left.-fn(\overline{u})\left(\overline{u}^2-\frac{\gamma\overline{pu}}{\dot{m}}\right)+2i\omega\overline{u}\right]\frac{dP}{dx}+\left[2\gamma\left(\frac{d\overline{u}}{dx}\right)^2\right. \\ &\left.+i\omega(\gamma-1)\overline{u}fn(\overline{u})+2i\omega(2+\gamma)\frac{d\overline{u}}{dx}-\omega^2\right]P \\ &=(\gamma-1)\left[\overline{u}\frac{dQ}{dx}+\left(i\omega+2\frac{d\overline{u}}{dx}-\overline{u}fn(\overline{u})\right)Q\right],\end{aligned} \tag{60}$$

where $fn(\overline{u})=(d^2\overline{u}/dx^2)/(i\omega+d\overline{u}/dx)$.

Unlike Eq. (18), Eq. (60) has variable coefficients. It is interesting to note that all the coefficients are functions of mean velocity, mean pressure, and mass flow rate. The latter two being constant along the length, it would be sufficient to specify the mean velocity profile to evaluate all the coefficients. Equation (60) can be written concisely as

$$\frac{d^2P}{dx^2}+C_1(x)\frac{dP}{dx}+C_2(x)P(x)=C_3(x)\frac{dQ}{dx}+C_4(x)Q(x), \tag{61}$$

where

$$\begin{aligned}C_1(x)=&\left(\left((3+\gamma)\overline{u}-\frac{\gamma\overline{p}}{m}\right)\frac{d\overline{u}}{dx}-fn(\overline{u})\left(\overline{u}^2-\frac{\gamma\overline{pu}}{m}\right)\right. \\ &\left.+2i\omega\overline{u}\right)\Big/\left(\overline{u}^2-\frac{\gamma\overline{pu}}{m}\right), \\ C_2(x)=&\left(2\gamma\left(\frac{d\overline{u}}{dx}\right)^2 i\omega(\gamma-1)\overline{u}fn(\overline{u})\right. \\ &\left.+2i\omega(2+\gamma)\frac{d\overline{u}}{dx}-\omega^2\right)\Big/\left(\overline{u}^2-\frac{\gamma\overline{pu}}{m}\right),\end{aligned}$$

$$C_3(x)=(\gamma-1)\overline{u}\Big/\left(\overline{u}^2-\frac{\gamma\overline{pu}}{m}\right),$$

$$C_4(x)=(\gamma-1)\left(i\omega+2\frac{d\overline{u}}{dx}-\overline{u}fn(\overline{u})\right)\Big/\left(\overline{u}^2-\frac{\gamma\overline{pu}}{m}\right).$$

The differential equation (61) is transformed into an integral equation by integrating the above equation from 0 to ξ and then from 0 to x and using the relations

$$C_1(x)\frac{dP}{dx}=\frac{d}{dx}(C_1(x)P)-P\frac{dC_1}{dx},$$

$$C_3(x)\frac{dQ}{dx}=\frac{d}{dx}(C_3(x)Q)-Q\frac{dC_3}{dx},$$

$$\begin{aligned}P(x)+\int_0^x\left(C_1(\xi)+(x-\xi)\left(C_2(\xi)-\frac{dC_1}{d\xi}\right)\right)P(\xi)d\xi\\ -(1+C_1(0)x)P(0)-xP'(0)\\ =\int_0^x\left(C_3(\xi)+\left((x-\xi)C_4(\xi)-\frac{dC_3}{d\xi}\right)\right)Q(\xi)d\xi\\ -xC_3(0)Q(0).\end{aligned}\tag{62}$$

Equation (62) is a Volterra integral equation for the unknown heat release distribution. It can be solved by the implicit least-squares method discussed in Sec. III B 2. Equation (62) can be written in a simpler form as

$$f[P(x),\overline{u}(x)]=\int_0^x k(x,\xi)Q(\xi)d\xi-xC_3(0)Q(0),\tag{63}$$

where the kernel $k(x,\xi)=C_3(\xi)+((x-\xi)C_4(\xi)-dC_3/d\xi)$. The left-hand side of Eq. (63) is a function of acoustic pressure amplitude and the mean velocity. Therefore, it is known in advance from the experimental values. Using Eq. (53) and the definition of MSE, minimization of E with respect to each of the unknown coefficient's a_l yields

$$\nabla E=\frac{1}{L}\int_0^L 2[f(x)_{\text{expt}}-f(x)_{\text{calc}}]\nabla f\,dx=0,\tag{64}$$

where the operator $\nabla=\partial/\partial a_l$ and $l=1$ *to* m.

Using Eq. (63), ∇f can be evaluated as

$$\nabla f=\int_0^x k(x,\xi)\Phi_l(\xi)d\xi-xC_3(0)\Phi_l(0).\tag{65}$$

Substituting for f and ∇f in Eq. (64) yields

$$\begin{aligned}\nabla E=\int_0^L\Bigg[f(x)_{\text{expt}}-\sum_{j=1}^m a_j\int_0^x k(x,\xi)\Phi_j(\xi)d\xi\\ -xC_3(0)\Phi_j(0)\Bigg]\Bigg(\int_0^x k(x,\xi)\Phi_l(\xi)d\xi\\ -xC_3(0)\Phi_l(0)\Bigg)dx=0.\end{aligned}\tag{66}$$

Equation (66) can be written in compact form as

$$\int_0^L f(x)_{\text{expt}}R_l(x)dx=\sum_{j=1}^m a_j\int_0^L R_j(x)R_l(x)dx,\quad 1\leqslant l\leqslant m,\tag{67}$$

where

$$R_l(x)=\int_0^x k(x,\xi)\Phi_l(\xi)d\xi-xC_3(0)\Phi_l(0).$$

In reality, the acoustic pressure and the mean velocity are known only at discrete locations along the length of the combustor. Therefore, the integrals in Eq. (67) can be approximated by the summation over all the x locations. This reduces Eq. (67) to

$$\begin{aligned}\sum_k\int_{x_{k-1}}^{x_k}f(x)_{\text{expt}}R_l(x)dx\\ =\sum_{j=1}^m a_j\sum_k\int_{x_{k-1}}^{x_k}R_l(x)R_j(x)dx,\quad 1\leqslant l\leqslant m,\end{aligned}\tag{68}$$

where the index k spans all the axial locations at which the experimental data are known. The piecewise integrals in Eq. (68) can be evaluated by using the trapezoidal rule. Therefore, Eq. (68) simplifies to

$$\begin{aligned}\sum_k 0.5(f(x_{k-1})_{\text{expt}}R_l(x_{k-1})+f(x_k)_{\text{expt}}R_l(x_k))\Delta x_k\\ =\sum_{j=1}^m a_j\sum_k 0.5(R_l(x_{k-1})R_j(x_{k-1})\\ +R_l(x_k)R_j(x_k))\Delta x_k,\quad 1\leqslant l\leqslant m.\end{aligned}\tag{69}$$

In Eq. (69), the coefficients R_l are also evaluated using the trapezoidal rule. With the aid of Eq. (69), all the unknown coefficients a_l can be determined from the m linear equations. Again, it should be noted that the choice of the basis functions is critical. As discussed earlier, higher powers of x and oscillatory functions are not advisable.

IV. RESULTS

The validity of the preceding theory is assessed with the aid of some numerical examples discussed in the following. In all the examples, the methodology adopted by Lieuwen *et al.*[6] is closely followed to construct the test problems. First, the mean flow quantities, viz., pressure, temperature, and mass flow rate together with the appropriate boundary conditions are specified in the combustor. In addition, the fluctuating heat release distribution $Q(x)$ is assumed in order to solve for the acoustic pressure. In all the cases, the governing differential equation for oscillatory pressure is solved using a fourth-order Runge–Kutta scheme. The resulting pressure distribution in the combustor is assumed to be exact. Random noise is then added to this exact pressure distribution to simulate the experimental values. The contaminated pressure data is fed as input to solve the integral equation for the unknown heat release distribution. In all the examples, the heat release distribution is approximated by a third-order polynomial and the integral equation is used to

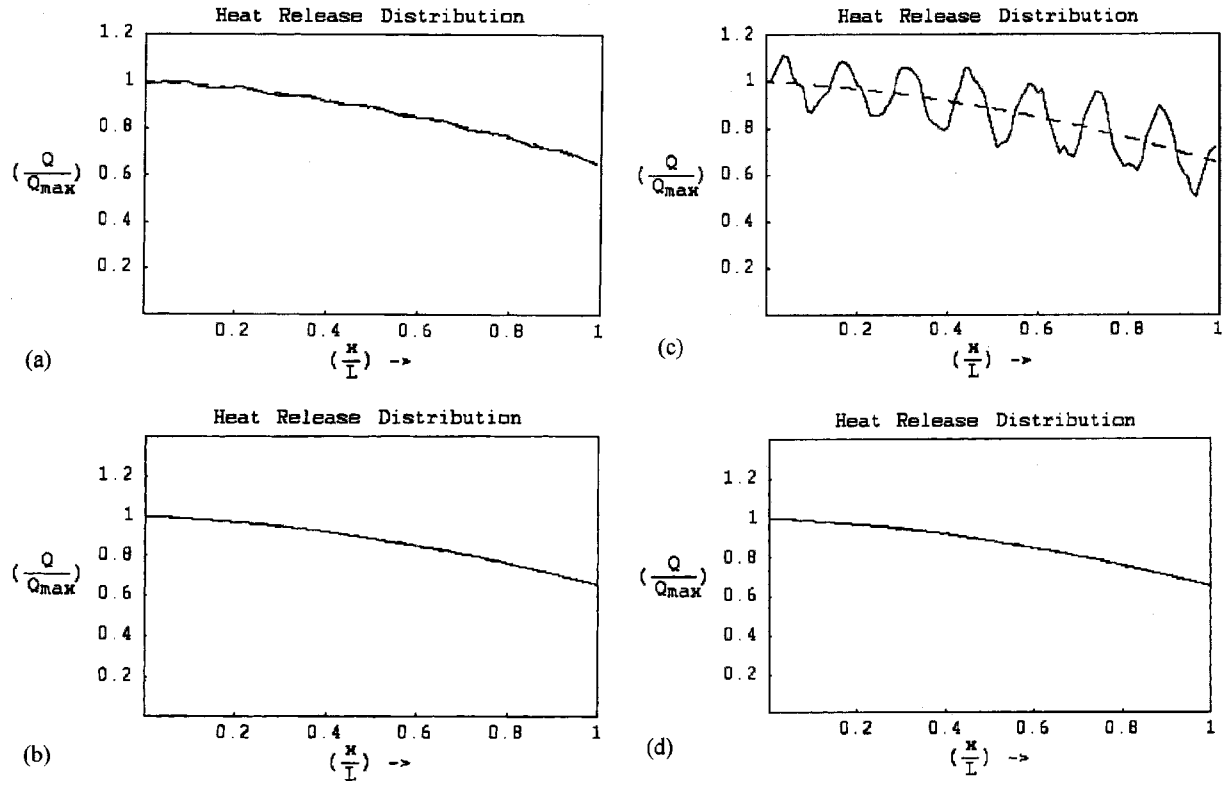


FIG. 1. (a) Recovery of heat release by DNM with 0.0% errors in the pressure; exact distribution (---) and recovered distribution (—). (b) Recovery of heat release by ILSM with 0.0% errors in the pressure; exact distribution (---) and recovered distribution (—). (c) Recovery of heat release by DNM with 0.01% errors in the pressure; exact distribution (---) and recovered distribution (—). (d) Recovery of heat release by ILSM with 0.01% errors in the pressure; exact distribution (---) and recovered distribution (—).

solve for the unknown coefficients. The difference between the recovered heat release and the exact heat release is a measure of the accuracy of the method.

Example 1: In this example the recovery of the heat release distribution from the direct numerical method (DNM) and the implicit least-squares method (ILSM) at the second natural frequency of the combustor is compared. The mean quantities assumed for the problem are: mass flux $(\dot{m}) = 50\ \mathrm{kg/(m^2\ s)}$; mean pressure $(\overline{p}) = 2\times 10^5$ Pa; mean temperature $(\overline{T}) = 700$ K (assumed constant along the length of the duct).

It is further assumed that the exact heat release distribution in the duct obeys the functional form

$$Q(x)/Q_{\max} = 1 - 0.1\left(\frac{x}{L}\right) - 0.25\left(\frac{x}{L}\right)^2,$$

where $Q_{\max} = 10^6\ \mathrm{J/(m^3\ s)}$ and L is the length of the combustor. The boundary conditions for the acoustic pressure amplitude are

$$P(0)/\overline{p} = 0.001, \quad P'(0)/(\overline{p}/L) = 0.$$

The sample calculations are done at a wave number of $k = 1.5\pi/L$. This corresponds to the second natural mode of the combustor. Figures 1(a) and (b) show the results obtained from DNM and ILSM, respectively. As expected, both methods yield good results in the absence of noise in the pressure data. However, when noise is added to the exact pressure data, the situation changed dramatically. Figures 1(c) and (d) show the results obtained from DNM and ILSM, respectively, when 0.01% random errors are added to the pressure. As expected, in the case of DNM, the errors as small as 0.01% resulted in fluctuations as large as 25% in heat release owing to the large sensitivity of the problem. The recovery of the heat release from ILSM is very good and the difference in exact and recovered heat release distribution is too small to be observed. It is evident from the results that the implicit least-squares method is far superior compared to the direct numerical method. It should be noted that the reformulation helped in the successful recovery of heat release even at the natural frequency. Extensive calculations show that the heat release can be recovered successfully at all frequencies.

Example 2: In this example, the recovery of heat release by ILSM is discussed. The mean quantities and the boundary conditions are same as those assumed in Example 1. The calculations are done at a frequency of $\omega = 300$ rad/s. Noise is added to the exact pressure data with varying amplitude according to

$$P(x_j)_{\mathrm{expt}} = P(x_j)_{\mathrm{exact}} + 0.01d(R_j - 0.5)P(0),$$

where the index j spans all the x locations, the noise "size" d takes the values 1, 5, and 10 and R_j is obtained from a random number generator which returns values uniformly distributed between 0 and 1. Figures 2(a)–(c) show the results obtained for different values of d. The results are quite satisfactory and the mean error in the recovered heat release is less than 10% even in the presence of noise levels as high as 10%. It is important to note that with such noise levels, DNM would yield completely erroneous results.

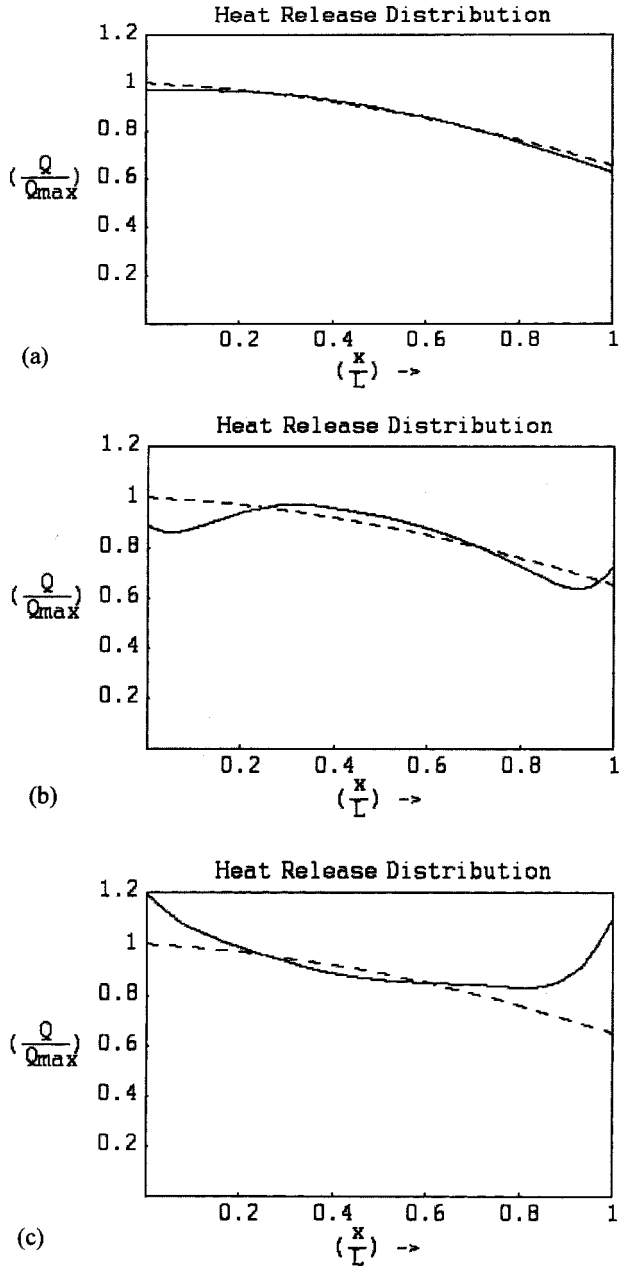


FIG. 2. (a) Recovery of heat release by ILSM with 1% errors in the pressure; exact distribution (---) and recovered distribution (—). (b) Recovery of heat release by ILSM with 5% errors in the pressure; exact distribution (---) and recovered distribution (—). (c) Recovery of heat release by ILSM with 10% errors in the pressure; exact distribution (---) and recovered distribution (—).

Example 3: In this example, the recovery of heat release in the presence of axial temperature gradients and mean flow is considered. As discussed earlier, the corresponding integral equations have to be numerically integrated to solve for the oscillatory heat release distribution. So, the numerical integration using trapezoidal rule will introduce additional errors in the process. This would lead to increased sensitivity of the solution to the actual pressure data. To assess the quality of the results, a model problem is solved. The following parameters for the problem are assumed: mass flux ($\dot{m}$) $=50\ \mathrm{kg/(m^2\ s)}$; mean temperature is assumed to obey the functional form

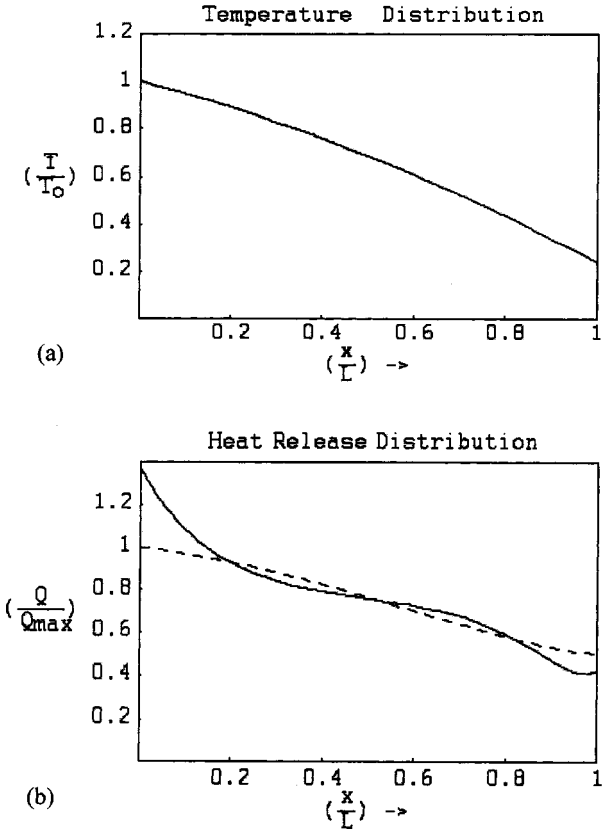


FIG. 3. (a) Temperature profile along the length of the combustor. (b) Recovery of heat release by ILSM with 10% errors in the pressure; exact distribution (---) and recovered distribution (—).

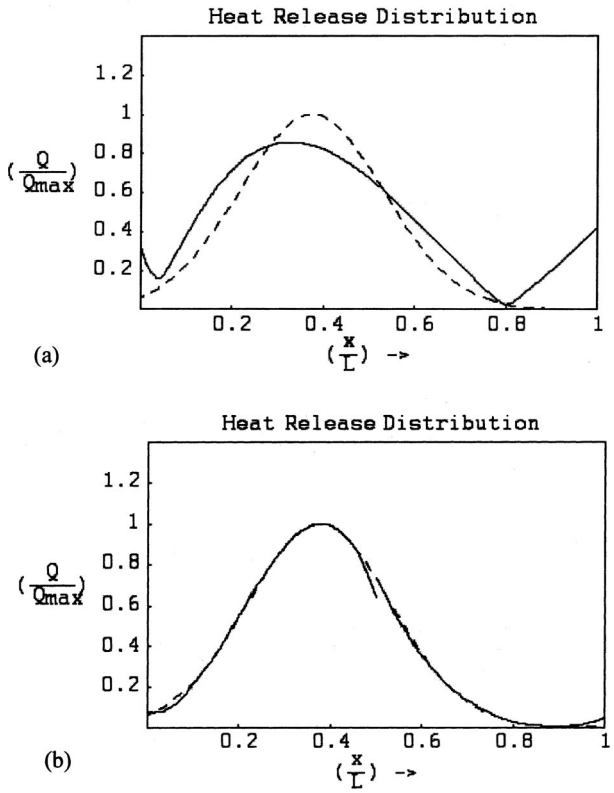


FIG. 4. (a) Recovery of heat release without domain decomposition; exact distribution (---) and recovered distribution (—). (b) Recovery of heat releases with domain decomposition; exact distribution (---) and recovered distribution (—).

$$\overline{T}(x)/T_0 = 1 - 0.1(x/L) - 0.25(x/L)^2 - 0.25(x/L)^3,$$

where $T_0 = 1000$ K and L the length of the combustor; mean pressure $(\overline{p}) = 2 \times 10^5$ Pa (assumed constant because $M^2 \ll 1$).

It is further assumed that the exact heat release distribution in the duct obeys the functional form

$$Q(x)/Q_{\max} = 1 - 0.25\left(\frac{x}{L}\right) - 0.5\left(\frac{x}{L}\right)^2 + 0.25\left(\frac{x}{L}\right)^4,$$

where $Q_{\max} = 10^6$ J/(m^3 s). The boundary conditions for the acoustic pressure amplitude are $P(0)/\overline{p} = 0.001$ and $P'(0)/(\overline{p}/L) = 0.01 + 0.01\,i$.

The sample calculations are performed at a frequency of $\omega = 500$ rad/s. Figure 3(a) shows the temperature distribution along the length of the combustor. Figure 3(b) shows the recovery of heat release in the presence of 10% errors in the pressure data. It should be noted that while the exact heat release distribution is a fourth-order polynomial, the present method approximates it by a third-order polynomial. Even in the presence of such approximation the results are quite satisfactory.

Example 4: In all the examples discussed so far, the pressure measurements are assumed to be confined to the combustion region, i.e., the region in which oscillatory heat release is nonzero. The mixing of fuel and air in a gas turbine combustor takes a finite amount of distance. Therefore, most of the combustion occurs in a region distributed about the center of the combustor. Ideally one would want to perform the pressure measurements at all axial locations along the length of the combustor. Therefore, a more appropriate profile of heat release distribution would be

$$Q(x)/Q_{\max} = \exp\left(-2\left(\frac{x}{L} - 0.375\right)^2\right),$$

where $Q_{\max} = 10^6$ J/(m^3 s). All the other parameters are the same as those assumed in Examples 1 and 2. As discussed earlier, in the implicit method, the exact heat release distribution (which is in the form of a Gaussian profile) is approximated by a third-order polynomial. Figure 4(a) shows the recovery of heat release distribution. It can be noticed that the results are quite unsatisfactory because the Gaussian profile cannot be approximated by a third-order polynomial over the entire domain. To overcome this difficulty, the domain is subdivided into two parts and the integral equation is solved separately. Figure 4(b) shows that the recovery is quite good when the domain decomposition is adopted.

V. CONCLUSIONS

In the existing literature, the inverse problem of determining the oscillatory heat release distribution from the knowledge of the acoustic pressure field within a combustor is formulated in terms of Fredholm integral equation. In this formulation, it is observed that the sensitivity of the solution to the exact pressure data is very high at the resonant frequencies of the combustor. Since the combustion instabilities usually occur when the combustion process excites one or more resonant modes, it is concluded that this technique in general is not feasible for ducted systems. To avoid this problem, the inverse problem is reformulated in terms of Volterra integral equation. The resulting Volterra integral equation is solved using the direct numerical method and the implicit least-squares method. The sensitivity analysis shows that the direct numerical method yields poor results in the presence of errors in the pressure data. Therefore, it is more appropriate to use implicit least-squares method to solve the inverse problem. The results show that the recovery of the heat release by implicit least-squares method is quite satisfactory even in the presence of errors as high as 10%. Therefore, it is concluded that this reformulation helps in recovering the heat release distribution successfully at all frequencies.

[1] K. Mcmanus, T. Poinsot, and S. M. Candel, "A review of active control of combustion instabilities," Prog. Energy Combust. Sci. **19**, 1–29 (1993).

[2] M. M. Lavrentev, V. G. Romanov, and S. P. Shishatskii, "Ill posed problems of Mathematical physics and analysis," *Translations of Mathematical Monographs*, Vol. 64 (1986).

[3] M. K. Ramachandra and W. C. Strahle, "Acoustic signature from flames as a combustion diagonistic tool," AIAA J. **21**, 1107–1112 (1983).

[4] Y.-C. Chao, "An implicit least-square method for the inverse problem of acoustic radiation," J. Acoust. Soc. Am. **81**, 1288–1292 (1987).

[5] T. N. E. Greville, "Some applications of the pseudo inverse of a matrix," SIAM Rev. **2**, 15–22 (1960).

[6] T. C. Lieuwen, Y. Neumeier, and B. T. Zinn, "Determination of unsteady heat release distribution in unstable combustor from acoustic pressure measurements," J. Propul. Power **15**, 613–616 (1999).

[7] B. Karthik, B. Manoj Kumar, and R. I. Sujith, "Exact solutions to one-dimensional acoustic fields with temperature gradient and mean flow," J. Acoust. Soc. Am. **108**, 38–43 (2000).

[8] F. R. Payne, "Direct Formal Integration (DFI): A global alternative to FDM/FEM," *Integral Methods in Science and Engineering* (Hemisphere, 1985), pp. 62–73.

[9] A. J. Jerri, *Introduction to Integral Equations with Applications*, 2nd ed. (Wiley, New York, 1999).

[10] G. Barton, *Elements of Green's Function and Propagation* (Clarendon, Oxford, 1989).

Comparison between ocean-acoustic fluctuations in parabolic-equation simulations and estimates from integral approximations

Stanley M. Flatté[a] and Michael D. Vera
Department of Physics, University of California at Santa Cruz, Santa Cruz, California 95064

(Received 2 May 2002; revised 25 April 2003; accepted 12 May 2003)

Line-integral approximations to the acoustic path integral have been used to estimate the magnitude of the fluctuations in an acoustic signal traveling through an ocean filled with internal waves. These approximations for the root-mean-square (rms) fluctuation and the bias of travel time, rms fluctuation in a vertical arrival angle, and the spreading of the acoustic pulse are compared here to estimates from simulations that use the parabolic equation (PE). PE propagations at 250 Hz with a maximum range of 1000 km were performed. The model environment consisted of one of two sound-speed profiles perturbed by internal waves conforming to the Garrett–Munk (GM) spectral model with strengths of 0.5, 1, and 2 times the GM reference energy level. Integral-approximation (IA) estimates of rms travel-time fluctuations were within statistical uncertainty at 1000 km for the SLICE89 profile, and in disagreement by between 20% and 60% for the Canonical profile. Bias estimates were accurate for the first few hundred kilometers of propagation, but became a strong function of time front ID beyond, with some agreeing with the PE results and others very much larger. The IA structure functions of travel time with depth are predicted to be quadratic with the form $\theta_v^2 c_0^{-2}(\delta z)^2$, where δz is vertical separation, c_0 is a reference sound speed, and θ_v is the rms fluctuation in an arrival angle. At 1000 km, the PE results were close to quadratic at small δz, with values of θ_v in disagreement with those of the integral approximation by factors of order 2. Pulse spreads in the PE results were much smaller than predicted by the IA estimates. Results imply that acoustic tomography of internal waves at ranges up to 1000 km can use the IA estimate of travel-time variance with reasonable reliability.
[DOI: 10.1121/1.1587732]

PACS numbers: 43.30.Cq, 43.30.Ft, 43.30.Re [WLS]

I. INTRODUCTION

It has long been known that the full wave equation is not necessary to describe propagation through sufficiently weak variability in a medium. Tatarskii, for example, used an approximate, parabolic wave equation to analytically characterize optical propagation through atmospheric turbulence in the 1960s.[1] The parabolic approximation also permits an accurate description of acoustic propagation through ocean internal waves.[2,3] Hardin and Tappert first applied what has come to be called the standard parabolic equation to the problems of ocean acoustics and presented a practical computer algorithm for solving it.[4,5] The use of the parabolic equation in numerical simulations of acoustic propagation through ocean internal waves was first described by Flatté and Tappert.[6] They investigated the fluctuations in a 100 Hz signal propagating to a range of 250 km.

The parabolic equation is analogous to the Schrödinger equation,[7] so it can be formally solved with the path-integral technique. It was shown in the 1970s that the path integral for propagation through a random media that is characterized by inhomogeneity, anisotropy, and a deterministic waveguide—as in the case of ocean acoustics—could be approximately solved in terms of line integrals along deterministic rays.[8,2,9] In order to determine the accuracy of the approximate integrals, they must be compared with computer-intensive parabolic-equation (PE) results. A systematic approach to this comparison has not been previously carried out. Our goal in this paper is to present a comparison of four quantities calculated from the integral approximations (IA) to corresponding results from simulations using the parabolic equation. The standard PE is used here because the integral expressions discussed below are approximations to the standard PE rather than the full wave equation. That is because the path integral can be done only if the quantity in the phase of the exponential is quadratic. The quantities compared include the following: the root-mean-square (rms) travel-time fluctuation (τ), the internal-wave-induced travel-time bias (τ_1), the induced spread in time of the acoustic pulse (τ_0), and the rms arrival-angle fluctuation, which characterizes the coherence of the acoustic signal at different depths (θ_v). An important motivation for an investigation into the accuracy of these approximations is their computational efficiency and the relative simplicity of subsequent analysis. Calculations using both techniques were performed with two different sound-speed profiles—a Munk Canonical profile[10] and a profile from the SLICE89 experiment[11]—for an acoustic frequency of 250 Hz. Three different internal-wave strengths—0.5, 1, and 2 times the reference Garrett–Munk (GM) level—were employed in both the integral approximations and PE simulations for each situation. For the parabolic

[a]Electronic mail: smf@pacific.ucsc.edu

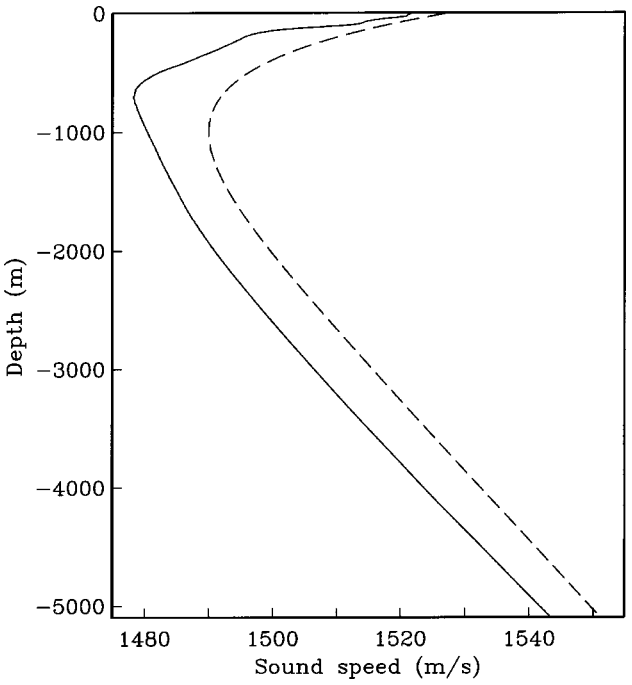


FIG. 1. The sound-speed profiles, $c(z)$, used in this paper. The SLICE89 (Canonical) profile is the solid (dashed) curve. At depth, their gradients are very close even though their sound speeds are different.

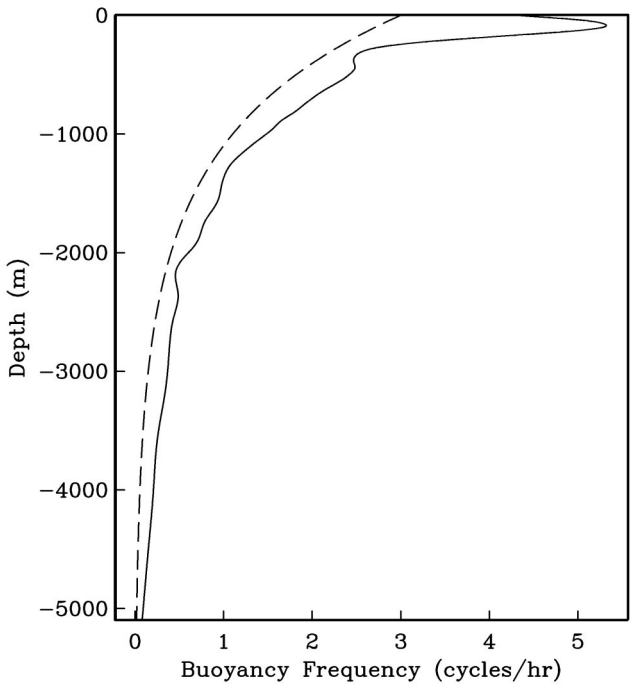


FIG. 2. Buoyancy frequency profiles. The SLICE89 (Canonical) profile is the solid (dashed) curve.

simulations, 50 different realizations of the stochastic internal-wave field were used for cases with 2 GM internal waves, 20 realizations were used for 1 GM, and 10 realizations were used for 0.5 GM. Information was recorded from all calculations at every 100 km interval to a maximum propagation range of 1000 km.

In Sec. II we present the ocean model used in the calculations. The PE simulations are discussed in Sec. III. In Sec. IV we describe the method of integral approximations. The comparison is made in Sec. V and the results are discussed in Sec. VI.

II. OCEAN MODEL

The simulations described in the following treat the ocean environment as an average sound-speed profile in depth perturbed by internal gravity waves that conform to the Garrett–Munk (GM) spectral model. Two different profiles were used, each possessing features appropriate to the temperate latitudes. The Munk Canonical profile[10] is an analytic model that takes the following form:

$$c(z)=c_{\text{axis}}[1+\epsilon(e^{\eta}-\eta-1)], \tag{1}$$

$$\eta=2(z-z_{\text{axis}})/B. \tag{2}$$

The parameters used here are $\epsilon=5.7\times10^{-3}$, $c_{\text{axis}}=1490$ m/s, $z_{\text{axis}}=-1000$ m, and $B=1000$ m. The other sound-speed profile used in this work was measured during the SLICE89 experiment that took place in the North Pacific.[11] The two sound-speed profiles are displayed in Fig. 1.

Both aspects of the simulated ocean—the deterministic profile and the stochastic perturbations—are expressed in the sound-speed equation,[2]

$$c(\mathbf{x},t)=c_0[1+U_0(z)+\mu(\mathbf{x},t)] \tag{3}$$

where $U_0(z)$ represents the dependence of the background profile $c(z)=c_0[1+U_0(z)]$ on depth and $\mu(\mathbf{x},t)$ contains the effect of internal waves on the speed. A useful definition of c_0 is the average of $c(z)$ over depth. For internal-wave vertical displacements, the value of μ can be expressed,[2]

$$\mu(\mathbf{x},t)=[\partial_z U_0(z)]_p\zeta(\mathbf{x},t), \tag{4}$$

where $\zeta(\mathbf{x},t)$ is the magnitude of the displacement and $[\partial_z U_0(z)]_p$ is the fractional potential gradient of the sound-speed profile.

The Garrett–Munk (GM) spectrum of internal-wave vertical displacement as a function of vertical mode number j and horizontal wave number k is

$$S_\zeta(j,k)=\frac{2B^2E}{\pi M}\frac{N_0}{N}\frac{1}{(j^2+j_*^2)}\frac{k_j k^2}{(k^2+k_j^2)^2}, \tag{5}$$

where $E=6.3\times10^{-5}$ (for the reference internal-wave energy referred to as one GM), $N_0=3$ cph and $j_*=3$ are empirical constants, $k_j\equiv\pi\omega_i j/N_0 B$, ω_i is the inertial frequency from the Earth's rotation at the latitude of the internal waves, and M is the normalization constant for the sum over mode number, $M=\Sigma_{j=1}^{\infty}(j^2+j_*^2)^{-1}$.[12–14] The latitude (used to define the inertial frequency) was taken as 30°. The Munk Canonical profile is derived using the assumption of an exponential buoyancy frequency,[10] $N(z)=N_s e^{z/B}$. The values used here are $N_s=3$ cycles per hour and $B=1000$ m. For simulations that use the SLICE89 sound-speed profile, a smoothed experimental buoyancy profile is used. Both buoyancy frequency functions are shown in Fig. 2.

For the analytic integrals, three derived moments of the spectrum are needed: variance of fractional sound-speed fluctuation $\langle\mu^2(z)\rangle$, with μ^2 from Eq. (4), and where $\langle\,\rangle$ means average over time; the correlation length parallel to a

given ray with depth z and horizontal angle θ, called $L_p(\theta,z)$; and $\{k_v^2\}$, where k_v is the vertical wave number, and $\{k_v^2\}$ is k_v^2 averaged over the internal-wave spectrum. (The internal-wave correlation length in the vertical is much shorter than in the horizontal.)

From Eq. (4),

$$\langle \mu^2 \rangle = \langle \zeta^2 \rangle [\partial_z U_0 - \gamma_A]^2 \approx \zeta_0^2 \frac{N_0}{N(z)} [\partial_z U_0 - \gamma_A]^2, \qquad (6)$$

where ζ_0 is the rms vertical fluctuation in an isodensity surface at the depth where the buoyancy frequency $N(z)$ is equal to the reference value N_0; and γ_A, called the adiabatic gradient, is the effect of pressure on the gradient of the sound channel, $U_0(z)$.

The following empirical expression for L_p has been shown to be sufficiently accurate:[15,16]

$$L_p(\theta,z) = L_{p0} \frac{[1-\exp[-(\sigma_c/\sigma_{\text{curv}})^p]]^q}{1+(\rho_s/\rho_c)^{q_a}(\sigma_c/\sigma_{\text{curv}})^{p_a}}, \qquad (7)$$

where $\rho_c = 3.5$, $\sigma_c = 0.0204$, $p = 0.385$, $p_a = 0.5$, $q = 1.3$, and $q_a = 2.0$ are dimensionless constants, and $L_{p0} = 12.7$ km. The quantity ρ_s is a measure of the anisotropy of the internal waves given by $\rho_s = N(z)\tan\theta/\omega_i$. The quantity σ_{curv} is a dimensionless measure of the ray curvature, obtained by dividing the real ray radius of curvature by about 1000 km. Typical internal-wave correlation lengths are a few kilometers in the horizontal direction and $O(100)$ m in the vertical.

The quantity $\{k_v^2\}$ follows directly from the GM spectrum.[14] Details are given in the reference; the result is

$$\{k_v^2\} = l_o^{-2} \left[\frac{N(z)}{N_0} \right]^2, \qquad (8)$$

and $l_o \approx 1000$ m.

III. PARABOLIC-EQUATION SIMULATIONS

The standard parabolic equation is used:

$$2iq_0 \frac{\partial \psi}{\partial r} + \frac{\partial^2 \psi}{\partial z^2} - 2\mu q_0^2 \psi = 0, \qquad (9)$$

where ψ is the acoustic field function (demodulated pressure, with the $1/\sqrt{r}$ from cylindrical spreading removed), $q_0 = \sigma/c_0$, and σ is the acoustic angular frequency. Any PE equation has the critical advantage that it can be advanced through a series of range steps without recourse to costly iterative or relaxation methods.[2,3,5,17]

Simulations with the Canonical profile had a source depth of 1000 m; the source in the SLICE89 simulations was located at a depth of 800 m. A total of 1024 frequencies spanned the 100 Hz bandwidth. The step size in range was 25 m and the number of vertical grid points spanning the water column from the surface to the bottom, $z_b = -5118.75$ m, was 2048. In order to implement the reflecting surface boundary condition, an image ocean was constructed above the surface for a total of 4096 vertical points. The bottom was flat with an absorbing boundary condition. The numerical internal-wave field for the parabolic simulations was constructed using a technique of Colosi and Brown.[18]

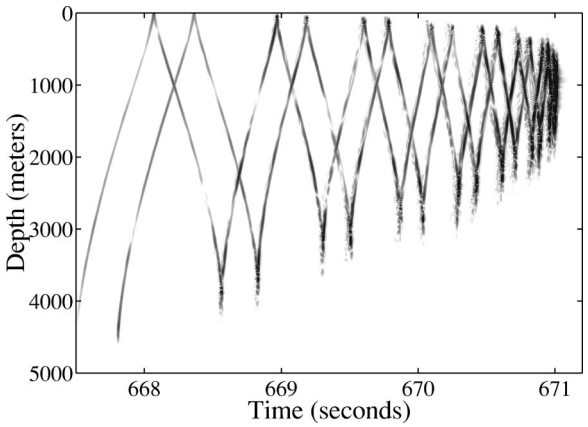


FIG. 3. A sample time front from a 250 Hz, Canonical-profile, parabolic-equation (PE) simulation with 2 GM internal waves. Significant front distortion is evident.

The simulated acoustic time fronts display the expected effects of internal-wave perturbations. The entire arrival, shown in Figs. 3 and 4, shows the reshaping of the transmission finalé.[19] Figures 5 and 6 show magnified sections of individual time fronts, showing the expected effects of internal-wave fluctuations.[11] A quantitative analysis of the simulation results requires the identification of the intensity peak for each examined depth and time front ID number. Only specified time front ID segments in the early-arriving portion of the front—sometimes referred to as "ray-like"—are considered in this analysis. The selected regions, in the SLICE89 case at a 1000 km range, have travel times between 673.7 and 674.6 s. The corresponding time interval in the Canonical case is from 668.6 to 669.6 s. Within each of these segments, a depth region was selected to avoid the caustic folds at the greatest and shallowest depths of each segment, and the points where segments cross. This was done in order to ensure an unambiguous connection between an intensity peak and a time front ID number.

The process of selecting intensity peaks begins with the construction of a time front using the rays of the geometrical-acoustics approximation. This series of ray arrivals was used as a template in locating intensity peaks in PE simulations for each profile in the absence of internal

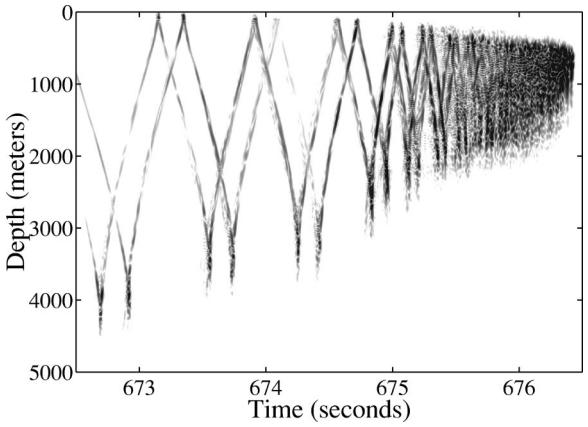


FIG. 4. A sample time front from a 250 Hz, SLICE89-profile, PE simulation with 2 GM internal waves. Significant distortion is evident, especially in the transmission finalé.

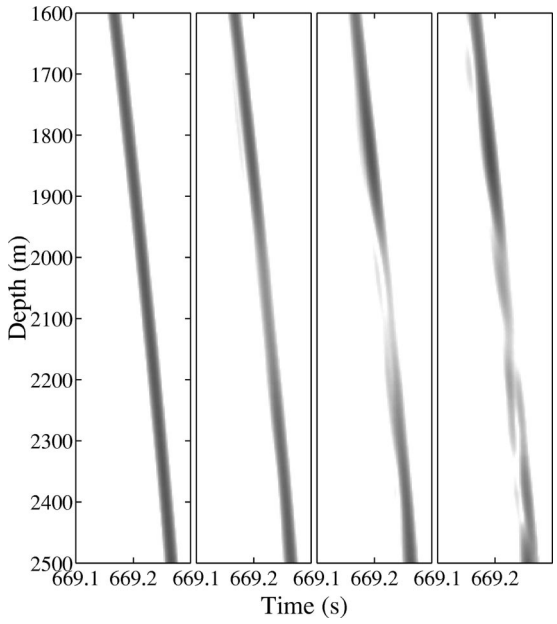


FIG. 5. The effect of internal waves on a time front segment (ID—36) from a 250 Hz, Canonical-profile, PE simulation. The internal-wave strengths are (from left to right) 0, 0.5, 1, and 2 GM.

waves. These peaks were then, in turn, used as a template in the analysis of the simulations performed with the internal-wave model in place.

Parabolic-equation simulations performed without internal-wave effects for each profile were compared with ray arrivals at each 100 km range interval. First, each local maximum in intensity greater than a small threshold intensity was examined for proximity to the ray arrivals. Whenever multiple candidate peaks existed for the same depth and ID—a relatively rare occurrence in the absence of internal waves—that candidate with the highest value of the following quality function was selected:[20]

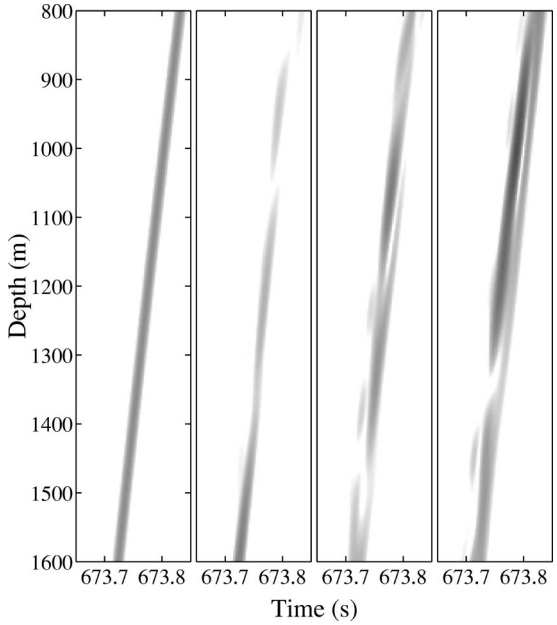


FIG. 6. The same as the previous figure, but for an ID of −37 and the SLICE89 profile.

$$Q = I_{\text{cand}}\left(1 - \sqrt{2\left|\frac{t_{\text{cand}} - t_{\text{tmplt}}}{t_Q}\right|}\right). \tag{10}$$

This function depends on both the intensity of the candidate peak, I_{cand}, and its travel time, t_{cand}. The difference between that travel time and the time on the line connecting adjacent ray arrivals at the depth of the candidate peak, t_{tmplt}, influences Q as a fraction of t_Q, where $t_Q = 160$ ms. This process is influenced both by a peak's intensity and its proximity to the template front in determining an intensity maximum that will represent the simulation results for a given depth and time front ID.

The simulated time fronts altered by internal waves were analyzed similarly, except for the following. No intensity threshold was implemented; all local maxima in intensity were considered. The value of t_Q was reduced to 80 ms, placing a greater emphasis on the proximity to the template. Because both sets of PE simulations were performed on the same depth grid, t_{tmplt} is simply the travel time of the template peak at the same depth and ID as the candidate. The need for the quality function, Q, arises much more often in this selection process as the presence of internal-wave disturbances frequently leads to multiple, local intensity maxima in the vicinity of the template segments; a few doubles can be seen, for example, in Figs. 5 and 6.

The selection of intensity peaks from the simulations both with and without intervening internal waves yields the travel time $T_i(z,\text{ID})$ for the acoustic signal at each relevant combination of realization i, depth z, and time front ID number.

The investigation into the fluctuations τ, τ_1, and θ_v involve the difference in travel times $T_i'(z,\text{ID})$ caused by the speed disturbances,

$$T_i'(z,\text{ID}) = T_i(z,\text{ID}) - T_{\text{noiw}}(z,\text{ID}), \tag{11}$$

where the subscript noiw indicates the time without internal waves. Values of $T_i'(z,\text{ID})$ were calculated at every 100 km range interval over the 1000 km propagation range. These values are then averaged over the examined depth interval according to $T_i'(\text{ID}) = (1/N_z)\Sigma_z T_i'(z,\text{ID})$. This process was repeated for each combination of sound-speed profile and internal-wave strength.

The internal-wave bias τ_1 is given by the average of these values,

$$\tau_1(\text{ID}) = \frac{1}{N_i}\sum_i T_i'(\text{ID}). \tag{12}$$

The rms travel-time fluctuation, τ, is given by

$$\tau(\text{ID}) = \left[\frac{\Sigma_i[T_i'(\text{ID}) - \tau_1(\text{ID})]^2}{N_i - 1}\right]^{1/2}. \tag{13}$$

When an average value of τ or τ_1 at a given range was necessary, the values corresponding to each ID segment were averaged.

The structure function $S_z(\delta z)$ of travel time with depth is determined from differences in travel time at depths separated by δz. The structure function is given by averages of differences in $T_i'(z,\text{ID})$,

$$S_z(\delta z,\mathrm{ID}) = \frac{1}{N_{\delta z}} \sum_i \sum_z (T_i'(z,\mathrm{ID}) - T_i'(z+\delta z,\mathrm{ID}))^2, \quad (14)$$

where $N_{\delta z}$ is the total number of available elements in the sum that are separated by δz. Finally, the structure function as used in this discussion is calculated by summing the $S_z(\delta z,\mathrm{ID})$ from the different ID segments,

$$S_z(\delta z) = \frac{1}{N_{\mathrm{ID}}} \sum_{\mathrm{ID}} S_z(\delta z,\mathrm{ID}). \quad (15)$$

Thus, the value of $S_z(\delta z)$ begins at 0 for $\delta z = 0$ and increases with increasing depth separation. The value of θ_v is determined by a quadratic fit to $S_z(\delta z)$. [See Eq. (25).] Note that previous work has discussed the vertical coherence of travel time in terms of the product of travel times at different depths.[20,19,11] In order to compare PE results with integral approximations, the structure function is used instead. (See Sec. V.)

The mean pulse shape was constructed for each time front ID at each range, using intensity information from a 200 ms window centered on each selected peak. Each such $I_i(t,z,\mathrm{ID})$ was translated by the time of the intensity maximum so that its highest value occurs at $t - T_i = 0$, thus forming $I_i(t - T_i, z, \mathrm{ID})$. Finally, all of these pulse shapes were averaged over depth and internal-wave realization number to generate the mean pulse shape, $\langle I(t - T_i, \mathrm{ID})\rangle$, for each time front ID number. This method of looking at the pulse shape does not have the extra broadening due to the convolution of a pulse with itself that occurs with the intensity autocorrelation function.[2] These mean pulse shapes have a greater temporal width than those generated by simulations without internal waves. The characteristic scale of this internal-wave-induced spread is τ_0.

Estimates for τ, τ_1, θ_v, and τ_0 are thus derived from the PE simulations using the above formulas. In Sec. IV, the same quantities are estimated by approximations generated by integrals along ray paths. In Sec. V, the PE and IA estimates are compared with each other.

IV. INTEGRAL-APPROXIMATION TECHNIQUE

The integral approximations presented here begin with the calculation of ray trajectories, subject only to the influence of the depth-dependent sound-speed profile. The results of the analytic approximations take the form of weighted integrals over the deterministic ray from the source to the receiver. It is our purpose in this work to investigate the accuracy of the statistical quantities calculated, relative to PE simulation results, so that the range of validity of these corrections to geometrical ray theory can be estimated for given accuracy requirements.

These integral expressions provide estimates of parameters in the second moments of the acoustic field function. The second moment for changes in depth can be expressed as

$$\langle \psi^*(\delta z)\psi(0)\rangle \approx \exp[-\tfrac{1}{2}(\theta_v^2 q^2 \delta z^2)], \quad (16)$$

where θ_v is a constant that is equal to the rms arrival-angle fluctuation if intensity variations can be neglected. (See Sec. V C.)

Similarly, for changes in frequency,

$$\langle \psi^*(\delta\sigma)\psi(0)\rangle \approx \exp[-\tfrac{1}{2}(\delta\sigma\tau)^2 + i\,\delta\sigma\tau_1 - \tfrac{1}{2}(\delta\sigma\tau_0)^2], \quad (17)$$

where σ is the acoustic angular frequency, τ^2 gives the variance of travel time, τ_1 is the internal-wave bias (the average difference in travel time caused by the presence of the speed disturbances), and τ_0 describes the spreading of an intensity peak in time due to internal waves.[21,9]

The variance of travel time, τ^2, is calculated from the properties of internal waves in the following way:[2]

$$\tau^2 = c_0^{-2} \int_0^R dx \langle \mu^2(z)\rangle L_p(\theta,z). \quad (18)$$

As previously stated, the integral is performed along the ray trajectory. The sound speed variance, $\langle \mu^2(z)\rangle$, is a profile in depth given in Eq. (6). The quantity $L_p(\theta,z)$ expresses the correlation length of the internal waves along the ray direction [see Eq. (7)]. This is the simplest of the integral approximations.

The integrals in the following expressions involve the depth separation of nearby unperturbed rays that differ by defined amounts at the source or receiver. For the ray, $z_{\mathrm{ray}}(x)$, and a nearby ray, $z(x)$, this separation is

$$\xi(x) = z(x) - z_{\mathrm{ray}}(x). \quad (19)$$

The behavior of the vertical separation in the sound channel, when assumed small, is governed by[2,15]

$$\partial_{xx}\xi(x) + U_0''\xi(x) = 0. \quad (20)$$

Of particular interest is the solution, ξ_1, that yields the vertical separation as a function of range for two rays that begin at the same source location and arrive at the receiver with unit separation. This function appears in the depth-coherence integral. The Green's function generated by the related equation, $\partial_{x'x'}g(x,x') + U_0''g(x,x') = \delta(x'-x)$, is another necessary component of the acoustic integrals to be performed.

The variance of the acoustic arrival angle can be expressed as an integral along the ray trajectory in the following way:

$$\theta_v^2 = \ln(\sigma\tau) \int_0^R dx \langle \mu^2\rangle L_p(\theta,z)\{k_v^2\}[\xi_1(x)]^2, \quad (21)$$

where the quantity $\{k_v^2\}$ is the average over the internal-wave spectrum of the internal-wave vertical wave number squared, from Eq. (8), with l_0 calculated as in Ref. 14. Note that θ_v^2 is nearly independent of frequency; the logarithm is a very slow function of frequency.

The internal-wave bias, τ_1, is the average change in travel time caused by the presence of sound-speed fluctuations. It is calculated according to

$$\tau_1 = \frac{\ln(\sigma\tau)}{2c_0} \int_0^R dx \langle \mu^2\rangle L_p(\theta,z)\{k_v^2\}g(x,x). \quad (22)$$

Analytic integral techniques also offer an estimate of the pulse spreading due to internal-wave fluctuations, τ_0. This quantity characterizes the additional width in time of the intensity peak for a time front ID segment, at the ray's arrival

depth, caused by the presence of internal-wave fluctuations. The pulse spreading is given by

$$\tau_0^2=\frac{1}{2}\left(\frac{\ln(\sigma\tau)}{c_0}\right)^2\int_0^R dx\langle\mu^2\rangle L_p(\theta,z)\{k_v^2\}\int_0^R dx'\langle\mu^2\rangle' \times L_p(\theta',z')\{k_v^2\}'[g(x,x')]^2. \quad (23)$$

These integral approximations were implemented using a computer code package referred to as "CAFI" (Computation of Acoustic Fluctuations from Internal waves) prepared by Flatté and Rovner.[15] The profiles of $\langle\mu^2\rangle$ that are needed by CAFI were generated according to Eq. (6). The weighting functions that enter into the integrals from the internal-wave spectrum depend on depth and not range. They are therefore periodic with range when evaluated along a ray. However, the weighting functions that are geometric in nature, that is $\xi_1(x)$ and $g(x,x')$, are not periodic in range. Nevertheless, the nature of these geometrical weighting functions is such that great gains in computational speed are achievable compared with straightforward integration.[15] Approximately 6500 ray paths were followed in each case at launch-angle intervals of 0.04°, and the integral approximations were performed for each one.

V. COMPARISON OF RESULTS FROM INTEGRAL EXPRESSIONS AND THE PARABOLIC EQUATION

In the following sections, four quantities calculated from integral approximations have been compared to corresponding results from simulations using the parabolic equation: τ, τ_1, τ_0, and θ_v.

A. Root-mean-square travel-time variability τ

Results for τ as a function of range, averaged over ID, are displayed in Fig. 7. The values for individual time front ID segments at a range of 1000 km are shown in Fig. 8. This quantity is among the most frequently used in the analysis of experimental data.[11] As can be seen in Eq. (18), the integral approximation is not a function of the acoustic frequency. The predicted magnitude of τ increases as the square root of range, increases with internal-wave strength, and also varies somewhat with the choice of sound-speed profile and time front ID number. The statistical uncertainty in the PE values was approximately 10% for the cases employing 50 realizations.

SLICE80 analytic values of τ were within statistical uncertainty of the 1000 km simulation results. Canonical values were within the statistical error of the simulation results in a few cases, at low time front ID numbers and internal-wave strengths. In other cases they overestimated by between 20% and 60%.

B. Internal-wave-induced travel-time bias τ_1

The calculated travel-time bias due to internal waves, τ_1, is displayed in Fig. 9. The integral approximation is a very slow function of acoustic frequency in this case (in fact logarithmic, and only as an overall factor that is a very slow function of range). The choice of sound-speed profile significantly affects the result, particularly for long ranges and

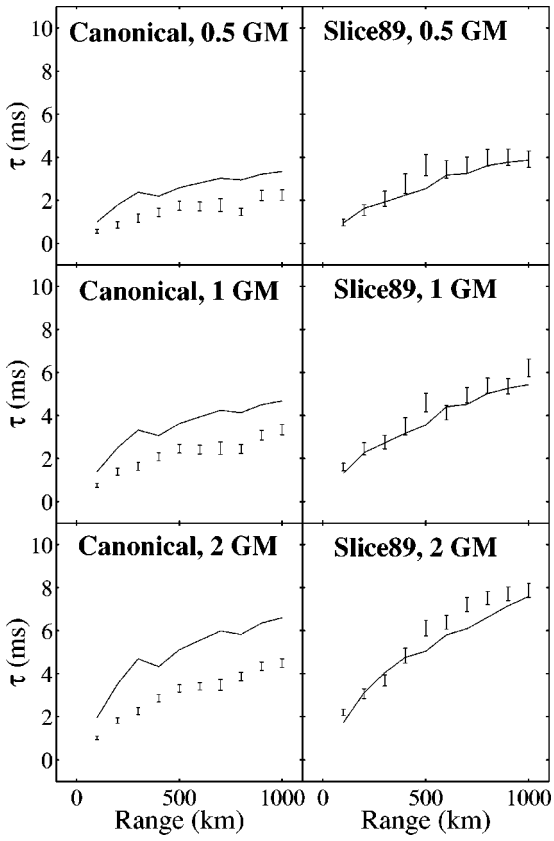


FIG. 7. The rms travel-time fluctuation τ for a 250 Hz acoustic signal as a function of range for different combinations of internal-wave strength and profile. Integral-approximation (IA) and PE-simulation results are shown as lines and points, respectively.

strong internal waves. The approximations and the PE results agree up to a range of 400–500 km. At 500 km, the SLICE89 analytic approximations are within the statistical uncertainty of the simulations. Except for the case with 2 GM internal waves, the Canonical approximation is within the simulation error at 400 km. At larger range the IA results for the Canonical profile diverge slowly from the PE results, differing at 1000 km by close to a factor of 10. At larger range for the SLICE89 profile the IA results begin to differ drastically between IDs. A few of the IDs have τ_1 very close to the PE result while others have τ_1 extremely large. Previous work in this area has included a comparison of the SLICE89 experimental results with integral approximations to τ and τ_1, showing reasonable agreement.[19] However, later work showed that the approximation in the 1994 paper used an incorrect formula for L_p; the later work was more in agreement with our results.[22,23]

C. Root-mean-square acoustic arrival-angle fluctuation θ_v

Equation (16) involves the product of two acoustic field functions. An appropriate approximation in cases in which the time front is easily identified, and the intensity varies slowly relative to the phase, is to approximate the product as

$$\langle\psi^*(\delta z)\psi(0)\rangle\approx\exp[-\tfrac{1}{2}\sigma^2 S_z(\delta z)], \quad (24)$$

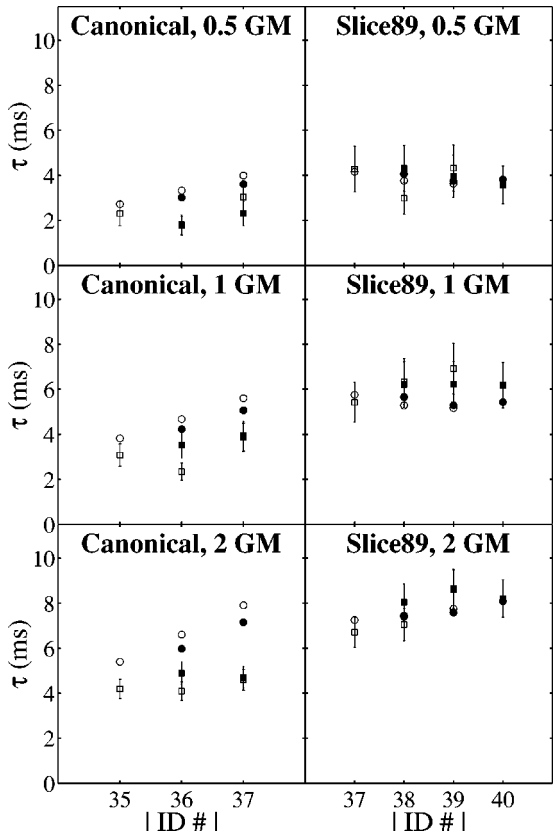


FIG. 8. The rms travel-time fluctuation τ for a 250 Hz acoustic signal as a function of time front ID number at 1000 km. The circles are the IA estimates and the squares result from the PE simulations. Solid (open) symbols correspond to positive (negative) ID numbers.

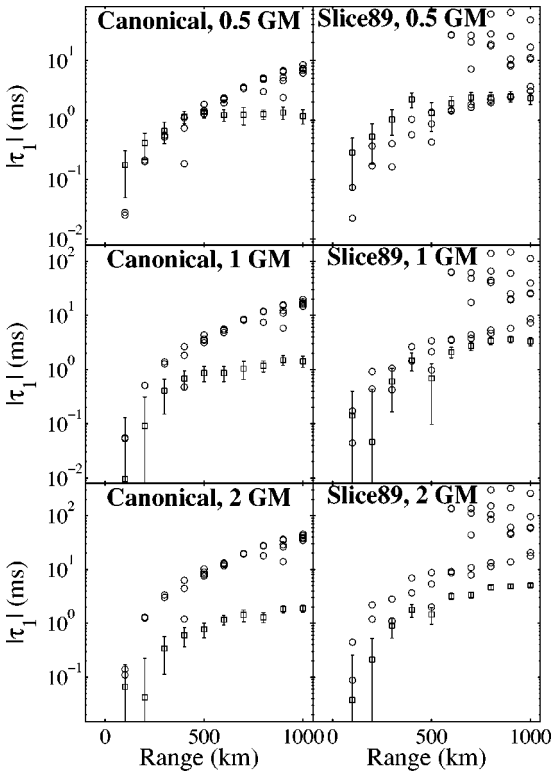


FIG. 9. The internal-wave bias τ_1 (the average difference from the travel time in the absence of internal waves) for each ID, for a 250 Hz acoustic signal, as a function of range. The IA values for each ID (circles) and the PE results (points) differ substantially at long ranges, particularly for high internal-wave strengths. The SLICE89 results at long range are highly variable between IDs.

where S_z is the structure function of the arrival time as a function of depth.

Hence, a comparison of Eq. (16) with Eq. (24) results in a relation between $S_z(\delta z)$ and θ_v^2:

$$S_z(\delta z) = \theta_v^2 c_0^{-2} \delta z^2. \tag{25}$$

Structure-function results from both the simulations and integral techniques are displayed in Fig. 10 for an acoustic frequency of 250 Hz and a range of 1000 km. The IA results use Eq. (25) with θ_v determined from Eq. (21) averaged over ID, which are shown in Table I. The PE points, equally spaced at 2.5 m, are shown averaged over all ID segments in each case; these are calculated as in Eq. (15). Also shown are quadratic fits to the PE points from each profile that have $\delta z < 20$ m; these fits are evidently good from the figure, and they yield the PE values of θ_v shown in Table I. Immediately beyond 20 m, the fits diverge from the points because the points continue nearly as straight lines, at least to $\delta z = 100$ m, rather than as continuations of the parabolic fits. The fact that the PE points behave quadratically near $\delta z = 0$ is in agreement with the prediction of the IA formula.

IA θ_v values are not greatly affected by the choice of speed profile; they are so close (within 10%) as to be indistinguishable, except for the 2 GM case, in Fig. 10. The IA values are also very close for the six different IDs that go into making the average. The internal-wave strength has a significant effect on θ_v, with values at 2 GM typically more than twice as large as the corresponding values at 0.5 GM.

The PE results for θ_v are typically of order twice the IA values; they are also dependent on the specific ID, with different IDs varying by as much as factors of order two from each other. This is the reason for the larger uncertainties in the θ_v values from the PE versus the IA calculations in Table I.

Reported results from the SLICE89 experiment (which used a 250 Hz source, and a range of 1000 km) include a range of coherence lengths from 60 m to 1 km (which translates to θ_v values of 0.05°–1°), though a different type of analysis was done and a different set of time front IDs was examined.[11] An analysis of the experiment suggests an internal-wave strength of 0.5 GM.[20,19] For those conditions, the PE estimate of θ_v is $2.0 \pm 0.7°$ and the IA estimate is $0.62 \pm 0.06°$.

D. Internal-wave-induced pulse spreading τ_0

The investigation into the accuracy of the integral approximation for the pulse spreading caused by the presence of internal waves, τ_0, is based on a comparison of the mean pulse shapes, $\langle I(t - T_i, \mathrm{ID}) \rangle$, from parabolic simulations with internal waves to those without internal waves. The construction of these mean pulses is discussed in Sec. III. Pulses are pictured in Fig. 11 for each speed profile, an acoustic frequency of 250 Hz, and an internal-wave strength of 1 GM at a range of 1000 km. The IA pulse-spread approximations, calculated according to Eq. (23) and averaged over each time front ID, are given in the figure caption.

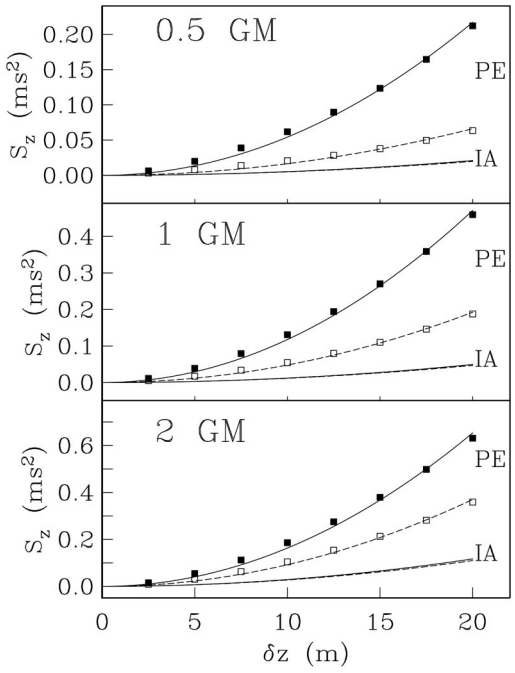


FIG. 10. The depth structure function of travel time $S_z(\delta z)$ as a function of vertical separation δz; IA and PE results are shown, averaged over ID, at 1000 km. The SLICE89 (Canonical) PE results are the closed (open) squares, with the solid (dashed) parabolic fits. The IA values of θ_v are so close that the IA curves for SLICE89 (solid) and Canonical (dashed) lie on top of each other except for the 2 GM case. The PE results behave close to quadratically at small δz as predicted by the integral approximations. The fitted PE values of θ_v are typically twice that of the IA values. Values of θ_v for both PE and IA calculations are shown in Table I.

In the PE simulations, internal-wave-influenced mean pulses are often nearly indistinguishable from those without internal-wave disturbances. The IA-predicted spreads are much greater than those seen in the PE curves. When the effects of internal waves on the mean pulse shape can be seen at all, it is intriguing that they often consist of an increase in intensity at early times. This early time signal is similar to that observed and predicted in the AFAR experiment at much shorter ranges.[2,24] The AFAR prediction required a calculation that could not be done at 1000 km. As mentioned above, even when the pulses at 1000 km display the sort of spreading that could potentially be explained in terms of a Gaussian function with a width increase given by a single τ_0 parameter, the IA value greatly exceeds the spreading seen in the PE results, sometimes by an order of magnitude.

TABLE I. Values of rms arrival-angle fluctuation θ_v, in degrees, calculated from PE simulations and IA calculations for both Canonical and SLICE89 sound-speed profiles. The first column is the internal-wave energy level in units of the Garrett–Munk reference level. The uncertainty values are equal to the rms spread over ID (*not* divided by the square root of the number of IDs, which is about 6).

	Canonical		SLICE89	
GM level	PE	IA	PE	IA
0.5	1.1±0.5	0.6±0.09	2.0±0.7	0.62±0.06
1.0	1.9±0.5	0.9±0.1	2.9±0.3	0.96±0.08
2.0	2.6±0.5	1.4±0.2	3.5±0.3	1.5±0.1

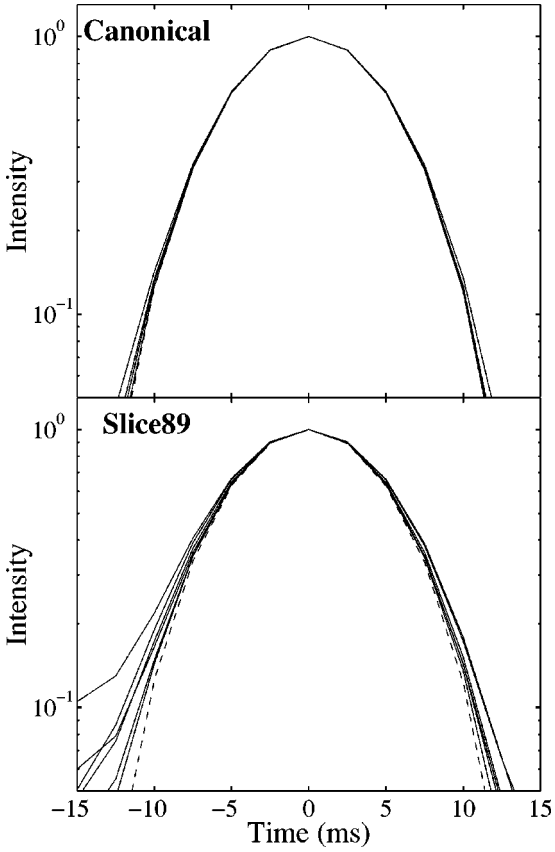


FIG. 11. Mean PE pulse shapes as a function of time for a 250 Hz acoustic signal at 1000 km for both the SLICE89 and Canonical profiles. The dashed curve in each panel is the pulse without internal waves present; the solid curves correspond to different IDs in simulations with 1 GM internal waves. The internal-wave-induced spread τ_0 from IA estimates for these IDs are much larger than seen in the PE curves; the IA values are between 9.4 and 12.7 ms for the Canonical cases and 12, 12, 16, 19, 25, and 71 ms for the SLICE89 cases.

Similar findings for the integral approximation of pulse spreading were reported for the 3250 km, 75 Hz Acoustic Engineering Test (AET), part of the Acoustic Thermometry of Ocean Climate (ATOC) project.[22,23] It was found that the τ_0 approximation was too large by more than an order of magnitude. It was also reported that the effects of internal waves on the mean pulse shape are more accurately characterized as increases in the mean intensity at times away from the maximum, rather than as spreading of the central peaks themselves.

VI. DISCUSSION AND CONCLUSIONS

A. Summary

The comparisons that have been made have serious implications for the usefulness of some of the analytic integral expressions in providing accurate estimates of their related fluctuation quantities *at long range*. Long-range acoustic tomography of internal waves has, up until now, depended on the comparison in the regime of identified rays of observed travel-time variance to a calculation of travel-time variance that depends on internal-wave strength. It has been suggested that the use of additional observed fluctuation quantities such as pulse spread, rms arrival-angle fluctuation in the vertical,

and travel-time bias may be useful in extending the power of long-range, internal-wave tomography, if they can be compared with calculations based on integral approximations.[25]

For rms travel-time fluctuation, the results in this paper confirm that calculations of the appropriate integral approximation are reliable up to 1000 km for the SLICE89 profile, but differ by 20%–60% for the Canonical profile. For the depth structure function, PE results show close to quadratic behavior, allowing the calculation of a rms arrival-angle fluctuation. The IA values are typically about one-half of the PE values of the rms arrival-angle fluctuation. For the bias τ_1, the approximation is accurate up to between 400 and 500 km; thereafter the IA results vary widely between IDs, while the PE results behave smoothly. No increase of pulse width due to internal waves could be observed in the PE results, while IA values should have been easily seen. It should be emphasized that previous work at a range of 35 km has shown remarkable agreement between these integral approximations and observed fluctuations.[9,24] The results in this paper extend this agreement for the bias to a few hundred kilometers; the results at 1000 km for rms arrival-angle fluctuation show factors of 2 disagreement, but for the pulse spread there is strong disagreement.

B. Reasons for success or failure

In this section, the reasons for the success or failure of each quantity are discussed. From the point of view of acoustic tomography of internal waves, the use of travel-time fluctuations for determining the distribution of internal-wave spectral strength is well borne out. These results offer additional support for the utility of τ as a fundamental parameter in tomographic measurements of internal-wave strength. There are several levels of acoustic tomography of τ: (1) The travel-time variance averaged over all identifiable timefront segments determines the internal-wave energy (E) averaged over depth and range; (2) a separate measurement for each identifiable time front segment provides some information about the distribution in depth of the range-averaged E; (3) the specific deterministic ray corresponding to a given depth on a timefront segment offers some information about the combined depth and range dependence of E. In the work described here, the analytic integral representation of τ *for each ID*, at 1000 km, is found to be within the statistical uncertainty ($\approx$10%) of PE simulation results for the SLICE89 profile, and between 20% and 60% for the Canonical profile, depending on which ray is estimated.

Thus, the fundamental quantity τ is found to be successful, supporting the correctness of the basic assumptions involved in that analytical integral expression. The quantity θ_v represents the behavior of the PE-calculated depth structure function of travel time; however, the PE values are about twice those of the IA values. Thus, they can be used only semi-quantitatively. The quantities τ_1 and τ_0 from the analytic integral expressions do not correspond to the PE results over the entire propagation range. The speculations below attempt to explain this failure and have a reasonable physical basis.

In doing these evaluations of success or failure, it has been assumed that the PE simulation results are accurate within the statistical measures of accuracy that are known. That assumption depends only on the accuracy of the parabolic equation, and on the numerical representation of the internal-wave field. The PE code accuracy is not in question. The particular internal-wave code used is appropriate in that it makes the same assumptions as the analytical integral expressions. Though the internal-wave models are sufficient for the purposes of this comparison, another approach might be more fruitful in an attempt to compare to actual experimental data. One example is the repeated ''patching'' of short-range (16 km) segments of sample internal waves.[19] This technique does not result in internal-wave components that are coherent over the entire propagation range; physically, damping and scattering of internal-wave components would be expected over shorter distances.

The rms arrival-angle fluctuation has the most simple analytic expression beyond the travel-time variance. The calculation of θ_v depends on a simple geometrical ray-tube quantity that to first order does not depend on frequency, as well as on $\{k_v\}^2$, which is an estimate of the curvature of the internal-wave sound-speed fluctuation in the vertical direction. Note that this curvature is dominated by internal-wave vertical modes with numbers of about 50, whose vertical wavelengths are of order 20 m.[25] When δz is above 20 m, approximating the internal-wave sound-speed fluctuation transverse to the ray by a curvature is not going to be accurate. This explains why the PE results do not follow a parabola above $\delta z = 20$ m. Why the accuracy is no better than a factor of 2 below 20 m is not clear.

Both τ_1 and τ_0 involve the Green's function, and therefore require that $\{k_v^2\}$ provide a valid estimate of the sound-speed curvature over a Fresnel length. For an acoustic frequency of 250 Hz, this distance is of order several hundred meters. From the discussion of θ_v above, these expressions are not expected to be accurate.

Finally, note that the analytic integral expressions have been derived under the assumption of a single acoustic frequency at the source. The PE simulations are done with a bandwidth of 100 Hz, for the simple reason that individual pulses cannot be distinguished in long-range propagation otherwise. Thus, single-frequency expressions are being compared with broadband numerical simulations. It is not clear how to test whether this is an important factor or not, but it seems likely this would affect the pulse-spread behavior.

C. Other methods

Experiments and PE simulation have already demonstrated that there is another useful measurement that does not depend on analytic integral estimates of the type considered here. The behavior of the finalé of the pulse, within which one cannot identify time fronts, has yielded clear measurements of the internal-wave strength.[19] Thus, there are at least two useful experimental measurements of internal-wave strength at a range of 1000 km (and presumably farther, for the reasons discussed above); some of the integral approximations discussed above for other quantities provide semi-quantitative results. There is always the possibility of using

PE simulations to compare with experiment, but this would be extremely computer intensive. Other analytic approaches to long-range internal-wave tomography may in the future lead to making the quantities discussed above, or others, more useful. The possibility of using intensity within the pulse has not been investigated.[23]

[1] V. Tatarskii, ''The effects of the turbulent atmosphere on wave propagation,'' translated by Israel Program for Scientific Translations, originally published in 1967 (U.S. Dept. of Commerce, National Technical Information Service, 1971).

[2] S. Flatté, R. Dashen, W. Munk, K. Watson, and F. Zachariasen, *Sound Transmission Through a Fluctuating Ocean, Mechanics and Applied Mathematics* (Cambridge University Press, Cambridge, 1979).

[3] F. Jensen, W. Kuperman, M. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, Woodbury, NY, 1994).

[4] R. Hardin and F. Tappert, ''Applications of the split-step Fourier method to the numerical solution of nonlinear and variable coefficient wave equations,'' SIAM Rev. **15**, 423 (1973).

[5] F. Tappert, ''The parabolic approximation method,'' in *Wave Propagation and Underwater Acoustics*, edited by J. Keller and J. Papadakis (Springer-Verlag, Berlin, 1977), pp. 224–287.

[6] S. Flatté and F. Tappert, ''Calculation of the effect of internal waves on oceanic sound transmission,'' J. Acoust. Soc. Am. **58**, 1151–1159 (1975).

[7] S. Flatté, ''The Schrödinger equation in classical physics,'' Am. J. Phys. **54**, 1088–1092 (1986).

[8] R. Dashen, ''Path integrals for waves in random media,'' J. Math. Phys. **20**, 894–920 (1979).

[9] R. Dashen, S. Flatté, and S. Reynolds, ''Path-integral treatment of acoustic mutual coherence functions for rays in a sound channel,'' J. Acoust. Soc. Am. **77**, 1716–1722 (1985).

[10] W. H. Munk, ''Sound channel in an exponentially stratified ocean, with application to SOFAR,'' J. Acoust. Soc. Am. **55**, 220–226 (1974).

[11] T. Duda, S. M. Flatté, J. Colosi, B. Cornuelle, J. Hildebrand, W. Hodgkiss, Jr., P. Worcester, B. Howe, J. Mercer, and R. Spindel, ''Measured wavefront fluctuations in 1000-km pulse propagation in the Pacific Ocean,'' J. Acoust. Soc. Am. **92**, 939–955 (1992).

[12] C. Garrett and W. Munk, ''Space–time scales of ocean internal waves,'' Geophys. Fluid Dyn. **2**, 225–264 (1972).

[13] C. Garrett and W. Munk, ''Space–time scales of internal waves: a progress report,'' J. Geophys. Res. **80**, 291–297 (1975).

[14] R. Esswein and S. Flatté, ''Calculation of the phase-structure function density from oceanic internal waves,'' J. Acoust. Soc. Am. **70**, 1387–1396 (1981).

[15] S. Flatté and G. Rovner, ''Calculations of internal-wave-induced fluctuations in ocean-acoustic propagation,'' J. Acoust. Soc. Am. **108**, 526–534 (2000).

[16] S. Flatté and G. Rovner, ''Path-integral expressions for fluctuations in acoustic transmission in the ocean waveguide,'' in *Methods of Theoretical Physics Applied to Oceanography*, Proceedings of the 9th ''Aha Huliko'a Hawaiian Winter Workshop,'' edited by P. Müller, 1997, pp. 167–174.

[17] S. Flatté, ''Wave propagation through random media: Contributions from ocean acoustics,'' Proc. IEEE **71**, 1267–1294 (1983).

[18] J. Colosi and M. Brown, ''Efficient numerical simulation of stochastic internal-wave-induced sound-speed perturbation fields,'' J. Acoust. Soc. Am. **103**, 2232–2235 (1998).

[19] J. Colosi, S. M. Flatté, and C. Bracher, ''Internal-wave effects on 1000-km oceanic acoustic pulse propagation: Simulation and comparison with experiment,'' J. Acoust. Soc. Am. **96**, 452–468 (1994).

[20] J. Colosi, Ph.D. thesis, University of California at Santa Cruz, 1993.

[21] S. Flatté and R. Stoughton, ''Predictions of internal-wave effects on ocean acoustic coherence, travel-time variance, and intensity moments for very long-range propagation,'' J. Acoust. Soc. Am. **84**, 1414–1424 (1988).

[22] J. A. Colosi, E. K. Scheer, and S. M. Flatté, ''Comparisons of measured and predicted acoustic fluctuations for a 3250-km propagation experiment in the eastern North Pacific Ocean,'' J. Acoust. Soc. Am. **105**, 3202–3218 (1999).

[23] J. Colosi, F. Tappert, and M. Dzieciuch, ''Further analysis of intensity fluctuations from a 3252-km acoustic propagation experiment in the eastern North Pacific,'' J. Acoust. Soc. Am. **110**, 163–169 (2001).

[24] S. Reynolds, S. Flatté, R. Dashen, B. Buehler, and P. Maciejewski, ''AFAR measurements of acoustic mutual coherence functions of time and frequency,'' J. Acoust. Soc. Am. **77**, 1723–1731 (1985).

[25] S. Flatté, ''Principles of acoustic tomography of internal waves,'' *Proc. Oceans '83*, 1983, Vol. 29, pp. 372–377.

High-frequency volume and boundary acoustic backscatter fluctuations in shallow water

Timothy C. Gallaudet and Christian P. de Moustier[a)]
Marine Physical Laboratory, Scripps Institution of Oceanography, Mail Code 0205, La Jolla, California 92037-0205

(Received 27 November 2001; revised 24 March 2003; accepted 21 April 2003)

Volume and boundary acoustic backscatter envelope fluctuations are characterized from data collected by the Toroidal Volume Search Sonar (TVSS), a 68 kHz cylindrical array capable of 360° multibeam imaging in the vertical plane perpendicular to its axis. The data are processed to form acoustic backscatter images of the seafloor, sea surface, and horizontal and vertical planes in the volume, which are used to attribute nonhomogeneous spatial distributions of zooplankton, fish, bubbles and bubble clouds, and multiple boundary interactions to the observed backscatter amplitude statistics. Three component Rayleigh mixture probability distribution functions (PDFs) provided the best fit to the empirical distribution functions of seafloor acoustic backscatter. Sea surface and near-surface volume acoustic backscatter PDFs are better described by Rayleigh mixture or log-normal distributions, with the high density portion of the distributions arising from boundary reverberation, and the tails arising from nonhomogeneously distributed scatterers such as bubbles, fish, and zooplankton. PDF fits to the volume and near-surface acoustic backscatter data are poor compared to PDF fits to the boundary backscatter, suggesting that these data may be better described by mixture distributions with component densities from different parametric families. For active sonar target detection, the results demonstrate that threshold detectors which assume Rayleigh distributed envelope fluctuations will experience significantly higher false alarm rates in shallow water environments which are influenced by near-surface microbubbles, aggregations of zooplankton and fish, and boundary reverberation.
[DOI: 10.1121/1.1588656]

PACS numbers: 43.30.Gv, 43.30.Vh, 43.30.Re [DLB]

I. INTRODUCTION

Validating statistical reverberation models is difficult because reverberation fluctuations are so strongly influenced by the sonar's beam pattern and spatial distribution of scatterers.[1–6] The former is usually known through system calibration, but the latter is more difficult to characterize. Acoustic and optical imaging methods have been used for this purpose, mostly for studies of the seafloor,[7–9] although some studies of the volume[10] and sea surface[11] have been performed. To the best of our knowledge, no such study has been conducted as a function of angle with a high resolution multibeam sonar measuring simultaneously seafloor, sea surface, and volume acoustic backscatter and reverberation. Such a study is warranted because rarely can reverberation be considered a single component process. Here, ''single component scattering process'' refers to a process dominated by acoustic backscatter from one type of scatterer, such as the sea floor, whereas ''two-component scattering process'' refers to a process dominated by acoustic backscatter from two types of scatterers, such as both boundaries, or a single boundary and biologic scatterers in the volume. Similarly, a three-component process refers to a process dominated by three types of scatterers, such as both boundaries and near-surface bubbles, etc.

A recent study of seafloor reverberation process was performed by Lyons and Abraham,[7] who found the three-component Rayleigh mixture distribution to be the most robust in describing observed fluctuations in seafloor acoustic backscatter amplitude data from a wide variety of seafloor types identified with optical and *in situ* sampling techniques. Here, we perform a similar study, but add to their results by (1) also including the log-normal probability distribution in the model-data comparisons, (2) analyzing data collected on a moving platform, thereby incorporating the influence of spatial variability on the backscatter amplitude fluctuations, (3) analyzing data from both boundaries and the volume, and (4) using coincident multibeam acoustic backscatter imagery to link the spatial distributions of various scatterers to the observed fluctuation statistics.

The data used in this study were collected by the Toroidal Volume Search Sonar (TVSS), a 68 kHz cylindrical array which was deployed on a towfish at a depth of 78 m in waters 200 m deep, 735 m astern of a towship during engineering tests conducted by the U.S. Navy's Coastal System Station (CSS), Panama City, Florida (Fig. 1). The multibeam acoustic data collected by the TVSS were processed to construct boundary[12,13] and volume[14] acoustic backscattering strength images in horizontal and vertical planes around the towfish (Fig. 2). Here, we examine the statistics of, and fit probability distributions to the backscatter amplitudes corresponding to these data. The multibeam acoustic backscatter imagery provides the means for discriminating between vari-

[a)]Present address: Center for Coastal and Ocean Mapping, University of New Hampshire, 24 Colovos Road, Durham, NH 03824.

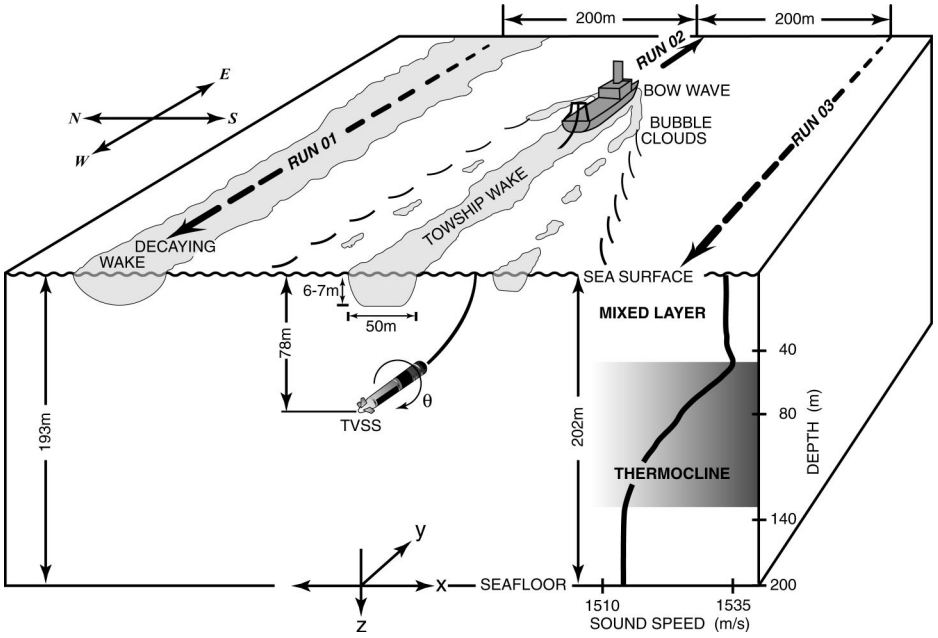


FIG. 1. Depiction of the TVSS deployment of 9 November 1994. Although each of the three parallel runs consisted of over 800 pings, the data presented in this paper are processed from only 100 pings in each of the three runs. The environmental conditions are summarized in the text, and more complete descriptions and analyses are presented in Refs. 12–14.

ous reverberation components and directly attributing nonhomogeneous spatial distributions of scatterers, such as bubbles, zooplankton, and multiple boundary interactions to non-Rayleigh backscatter amplitude distributions.

A useful model for understanding the statistical properties of acoustic reverberation is the point scattering model,[15–20] which assumes that the total backscattered signal is the sum of n replicas of the transmitted signal $s(t)$ backscattered from a homogeneous distribution of point reflectors

$$F(t)=\sum_{i=1}^{n} a_i B(t_i) s(t-t_i,\xi_i), \tag{1}$$

where t_i is the time of arrival from the ith scatterer, a_i is the stochastic amplitude which corresponds to that scatterer's acoustic cross section, $B(t_i)$ describes the sonar's acoustic geometry and gain, and ξ_i is a set of stochastic parameters defining the characteristics of the scattered signals, which may depend upon the relative motion between the acoustic array and the scatterers, their physical properties, and their spatial distribution.

In general, $F(t)$ will fluctuate around some time-varying mean value, and the quadrature components of the fluctuating part may be expressed as[16]

$$V_{I,Q}(t)=F(t)/g(t), \tag{2}$$

where $g(t)$ is the transient function whose reciprocal transforms the nonstationary reverberation sum in Eq. (1) to the stationary form $V_{I,Q}(t)$. This fluctuating signal, and its corresponding envelope are important because their probability density functions (PDFs) are used as the noise models against which target detection algorithms must operate.[21]

The model in (1) and (2) assumes that the number n is governed by a Poisson distribution, where the scatterers producing the resulting reverberation are discrete, statistically independent in position, and homogeneously distributed within the sonar's resolution cell. If the number of scatterers in a single resolution cell is very large, and their scattering coefficient distribution (a_i) is such that no small number of them contributes significantly to the reverberation energy, application of the central limit theorem results in a Gaussian distribution for $V_{I,Q}(t)$, with a Rayleigh distributed envelope and uniformly distributed phase.

In typical shallow water environments, the distributions of scatterers can rarely be assumed to be homogeneous, and different types of scatterers distributed on different spatial scales tend to produce more extreme reverberation values, depending upon the density of scatterers relative to the sonar's resolution cell size. For envelope fluctuation distributions, these may appear as multiple modes and/or large tails, deviating significantly from the traditional Rayleigh PDF.[7,8,22–24]

The distribution models considered in this study are the Rayleigh, K, Weibull, log-normal, and Rayleigh-mixture distributions. We chose these because (1) they are commonly used in underwater acoustics, (2) they have been observed in previous studies of volume and boundary backscatter and reverberation, and (3) some have been analytically related to the physical scattering mechanisms which produce them. Although a number of probability distribution models have been developed for specific boundary or volume reverberation conditions,[1,3,10,25,26] our objective is to determine whether there is a common model flexible enough to describe both boundary and volume backscatter arising from nonhomogeneous, or patchy scatterer distributions that are typical in shallow water.

We begin in Sec. II with a description of the PDF models used in this study. Section III describes the TVSS signal processing methods and the data preparation steps. The results are described in Sec. IV, and we assess in Sec. V the physical mechanisms influencing these results and their implications for target detection.

II. PROBABILITY DISTRIBUTION MODELS

Each of the distribution functions discussed here may be represented as a function of one or several parameters that

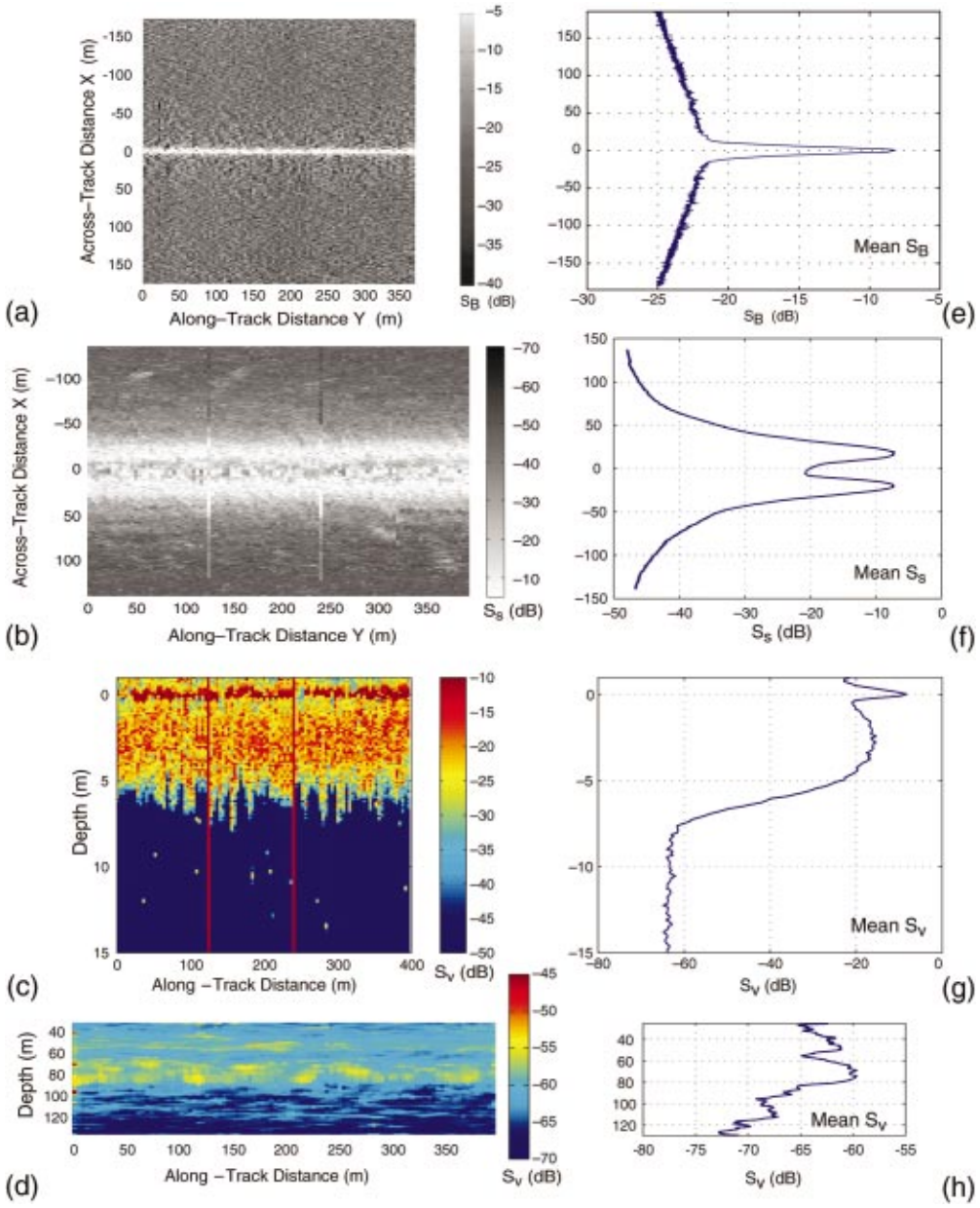


FIG. 2. TVSS-derived acoustic backscattering strength images displayed in coordinates relative to the towfish (a)–(d) and their corresponding along-track averages (e)–(h). (a) Bottom acoustic backscattering strength (S_B): The normal incidence return extends along-track near the center in the seafloor image and results from the natural angular dependence function of the silt and sand sediments in the region (Ref. 12). (b) Sea surface acoustic backscattering strength (S_S): The feature near the track center extending along-track in the sea surface backscattering strength image is influenced by vertical attenuation through bubbles in the township's wake (Fig. 1) (Ref. 13). The moderately high backscattering strength features 50–100 m to the right and left of the track centerline are due to resonant scattering from bubble clouds generated by breaking ship waves. The two across-track lines near $Y=120$ m and 235 m in this image and the vertical lines in (c) at the same along-track locations are corrupted data and were excluded from the analysis. (c), (d) Volume acoustic backscattering strength (S_V): The vertical volume backscattering strength image (c) formed by using the upward looking beams shows that the bubble layer associated with the township's wake varies in scattering strength and depth along-track. The vertical volume image (d) formed in the vertical plane 47 m to the right of the TVSS shows the presence of volume scattering layers in the mixed layer and upper thermocline (Ref. 14).

must be estimated from the observed amplitude data, $A=\{A_1A_2\cdots A_N\}$, whose samples are assumed to be independent and identically distributed. For parameter estimation, we use maximum likelihood estimation and the method of moments as described and implemented by Abraham.[27]

We start with the Rayleigh PDF for acoustic reverberation of amplitude $A\geqslant 0$:

$$p_R(A)=\frac{2A}{\lambda_R}e^{-A^2/\lambda_R}, \tag{3}$$

and its cumulative distribution function (CDF)

$$P_R(A)=1-e^{-A^2/\lambda_R}, \tag{4}$$

where $\lambda_R=\langle A^2\rangle$, with $\langle\ \rangle$ representing the expected value. It describes reverberation whose in-phase and quadrature components are normally distributed with zero mean, and results from enough scatterers in the sonar's resolution cell for the central limit theorem to hold.[16] The Rayleigh distribution has been observed for high frequency backscatter and reverberation from the seafloor,[7] sea surface,[28] and volume,[10,29] and is

a limiting case of the Ricean distribution in which scattering is primarily incoherent.[30] Stanton[30] has related the Rayleigh PDF of seafloor acoustic backscatter amplitude to the rms roughness and correlation area of the bottom.

The *K* distribution may be represented as the product of a rapidly fluctuating, Rayleigh-distributed random variable, and a slowly varying, chi-distributed variable.[31–33] Its PDF is[27]

$$p_K(A)=\frac{4}{\sqrt{\alpha}\Gamma(\nu)}\left(\frac{A}{\sqrt{\alpha}}\right)^{\nu}K_{\nu-1}\left(\frac{2A}{\sqrt{\alpha}}\right), \tag{5}$$

and its CDF is

$$P_K(A)=1-\frac{1}{\Gamma(\nu)2^{\nu-1}}\left(2\frac{A}{\sqrt{\alpha}}\right)^{\nu}K_{\nu}\left(\frac{2A}{\sqrt{\alpha}}\right), \tag{6}$$

with $A\geqslant 0$. $K_{\nu-1}$ is the $\nu-1$ order modified Bessel function and $\Gamma(\ )$ is the gamma function. When the scale $1/\sqrt{\nu}$ is applied to *A*, the Rayleigh distribution with power α is obtained, in the limit as ν tends to infinity, from the *K* distribution.[27] The *K* distribution has been used to describe radar sea surface clutter because it has a direct physical interpretation: the Rayleigh component, with relatively short correlation widths, results from the many scattering contributions within the resolution cell that arise from small scale facets on the sea surface, whereas the chi-distributed component, with relatively long correlation widths, arises from the larger scale, mean sea surface tilt (e.g., swell). The *K* distribution also has been used to describe signal envelope fluctuations in wireless channels[34] and seafloor acoustic backscatter in sidescan sonar images.[8,22]

The Weibull distribution also is related to the Rayleigh distribution and has been used to describe seafloor backscatter amplitude distributions.[7] The two-parameter Weibull PDF is[27]

$$p_W(A)=\alpha\beta A^{\beta-1}e^{-\alpha A^{\beta}} \tag{7}$$

for $A\geqslant 0$, with its CDF given by

$$P_W(A)=1-e^{-\alpha A^{\beta}}, \tag{8}$$

where it can be seen that the Rayleigh distribution results when $\beta=2$ and $\alpha=1/\lambda_R$.

Whereas the *K* and Weibull distributions may be related to physical scattering mechanisms through their relationships with the Rayleigh distribution, the log-normal distribution has yet to reveal such analytical connections. Nevertheless, the log-normal distribution has been observed in studies of underwater acoustic backscatter and propagation,[23,35,36] radar clutter from the sea surface,[37] and signal envelope fluctuations in wireless channels.[34,38] The two-parameter log-normal PDF is[39]

$$p_{LN}(A)=\frac{1}{\sqrt{2\pi}\alpha A}e^{-(\ln A-\beta)^2/2\alpha^2} \tag{9}$$

for $A>0$. It has the property that $\ln(A)$ is normally distributed with mean β and variance α^2. The log-normal CDF is

$$P_{LN}(A)=\Phi\left(\frac{\ln A-\beta}{\alpha}\right), \tag{10}$$

where

$$\Phi(u)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u}e^{-w^2/2}\,dw \tag{11}$$

is the CDF of a standard normal random variable *u*. Another property of the log-normal distribution is that if *A* is log-normally distributed, so is A^2; i.e., if the echo amplitude PDF has the form of Eq. (9), so will the PDF of the echo intensity.[39]

In typical shallow water environments, acoustic backscatter and reverberation result from several independent scattering mechanisms, such as bubbles, bioacoustic scatterers, and boundary roughness, and each of these may be characterized by different spatial scales. For high resolution, narrow beam sonars used in bioacoustic studies, multibeam bathymetric surveys, studies of near surface physical processes, and mine-countermeasures, the echo from a given resolution cell typically, though not necessarily, contains only one type of scatterer. Therefore, it is reasonable to consider that reverberation in such a scenario might be represented by a mixture of *m* Rayleigh random variables, each with a component probability ε_i and power $\lambda_{R,i}$. The resulting Rayleigh mixture PDF is[27]

$$p_{RM}(A)=\sum_{i=1}^{m}\varepsilon_i\frac{2A}{\lambda_{R,i}}e^{-A^2/\lambda_{R,i}}, \tag{12}$$

and its CDF is

$$P_{RM}(A)=1-\sum_{i=1}^{m}\varepsilon_i e^{-A^2/\lambda_{R,i}}, \tag{13}$$

where

$$\sum_{i=1}^{m}\varepsilon_i=1 \tag{14}$$

is required to ensure a valid CDF.

Although the component densities in a mixture distribution need not be Rayleigh, or even members of the same parametric family,[40] Rayleigh-mixture distributions have been fit successfully to seafloor acoustic backscatter.[7,24,27] Because mixture distributions have yet to be evaluated for reverberation from both boundaries and the volume, we shall test them below with data collected by the TVSS. We begin by describing the TVSS, the data, and aspects of the acoustic geometry that help in understanding the results.

III. TVSS DATA

A. TVSS data collection

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° apart around the cylinder and designed to produce a ''toroidal'' beam pattern that is meant to be omni-directional in the plane perpendicular to the cylinder's axis (usually across-track) and 3.7° wide at −3

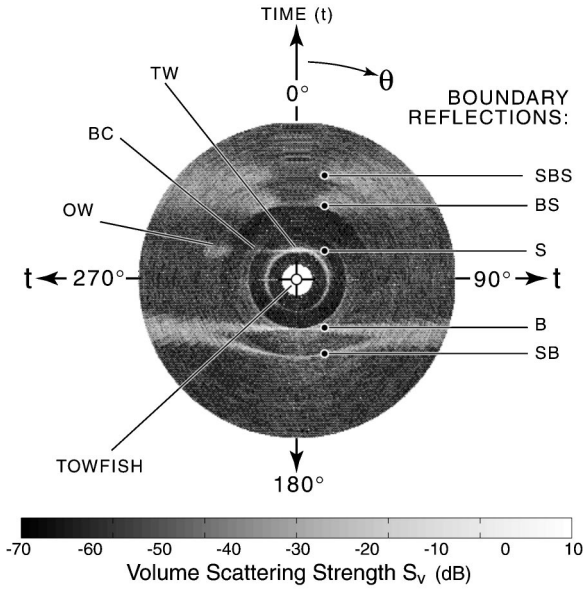


FIG. 3. Center locations of the analysis regions used in this study for a single TVSS ping. The data sets for each region consisted of 100 pings and spanned the horizontal and vertical dimensions listed in Table I.

dB in any plane containing the cylinder's axis (usually along-track). The hydrophone array consists of 120 elements equally spaced every 3° around the cylinder. In the work presented here, split aperture beamforming of the hydrophone array yielded 120 receive beams, each 4.95° wide at −3 dB and spaced 3° apart to cover the full 360° around the array in the plane perpendicular to the array's axis. Details of the data processing are available in Refs. 12–14 and 41, 42.

The acoustic data were collected by the TVSS in a 2 nm^2 area 65 nm southeast of Panama City, Florida, in the northeastern Gulf of Mexico. The TVSS was towed approximately 735 m aft of the towship MR. OFFSHORE at a nearly constant depth of 78 m (Fig. 1). Three runs of 100 consecutive pings of acoustic backscatter data, from 200 μs CW pulses of 68 kHz transmitted once a second, were obtained while the towship speed was nearly constant at 4.1 m/s. Towfish attitude and motion data were sampled at 1 Hz (once per ping) and included roll, roll rate, pitch, heading, speed, and depth. The environmental data collected during the experiment included a single CTD cast, which revealed the presence of an isothermal mixed layer with a temperature of 24.8 °C extending to a depth of 49 m, a thermocline between 49 m–150 m depth, and a nearly isothermal layer above the bottom with a temperature of 15.6 °C. The surface salinity was 35.1 ppt, and the surface sound speed was 1534 m/s. The wind speed recorded at 0658AM onboard MR. OFFSHORE was 6 knots (3 m/s), and the sea state was 1.5.

B. TVSS acoustic geometry

The statistical results are best interpreted with an understanding of the TVSS acoustic geometry, which may be obtained from Figs. 1–3. Figure 3 depicts a vertical slice of volume scattering strength (S_V) perpendicular to the towfish axis, formed by displaying the acoustic data in each of the 120 TVSS receive beams in a single ping around the TVSS in coordinates of depth vs horizontal range. In this representation, echoes from the sea surface and seafloor appear as the high backscatter, horizontal features above and below the towfish. 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. The circular features result from boundary reflections received in the sidelobes of beams directed away from the boundary.

The angular sample spacing in this figure is the spacing between maximum response axes of adjacent beams: $\theta_s = 3°$. The quadrature sampling time increment of $\tau_s = 160$ μs results in a 12 cm slant range sample spacing assuming a sound speed in seawater $c = 1500$ m/s. With the TVSS pulse length $\tau_p = 200$ μs, the bandwidth is $W = 0.88/\tau_p = 4.4$ kHz, which yields a range resolution $\Delta R = c/2W = 17$ cm.

The volumetric resolution in each ping is determined by the spatial dimensions of the volume ensonified by the TVSS transmit pulse within each receive beam. We approximated the ensonified volumes (V) in Table I as the ellipsoidal shell formed from the intersection of the transmitted pulse bound by the transmit beam pattern, and the receive beam. Thus, the dimensions of V increase with slant range from the TVSS, and ensonified volumes at equal ranges from the TVSS in adjacent beams overlap by 39.4%. The towfish's speed, $V_{\mathrm{TVSS}} = 4.1$ m/s, and the relatively narrow fore–aft transmit beamwidth resulted in overlap between ensonified volumes in the same beam angle for consecutive pings, which increased with range beyond 62 m.

On the boundaries, resolution is defined by the area (A) ensonified by the transmitted pulse within each receive beam. The area is approximated by an ellipse near normal incidence, and by an annulus sector away from normal incidence. Thus, the maximum ensonified areas on the boundaries are at the towfish's zenith and nadir (Table I), where the horizontal resolution is poorest. Expressions for these and other characteristics of the TVSS acoustic geometry are given in Refs. 12–14.

C. Data partitioning and description

Applying sidescan imaging techniques to the TVSS data collected over multiple pings, we constructed seafloor, sea surface, and horizontal and vertical volume backscattering strength images, which are analyzed in Refs. 12–14. Four of these images are shown in Fig. 2 with their along-track averages. Whereas the seafloor image appears fairly homogeneous away from the track centerline [Fig. 2(a)], the sea surface and volume images exhibit significant spatial variability due to bubbles and bubble clouds [Figs. 2(b) and (c)] and aggregations of volume scatterers [Fig. 2(d)].

The acoustic backscatter amplitude data corresponding to these and other images were then partitioned into data sets which encompassed the analysis regions defined in Table I. The locations of the centers of these regions are indicated in Fig. 3. For 14 of the 15 analysis regions in Table I, three separate runs of 100 pings were used, and for one region

TABLE I. Analysis regions for the TVSS data set. Negative across-track distances are left of the towfish's track. Grazing angles in regions VL1 and VL2 are defined with respect to the vertical along-track plane 47 m to the left of the towfish's track. Grazing angles in regions NS1–NS5 are defined with respect to the horizontal plane at 3 m depth. Ensonified areas (boundary regions SF, SS) and ensonified volumes (volume regions NS, VL) are listed in the last column.

Analysis region	Primary acoustic scattering and reverberation features	Across-track distance(s) (m)	Depth(s) (m)	Grazing angles	Areas or volumes (m^2,m^3)
SF1	seafloor backscatter	−35–+35	192–202	72°–90°	4.5–73
SF2	seafloor backscatter+surface reverberation after first surface echo	−50–−100	192–202	48°–66°	2.6–4.0
SF3	seafloor backscatter+surface and bottom reverberation after surface-bottom multiple	−150–−200	192–202	29°–37°	2.6–3.1
SS1	sea surface backscatter+attenuation from bubbles in towship's wake	−30–+30	0	68°–90°	2–40
SS2	sea surface backscatter+backscatter from bubble clouds generated by ship and ambient waves	40–80	0	44°–66°	2
SS3	sea surface acoustic backscatter+backscatter from bubble clouds generated by ship and ambient waves+bottom reverberation from first bottom echo	100–150	0	27°–38°	2
NS1	near-surface volume and sea surface backscatter+backscatter from bubbles in towship's wake generated during previous runs+surface and bottom reverberation after bottom-surface multiple	180–220 and −180–−220	3	18°–23°	37–52
NS2	near-surface volume backscatter from bubbles within the towship's wake+surface reverberation after first surface echo	−30–+30	3	68°–90°	5–6
NS3	near-surface volume and surface backscatter from bubble clouds generated by ship waves+surface reverberation after first surface echo	40–80	3	42°–62°	6–12
NS4	near-surface volume backscatter from bubble clouds generated by ship and ambient waves+surface and bottom reverberation after first surface and bottom echoes	100–150	3	26°–37°	15–27
NS5	surface and bottom reverberation after bottom-surface multiple	200–250 and −200–−250	3	16°–20°	44–66
VL1	volume backscatter from densely distributed zooplankton in mixed layer and upper thermocline	−47	40–70	50°–81°	2–4
VL2	volume backscatter from sparsely distributed zooplankton in middle and lower thermocline	−47	90–120	47°–76°	2–4
VL3	volume backscatter from sparsely distributed zooplankton in lower thermocline	0	125–140	89°–90°	2–4
VL4	volume backscatter below thermocline from sparsely distributed fish +surface reverberation after first surface echo	0	165–180	89°–90°	7–10

(NS1), two runs of 100 pings were used. Thus, the partitioning formed a total of 44 data sets.

Ideally, we would analyze the data collected in each grazing angle/depth/across-track distance location separately. However, this would have resulted in less than 100 samples per analysis region, and the PDF models and parameter estimation methods used here require much larger sample sizes to perform well.[27] Therefore, we grouped data into the regions defined in Table I. To ensure that the data did not vary significantly over the range of grazing angles within each region, they were tested for homogeneity across both grazing angles and pings, as discussed below.

The partitioned data corresponding to the seafloor analysis regions span three different grazing angle regimes: normal and near normal incidence (SF1), moderate to high grazing angles (SF2), and moderate to low grazing angles (SF3). Bathymetry constructed from the TVSS backscatter data revealed a relatively flat bottom, with a 3 m/km south west slope, and an average depth of 198 m. Seafloor acoustic backscattering strength imagery indicated a homogeneous spatial distribution of sediments, and the angular dependence function estimated from the acoustic backscattering strength is consistent with the silt–sand mixture of sediments previously surveyed in the region [e.g., Fig. 2(a)].[12]

The sea surface analysis regions were influenced only slightly by sea surface roughness produced by the ambient 3 m/s winds. Because of the vertical extent of the transmitted acoustic pulse intersecting the sea surface, the sea surface data were more strongly influenced by clouds of resonant mircobubbles which were characterized by different spatial dimensions and scattering characteristics that depended upon their generating mechanisms. These included (1) very dense bubble clouds generated primarily by propeller cavitation within the towship's wake (SS1), (2) large-scale [$O(10^2)$ to $O(10^3)$ m^2] bubble clouds generated by breaking ship waves (SS2), and (3) sparsely distributed, small scale [O(1) to O(10) m^2] bubble clouds generated by the ambient sea (SS3) [e.g., Fig. 2(b)]. The SS3 region also was influenced strongly by bottom reverberation received in the sidelobes after the first bottom echo arrival. Although we did not have *in-situ* bubble size and density data, we used the resonant bubble approximation to estimate the densities of bubbles in the analysis regions from the surface and near-surface acoustic backscattering strength data in Ref. 13.

The near-surface volume regions were influenced by the same processes that influenced the sea surface backscatter.

Clouds of resonant microbubbles in the towship's wake contributed to the backscatter in both the NS1 and NS2 regions, but these clouds were denser in NS1 than in NS2 because the wake in NS2 was about 20 minutes old, whereas the wake in NS1 was only 3 minutes old [Fig. 2(c)]. In the same across-track location as the SS2 region, the NS3 region at 3 m depth was also affected by large scale bubble clouds generated by breaking towship waves. Similarly, the NS4 region was in the same across-track location as the SS3 region, and was also influenced by bottom reverberation and smaller scale bubble clouds generated by the ambient sea. The NS5 region was influenced by both near-surface bubbles and multiple boundary reflections occurring after the first bottom-surface multiple arrival. The NS1, NS3, NS4, and NS5 regions were influenced somewhat by surface roughness, due to the vertical extent of the ensonified volume.

Three of the volume regions were influenced by aggregations of zooplankton whose density generally decreased with depth from the base of the mixed layer (VL1), through the upper (VL2) and lower thermocline (VL3) [e.g., Fig. 2(d)].[14] The VL4 region near the bottom was influenced slightly by a sparse distribution of small fish, but more strongly by surface reverberation received in the sidelobes after the first surface echo arrival. As with the near-surface data, we lacked the *in-situ* data to characterize absolute densities and sizes of organisms in the volume, so we have inferred the relative densities from the corresponding volume acoustic backscattering strength data in Ref. 14. Although several dense fish schools were observed near the bottom, the backscatter data in these regions could not pass statistical independence tests, so they were not included in the analysis.

D. Data preparation

After grouping the TVSS acoustic backscatter amplitude data according to the analysis regions in Table I, data contaminated by noise spikes were removed. Because statistical analyses require independent and identically distributed data, the amplitudes were decimated by taking only those samples separated by at least a correlation width across grazing angles and pings. The correlation widths were estimated as the horizontal or vertical lags corresponding to the first null of the normalized spatial autocovariance. In cases where the autocovariance dropped sharply to a low value (<0.1), and then fell gradually to zero, we used the distance for which it decreased to 0.1.

As we are interested in reverberation fluctuations, we removed nonstationarities resulting from backscatter angular dependence and angular variations in the TVSS transmit and receive beam patterns by grouping the amplitude data in each analysis region into bins 1° wide according to grazing angle and angle with respect to the TVSS, and then normalizing by the mean in each group. The normalized data were then regrouped into each analysis region (Table I), and inspected to ensure that all nonstationarities due to beam pattern variations and grazing angle dependence were adequately removed.

To ensure that the samples in each analysis region were statistically independent and identically distributed across pings and grazing angles, we performed the one sample runs test[43] for randomness and the Mann–Whitney U test for homogeneity:[44] (1) to the normalized samples in each grazing angle across pings, and (2) to the normalized samples in each ping across grazing angles. Most of the data in the seafloor (SF) and volume (VL) regions passed the tests at the 95% confidence level, but 20%–50% of the sea surface (SS) and near-surface (NS) data failed the tests. In studies of data collected on fixed platforms, the approach is to simply remove data which do not pass the tests at the specified confidence level.[21] Doing so in our study was not possible because the TVSS data were collected from a moving platform. Therefore, for each analysis region, we selected only those samples within the largest contiguous regions (across pings and grazing angles) which passed both tests at the 95% confidence level. We verified that the retained samples included contributions from the various backscattering and reverberation features in Table I by analyzing backscattering strength images formed from these data (e.g., Fig. 2).

IV. RESULTS

A. Backscattering strength, amplitude and intensity statistics

The data in each of the analysis regions depicted in Fig. 3 were first characterized by averaging statistical estimates of the corresponding backscattering strength ($S_{B,S,V}$), amplitude (A), and intensity (A^2) over the three TVSS runs (Table II). Expressions for the mean (μ_A), variance (σ_A^2), skewness ($\gamma_{3,A}$), and kurtosis ($\gamma_{4,A}$) are given in the Appendix. The scintillation index, which is the variance of the intensity fluctuation scaled by the square of the mean intensity, was computed as

$$\sigma_{A^2}^2 = \frac{\langle (A^2 - \lambda_R)^2 \rangle}{\lambda_R^2}. \tag{15}$$

We include this quantity because it generally indicates the extent to which the data depart from a Rayleigh distribution, as Rayleigh-distributed amplitudes result in a scintillation index of one.

Table II shows that the scintillation indices for the seafloor regions are the closest to one, suggesting that they depart the least from Rayleigh distributions. In addition, the amplitude variance, skewness, and kurtosis values are lower for the seafloor regions. Mean backscattering strengths decrease away from the nadir region (SF1) [e.g., Fig. 2(a)], which is consistent with composite roughness model predictions for the silt-sand sediment type in the region and expected for rough-surface models of relatively smooth seafloors.[12] The region at nadir also exhibits the highest variance, skewness, and kurtosis of the three seafloor regions.

Statistics for the sea surface regions differ significantly from those for the seafloor regions. The region at zenith (SS1) has the highest mean backscattering strengths of all regions, but these are attenuated approximately 22 dB below model predictions by resonant microbubbles in the towship's wake [e.g., Figs. 2(b) and (f)].[13] Backscattering strength decreases with grazing angle, but scintillation indices, skewness, and kurtosis increase with decreasing grazing angle. This trend is opposite that of the bubble densities inferred

TABLE II. Average TVSS backscatter amplitude statistics. We calculated backscattering strengths using expressions in Refs. 12–14. The scintillation index is computed from Eq. (15) in the text. The range is the maximum minus the minimum amplitude, and all other terms are computed from expressions in the Appendix. Because they were computed from the normalized amplitudes, all quantities except the backscattering strength are dimensionless. The statistics have been averaged over runs 1–3, except those for the NS1 region, which were averaged over runs 2–3.

Analysis region	Mean number of samples N	Mean backscattering strength $S_{B,S,V}$ (dB)	Mean normalized amplitude μ_A	Range	Variance σ_A^2	Skewness $\gamma_{3,A}$	Kurtosis $\gamma_{4,A}$	Scintillation index $\sigma_{A^2}^2$
SF1	1505	−19.4	1.0292	3.2242	0.3115	0.9022	1.4171	1.2262
SF2	3840	−22.6	1.0037	3.1795	0.2794	0.6447	0.1927	1.0156
SF3	4406	−28.7	0.9962	3.4766	0.2793	0.7185	0.4448	1.0646
SS1	1330	−12.9	1.0108	4.4545	0.4469	1.1528	1.6478	1.8328
SS2	2725	−41.5	0.9943	7.6032	0.4834	2.0793	10.329	3.6730
SS3	3951	−51.3	0.9981	13.3277	0.3479	6.2665	119.56	14.358
NS1	2500	−51.7	1.0604	3.5567	0.2902	0.9185	0.9124	1.0575
NS2	1823	−28.4	1.0139	5.1983	0.3963	1.6726	4.7479	2.2058
NS3	2350	−50.7	1.0159	8.0735	0.5233	2.4721	14.042	4.6218
NS4	4985	−60.7	0.9977	9.0243	0.2773	3.6466	38.646	4.8148
NS5	7296	−59.4	1.0010	9.2410	0.1785	4.2419	61.760	3.1274
VL1	2835	−65.5	1.0001	3.8499	0.2687	1.0841	1.4969	1.0282
VL2	2522	−73.4	1.0094	5.1744	0.5285	1.6516	3.0865	2.5656
VL3	2732	−76.3	1.0015	5.3451	0.4954	2.3111	6.9617	3.2434
VL4	2591	−72.2	1.0167	3.6934	0.2919	1.3342	2.3139	1.3993

from backscattering strength imagery [Fig. 2(b)],[13] as the highest densities occurr near the towfish zenith (SS1), lower densities occurr in the regions influenced by large scale bubble clouds produced by breaking ship waves (SS2), and the lowest bubble densities are observed farther across track (SS3) where the near-surface bubble population consists primarily of bubbles generated by small scale breaking of the ambient sea waves.

Statistics for three of the near-surface regions (NS2–NS4) exhibit a grazing angle dependence similar to that of the sea surface regions (SS1–SS3), with mean backscattering strength decreasing, and skewness, kurtosis, and scintillation index increasing away from the towfish's zenith. For the NS1 and NS5 regions, backscattering strength increases with decreasing grazing angle as a result of the bottom-surface multiple echo. For the NS1 region, backscattering from microbubbles in the decaying towship's wake [Figs. 1 and 2(b)] also increases the mean volume backscattering strength above that for the higher grazing angle region (NS4). Although the NS1 and NS5 regions are at similar across-track locations and have the largest ensonified volumes of all the analysis regions (Table I), their statistics are dramatically different. The skewness, kurtosis, and the scintillation index values for the NS5 region are among the highest values of all the analysis regions, and result from sparsely distributed bubbles generated by small scale breaking of ambient sea waves. The corresponding values for the NS1 region are significantly lower, and are the result of scattering from the denser distribution of bubbles in the towship's wake from the previous run.

The statistics for the volume regions are similar to those for the near-surface and sea surface regions in that they are mostly influenced by the density of scatterers. For the VL1–VL3 regions, mean backscattering strength decreases, and skewness, kurtosis, and scintillation index increase with depth, resulting from the decrease in zooplankton density with depth [e.g., Fig. 2(d)]. Statistics for the VL4 region depart from this trend, and this may be related to the influence of surface reverberation after the first surface echo.

Before fitting the various PDF models to the TVSS data, we evaluated their potential suitability for describing backscatter fluctuations by comparing plots of the skewness and kurtosis descriptors (β_1, β_2) of the normalized backscatter amplitude data with the possible values for each PDF family (Fig. 4), where $\beta_1 = \gamma_{3,A}^2$, and $\beta_2 = \gamma_{4,A} + 3$. The Appendix describes the basis for this figure, which is taken from Abraham,[27] and Johnson *et al.*[45] Although matching skewness and kurtosis does not imply that distributions are identical or even a good approximation to one another, estimates of β_1 and β_2 from data can provide an indication of which PDF families are appropriate to consider. Except for the log-normal distribution, all PDF models appear suitable for describing the seafloor backscatter data, which is closer to being Rayleigh distributed than the amplitude data in the other regions. Skewness and kurtosis descriptors estimated from the sea surface, near-surface, and volume backscatter amplitude data are spread among all the PDF models, but only the Rayleigh mixture model is flexible enough to encompass all the measurements.

B. Probability distribution functions

Rayleigh, *K*, Weibull, log-normal, and Rayleigh mixture distributions were fit to the empirical distribution functions corresponding to the backscatter amplitude data in each run and TVSS analysis region. Figures 5–8 show results of representative runs for the seafloor, sea surface, near-surface, and volume displayed as probabilities of false alarm (PFA $=1-$CDF). PFA is the probability that the amplitude will be higher than or equal to a given value, and we use it to display

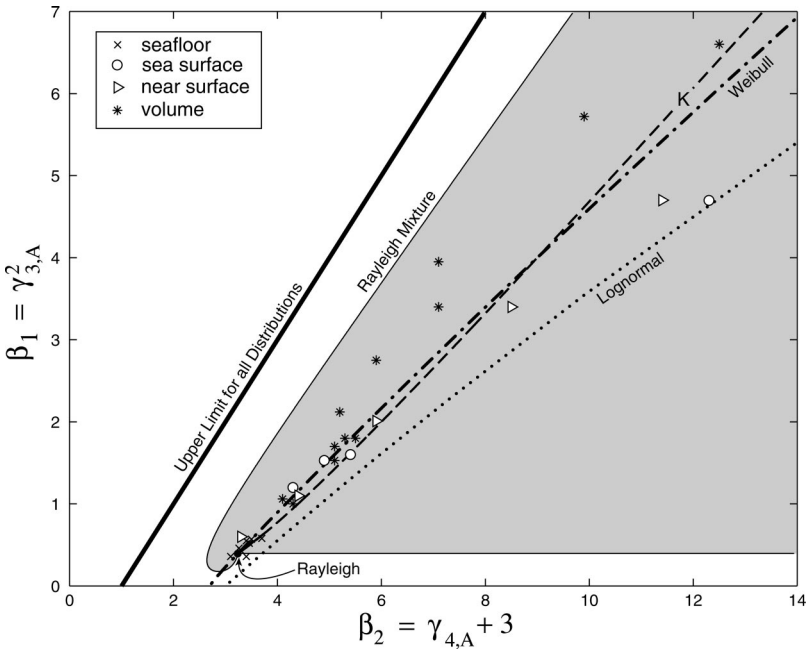


FIG. 4. Plot of skewness descriptor (β_1) vs kurtosis descriptor (β_1) values [defined in the Appendix, Eqs. (A1) and (A2)] computed from the backscattered amplitude data collected in each run and analysis region in Table I. The Rayleigh distribution is represented by a point, whereas the *K*, Weibull, and log-normal distributions are represented by lines. The two-component Rayleigh mixture model is represented by the shaded region. The basis for this figure is presented in the Appendix.

the results because it illustrates best the non-Rayleigh nature of the data, which is mostly seen in the tails of the distributions. Although backscatter amplitude is displayed in decibels on the abscissa of each plot, the distribution fits were obtained from the data in linear units.

We assessed the goodness of fit between the model distributions and the empirical distributions derived from the TVSS data with the nonparametric Kolmogorov–Smirnoff test statistic, defined as the maximum absolute difference between the theoretical CDF $[P(A)]$ and that formed from the N data samples:[46]

$$D_{ks} = \max|P(A) - F_N(A)|, \quad -\infty < A < \infty. \tag{16}$$

When the empirical data $[F_N(A)]$ are drawn from a population in which the random variable A has a continuous distribution function $P(A)$, the limiting distribution of D_{ks} derived by Kolmogorov is[46]

$$\lim_{N\to\infty} P\left(D_{ks} < \frac{h}{\sqrt{N}}\right) = Q(h), \tag{17}$$

where, for $h > 0$,

$$Q(h) = \sum_{i=-\infty}^{\infty} (-1)^i e^{-2i^2h^2}. \tag{18}$$

The KS value $p = (1 - Q(h))$ represents the probability from 0 to 1 of observing a more extreme value of D_{ks} under the null hypothesis that the data are distributed according to $P(A)$. The closer p is to one, the more likely that the observed data follow the model CDF. Although this test is widely used to fit theoretical CDFs to empirical data,[7,27] Eq. (18) is not strictly valid when parameters for the theoretical distribution are estimated from the empirical data.[46] Therefore, we assessed also the relative goodness of fit for the different PDF models by computing the root mean square difference between the model and the empirical distribution functions

$$D_{\text{rms}} = \left[\frac{1}{N}\sum_{i=1}^{N} (P(A_i) - F_N(A_i))^2\right]^{1/2} \tag{19}$$

and averaged values over the three TVSS runs for each region (Table IV). In addition, rms differences were computed and averaged only for the samples in the distributions for which the PFA was less than 10^{-2} (Table V) in order to evaluate how well the model CDFs fit the TVSS data in the tails of the distributions. This ''tail rms difference'' was calculated because relatively high kurtosis values in the near-surface and surface data suggested that large tails would be present in the distributions of these data (Table II).

Among all analysis regions and PDF model types, the seafloor amplitude data has the lowest rms differences and best statistical fits (KS p values). All ranges of grazing angles (SF1–SF3) are non-Rayleigh, but the moderate grazing angle region (SF2) is relatively close to Rayleigh (Fig. 5). However, *K*, Weibull, and Rayleigh mixture distributions provide good fits to the distributions (Tables III and IV) and the tails (Table V). The Rayleigh mixture distributions show the best overall performance. In addition, rms differences and KS p values indicate that no significant advantage is gained by using more than three-components in the Rayleigh mixture. These results are generally consistent with those in Lyons and Abraham[7] for backscatter amplitude data from mud bottom types in the 40°–60° and 60°–80° grazing angle regimes. The KS p values in Table III are slightly lower than theirs, probably because of spatial variations in the bottom. Such variations were not present in their data because they were collected from fixed platforms.

Backscatter amplitude fluctuations from the sea surface (Fig. 6) are more non-Rayleigh than those from the seafloor, and depend mostly upon grazing angle and the density of bubbles relative to the vertical extents of the ensonified volumes adjacent to the sea surface. For the zenith region (SS1), where very high densities of bubbles in the towship's wake attenuated the acoustic backscatter, only the Rayleigh mix-

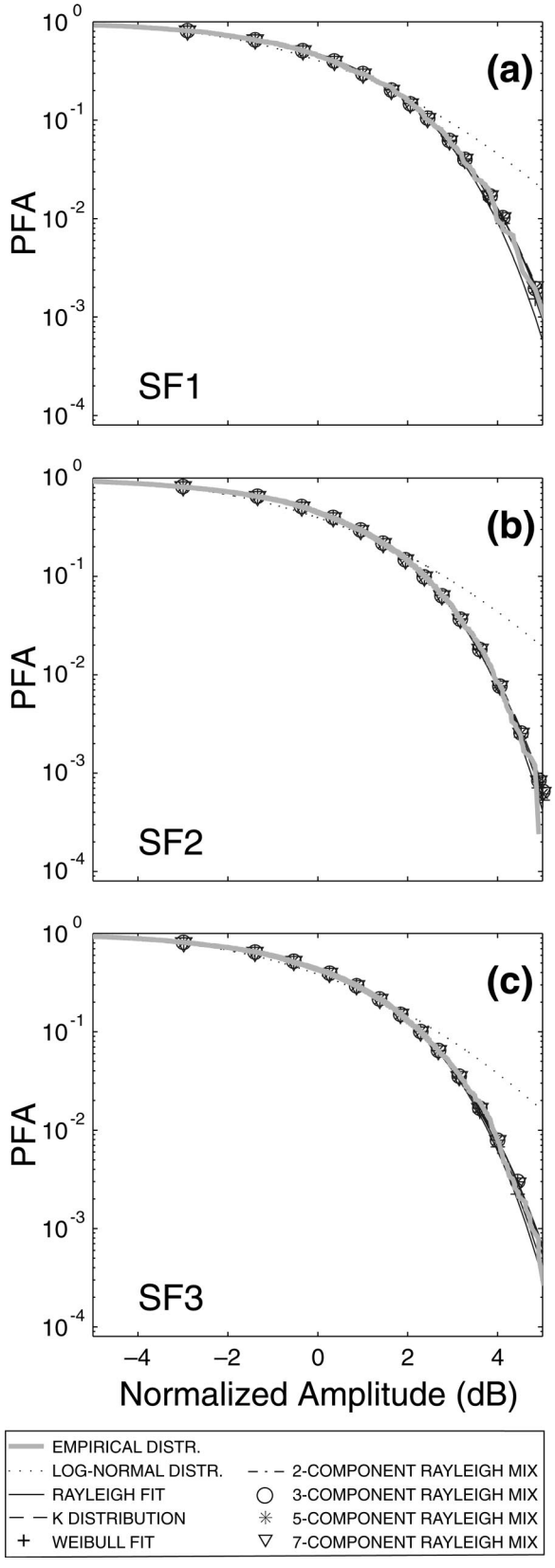


FIG. 5. Acoustic backscattered amplitude distributions displayed as probability of false alarm (PFA) for the three different seafloor regions in run 1.

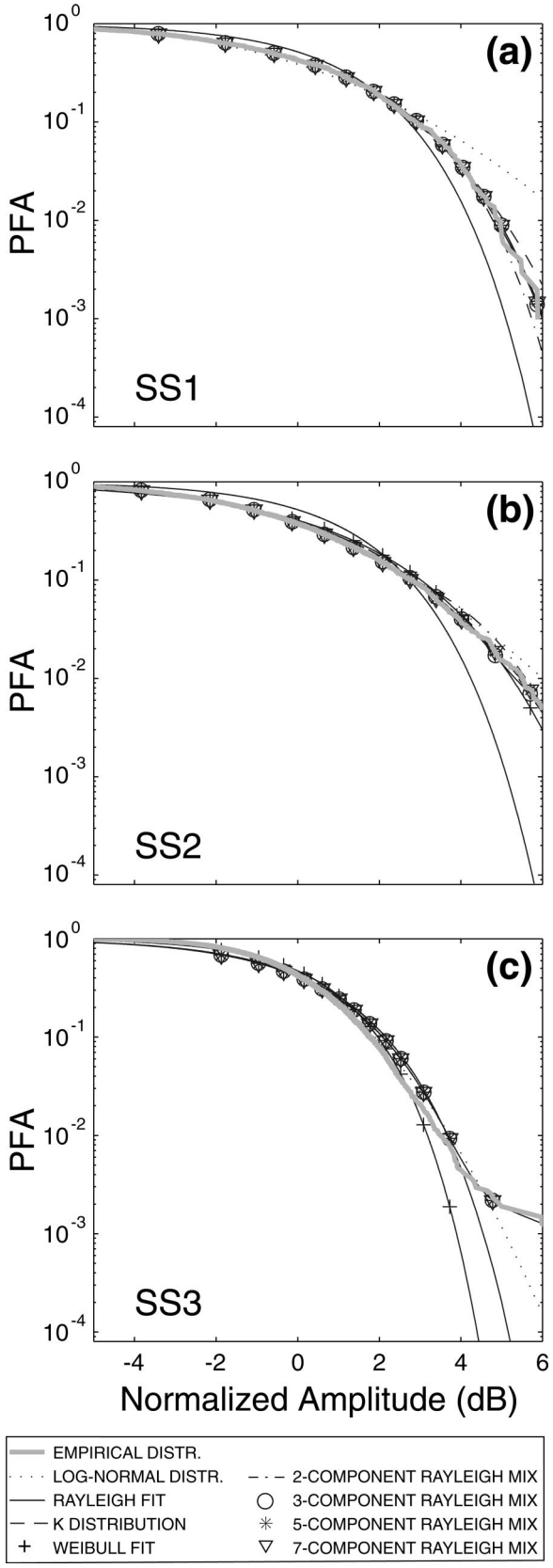


FIG. 6. PFA plots corresponding to the backscattered amplitude data from the sea surface regions in run 3.

ture models provide statistically good fits to the observed data. Outside the wake, where bubble densities resulting from breaking waves generated by the towship were significantly lower than at zenith, KS p values are lower and rms differences are higher, with Rayleigh mixture models again showing superior overall performance.

Backscatter amplitude distributions in the lowest grazing angle region [SS3, Fig. 6(c)] appear to be multimodal (cf.

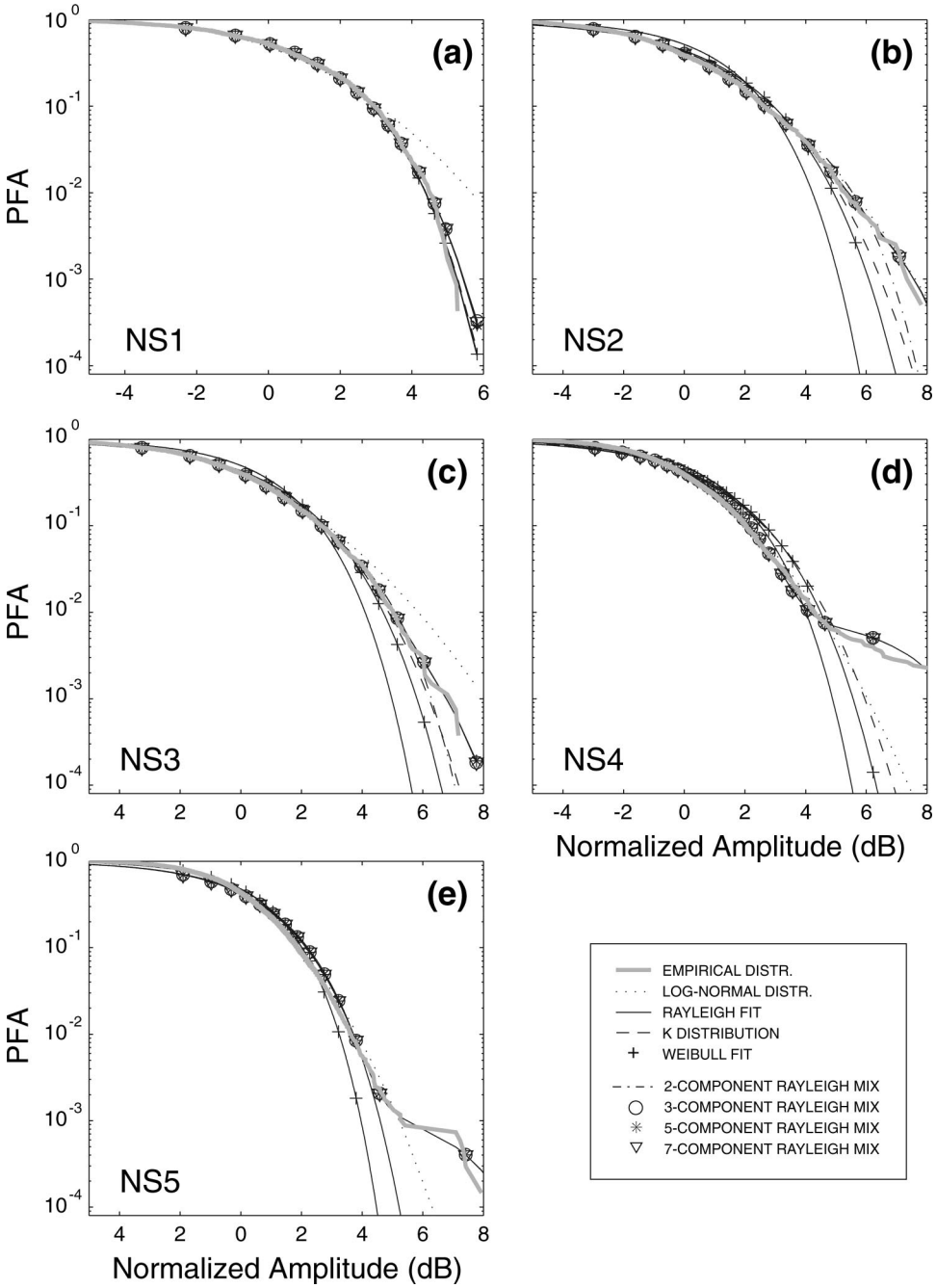


FIG. 7. PFA plots corresponding to the backscattered amplitude data from the near-surface regions in run 2.

Ref. 40, Fig. 4.1.4). Analysis of Fig. 3 and the sea surface backscattering strength imagery corresponding to these data [e.g., Fig. 2(b)] indicates that the centers of the distributions for the SS3 regions are dominated by bottom reverberation, and the tails are dominated by randomly distributed bubbles within a meter of the sea surface. The log-normal model provides the best overall fits to the data in the SS3 regions (Tables III and IV) in terms of the KS p value and rms difference which emphasize samples near the center of the distribution. As indicated by Fig. 6(c) and the rms differences for PFA values less than 10^{-2} (Table V), the Rayleigh mixture models provide the best fits to the tails in the SS3 regions.

Similar to the SS3 region, most of the near-surface data are best described by the log-normal model (Table III), but the Rayleigh mixture models provide the best fits to the tails (Fig. 7; Table V). Results for the NS1 and NS3 regions are split, with the Rayleigh mixture and log-normal models both providing the best fits for different data runs (Table III). Overall, the model-data fits are statistically poor, and characterized by the highest rms differences and lowest KS p values of all the analysis regions (Tables III and IV). Sidelobe returns from the bottom-surface multiple occur in the NS1 and NS5 regions, and sidelobe returns from the bottom echo are evident in the NS4 region. The outer edges of the NS1 region also are influenced by sidelobe returns from the first surface echo at the towfish's zenith, and the NS3 region is influenced by sidelobe returns from the first surface echo. The best fits for the log-normal distribution are in the regions where the boundary reverberation is the strongest (i.e., NS2, after the first surface echo; and NS4, after the first bottom echo).

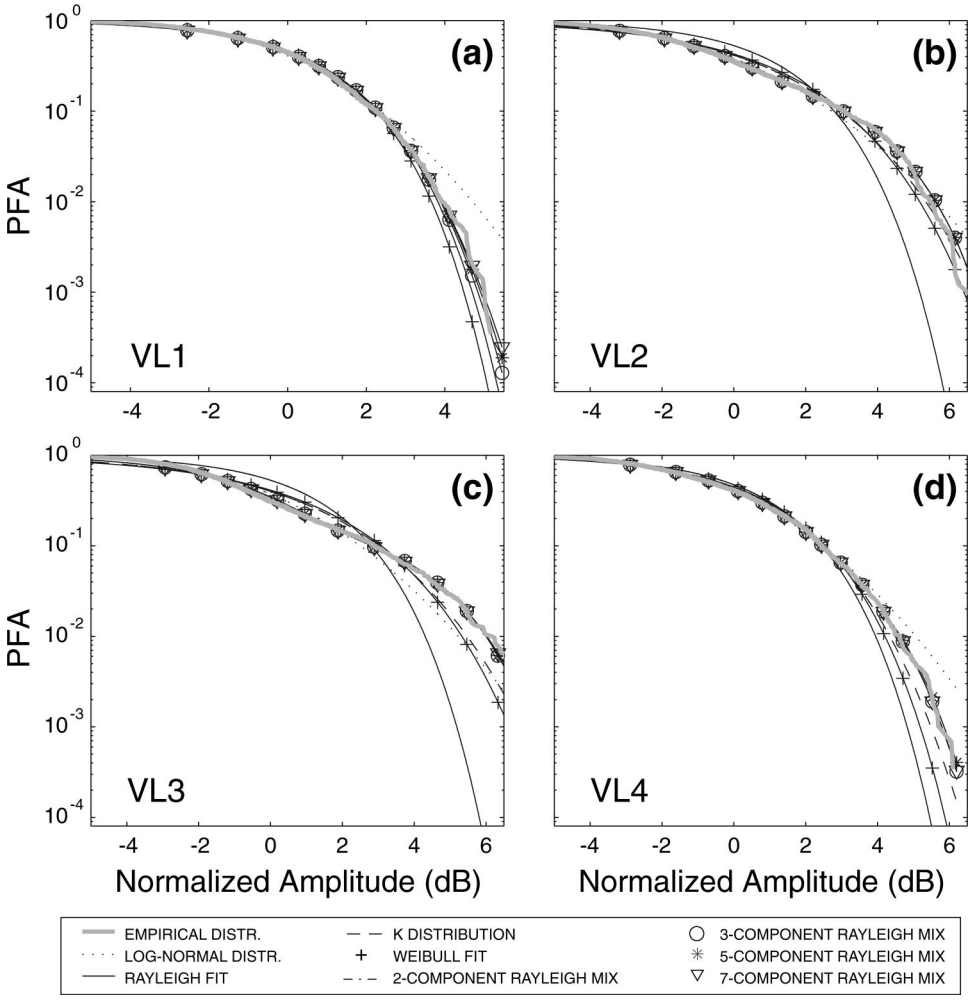


FIG. 8. PFA plots corresponding to the backscattered amplitude data from the volume regions in run 3.

Analysis of backscattering strength imagery[13] indicated that the dominant mechanisms contributing to the tails of the near-surface distributions are scattering from resonant microbubbles and bubble clouds, with bubble density controlling the tail shape. When the bubbles are sparsely distributed, such as those generated by the breaking ambient sea waves, the tail is well-separated from the distribution center, resulting in what appears to be a multi-modal distribution [NS4, Fig. 7(d); NS5, Fig. 7(c)]. When the bubbles are more densely packed, such as in the towship's wake (NS1,2) and in the region affected by large scale bubble clouds generated by ship-waves (NS3), the distributions appear unimodal, with the largest tails occurring in the regions with the highest bubble densities [NS2, Fig. 7(b)]. As with the sea surface results, little or no improvement in fitting the tails of the distributions occurred when we increased the number of components in the Rayleigh mixture above 3, and 2 components were sufficient in most cases.

Results for the volume backscatter amplitude data (Fig. 8) are generally similar to those for the near-surface data: they are best fit by the log-normal model over the center of the distribution and the Rayleigh mixture distributions in the tails. In addition, the fits are not statistically good, with relatively low KS p values and high rms differences (Tables III–V). The best fits to the log-normal model are obtained when boundary reverberation is present, i.e., in data for the VL4 region [Fig. 8(d)] that contain sidelobe returns after the first surface echo (Fig. 3). When boundary reverberation is absent, the shapes and tails of the distributions are affected by the density of the scatterers (zooplankton). For the highest scatterer densities [VL1, Fig. 8(a)], the distributions appear to be unimodal with lower tails. As the density of scatterers decreases, the distributions become multimodal, with heavier tails [VL2, Fig. 8(b); VL3, Fig. 8(c)]. These observations are somewhat consistent with those for the surface and near-surface regions, in that a sparse, nonhomogeneous spatial distribution of scatterers (bubbles) results in distributions with more complex (multimodal) shapes [cf. SS3, Fig. 6(c); NS4, Fig. 7(d)].

V. DISCUSSION

A. Nonstationarity of shallow water reverberation fluctuations

Before offering physical arguments for the observed results, we address the fact that none of the regions could be considered stationary across all pings and grazing angles. The primary factors contributing to the observed nonstationarities are the towfish's motion through the generally nonhomogeneous spatial distribution of scatterers in each region, and boundary reverberation received in the sidelobes. This is evident in Table VI, which lists the samples sizes and percent

TABLE III. KS statistic *p*-values computed from the model-data PDF fits to the TVSS acoustic backscatter amplitude data in each of the analysis regions. The highest value for each run is in bold and corresponds to the best fit.

Region run/	Rayleigh	*K*	Weibull	Log-normal	Rayleigh 2-mixture	Rayleigh 3-mixture	Rayleigh 5-mixture	Rayleigh 7-mixture
SF1/01	0.466	0.886	0.708	3.01×10^{-4}	0.936	0.932	0.936	**0.940**
/02	0.015	0.765	0.263	2.09×10^{-4}	0.890	0.890	0.982	**0.984**
/03	0.689		**0.928**	9.25×10^{-4}	0.689	0.518	0.503	0.500
SF2/01	0.442	0.878	0.907	1.94×10^{-21}	0.443	0.901	**0.939**	**0.939**
/02	0.867	0.989	**0.993**	8.18×10^{-17}	0.869	0.907	0.890	0.878
/03	0.868		**0.913**	6.02×10^{-17}	0.861	0.859	0.855	0.854
SF3/01	0.055	0.329	0.241	1.16×10^{-13}	**0.568**	0.520	0.489	0.541
/02	0.298	0.973	0.969	3.26×10^{-15}	**0.987**	0.978	0.967	0.960
/03	0.419	0.724	0.587	3.65×10^{-25}	0.432	0.769	0.424	**0.788**
SS1/01	2.19×10^{-9}	0.739	0.392	0.002	0.819	0.942	**0.973**	0.969
/02	1.05×10^{-25}	0.358	0.322	3.54×10^{-5}	0.267	**0.999**	**0.999**	**0.999**
/03	6.67×10^{-13}	0.666	0.764	0.013	0.901	**0.998**	**0.998**	**0.998**
SS2/01	2.18×10^{-67}	2.48×10^{-4}	7.41×10^{-9}	5.17×10^{-4}	0.131	0.840	0.975	**0.983**
/02	3.12×10^{-16}	0.395	0.011	2.69×10^{-4}	**0.997**	**0.997**	0.996	0.992
/03	5.81×10^{-63}	1.97×10^{-7}	7.27×10^{-14}	**0.370**	0.034	0.311	0.234	0.235
SS3/01	4.53×10^{-23}	3.14×10^{-23}	3.64×10^{-29}	**0.342**	7.53×10^{-11}	7.52×10^{-11}	7.44×10^{-11}	1.56×10^{-12}
/02	7.96×10^{-60}	2.36×10^{-53}	4.45×10^{-70}	**0.260**	7.28×10^{-19}	7.16×10^{-19}	1.33×10^{-22}	1.39×10^{-22}
/03	3.10×10^{-64}		3.58×10^{-14}	**0.045**	5.21×10^{-73}	5.21×10^{-73}	5.21×10^{-73}	5.20×10^{-73}
NS1/02	0.031	0.043	0.038	1.13×10^{-4}	0.063	**0.069**	0.057	0.063
/03	3.03×10^{-9}		2.58×10^{-4}	**0.009**	2.93×10^{-9}	6.57×10^{-10}	3.69×10^{-10}	3.13×10^{-10}
NS2/01	5.25×10^{-10}	1.40×10^{-4}	1.29×10^{-5}	**0.113**	1.43×10^{-4}	1.44×10^{-4}	1.44×10^{-4}	1.73×10^{-4}
/02	3.15×10^{-32}	2.21×10^{-6}	3.02×10^{-11}	**0.649**	0.016	6.12×10^{-4}	6.26×10^{-4}	6.26×10^{-4}
/03	7.97×10^{-11}	0.022	8.22×10^{-4}	**0.433**	0.083	0.081	0.082	0.079
NS3/01	5.65×10^{-29}	0.033	2.55×10^{-5}	1.95×10^{-4}	0.266	0.943	0.924	**0.958**
/02	5.29×10^{-25}	0.029	1.11×10^{-5}	0.004	0.485	**0.632**	0.627	0.631
/03	3.13×10^{-53}	2.37×10^{-14}	6.76×10^{-21}	**0.277**	0.031	0.100	0.093	0.088
NS4/01	2.63×10^{-35}		5.64×10^{-28}	**0.122**	1.44×10^{-49}	1.43×10^{-49}	1.44×10^{-49}	1.43×10^{-49}
/02	1.01×10^{-41}	1.53×10^{-44}	1.07×10^{-54}	**0.167**	3.04×10^{-25}	3.03×10^{-25}	3.02×10^{-25}	2.92×10^{-25}
/03	1.22×10^{-59}		5.57×10^{-13}	**0.003**	1.29×10^{-70}	1.29×10^{-70}	1.29×10^{-70}	1.29×10^{-70}
NS5/01	1.87×10^{-91}		4.23×10^{-30}	**1.03**$\times10^{-4}$	1.25×10^{-108}	1.24×10^{-108}	1.25×10^{-108}	1.25×10^{-108}
/02	3.29×10^{-77}		2.17×10^{-19}	**1.94**$\times10^{-4}$	1.86×10^{-84}	1.86×10^{-84}	1.86×10^{-84}	1.86×10^{-84}
/03	3.26×10^{-145}		5.02×10^{-13}	**7.58**$\times10^{-6}$	3.26×10^{-145}	3.26×10^{-145}	3.26×10^{-145}	3.26×10^{-145}
VL1/01	9.75×10^{-8}		7.59×10^{-6}	**0.023**	3.59×10^{-9}	9.74×10^{-8}	4.21×10^{-8}	5.08×10^{-9}
/02	1.56×10^{-10}	6.55×10^{-6}	1.15×10^{-7}	**0.303**	1.21×10^{-7}	1.48×10^{-7}	1.79×10^{-7}	1.61×10^{-9}
/03	1.11×10^{-7}		9.67×10^{-4}	**5.88**$\times10^{-4}$	1.11×10^{-7}	1.61×10^{-8}	1.07×10^{-8}	9.76×10^{-9}
VL2/01	5.22×10^{-68}	3.44×10^{-7}	8.16×10^{-14}	**0.029**	7.16×10^{-5}	7.82×10^{-5}	6.76×10^{-5}	6.62×10^{-5}
/02	7.32×10^{-89}	6.25×10^{-12}	2.91×10^{-19}	**2.20**$\times10^{-4}$	1.84×10^{-5}	1.88×10^{-5}	1.07×10^{-5}	9.06×10^{-6}
/03	2.20×10^{-70}	1.58×10^{-14}	1.05×10^{-22}	**2.29**$\times10^{-4}$	1.28×10^{-7}	1.28×10^{-7}	1.37×10^{-7}	1.36×10^{-7}
VL3/01	1.41×10^{-104}	3.09×10^{-28}	2.98×10^{-39}	**4.96**$\times10^{-11}$	1.36×10^{-18}	1.39×10^{-18}	1.52×10^{-18}	1.44×10^{-18}
/02	3.21×10^{-26}	1.14×10^{-26}	3.36×10^{-35}	**2.04**$\times10^{-6}$	2.68×10^{-25}	2.68×10^{-25}	2.69×10^{-25}	2.71×10^{-25}
/03	4.20×10^{-121}	1.70×10^{-38}	2.65×10^{-48}	**2.04**$\times10^{-11}$	3.87×10^{-19}	3.89×10^{-19}	4.32×10^{-19}	2.46×10^{-19}
VL4/01	8.27×10^{-10}		2.66×10^{-8}	**0.323**	2.38×10^{-13}	2.40×10^{-13}	2.44×10^{-13}	2.86×10^{-13}
/02	4.30×10^{-9}		3.01×10^{-8}	**0.848**	1.78×10^{-14}	1.78×10^{-14}	1.80×10^{-14}	1.86×10^{-14}
/03	1.13×10^{-15}	1.39×10^{-7}	9.09×10^{-11}	**0.092**	2.92×10^{-8}	2.98×10^{-8}	3.62×10^{-8}	3.00×10^{-8}

of data in each analysis region that were validated across pings and grazing angles as stationary and homogeneous. The lowest percentages occur in regions where multiple reverberation components with widely varying characteristics are present. For example, the NS1 region has the smallest percentage of contiguous samples satisfying stationarity and is influenced by microbubbles within the decaying ship's wake, bubbles associated with breaking waves in the ambient sea, and multiple boundary reflections received in the sidelobes. Similarly, the NS3 region has a small percentage of contiguous samples satisfying stationarity, and it is influenced by large and small scale bubble clouds, sea surface backscatter, and sea surface reverberation in the sidelobes. On the other hand, the VL1 region has the largest percentage of contiguous samples satisfying stationarity, and it is influenced almost entirely by scattering from zooplankton.

Although the TVSS data set is unique in that it contains a wide variety of backscatter and reverberation processes received in narrow beams simultaneously, it is consistent with data in other studies which have observed that shallow water

TABLE IV. RMS differences (D_{rms}) between model and TVSS-derived empirical CDFs averaged over runs 1–3. The lowest value for each region is displayed in bold and corresponds to the model with the best overall fit to the empirical distribution function.

Analysis region	Rayleigh	*K*	Weibull	Log-normal	Rayleigh 2-mixture	Rayleigh 3-mixture	Rayleigh 5-mixture	Rayleigh 7-mixture
SF1	0.0132	0.0070	0.0072	0.0359	**0.0060**	0.0063	0.0062	0.0063
SF2	0.0045	**0.0029**	0.0032	0.0425	0.0045	0.0036	0.0034	0.0034
SF3	0.0067	0.0041	0.0047	0.0395	0.0041	**0.0039**	**0.0039**	**0.0039**
SS1	0.0694	0.0091	0.0102	0.0287	0.0074	0.0040	**0.0039**	0.0040
SS2	0.0836	0.0191	0.0300	0.0180	0.0086	0.0045	**0.0041**	**0.0041**
SS3	0.0685	0.0649	0.0611	**0.0096**	0.0490	0.0490	0.0490	0.0490
NS1	0.0195	0.0155	0.0200	**0.0192**	0.0213	0.0219	0.0219	0.0219
NS2	0.0586	0.0246	0.0341	**0.0091**	0.0154	0.0153	0.0153	0.0153
NS3	0.0863	0.0282	0.0387	0.0164	0.0105	**0.0061**	**0.0061**	**0.0061**
NS4	0.0592	0.0601	0.0503	**0.0099**	0.0542	0.0542	0.0542	0.0542
NS5	0.0713		0.0348	**0.0125**	0.0744	0.0744	0.0744	0.0744
VL1	0.0291	0.0265	0.0258	**0.0130**	0.0237	0.0243	0.0240	0.0240
VL2	0.1142	0.0376	0.0476	**0.0179**	0.0203	0.0203	0.0203	0.0203
VL3	0.1164	0.0670	0.0766	**0.0317**	0.0459	0.0459	0.0459	0.0459
VL4	0.0411	0.0258	0.0369	**0.0082**	0.0322	0.0322	0.0322	0.0322

acoustic reverberation fluctuations received by moving and fixed platforms are often nonstationary.[6,23] Consequently, our results regarding probability distribution functions pertain to locally stationary reverberation fluctuations embedded in larger scale, nonstationary processes.

B. Suitability of Rayleigh mixture distributions

The Rayleigh mixture distributions provide the best fits to the observed data in most of the boundary regions (SF1-3, SS1-2), and some of the near-surface regions (NS1,3). Thus, the Rayleigh mixture model adequately handles patchy, nonhomogeneously (non-Poisson) distributed scatterers. For example, the seafloor in the area of the TVSS experiment has a bimodal sediment composition of sand and silt,[12] making the mixture of two Rayleigh random variables, hence a two-component Rayleigh mixture distribution a logical model for the observed backscatter amplitude distributions. Three, five, and seven component Rayleigh mixtures sometimes provided better fits probably because bottom roughness and the spatial variability of the bottom across each run introduced additional components.

For the sea surface and near-surface backscatter data in the SS1, SS2, and NS3 regions, bubble clouds with varying spatial scales are the most likely sources for the different components of the Rayleigh mixture distributions that were fit to the observed data. Previous analyses[13] indicate that the bubble densities in the township's wake (SS1) and in the bubble clouds generated by breaking ship waves (SS2,NS3) are relatively high, suggesting that the empirical distribution of backscatter in these regions could be modeled by a mixture of Rayleigh variables, each accounting for the different scales of the bubble clouds and scattering from the sea sur-

TABLE V. RMS differences (D_{rms}) between model and TVSS-derived empirical CDFs for PFA values below 10^{-2}, averaged over runs 1–3. The lowest value for each region is displayed in bold and corresponds to the model with the best overall fit to the empirical distribution function.

Analysis region	Rayleigh	*K*	Weibull	Log-normal	Rayleigh 2-mixture	Rayleigh 3-mixture	Rayleigh 5-mixture	Rayleigh 7-mixture
SF1	0.0020	0.0012	0.0015	0.0220	0.0011	0.0010	**0.0009**	**0.0009**
SF2	0.0007	0.0006	**0.0005**	0.0320	0.0007	0.0006	0.0006	0.0006
SF3	0.0011	**0.0004**	0.0007	0.0272	**0.0004**	0.0006	0.0005	0.0005
SS1	0.0050	0.0018	0.0014	0.0200	0.0020	0.0014	**0.0013**	**0.0013**
SS2	0.0054	0.0009	0.0024	0.0087	0.0007	0.0007	**0.0005**	**0.0005**
SS3	0.0077	0.0043	0.0034	0.0015	0.0017	0.0017	0.0011	**0.0008**
NS1	**0.0005**	**0.0005**	0.0007	0.0100	0.0009	0.0008	0.0007	0.0008
NS2	0.0011	0.0006	0.0007	0.0031	**0.0006**	0.0007	**0.0006**	0.0007
NS3	0.0010	0.0007	0.0009	0.0026	0.0006	**0.0002**	**0.0002**	**0.0002**
NS4	0.0011	0.0012	0.0011	0.0010	**0.0002**	**0.0002**	**0.0002**	**0.0002**
NS5	0.0017		0.0011	**0.0007**	0.0009	0.0009	0.0009	0.0009
VL1	0.0027	0.0015	0.0030	0.0091	0.0013	0.0016	0.0011	**0.0010**
VL2	0.0055	**0.0010**	0.0020	0.0048	0.0014	0.0013	0.0012	0.0013
VL3	0.0057	0.0030	0.0041	0.0028	0.0020	0.0020	0.0020	**0.0019**
VL4	0.0046	0.0019	0.0041	0.00382	**0.0007**	0.0008	0.0008	0.0008

TABLE VI. Sample sizes for the TVSS analysis regions. The total number of samples (column 2) corresponds to each region defined in Table I (column 1), averaged across the 3 runs. The validated samples (column 3) correspond to the data used to form the empirical distribution functions displayed in Figs. 5–8, and are the largest contiguous subsets of the regions validated as stationary and homogeneous across-pings and grazing angles.

Analysis region	Total number of samples	Number of validated samples	Percent of validated samples
SF1	2185	1505	69
SF2	5529	3840	69
SF3	9296	4406	47
SS1	2616	1330	51
SS2	6537	2725	42
SS3	8374	3951	47
NS1	10036	2500	25
NS2	2829	1823	64
NS3	6672	2350	35
NS4	11383	4985	44
NS5	12510	7296	58
VL1	3412	2835	83
VL2	3914	2522	64
VL3	3974	2732	69
VL4	4020	2591	64

face. This is supported by backscattering strength images constructed from the data in these regions [e.g., Fig. 2(b)] which reveal a patchy distribution of bubble clouds with varying length scales along and across track.[13]

Although the Rayleigh mixture models do not fit the observed near-surface and volume backscatter data well, they are effective in fitting the distribution tails for every type of analysis region (Table V). This is most noticeable for the regions in which bubbles contributed to the tails of the distributions [e.g., SS3, Fig. 6(d); NS2, Fig. 7(b); NS4], indicating that scattering from bubbles and bubble clouds could be described as a mixture of Rayleigh distributed random variables. These figures also suggest that acoustic backscatter and reverberation from these regions is probably best described by a mixture of both Rayleigh and non-Rayleigh random variables, and that the Rayleigh mixture model is flexible enough to describe that portion of the empirical distribution resulting from a mixture of Rayleigh random variables, even if the entire distribution does not fit such a model.

The flexibility of the Rayleigh mixture model comes from its unification of a number of physically based models. On one end, a Rayleigh mixture distribution dominated by a single component can approximate a single Rayleigh distribution, a Weibull distribution with $\beta=2$, and a K distribution in the limit as ν tends to infinity when the scale $1/\sqrt{\nu}$ is applied to the data.[27] A Rayleigh mixture distribution can also approximate the K distribution by quantizing the Rayleigh speckle and chi-distributed components, and equating the latter with the mixture proportions.[7] This explains why the Rayleigh mixture model fits the data well when the Weibull or K distributions do (e.g., SF2, Tables III and IV). In addition, Lyons and Abraham[7] have related the Rayleigh mixture model to other models based on physical processes, such as Crowther's[1] and McDaniel's[3] for seafloor backscatter.

C. Scattering processes approximated by log-normal distributions

The log-normal distribution provides the best fits to most of the near-surface and volume backscatter data, but the fits are only good in the centers of the distributions where boundary reverberation in the sidelobes (NS1,2,4,5; VL4) or scattering from patchy aggregations of zooplankton (VL1,2,3) dominated the acoustic backscatter. Here, we offer several explanations for these results in terms of approximately log-normal distributions. Approximate results are sufficient because the model-data fits were never statistically good since KS p values were always below 0.85.

First, we consider the near surface and volume data in which boundary reverberation in the sidelobes dominated the distribution centers (NS1,2,4,5,VL4). Because the amplitude data are validated as stationary, independent, and identically distributed across-grazing angles and pings, we can use the model in (1) and (2) to express the corresponding in-phase and quadrature components as[16]

$$V_{I,Q}(r)=\frac{1}{g(r)}\sum_{i=1}^{n} a_i B(r_i) s(r-r_i,\xi_i), \tag{20}$$

where the generalized variable r is used in place of t to represent the ranges of grazing angle, depth, along-track, or across-track distance in Table I. For the TVSS data, $g(r)$ represents the normalization procedure. Equation (20) can be rewritten as

$$V_{I,Q}(r)=\sum_{i=1}^{n} \nu_i(r), \tag{21}$$

where $\nu_i(r)$ is the reverberation component in the sum of (20) made stationary by $g(r)$. If we assume that each $\nu_i(r)$ can be expressed in terms of a random proportion of the preceding term $\nu_{i-1}(r)$, then

$$\nu_i(r)=\nu_{i-1}(r)+\zeta_i \nu_{i-1}(r), \tag{22}$$

where the random set $\{\zeta_i\}$ is mutually independent and independent of the set $\{\nu_i(r)\}$. This assumption is reasonable if the scattered amplitudes a_i and stochastic parameters ξ_i are random, as assumed in (1) and (20). Rearranging (22), we have

$$\frac{\nu_i(r)-\nu_{i-1}(r)}{\nu_{i-1}(r)}=\zeta_i, \tag{23}$$

so that

$$\sum_{i=1}^{n} \frac{\nu_i(r)-\nu_{i-1}(r)}{\nu_{i-1}(r)}=\sum_{i=1}^{n} \zeta_i. \tag{24}$$

Now, supposing the difference between successive reverberation components $\nu_i(r)$ is small,

$$\sum_{i=1}^{n} \frac{\nu_i(r)-\nu_{i-1}(r)}{\nu_{i-1}(r)} \approx \int_{\nu_1(r)}^{\nu_n(r)} \frac{d\nu}{\nu}=\ln(\nu_n(r))-\ln(\nu_1(r)), \tag{25}$$

which, from (24), becomes

$$\ln(\nu_n(r)) = \ln(\nu_1(r)) + \zeta_1 + \cdots + \zeta_n . \tag{26}$$

For large n, the central limit theorem implies that $\ln(\nu_n(r))$ is normally distributed, so $\nu_n(r)$ is log-normally distributed by the properties of the log-normal distribution discussed in Sec. II. If the minimum number for which (26) converges to log-normal is much less than the total number of scatterers (n), then $V_{I,Q}(r)$ in Eq. (20) will be approximately log-normal because the sum of log-normal variables is approximately log-normal.[38] By the same virtue, the TVSS amplitude data that are related to the in-phase and quadrature components by

$$A(r) = \sqrt{V_I^2(r) + V_Q^2(r)} \tag{27}$$

are approximately log-normal because a log-normal variable raised to a power is also log-normally distributed.[39]

From this development one might wonder why Eq. (26) converges to a normal distribution, but not Eq. (20). The central limit theorem may be invoked for Eq. (20) when n is large, but the value of n for which Eq. (20) converges to a normal distribution will be smaller if the variables $\nu_i(t)$ are from the same underlying distributions. This can be expected when the total scattered signal arises from numerous scatterers of the same type. The value of n for which Eq. (20) converges to a normal distribution will be larger when each $\nu_i(r)$ arises from a different underlying distribution, especially those which are highly skewed or have large tails.[44,47] This is likely to occur for volume backscatter received by a moving platform and dominated by boundary sidelobe returns, because each sidelobe is directed towards a different grazing angle with respect to the boundary (see Ref. 41 for the receive beam patterns of the TVSS). The total reverberation will be the sum of the components arriving in each sidelobe, where each component follows a different parent distribution. Thus, a log-normal distribution may approximate the observed data better than a normal distribution when the underlying distributions for the proportions ζ_i are less skewed, with smaller tails, than those for $\nu_i(r)$.

Perhaps the log-normal model best fits the volume backscatter data in and above the thermocline (VL1,2,3) because several biological and physical factors affecting the distribution of zooplankton are log-normally distributed. In observations made by Dugan *et al.*,[48] the distribution of horizontal temperature fluctuations in the seasonal thermocline followed a log-normal distribution on scales from 10 cm to 1 km, whereas Campell's[49] analysis showed that a variety of factors related to phytoplankton, such as chlorophyll concentration and cell size, are log-normally distributed. The spatial distribution of zooplankton in the northeastern Gulf of Mexico can be related to both of these. Zooplankton have been shown to be concentrated near the mixed layer depth, which is influenced by both weather and mixing processes, and also near the depth of the primary productivity maximum, which is related to phytoplankton, hence chlorophyll, distribution.[14] Because these factors influence the number n of scatterers in (20), they directly influence the empirical distributions of the backscatter amplitude fluctuations.

It is interesting to note that the log-normal distribution has been used to describe a wide variety of physical phenomena which may be indirectly related to volume acoustic backscatter fluctuations. In theoretical biology, for example, species abundance has been successfully described by the log-normal model, and organism growth was the first application that used the development in Eqs. (20)–(26), which is known as the law of proportional effect.[50] The TVSS volume backscatter fluctuations are influenced by the size and species abundance distribution of sound scattering zooplankton through their backscattering cross sections, which are represented by the a_i terms in (20). Although backscattering strength imagery indicated that the spatial distribution of zooplankton was nonhomogeneous [e.g., Fig. 2(d)], we did not collect net tow or trawl samples to verify whether the approximately log-normal distributions of the volume backscatter fluctuations were related to log-normal size or species abundance distributions of volume scatterers. Nevertheless, such observations are not likely to be purely coincidental, and deserve further investigation, in view of other studies in which high frequency volume acoustic backscatter from biological sound scatterers was approximately log-normal.[29,36,51]

A final observation for approximately log-normal acoustic backscatter and reverberation distributions is that a log-normal distribution can be expressed as a mixture of several physically relevant distributions. Although we do not think this is the reason for our observations, this might explain others' (e.g., Refs. 35 and 23). For example, Titterington *et al.*[40] show an example in which two Gaussian distributions are used to approximate a two-parameter log-normal distribution.[40] Thus, two Ricean distributed scattering processes, each dominated by a coherent scattering component, might yield a distribution which is approximately log-normal. Similarly, the log-normal distribution may be closely approximated by the gamma distribution,[52] which has been directly related to a variety of scattering processes by Middleton.[53] This is appropriate for shallow water reverberation, which typically includes multiple components from the boundaries and volume. These considerations are consistent with our previous conclusions that mixture distributions with component densities from different parametric families may better describe fluctuations of acoustic backscatter and reverberation in shallow water.

D. Implications for target detection

This study has several implications for undersea target detection. The non-Rayleigh nature of envelope fluctuations that arise from nonhomogeneous spatial distributions of scatterers is seen in large tails and/or multimodality in the distributions. For a predetermined probability of false alarm (PFA), this implies that threshold detectors which assume Rayleigh-distributed envelope fluctuations will experience significantly higher false alarm rates. However, even with the appropriate PDF model for envelope fluctuations due to the environment, target detection is difficult for data within the tails of the distributions. For these data, combining statistical

techniques with analyses of multibeam imagery, as we have done here, may be a more effective method for discriminating between targets and noise.

VI. SUMMARY AND CONCLUSIONS

In this study, we have analyzed the fluctuations of seafloor, sea surface, and volume acoustic backscatter data simultaneously collected by the Toroidal Volume Search Sonar (TVSS) while it was towed in a shallow water region in the northeastern Gulf of Mexico. The 68 kHz acoustic backscatter data were grouped according to 15 analysis regions in which scattering contributions from the volume and/or boundaries were present. After normalizing for backscatter variations due to grazing angle dependence and nonuniformity in the TVSS's beam patterns, the data were validated as independent and identically distributed. Various moments and statistics were estimated for the data in each region, and Rayleigh, *K*, Weibull, log-normal, and Rayleigh mixture probability distributions were fit to the empirical distribution functions in each region. We used previously published volume and boundary acoustic backscattering strength images constructed from the multibeam data collected by the TVSS to interpret the observed backscatter and reverberation statistics in terms of the spatial distribution of scatterers.

Rayleigh mixture models provided the best fits to the backscatter data collected from both boundaries, and in most cases, three-component mixtures adequately described the observed data. For the near-surface and volume regions, none of the models considered yielded statistically good fits. The Rayleigh mixture distributions provided the best fits to the larger tails in the data for these regions, which were mostly due to sparse distributions of bubbles near the surface or zooplankton in the mixed layer and thermocline. The log-normal distribution best fit the centers of the distributions for the near-surface and volume regions, particularly when single and multiple boundary interactions were received in the sidelobes. Together, these observations suggest that mixture distributions with component densities from different parametric families might better describe the multiple-component reverberation that is typical of most shallow water environments.

With mixture distributions, it is difficult to determine the number of components required to represent the data, or what PDF families are appropriate, particularly since several different mixtures can be used to approximate the same distribution. This guess work can be reduced by identifying candidate mixture components based on the spatial distribution of scattering features observed in coincident acoustic backscatter imagery.

The results were displayed as probabilities of false alarm (PFAs) in order to emphasize the larger tails of the non-Rayleigh backscatter statistics. The tails corresponding to data in the near-surface, sea-surface, and volume were much larger than those for the seafloor. Large tails resulted mostly from nonhomogeneous spatial distributions of bubbles near the sea surface, and zooplankton and small fish at the base of the mixed layer and in the thermocline. Multimodal distributions with extended tails were observed when the data were influenced by both discrete scatterers and multiple boundary interactions received in the sidelobes. The results demonstrate that the dominant environmental sources of noise in shallow water target detection applications are likely to be resonant microbubbles near the surface, aggregations of zooplankton and fish in the mixed layer and upper thermocline, and boundary reverberation throughout the entire water column.

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 (ret) (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. Special thanks are due to Doug Abraham for providing Ref. 27, which was essential in completing this work. Thanks are due to Jo Griffith (MPL) for helping with the figures.

APPENDIX: SKEWNESS AND KURTOSIS DESCRIPTORS

Expressions for the skewness and kurtosis descriptors in Fig. 4 are given by

$$\beta_1=\gamma_{3,A}^2, \quad \beta_2=\gamma_{4,A}+3, \tag{A1}$$

which are obtained from the moments

$$\begin{aligned}
&\mu_A=E[A] \quad \text{(mean)},\\
&\sigma_A^2=E[(A-\mu_A)^2] \quad \text{(variance)},\\
&\gamma_{3,A}=\frac{E[(A-\mu_A)^3]}{\sigma_A^3} \quad \text{(skewness)},\\
&\gamma_{4,A}=\frac{E[(A-\mu_A)^4]}{\sigma_A^4} \quad \text{(kurtosis)}.
\end{aligned} \tag{A2}$$

Skewness and kurtosis are measures of departure from normality. Skewness represents asymmetry in the PDF, and high kurtosis indicates a relatively large number of values near the mean of the distribution.

For the TVSS data, the sample moments were calculated using (A2) with

$$E[u]=\frac{1}{N}\sum_{i=1}^{N} u_i, \tag{A3}$$

for the expected value.

For the PDF models, expressions for β_1 and β_2 in terms of the noncentral moments

$$\alpha_i=E[A^k]=\int_{-\infty}^{\infty} A^k p(A)\,dA, \tag{A4}$$

where $p(A)$ is the probability density function of A, can be obtained by using the k central moments

$$E[(A-\mu_A)^k]=\int_{-\infty}^{\infty} (A-\mu_A)^k p(A)\,dA, \tag{A5}$$

yielding[27,54]

$$\beta_1=\frac{(\alpha_3-3\alpha_1\alpha_2+2\alpha_1^3)^2}{(\alpha_2-\alpha_1^2)^3} \tag{A6}$$

and

$$\beta_2=\frac{\alpha_4-4\alpha_1\alpha_3+6\alpha_1^2\alpha_2-3\alpha_1^4}{(\alpha_2-\alpha_1^2)^2}. \tag{A7}$$

From these expressions and knowledge of the noncentral moments of the various PDF models, one can obtain the skewness and kurtosis descriptors plotted in Fig. 4. The kth noncentral moment of the Rayleigh distribution [Eq. (3)] is[45]

$$E[A^k]=\lambda_R^{k/2}\Gamma\left(1+\frac{k}{2}\right), \tag{A8}$$

which can be used with (A6) and (A7) to show that it is represented by a single point in the (β_1,β_2) plane

$$\beta_1=\frac{4\pi(\pi-3)^2}{(4-\pi)^3} \tag{A9}$$

and

$$\beta_2=\frac{32-3\pi^2}{(4-\pi)^2}, \tag{A10}$$

because the Rayleigh distribution is fully represented by a single scale parameter (λ_R), and skewness and kurtosis are scale invariant descriptors.

One can use the kth noncentral moments of the two-parameter K, Weibull, and log-normal distributions to show that each of these is represented by a line in the (β_1,β_2) plane (Fig. 4). The noncentral moments of the K distribution [Eq. (5)] are[27]

$$E[A^k]=\alpha^{k/2}\Gamma\left(1+\frac{k}{2}\right)\frac{\Gamma\left(\nu+\frac{k}{2}\right)}{\Gamma(\nu)}, \tag{A11}$$

those for the Weibull distribution [Eq. (7)] are[45]

$$E[A^k]=\alpha^{-k/\beta}\Gamma\left(1+\frac{k}{\beta}\right), \tag{A12}$$

and the moments of the log-normal distribution [Eq. (9)] are[39]

$$E[A^k]=e^{(k\beta+(1/2)k^2\alpha^2)}, \tag{A13}$$

where α, β, and ν are the parameters of the various distributions as defined in the text.

The m-component Rayleigh mixture distribution [Eq. (12)] is represented by a region in the (β_1,β_2) plane, which can be seen by using the equation for the kth noncentral moment:[27]

$$E[A^k]=\sum_{i=1}^{m}\varepsilon_i\lambda_{R,i}^{k/2}\Gamma\left(1+\frac{k}{2}\right) \tag{A14}$$

in (A6) and (A7) and by varying the proportions (ε_i) over the interval [0, 1] and the powers $\lambda_{R,i}$ over $[0,\infty]$. A two component Rayleigh mixture distribution taken from Abraham[27] is shown in Fig. 4.

[1] P. A. Crowther, "Fluctuation statistics of sea-bed acoustic backscatter," in *Bottom-Interacting Ocean Acoustics*, edited by W. A. Kuperman and F. B. Jensen (Plenum, New York, 1980), pp. 609–622.
[2] S. Stanic and E. Kennedy, "Fluctuations of high-frequency shallow-water seafloor reverberation," J. Acoust. Soc. Am. **91**, 1967–1973 (1992).
[3] S. T. McDaniel, "Seafloor reverberation fluctuations," J. Acoust. Soc. Am. **88**, 1530–1535 (1990).
[4] T. D. Plemons, J. A. Shooter, and D. Middleton, "Underwater acoustic scattering from lake surfaces. I. theory, experiment, and validation of the data," J. Acoust. Soc. Am. **52**, 1487–1502 (1972).
[5] T. D. Plemons, J. A. Shooter, and D. Middleton, "Underwater acoustic scattering from lake surfaces. II. covariance functions and related statistics," J. Acoust. Soc. Am. **52**, 1503–1515 (1972).
[6] N. P. Chotiros, H. Boehme, T. G. Goldsberry, S. P. Pitt, R. A. Lamb, A. L. Garcia, and A. Altenburg, "Acoustic backscattering at low grazing angles from the ocean bottom: Part II. Statistical characteristics of bottom backscatter at a shallow water site," J. Acoust. Soc. Am. **77**, 975–982 (1985).
[7] A. P. Lyons and D. A. Abraham, "Statistical characterization of high-frequency shallow-water seafloor backscatter," J. Acoust. Soc. Am. **106**, 1307–1315 (1999).
[8] J. Dunlop, "Statistical modeling of sidescan sonar images," Proceedings of the MTS/IEEE Oceans '97, 1997, Vol. 1, pp. 33–38.
[9] C. de Moustier, "Beyond bathymetry: mapping acoustic backscattering from the deep seafloor with Sea Beam," J. Acoust. Soc. Am. **79**, 316–331 (1986).
[10] T. K. Stanton and C. S. Clay, "Sonar echo statistics as a remote sensing tool: volume and seafloor," IEEE J. Ocean. Eng. **OE-11**, 79–96 (1986).
[11] T. K. Stanton, "Sea surface scattering: Echo peak PDF," J. Acoust. Soc. Am. **77**, 1367–1369 (1985).
[12] T. C. Gallaudet, "Using environmental information to estimate and correct for errors in bathymetry and seafloor acoustic imagery," Chapter 3 of "Shallow water acoustic backscatter and reverberation measurements using a 68-kHz cylindrical array," Ph.D. thesis, University of California, San Diego, 2001.
[13] T. C. Gallaudet and C. P. de Moustier, Sea surface and volume acoustic backscatter imagery of the microbubble field of a ship's wake J. Acoust. Soc. Am. (to appear).
[14] T. C. Gallaudet and C. P. de Moustier, "Multibeam volume acoustic backscatter imagery and reverberation measurements in the Northeastern Gulf of Mexico," J. Acoust. Soc. Am. **112**, 489–502 (2002).
[15] P. Faure, "Theoretical model of reverberation noise," J. Acoust. Soc. Am. **36**, 259–266 (1964).
[16] V. V. Ol'shevskii, *Characteristics of Sea Reverberation* (Consultants Bureau, New York, 1967).
[17] D. Middleton, "A statistical theory of reverberation and similar first-order scattered fields: Part I: Waveforms and the general process," IEEE Trans. Inf. Theory **IT-13**, 372–392 (1967).
[18] D. Middleton, "A statistical theory of reverberation and similar first-order scattered fields: Part II: Moments, spectra, and special distributions," IEEE Trans. Inf. Theory **IT-13**, 393–414 (1967).
[19] D. Middleton, "A statistical theory of reverberation and similar first-order scattered fields: Part III: Waveforms and fields," IEEE Trans. Inf. Theory **IT-18**, 35–67 (1967).
[20] D. Middleton, "A statistical theory of reverberation and similar first-order scattered fields: Part IV: Statistical models," IEEE Trans. Inf. Theory **IT-18**, 68–90 (1967).
[21] S. Stanic and E. G. Kennedy, "Reverberation fluctuations from a smooth seafloor," IEEE J. Ocean. Eng. **OE-18**, 95–99 (1993).
[22] H. Griffiths, J. Dunlop, and R. Voles, "Textural analysis of sidescan sonar imagery using statistical scattering models," in *High Frequency Acoustics in Shallow Water*, edited by N. G. Pace, E. Pouliquen, O. Bergem, and A. P. Lyons (NATO SACLANT Undersea Research Center, La Spezia, 1997), pp. 187–194.
[23] M. Gensane, "A statistical study of acoustic signals backscattered from the sea bottom," IEEE J. Ocean. Eng. **14**, 84–93 (1989).
[24] W. K. Stewart, D. Chu, S. Malik, S. Lerner, and H. Singh, "Quantitative seafloor characterization using bathymetric sidescan sonar," IEEE J. Ocean. Eng. **19**, 599–610 (1994).
[25] S. T. McDaniel, "Sea surface reverberation fluctuations," J. Acoust. Soc. Am. **94**, 1551–1559 (1993).
[26] P. H. Dahl and W. J. Plant, "The variability of high-frequency acoustic backscatter from the region near the sea surface," J. Acoust. Soc. Am. **101**, 2596–2602 (1997).
[27] D. A. Abraham, *Modeling non-Rayleigh reverberation*, SACLANT Under-

sea Research Center Report SR-266 (SACLANT Undersea Research Center, La Spezia, June 1997). (Note: Although limited in distribution, this is available from the author upon request via email: dabraham@ieee.org).
[28] M. E. Frazer, ''Some statistical properties of lake surface reverberation,'' J. Acoust. Soc. Am. **64**, 858–868 (1978).
[29] V. C. Anderson, ''Frequency dependence of reverberation in the ocean,'' J. Acoust. Soc. Am. **41**, 1467–1474 (1967).
[30] T. K. Stanton, ''Sonar estimates of seafloor microroughness,'' J. Acoust. Soc. Am. **75**, 809–817 (1984).
[31] K. D. Ward, ''Compound representation of high resolution sea clutter,'' Electron. Lett. **17**, 561–563 (1981).
[32] S. Watts, ''Radar detection prediction in sea clutter using the compound *K* distribution model,'' IEE Proc. F, Commun. Radar Signal Process. **7**, 613–620 (1985).
[33] E. Jakeman and P. N. Pusey, ''A model for non-Rayleigh sea echo,'' IEEE Trans. Antennas Propag. **24**, 806–814 (1976).
[34] A. Abdi and M. Kaveh, ''*K* distribution: an approximate substitue for Rayleigh-lognormal distribution in fading shadowing wireless channels,'' Electron. Lett. **34**, 851–852 (1998).
[35] G. V. Frisk, ''Intensity statistics for long-range acoustic propagation in the ocean,'' J. Acoust. Soc. Am. **N64**, 257–259 (1978).
[36] B. Castile, ''Characterization of acoustic reverberation in the ocean for high frequency, high resolution, sonar systems,'' Ph.D. thesis, University of California, San Diego, 1978.
[37] G. V. Trunk and S. F. George, ''Detection of targets in non-Gaussian clutter,'' IEEE Trans. Aerosp. Electron. Syst. **AES-6**, 620–628 (1970).
[38] N. C. Beaulieu, A. A. Abu-Dayya, and P. J. McLane, ''Estimating the distribution of a sum of independent lognormal random variables,'' IEEE Trans. Commun. **43**, 2869–2873 (1995).
[39] K. Shimizu and E. L. Crow, ''History, genesis, and properties,'' in *The Lognormal Distribution: Theory and Applications*, edited by E. L. Crow and K. Shimizu (Marcel Dekker, New York, 1988), Chap. 1, pp. 1–26.
[40] D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions* (Wiley, New York, 1985).
[41] T. C. Gallaudet and C. P. de Moustier, ''On optimal amplitude shading for arrays of irregularly spaced or noncoplanar elements,'' IEEE J. Ocean. Eng. **25**, 553–567 (2000).
[42] T. C. Gallaudet and C. P. de Moustier, ''Corrections to an optimal shading for arrays of irregularly-spaced or noncoplanar ends'' IEEE J. Ocean. Eng. **26-3**, 416–419 (2001).
[43] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*, 2nd ed. (Wiley, New York, 1986).
[44] J. A. Rice, *Mathematical Statistics and Data Analysis* (Duxbury Press, Belmont, CA, 1995), Chap. 5, pp. 163–176.
[45] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, 2nd ed. (Wiley, New York, 1994), Vol. 1.
[46] M. Fisz, *Probability Theory and Mathematical Statistics*, 3rd ed. (Wiley, New York, 1963).
[47] I. T. Joliffe, ''Sample sizes and the central limit theorem: The Poisson distribution as an illustration,'' Am. Stat. **49**, 269 (1995).
[48] J. P. Dugan, B. W. Stalcup, and R. L. DiMarco, ''Small scale activity in the upper ocean,'' J. Geophys. Res. Oceans **97**, 5665–5675 (1992).
[49] J. Campbell, ''The lognormal distribution as a model for bioptical variability in the sea.'' J. Geophys. Res. Oceans **100**, 13237–13254 (1995).
[50] J. Aitchison and J. A. C. Brown, *The Lognormal Distribution* (Cambridge University Press, London, 1969).
[51] P. Greenblatt, ''Distributions of volume scattering observed with an 87.5 kHz sonar,'' J. Acoust. Soc. Am. **71**, 879–885 (1982).
[52] J. R. Clark and S. Karp, ''Approximations for lognormally fading optical signals,'' Proc. IEEE **58**, 1964–1965 (1970).
[53] D. Middleton, ''New physical-statistical methods and models for clutter and reverberation: The KA-distribution and related probability structures,'' IEEE J. Ocean. Eng. **24**, 261–284 (1999).
[54] The expressions for β_1 and β_2 in Ref. 27 contain typographical errors. Equations (A6) and (A7) are the correct forms.

On the accuracy of perturbative solutions to wave scattering from rough closed surfaces

Nikolaos C. Skaropoulos[a] and Dimitrios P. Chrissoulidis[b]
Department of Electrical and Computer Engineering, Faculty of Engineering, Aristotle University of Thessaloniki, P.O. Box 1562, GR-54124 Thessaloniki, Greece

(Received 3 July 2002; revised 25 April 2003; accepted 19 May 2003)

The method of small perturbations is applied to the problem of plane-wave scattering from a soft circular surface with sinusoidal roughness. Rayleigh-theory and extinction-theorem perturbative solutions of arbitrary order are developed, and they are compared to each other as well as to an exact solution. A numerical study yields quantitative information about effects associated with the use of the Rayleigh hypothesis, about the merit of higher-order solutions, and about the reliability of the error criteria that measure *a posteriori* the conformity of perturbative solutions to the boundary condition. It is shown that the accuracy of perturbative solutions depends not only on the height, but also on the slope, of surface corrugations. Moreover, it is shown that the Rayleigh hypothesis does not affect the accuracy of far-field perturbative calculations. The mean extinction error is proposed as a measure for the error of the extinction-theorem perturbative solution. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1591774]

PACS numbers: 43.30.Hw, 43.20.Fn [LLT]

I. INTRODUCTION

The method of small perturbations is commonly applied to problems of wave scattering from rough closed surfaces with dimensions comparable to the incident wavelength.[1–15] Typically, such surfaces are obtained by deterministic or random deformation of a smooth reference surface, which is either cylindrical[1–6] or spherical,[7–15] and perturbative solutions are accurate to second order in a smallness parameter that refers to the degree of deformation of the reference surface.

Despite the widespread use of second-order perturbative solutions, fundamental issues that pertain to their applicability are still unresolved. On the one hand, there is no consensus[16,17] on whether formulations based on the extinction theorem are superior or equivalent to formulations based on the Rayleigh hypothesis.[18,19] On the other hand, it is generally not known whether and how perturbation series converge[20,21] and even if it is useful to pursue higher-order solutions, which require considerably more effort than second-order ones. Furthermore, exact solutions, in particular for randomly rough surfaces, are scarce and computationally demanding[14] and quantitative error criteria that would allow the determination of the domain of applicability of perturbative solutions are not available.

This paper attempts to address some of the above issues through the study of plane-wave scattering from a simple rough closed surface, namely a soft circular surface with deterministic sinusoidal roughness. The simplicity of the aforesaid problem allows formulating an exact solution as well as Rayleigh-theory and extinction-theorem perturbative solutions of any order. Comparisons between these solutions provide quantitative information about (a) the effect of the Rayleigh hypothesis on the accuracy of perturbative solutions, (b) the usefulness of higher-order terms, and (c) the reliability of error criteria that measure the conformity of perturbative solutions to the boundary condition used in their formulation.

It is worth noting that related topics have been studied extensively in the context of planar surfaces, in which case a vast literature exists (see Refs. 16–25, 29–31, and references cited therein). For instance, second- and fourth-order solutions were compared by Soto-Crespo *et al.*,[22] extinction-theorem and Rayleigh-theory perturbative formulations were compared by Jackson *et al.*[20] and Roginsky;[21] and the connection between near field and far field error criteria was discussed by Wirgin,[23] Hugonin *et al.*,[24] and Van den Berg.[25] Nonetheless, the techniques developed in the above works have not yet been applied to the study of rough closed surfaces, such as randomly rough spheres and cylinders, which have important applications. The work presented here represents a first step in this direction: The sinusoisally rough circular surface can be considered as a ''canonical'' rough closed surface; conclusions drawn from its study may apply qualitatively to more complex geometric configurations and boundary conditions.

The analytical formulation of this paper makes use of expansions of the incident and scattered fields into cylindrical wavefunctions, which are regular and nonregular at the origin of coordinates, respectively. These wavefunctions are arranged in invariant subspaces by means of a group-theoretic consideration for the periodicity of the scattering surface. The coefficients of the outgoing-wave expansion of the scattered field (i.e., the scattered wave amplitudes) are the unknowns which are determined in three different ways as described below.

The exact formulation makes use of the Helmholtz integral formula to relate the scattered wave to the normal derivative of the field on the scattering surface. The latter field

[a]Currently with IRCTR, Delft University of Technology, 2628CD Delft, The Netherlands; electronic mail: N.Skaropoulos@irctr.tudelft.nl
[b]Electronic mail: dpchriss@eng.auth.gr; phone/fax: +30-2310-996334.

quantity is expanded into a Fourier series and the coefficients of that series are obtained by application of the extinction theorem. The end-result is an infinite set of linear equations that can be solved numerically. The methodology is analogous to the one applied by Masel *et al.*[26] to the problem of wave scattering from a sinusoidal planar surface.

The perturbative formulations express the scattered wave amplitudes as power series in a smallness parameter. The Rayleigh-theory formulation yields the coefficients of that series directly by application of the boundary condition on the scattering surface; this solution invokes the assumption, known as the Rayleigh hypothesis, that the outgoing-wave expansion of the scattered field is valid even on the scattering surface. The need for the aforesaid assumption is circumvented by the extinction-theorem formulation, which makes use of the Helmholtz integral formula and of the extinction theorem. In either case, the end-result is a recursive relation for the general-order term of the power series expansion of the scattered wave amplitudes.

The numerical study included in this paper is aimed at assessing the Rayleigh-theory and extinction-theorem perturbative solutions. First, the numerical stability of the exact solution is investigated. This is done by means of two error measures, namely, the boundary error, which measures the conformity of the exact solution to the boundary condition at a protrusion of the scattering surface, and the energy error, which measures the conformity of this solution to the forward-scattering theorem. Next, the expressions for the scattered wave amplitudes of the extinction-theorem and Rayleigh-theory perturbative solutions are compared to each other; this is done analytically up to the third-order and numerically for higher orders. Higher-order perturbative solutions are also compared to second-order ones through use of the far-field error, which is defined by reference to the exact solution. Finally, the possibility of setting up quantitative error criteria for Rayleigh-theory or extinction-theorem perturbative solutions is probed. In either case, the error measure indicates the degree to which the perturbative solution *a posteriori* satisfies the boundary condition that was used for its derivation.

The paper is organized as follows. The scattering problem is formulated in Sec. II. The group-theoretical formulation of the solution is presented in Sec. III. The exact solution, the extinction-theorem perturbative solution, and the Rayleigh-theory perturbative solution are developed in Secs. IV A, IV B, and IV C, respectively. Numerical results and discussions follow in Sec. V. Finally, concluding remarks are made in Sec. VI.

II. DESCRIPTION OF THE SCATTERING PROBLEM

Let S be a sinusoidally rough circular surface in a 2-D unbounded homogeneous medium (Fig. 1). With regard to the polar coordinate system $(O;r\theta)$, any point on S is determined by

$$r_S(\theta)=a[1+\varepsilon f(\theta)], \tag{1}$$

where a is the mean radius of S, ε is the smallness parameter, $f(\theta)=\sin(j\theta)$, and $j\in N$ is the order of sinusoidal roughness. S is periodic with (angular) period $2\pi/j$; maxima of r_S occur

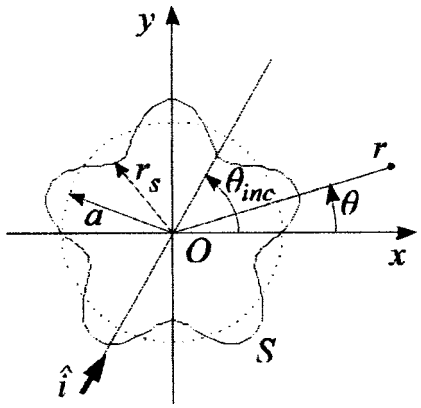


FIG. 1. Geometric configuration.

at $\theta=\theta_\nu^+=\pi(1+4\nu)/2j$, whereas minima occur at $\theta=\theta_\nu^-=\pi(3+4\nu)/2j$; $\nu=0,1,...,j-1$. Examples of S for various values of j are shown in Fig. 2.

A plane wave is incident upon S from the direction $\hat{i}=\hat{x}\cos\theta_{\rm inc}+\hat{y}\sin\theta_{\rm inc}$ (Fig. 1); this excitation is represented by the scalar potential $\psi^{\rm inc}(r,\theta)=\exp[ikr\cos(\theta-\theta_{\rm inc})]$, where k is the wave number of the surrounding medium. The time dependence of the incident wave is of the type $e^{-i\omega t}$ and it is suppressed throughout the analysis. The total field

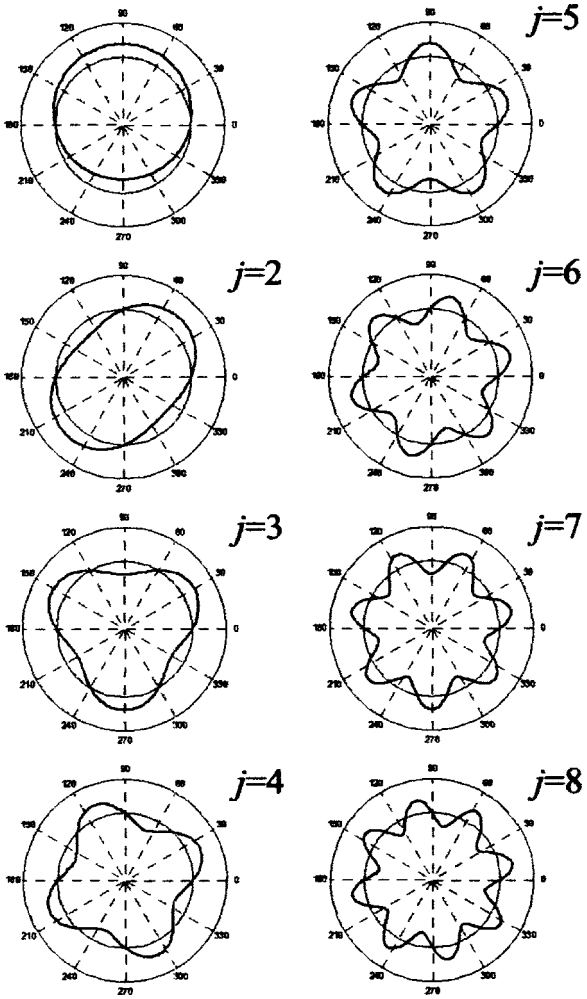


FIG. 2. The sinusoidally rough circular surface $r_S(\theta)=a[1+\varepsilon\sin(j\theta)]$ for $a=1$, $\varepsilon=0.2$, and $1\leq j\leq 8$.

$\psi=\psi^{\text{inc}}+\psi^{\text{sca}}$ outside S (i.e., for $r\geqslant r_s$) is the sum of the incident and scattered waves. The scalar potential ψ obeys the 2-D Helmholtz equation $(\nabla^2+k^2)\psi=0$. Furthermore, ψ is assumed to satisfy a boundary condition of the Dirichlet type on S; hence, $[\psi]_{r=r_s(\theta)}=0$.

III. GROUP-THEORETIC FORMULATION

It is convenient to express the scalar potentials of the field in terms of a suitable set of basis functions; the latter should be solutions to the Helmholtz equation and they should reflect the symmetry of the scattering problem in hand. In the remainder of this section, the basis functions are selected by means of a group-theoretic consideration for the periodicity of the scattering surface.

The pertinent symmetry group is the cyclic group of order j, which is denoted as C_j. The elements of C_j are $E,A,A^1,\ldots,A^{j-1}$; E denotes the identity element and A denotes rotation by angle $2\pi/j$ about an axis, normal to S, through the origin of coordinates. The group C_j is abelian and it has j irreducible representations, the μ-th of which has the characters $\chi_\mu(A^\nu)=e^{i2\pi\nu\mu/j}$, where $\mu,\nu=1,2,\ldots,j$.

Let V be the vector space of square-integrable functions on the circle. Given any $f\in V$, the set of operators $U(A^\nu)f(\theta)=f(\theta+2\pi\nu/j)$ defines a representation of C_j on V, which can be used to decompose V into a sum of subspaces $V=\Sigma_{\mu=1}^{j}V_\mu$; V_μ is the irreducible invariant subspace of V with respect to the μ-th irreducible representation of C_j on V. The projection operator to V_μ, which is denoted by P_μ, is defined as:[27]

$$P_\mu=\frac{1}{j}\sum_{\nu=1}^{j}\chi_\mu^{-1}(A^\nu)U(A^\nu). \tag{2}$$

It follows that any function $f\in V$ may be written in the form

$$f(\theta)=\sum_{\mu=1}^{j}f_\mu(\theta), \tag{3}$$

where $f_\mu(\theta)\equiv P_\mu f(\theta)$. Furthermore, it can be shown through use of Eqs. (2) and (3) that $e^{-i\mu\theta}f_\mu(\theta)$ is periodic with period $2\pi/j$. Hence, f_μ can be expanded in a Fourier series:

$$f_\mu(\theta)=\sum_{n=-\infty}^{\infty}F_{\mu n}e^{i(\mu+jn)\theta}. \tag{4}$$

The scalar potential of the incident wave can thus be decomposed into irreducible components

$$\psi^{\text{inc}}=\sum_{\mu=1}^{j}\psi_\mu^{\text{inc}}, \tag{5}$$

where $\psi_\mu^{\text{inc}}\in V_\mu$. If Eqs. (2) and (4) are applied, Eq. (5) yields

$$\psi_\mu^{\text{inc}}(\rho,\theta)=\sum_{n=-\infty}^{\infty}i^{\mu+jn}J_{\mu+jn}(\rho)e^{i(\mu+jn)(\theta-\theta_{\text{inc}})}, \tag{6}$$

where $\rho=kr$ and $J_n(\cdot)$ is the nth order cylindrical Bessel function of the first kind.[28]

The scalar potential of the scattered wave can be formulated accordingly:

$$\psi^{\text{sca}}=\sum_{\mu=1}^{j}\psi_\mu^{\text{sca}}. \tag{7}$$

On the one hand, the azimuthal dependence of ψ_μ^{sca} is identical to that of ψ_μ^{inc} since $\psi_\mu^{\text{sca}}\in V_\mu$. On the other hand, the radial dependence of ψ_μ^{sca} is determined by the requirement that it should satisfy the Helmholtz equation and the radiation condition at infinity. It follows that ψ_μ^{sca} can be expressed as a sum of outgoing cylindrical waves in the region of space beyond the outermost point on S:

$$\psi_\mu^{\text{sca}}(\rho,\theta)=\sum_{n=-\infty}^{\infty}i^{\mu+jn}A_{\mu n}H_{\mu+jn}^{(1)}(\rho)e^{i(\mu+jn)(\theta-\theta_{\text{inc}})}. \tag{8}$$

Equation (8) is valid for $\rho\geqslant\max\{\rho_s(\theta)\}$, where $\rho_s(\theta)=kr_s(\theta)$; $H_n^{(1)}(\rho)$ is the nth order cylindrical Hankel function of the first kind,[28] which acquires the form of an outgoing cylindrical wave for $\rho\gg1$. The wave amplitudes $A_{\mu n}$ are the unknowns of the scattering problem; they will be determined in the following section.

It is important to note that excitation of the scattering surface by ψ_μ^{inc} only gives rise to ψ_μ^{sca}, which follows from the invariance of the Helmholtz equation and of the boundary condition under rotations by multiples of $2\pi/j$ about an axis normal to S. Hence, it is convenient to treat the scattering problem on a term-by-term basis.

IV. DETERMINATION OF THE SCATTERED WAVE

A. Exact solution

The Helmholtz equation can be integrated by application of Green's second identity.[29] Thus, it may be proven that the μth component of the scalar potential of the scattered wave is given by the following integral

$$\psi_\mu^{\text{sca}}(\rho,\theta)=-\int_0^{2\pi}H_0^{(1)}(|\vec{\rho}-\vec{\rho}_s(\theta')|)\varphi_\mu(\theta')d\theta', \tag{9}$$

where

$$\varphi_\mu(\theta)=\frac{1}{4i}\left[\frac{\partial\psi_\mu}{\partial n}\frac{ds}{d\theta}\right]_{\rho=\rho_s(\theta)}, \tag{10}$$

$\psi_\mu=\psi_\mu^{\text{inc}}+\psi_\mu^{\text{sca}}$, and $\partial/\partial n$ is the derivative along the outward normal to S. Equation (9) holds outside the scattering surface, namely for $\rho>\rho_s(\theta)$. If the Helmholtz equation is integrated inside the scattering surface, the following integral equation is obtained:

$$\psi_\mu^{\text{inc}}(\rho,\theta)=\int_0^{2\pi}H_0^{(1)}(|\vec{\rho}-\vec{\rho}_s(\theta')|)\varphi_\mu(\theta')d\theta'. \tag{11}$$

Equation (11) holds for $\rho<\rho_s(\theta)$ and it is the mathematical expression of the extinction theorem.

The wave amplitudes $A_{\mu n}$ can be determined from Eqs. (9) and (11) in two steps. First, $\varphi_\mu(\theta)$ is determined by

application of Eq. (11) for $\rho \leqslant \rho_{\min} \equiv \rho_s(\theta_\nu^-)$ and, subsequently, $A_{\mu n}$ are determined by application of Eq. (9) for $\rho \geqslant \rho_{\max} \equiv \rho_s(\theta_\nu^+)$.

On the one hand, handling of Eq. (11) requires use of the identity

$$H_0^{(1)}(|\vec{\rho}-\vec{\rho}\,'|)=\sum_{n=-\infty}^{\infty} e^{in(\theta-\theta')}J_n(\rho)H_n^{(1)}(\rho'), \qquad (12)$$

which holds for $\rho \leqslant \rho'$. Substitution of Eqs. (6) and (12) into the left- and right-hand sides of Eq. (11), respectively, and use of the orthogonality of $\{e^{in\theta}, n \in Z\}$ yields:

$$i^{\mu+jn}e^{-i(\mu+jn)\theta_{\rm inc}}=\int_0^{2\pi} e^{-i(\mu+jn)\theta}H_{\mu+jn}^{(1)}[\rho_s(\theta)]\varphi_\mu(\theta)d\theta. \qquad (13)$$

Since $\phi_\mu(\theta) \in V_\mu$, it can be expanded into a Fourier series

$$\phi_\mu(\theta)=\sum_{n=-\infty}^{\infty} F_{\mu n}e^{i(\mu+jn)\theta}, \qquad (14)$$

by use of which Eq. (13) yields the following set of infinite linear equations:

$$i^{\mu+jn}e^{-i(\mu+jn)\theta_{\rm inc}}=\sum_{\nu=-\infty}^{\infty} F_{\mu\nu}\mathbf{H}_{\mu+jn,\mu+j\nu}, \qquad (15)$$

where:

$$\mathbf{H}_{n,\nu}=\int_0^{2\pi} e^{i(\nu-n)\theta}H_n^{(1)}[\rho_0(1+\varepsilon\sin(j\theta))]d\theta \qquad (16)$$

and $\rho_0 = k\alpha$. Equation (15) can be solved for $F_{\mu\nu}$ by truncation and matrix-inversion.

On the other hand, handling of Eq. (9) requires use of the identity

$$H_0^{(1)}(|\vec{\rho}-\vec{\rho}\,'|)=\sum_{n=-\infty}^{\infty} e^{in(\theta-\theta')}H_n^{(1)}(\rho)J_n(\rho'), \qquad (17)$$

which holds for $\rho \geqslant \rho'$. Substitution of Eqs. (8) and (17) into the left- and right-hand sides of Eq. (9), respectively, and use of the orthogonality of $\{e^{in\theta}, n \in Z\}$ yields:

$$i^{\mu+jn}e^{-i(\mu+jn)\theta_{\rm inc}}A_{\mu n} = -\int_0^{2\pi} e^{-i(\mu+jn)\theta}J_{\mu+jn}[\rho_s(\theta)]\varphi_\mu(\theta)d\theta. \qquad (18)$$

Finally, by use of Eq. (14), Eq. (18) produces the following result:

$$i^{\mu+jn}e^{-i(\mu+jn)\theta_{\rm inc}}A_{\mu n}=-\sum_{\nu=-\infty}^{\infty} F_{\mu\nu}\mathbf{J}_{\mu+jn,\mu+j\nu}, \qquad (19)$$

where:

$$\mathbf{J}_{n,\nu}=\int_0^{2\pi} e^{i(\nu-n)\theta}J_n[\rho_0(1+\varepsilon\sin(j\theta))]d\theta. \qquad (20)$$

The wave amplitudes $A_{\mu n}$ can be determined from Eq. (19).

B. Extinction-theorem perturbative solution

The extinction-theorem perturbative formulation makes use of the Helmholtz integral formulas of Eqs. (9) and (11). Both $A_{\mu n}$ and $\phi_\mu(\theta)$ are expressed as asymptotic power series in the smallness parameter ε

$$A_{\mu n}=\sum_{p=0}^{\infty} \varepsilon^p A_{\mu n}^p, \qquad (21)$$

$$\phi_\mu(\theta)=\sum_{p=0}^{\infty} \varepsilon^p \phi_\mu^p(\theta)=\sum_{p=0}^{\infty} \varepsilon^p \sum_{n=-\infty}^{\infty} F_{\mu n}^p e^{i(\mu+jn)\theta}, \qquad (22)$$

and $A_{\mu n}^p$ are obtained by a two-step procedure similar to that of Sec. IV A.

The first step yields $F_{\mu n}^p$. By use of Eq. (13), Eq. (22), and of the Taylor series

$$H_{\mu+jn}^{(1)}[\rho_s(\theta)]=\sum_{p=0}^{\infty} \varepsilon^p \frac{[\rho_0 f(\theta)]^p}{p!}\tilde{H}_{\mu+jn}^p, \qquad (23)$$

the following equation is obtained:

$$i^{\mu+jn}e^{-i(\mu+jn)\theta_{\rm inc}} = \sum_{p=0}^{\infty} \varepsilon^p \frac{\rho_0^p}{p!}\tilde{H}_{\mu+jn}^p \sum_{\nu=-\infty}^{\infty} I_{j(\nu-n)}^p \sum_{q=0}^{\infty} \varepsilon^q F_{\mu\nu}^q. \qquad (24)$$

The abbreviations $\tilde{H}_\nu^p = [d^p H_\nu^{(1)}(\rho)/d\rho^p]_{\rho=\rho_0}$ and $I_\nu^p = \int_0^{2\pi} f^p(\theta)e^{i\nu\theta}d\theta$ appear in Eq. (24). It can be proven that $I_\nu^p = 2\pi(2i)^{-p}\Sigma_{q=0}^p(-1)^{p-q}\binom{p}{q}\delta_{-\nu,j(2q-p)}$, where $\binom{p}{q} = p!/q!(p-q)!$ is the binomial coefficient. By use of the identity

$$\sum_{p=0}^{\infty} \varepsilon^p A_p \sum_{q=0}^{\infty} \varepsilon^q B_q = \sum_{p=0}^{\infty} \varepsilon^p \sum_{q=0}^{p} A_q B_{p-q}, \qquad (25)$$

Eq. (24) acquires the following form:

$$i^{\mu+jn}e^{-i(\mu+jn)\theta_{\rm inc}} = \sum_{p=0}^{\infty} \varepsilon^p \sum_{q=0}^{p} \frac{\rho_0^q}{q!}\tilde{H}_{\mu+jn}^q \sum_{\nu=-\infty}^{q} I_{j(\nu-n)}^q F_{\mu\nu}^{p-q}. \qquad (26)$$

Coefficients of the same power of ε on either side of Eq. (26) can be equated. Thus, Eq. (26) yields

$$F_{\mu n}^0 = \frac{i^{\mu+jn}e^{-i(\mu+jn)\theta_{\rm inc}}}{2\pi\tilde{H}_{\mu+jn}^0}, \qquad (27)$$

if $p=0$, and

$$F_{\mu n}^p = -\frac{1}{2\pi\tilde{H}_{\mu+jn}^0}\sum_{q=1}^{p} \frac{\rho_0^q}{q!}\tilde{H}_{\mu+jn}^q \sum_{\nu=-\infty}^{\infty} I_{j(\nu-n)}^q F_{\mu\nu}^{p-q} \qquad (28)$$

if $p \geqslant 1$. The right-hand side of Eq. (28) involves known quantities and the coefficients $F_{\mu n}^{p'}$ of order $0 \leqslant p' \leqslant p-1$.

The second step yields $A_{\mu n}^p$ in terms of $F_{\mu n}^p$. By use of Eq. (21), Eq. (22), and of the Taylor series

$$J_{\mu+jn}[\rho_s(\theta)]=\sum_{p=0}^{\infty} \varepsilon^p \frac{[\rho_0 f(\theta)]^p}{p!}\tilde{J}_{\mu+jn}^p, \qquad (29)$$

Eq. (18) yields

$$i^{\mu+jn}e^{-i(\mu+jn)\theta_{\rm inc}}\sum_{p=0}^{\infty}\varepsilon^p A_{\mu n}^p = -\sum_{p=0}^{\infty}\varepsilon^p\frac{\rho_0^p}{p!}\tilde{J}_{\mu+jn}^p\sum_{\nu=-\infty}^{\infty}I_{j(\nu-n)}^p\sum_{q=0}^{p}\varepsilon^p F_{\mu\nu}^p, \tag{30}$$

wherein the abbreviation $\tilde{J}_\nu^p=[d^pJ_\nu(\rho)/d\rho^p]_{\rho=\rho_0}$ appears. Finally, by use of Eq. (25) and after coefficients of the same power of ε on either side of Eq. (30) have been equated, the following expression for $A_{\mu n}^p$ is obtained:

$$A_{\mu n}^p = -i^{-(\mu+jn)}e^{i(\mu+jn)\theta_{\rm inc}} \times\sum_{q=0}^{p}\frac{\rho_0^q}{q!}\tilde{J}_{\mu+jn}^q\sum_{\nu=-\infty}^{\infty}I_{j(\nu-n)}^q F_{\mu\nu}^{p-q}. \tag{31}$$

C. Rayleigh-theory perturbative solution

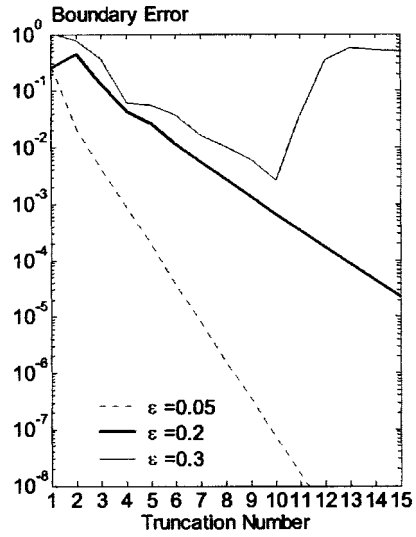
The Rayleigh-theory perturbative formulation makes use of the Rayleigh hypothesis, which allows use of Eq. (8), whereby the scattered field is represented by a sum of outgoing waves, not only beyond the outermost point on S, but everywhere outside S, even at the innermost point on S. The scattered wave amplitudes $A_{\mu n}$ are expressed as in Eq. (21) and the coefficients $A_{\mu n}^p$ of that asymptotic power series in the smallness parameter ε are obtained by enforcement of the boundary condition $[\psi_\mu^{\rm inc}+\psi_\mu^{\rm sca}]_{\rho=\rho_S(\theta)}=0$. The latter is expanded into a Taylor series around $\rho_0=ka$ through use of Eqs. (6), (8), (21), (23), and (29):

$$\sum_{n=-\infty}^{\infty}i^{\mu+jn}e^{i(\mu+jn)(\theta-\theta_{\rm inc})}\left[\sum_{p=0}^{\infty}\varepsilon^p\frac{[\rho_0 f(\theta)]^p}{p!}\tilde{J}_{\mu+jn}^p + \sum_{q=0}^{\infty}\varepsilon^q A_{\mu n}^q\sum_{p=0}^{\infty}\varepsilon^p\frac{[\rho_0 f(\theta)]^p}{p!}\tilde{H}_{\mu+jn}^p\right]=0. \tag{32}$$

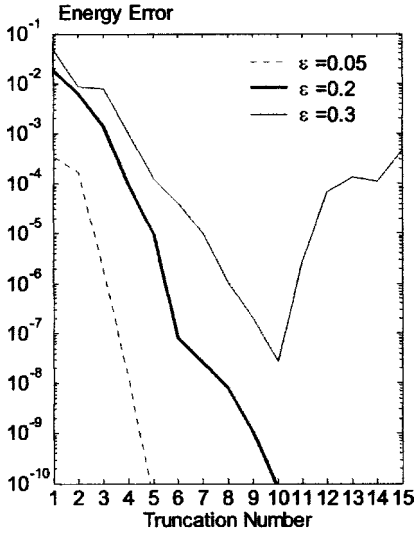
The identity of Eq. (25) is invoked and, subsequently, coefficients of the same power of ε on either side of Eq. (32) are equated to obtain:

$$\sum_{n=-\infty}^{\infty}i^{\mu+jn}e^{i(\mu+jn)(\theta-\theta_{\rm inc})}A_{\mu n}^p\tilde{H}_{\mu+jn}^0 = -\sum_{n=-\infty}^{\infty}i^{\mu+jn}e^{i(\mu+jn)(\theta-\theta_{\rm inc})} \times\left[\frac{[\rho_0 f(\theta)]^p}{p!}\tilde{J}_{\mu+jn}^p+\sum_{q=1}^{p}\frac{[\rho_0 f(\theta)]^q}{q!}A_{\mu n}^{p-q}\tilde{H}_{\mu+jn}^q\right]. \tag{33}$$

Both sides of Eq. (33) are multiplied by $e^{-i(\mu+j\nu)\theta}$, $\nu\in Z$, and they are subsequently integrated with respect to θ in the range $[0,2\pi]$. The end-result is the recurrence relation



(a)



(b)

FIG. 3. Accuracy of the exact solution ($a=1, j=5, \theta_{\rm inc}=0°$) vs truncation number N: (a) boundary error B (%) and (b) energy error E (%).

$$A_{\mu\nu}^p = -\frac{i^{-(\mu+j\nu)}e^{i(\mu+j\nu)\theta_{\rm inc}}}{2\pi\tilde{H}_{\mu+j\nu}^0}\sum_{n=-\infty}^{\infty}i^{\mu+jn}e^{-i(\mu+jn)\theta_{\rm inc}} \times\left[\frac{\rho_0^p}{p!}I_{j(n-\nu)}^p\tilde{J}_{\mu+jn}^p+\sum_{q=1}^{p}\frac{\rho_0^q}{q!}I_{j(n-\nu)}^q A_{\mu n}^{p-q}\tilde{H}_{\mu+jn}^q\right], \tag{34}$$

which is initiated with $A_{\mu\nu}^0=-\tilde{J}_{\mu+j\nu}^0/\tilde{H}_{\mu+j\nu}^0$.

V. NUMERICAL RESULTS AND DISCUSSION

A. Numerical checks of the exact solution

Use of the exact solution of Sec. IV A requires truncation and matrix-inversion of the infinite set of linear equations represented by Eq. (15); because of eventual ill-conditioning of the coefficient matrix of that set, the exact

solution may be numerically unstable. Instability may occur for large values of the smallness parameter ε, which are associated with deep corrugations on S. Full discussion of this issue, which has been addressed thoroughly within the context of wave scattering from planar sinusoidal surfaces,[29–31] is beyond the scope of this paper, where the focus is on the perturbative solutions of Secs. IV B and IV C. Still, some numerical checks of the exact solution have been made by use of the error measures defined below.

Conformity of the exact solution to the (Dirichlet) boundary condition on the scattering surface is checked by use of the boundary error

$$B=\left|\left[\frac{\psi^{\mathrm{sca}}+\psi^{\mathrm{inc}}}{\psi^{\mathrm{inc}}}\right]_{\vec{r}=\vec{r}_S(\theta_0^+)}\right|, \tag{35}$$

which is measured at a protrusion of S. It is worth noting that the outgoing-wave expansion of Eq. (8) for the scalar potential of the scattered wave is valid unconditionally at $r_S(\theta_0^+)=r_S(\pi/2j)$. The scalar potential of the incident wave at $r_S(\theta_0^+)$ is calculated by use of the exact, closed-form expression $[\psi^{\mathrm{inc}}]_{\vec{r}=\vec{r}_S(\theta_0^+)}=e^{i\rho_{\max}\cos(\theta_0^+-\theta_{\mathrm{inc}})}$, whereas the scalar potential of the scattered wave at the very same point on S is calculated by use of the truncated, hence approximate, expansion:

$$[\psi^{\mathrm{sca}}]_{\vec{r}=\vec{r}_S(\theta_0^+)}=\sum_{\mu-1}^{j}\sum_{n=-N}^{N} i^{\mu+jn}A_{\mu n}H^{(1)}_{\mu+jn}(\rho_{\max}) \times e^{i(\mu+jn)(\theta_0^+-\theta_{\mathrm{inc}})}. \tag{36}$$

The truncation number N is the number of equations that are retained in the numerical solution of Eq. (15).

Conformity of the exact solution with the forward-scattering theorem, can be checked by use of the following equation:[6]

$$4\sum_{\mu=1}^{j}\sum_{n=-N}^{N}|A_{\mu n}|^2=-4\sum_{\mu=1}^{j}\sum_{n=-N}^{N}\mathrm{Re}\,A_{\mu n}. \tag{37}$$

If *LHS* and *RHS* is short-hand notation for the left- and right-hand side of Eq. (37), respectively, the energy error

$$E=\left|\frac{RHS-LHS}{RHS}\right| \tag{38}$$

provides a means to assess the energy consistency of the exact solution.

Numerical results for B and E are presented in Fig. 3. Both error measures have been plotted against the truncation number N for several values of the smallness parameter ε. Figure 3 manifests that the numerical behavior of the exact solution is strongly dependent on ε. Small values of ε ensure that the exact solution is well behaved, which implies that the error, either boundary or energy, decreases with increasing N. This is not true for large values of ε, since the aforesaid errors decrease with increasing N only up to a particular value of the latter; if N is increased further, the exact solution becomes numerically unstable and both errors increase rapidly. Figure 3 suggests that the exact solution of Sec. IV A should be used with caution. Hereinafter, only numerical results from the exact solution that meet the, more or less arbitrary, requirements $B\leqslant 10^{-4}$ and $E\leqslant 10^{-5}$ are used.

B. Comparisons between perturbative solutions

The extinction-theorem and Rayleigh-theory perturbative solutions, which were derived in Secs. IV B and IV C, respectively, can be compared analytically as well as numerically. Analytical comparisons of $A^p_{\mu n}$, as obtained by the aforesaid perturbative formulations, are herein limited to $p\leqslant 3$; the expressions for the wave amplitudes are cumbersome for higher values of p.

On the one hand, the wave amplitudes that pertain to the extinction-theorem perturbative formulation are designated as ${}_XA^p_{\mu n}$, and the expressions given below have been obtained by use of Eqs. (27), (28), and (31):

$${}_XA^0_{\mu n}=-\frac{\tilde{J}^0_{\mu+jn}}{\tilde{H}^0_{\mu+jn}}, \tag{39a}$$

$${}_XA^1_{\mu n}=-\frac{\rho_0}{2\pi}\sum_{\nu=-\infty}^{\infty} i^{j(\nu-n)}e^{-ij(\nu-n)\theta_{\mathrm{inc}}}I^1_{j(\nu-n)}\frac{\tilde{W}^1_{\mu+jn}}{\tilde{H}^0_{\mu+j\nu}}, \tag{39b}$$

$${}_XA^2_{\mu n}=-\frac{\rho_0^2}{4\pi}\sum_{\nu=-\infty}^{\infty} i^{j(\nu-n)}e^{-ij(\nu-n)\theta_{\mathrm{inc}}}I^2_{j(\nu-n)}\frac{\tilde{W}^2_{\mu+jn}}{\tilde{H}^0_{\mu+j\nu}} + \frac{\rho_0^2}{4\pi^2}\sum_{\nu=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} i^{j(l-n)}e^{-ij(l-n)\theta_{\mathrm{inc}}} \times I^1_{j(l-\nu)}I^1_{j(\nu-n)}\frac{\tilde{W}^1_{\mu+jn}}{\tilde{H}^0_{\mu+jl}}\frac{\tilde{H}^1_{\mu+j\nu}}{\tilde{H}^0_{\mu+j\nu}}, \tag{39c}$$

$${}_XA^3_{\mu n}=-\frac{\rho_0^3}{12\pi}\sum_{\nu=-\infty}^{\infty} i^{j(\nu-n)}e^{-ij(\nu-n)\theta_{\mathrm{inc}}}I^3_{j(\nu-n)}\frac{\tilde{W}^3_{\mu+jn}}{\tilde{H}^0_{\mu+j\nu}} + \frac{\rho_0^3}{8\pi^2}\sum_{\nu=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} i^{j(l-n)}e^{-ij(l-n)\theta_{\mathrm{inc}}}I^1_{j(l-\nu)}I^2_{j(\nu-n)}\frac{\tilde{W}^2_{\mu+jn}}{\tilde{H}^0_{\mu+jl}}\frac{\tilde{H}^1_{\mu+j\nu}}{\tilde{H}^0_{\mu+j\nu}} + \frac{\rho_0^3}{8\pi^2}\sum_{\nu=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} i^{j(l-n)}e^{-ij(l-n)\theta_{\mathrm{inc}}}I^2_{j(l-\nu)}I^1_{j(\nu-n)}\frac{\tilde{W}^1_{\mu+jn}}{\tilde{H}^0_{\mu+jl}}\frac{\tilde{H}^2_{\mu+j\nu}}{\tilde{H}^0_{\mu+j\nu}} - \frac{\rho_0^3}{8\pi^3}\sum_{\nu=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\sum_{s=-\infty}^{\infty} i^{j(s-n)}e^{-ij(s-n)\theta_{\mathrm{inc}}}I^1_{j(s-l)}I^1_{j(l-\nu)}I^1_{j(\nu-n)}\frac{\tilde{W}^1_{\mu+jn}}{\tilde{H}^0_{\mu+js}}\frac{\tilde{H}^1_{\mu+jl}}{\tilde{H}^0_{\mu+jl}}\frac{\tilde{H}^1_{\mu+j\nu}}{\tilde{H}^0_{\mu+j\nu}}. \tag{39d}$$

The abbreviation $\tilde{W}^p_\nu = \tilde{J}^p_\nu + {}_X A^0_{\mu n} \tilde{H}^p_\nu$ has been used in Eqs. (39).

On the other hand, the wave amplitudes that pertain to the Rayleigh-theory perturbative formulation are designated as ${}_R A^p_{\mu n}$, and the expressions given below have been obtained by use of Eq. (34):

$$ {}_R A^0_{\mu n} = -\frac{\tilde{J}^0_{\mu+jn}}{\tilde{H}^0_{\mu+jn}}, \tag{40a} $$

$$ {}_R A^1_{\mu n} = -\frac{\rho_0}{2\pi} \sum_{\nu=-\infty}^{\infty} i^{j(\nu-n)} e^{-ij(\nu-n)\theta_{\rm inc}} I^1_{j(\nu-n)} \frac{\tilde{W}^1_{\mu+j\nu}}{\tilde{H}^0_{\mu+jn}}, \tag{40b} $$

$$ {}_R A^2_{\mu n} = -\frac{\rho_0^2}{4\pi} \sum_{\nu=-\infty}^{\infty} i^{j(\nu-n)} e^{-ij(\nu-n)\theta_{\rm inc}} I^2_{j(\nu-n)} \frac{\tilde{W}^2_{\mu+j\nu}}{\tilde{H}^0_{\mu+jn}} + \frac{\rho_0^2}{8\pi^2} \sum_{\nu=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} i^{j(l-n)} e^{-ij(l-n)\theta_{\rm inc}} I^1_{j(l-\nu)} I^1_{j(\nu-n)} \frac{\tilde{W}^1_{\mu+jl}}{\tilde{H}^0_{\mu+jn}} \frac{\tilde{H}^1_{\mu+j\nu}}{\tilde{H}^0_{\mu+j\nu}}, \tag{40c} $$

$$ \begin{aligned} {}_R A^3_{\mu n} = & -\frac{\rho_0^3}{12\pi} \sum_{\nu=-\infty}^{\infty} i^{j(\nu-n)} e^{-ij(\nu-n)\theta_{\rm inc}} I^3_{j(\nu-n)} \frac{\tilde{W}^3_{\mu+j\nu}}{\tilde{H}^0_{\mu+jn}} + \frac{\rho_0^3}{8\pi^2} \sum_{\nu=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} i^{j(l-n)} e^{-ij(l-n)\theta_{\rm inc}} I^2_{j(l-\nu)} I^1_{j(\nu-n)} \frac{\tilde{W}^2_{\mu+jl}}{\tilde{H}^0_{\mu+jn}} \frac{\tilde{H}^1_{\mu+j\nu}}{\tilde{H}^0_{\mu+j\nu}} \\ & + \frac{\rho_0^3}{8\pi^2} \sum_{\nu=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} i^{j(l-n)} e^{-ij(l-n)\theta_{\rm inc}} I^1_{j(l-\nu)} I^2_{j(\nu-n)} \frac{\tilde{W}^1_{\mu+jl}}{\tilde{H}^0_{\mu+jn}} \frac{\tilde{H}^2_{\mu+j\nu}}{\tilde{H}^0_{\mu+jn}} \\ & - \frac{\rho_0^3}{8\pi^3} \sum_{\nu=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} i^{j(s-n)} e^{-ij(s-n)\theta_{\rm inc}} I^1_{j(s-l)} I^1_{j(l-\nu)} I^1_{j(\nu-n)} \frac{\tilde{W}^1_{\mu+js}}{\tilde{H}^0_{\mu+jn}} \frac{\tilde{H}^1_{\mu+jl}}{\tilde{H}^0_{\mu+jl}} \frac{\tilde{H}^1_{\mu+j\nu}}{\tilde{H}^0_{\mu+j\nu}}. \end{aligned} \tag{40d} $$

Equations (39a) and (40a) are identical, but, at first glance, Eqs. (39b), (39c), and (39d) are not identical to Eqs. (40b), (40c), and (40d), respectively; for instance, Eq. (39b) involves the term $\tilde{W}^1_{\mu+jn}/\tilde{H}^0_{\mu+j\nu}$, whereas Eq. (40b) involves the term $\tilde{W}^1_{\mu+j\nu}/\tilde{H}^0_{\mu+jn}$. Nevertheless, it can be proven[28] that $\tilde{W}^1_\nu = -2i/\pi\rho_0\tilde{H}^0_\nu$ and $\tilde{W}^2_\nu = 2i/\pi\rho_0^2\tilde{H}^0_\nu$; by use of these identities it can readily be proven that Eqs. (39b) and (39c) are equivalent to Eqs. (40b) and (40c), respectively. The equivalence of Eq. (39d) to Eq. (40d) has been proven, too, albeit through more complicated, yet straightforward, algebra.

Higher-order expressions for ${}_X A^p_{\mu n}$ and ${}_R A^p_{\mu n}$ have been compared numerically and they have been found to be identical in all cases studied. This finding may seem paradoxical at first glance; the perturbative formulations of this paper use different starting points, namely the extinction-theorem and the Rayleigh hypothesis, the latter being applicable[32] to a narrower range of scattering surfaces than the former. Nonetheless, this finding is not unique to the scattering problem studied herein; similar conclusions have been stated[20,21] in comparisons of extinction-theorem and Rayleigh-theory solutions to wave scattering from planar sinusoidal surfaces.

C. Comparison of second- to higher-order perturbative solutions

The merit of perturbative solutions of various orders is assessed through comparisons with the exact solution of Sec. IV A. Interest is focused on the accuracy of the perturbative solutions in the far field, which is herein measured by use of the far-field error:

$$ F = \frac{1}{2\pi} \int_0^{2\pi} |F(\theta)| d\theta; $$

$$ F(\theta) = \lim_{\rho\to\infty} \frac{\tilde{\psi}^{\rm sca}(\rho,\theta) - \psi^{\rm sca}(\rho,\theta)}{\psi^{\rm sca}(\rho,\theta)}. \tag{41} $$

$\psi^{\rm sca}$ is the scalar potential of the scattered wave as obtained by use of the exact solution of Sec. IV A, whereas

$$ \tilde{\psi}^{\rm sca}(\rho,\theta) = \sum_{\mu=1}^{j} \sum_{n=-N}^{N} \sum_{p=0}^{P} i^{\mu+jn} \varepsilon^p A^p_{\mu n} H^{(1)}_{\mu+jn}(\rho) \times e^{i(\mu+jn)(\theta-\theta_{\rm inc})} \tag{42} $$

is an approximation to $\psi^{\rm sca}$ obtained by use of a Pth-order perturbative solution, either the one of Sec. IV B or that of Sec. IV C.

The dependence of F on the smallness parameter ε as well as on the order P of the perturbative solution is investigated through Fig. 4. Three plausible conclusions are reached in the case of a scattering surface with gently sloping corrugations: (a) F increases with increasing ε for any value of P, (b) the perturbative solution is more accurate the higher the order P, and (c) inclusion of higher-order terms in the perturbation series for the scalar potential of the scattered wave reduces the far-field error F by a factor that becomes more significant with increasing ε. A somewhat unexpected result occurs in the case of a scattering surface with steep corrugations: the perturbation series behaves as in the previous case only up to $\varepsilon = 0.28$ and, from that point onward, the error of the sixth- and eighth-order solutions increase rapidly, exceeding the error of second- and fourth-order solutions.

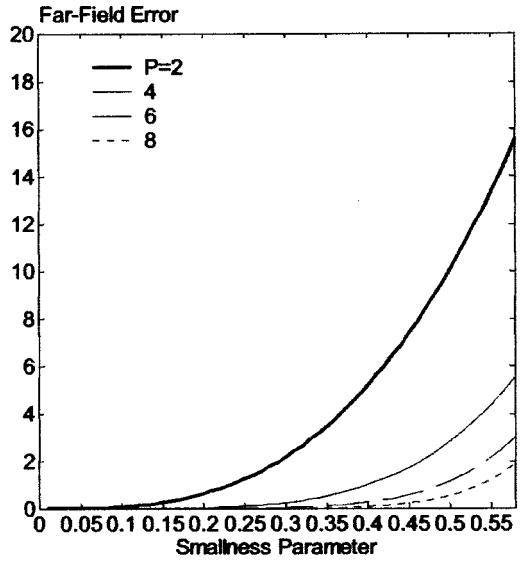


FIG. 4. Far-field error F (%) vs smallness parameter ε of Pth-order perturbative solutions: (a) gently sloping ($j=2$) and (b) steep ($j=5$) corrugations.

Furthermore, F increases more rapidly with increasing ε in Fig. 4(b) than in the case of Fig. 4(a). These remarks suggest that the accuracy as well as the convergence of perturbative solutions depend not only on the height of corrugations, which is determined by the smallness parameter ε, but also on the slope of corrugations, which is determined in this paper by the order j of the sinusoidal roughness on S.

Further information about the effect of the shape of corrugations is obtained by use of Fig. 5, which depicts regions of the $j-\varepsilon$ plane wherein the second- and fourth-order perturbative solutions meet the requirement $F\leqslant 1\%$. It may be seen that the range of values of the smallness parameter ε that can be handled by the second-order perturbative solution decreases with increasing j; the rate of decrease changes

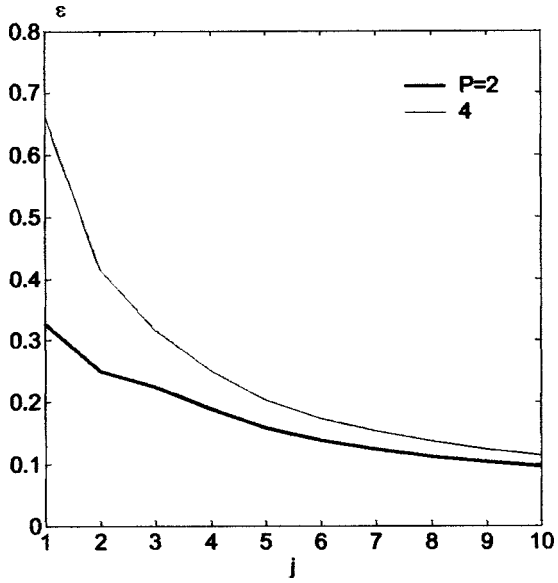


FIG. 5. Applicability limit of second- and fourth-order solutions on $j-\varepsilon$ plane ($a=1,\theta_{\mathrm{inc}}=0°$); the limit corresponds to $F\leqslant 1\%$.

from fast to slow from the leftmost to the rightmost part of the plot. The same apply to the fourth-order perturbative solution, which has a wider range of values of ε. Yet, the advantage of the fourth-order solution over the second-order one becomes almost insignificant for large values of j, which agrees with the comments of Fig. 4 that higher-order solutions are more useful for scattering surfaces with gently sloping corrugations.

D. Error measures for perturbative solutions

Up to this point, the investigation has benefited from the existence of an exact solution to the scattering problem in hand. Nonetheless, the method of small perturbations is most commonly applied to problems of rough surface scattering where an exact solution is not available. It is, therefore, of interest to pursue the investigation of this paper further by introduction of error measures that can be applied to perturbative solutions in the absence of an exact solution. The key idea is to examine whether the extinction-theorem or Rayleigh-theory perturbative solution *a posteriori* satisfies the boundary condition that was used in its formulation.

Conformity of the Rayleigh-theory perturbative solution to the (Dirichlet) boundary condition on S can be checked by use of the mean boundary error:

$$\overline{B}=\frac{1}{2\pi}\int_0^{2\pi}|B(\theta)|d\theta;B(\theta)=\left[\frac{\tilde{\psi}^{\mathrm{sca}}+\psi^{\mathrm{inc}}}{\psi^{\mathrm{inc}}}\right]_{r=r_S(\theta)}. \quad (43)$$

$\overline{B}$ is similar to the boundary error B, defined by Eq. (35), but (a) it is averaged over the scattering surface and (b) it involves $\tilde{\psi}^{\mathrm{sca}}$, defined by Eq. (42), instead of ψ^{sca}, which is seldom available.

The merit of $\overline{B}$ as error measure for Rayleigh-theory perturbative solutions can be assessed through comparisons of $\overline{B}$ to the far-field error F, defined in Sec. V C. The results of Fig. 6 expose some problems related to the use of $\overline{B}$. As ε

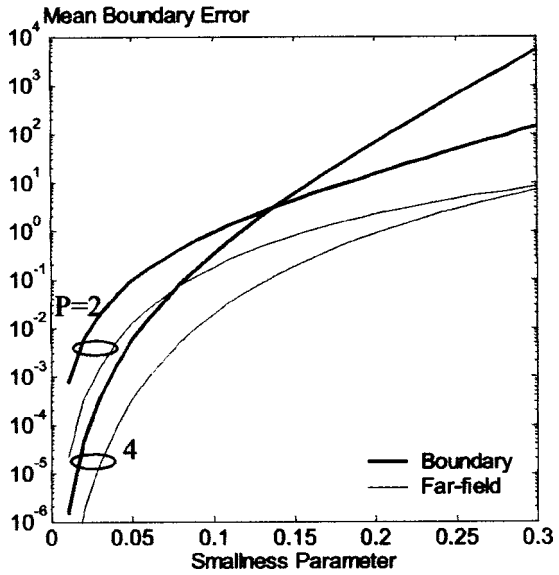


FIG. 6. Mean boundary error $\overline{B}$ (%) vs smallness parameter ε of second- and fourth-order Rayleigh-theory perturbative solutions ($a=1$, $j=5, \theta_{\text{inc}}=45°$); the far-field error F (%) serves as reference.

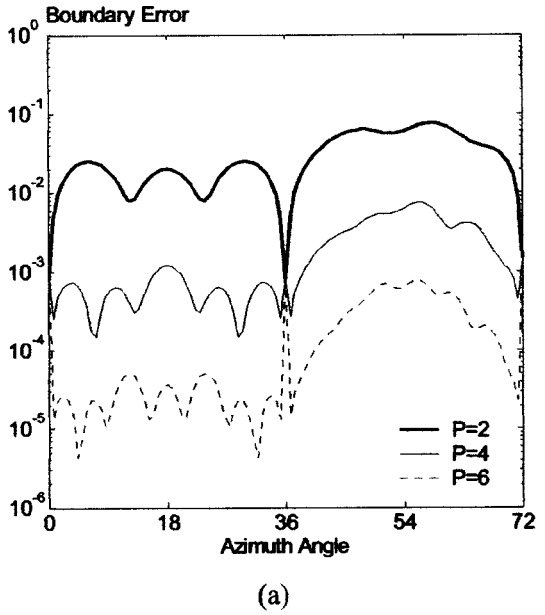
increases, $\overline{B}$ acquires unrealistic values; those of the fourth-order perturbative solution, in particular, can exceed the corresponding values of F by as much as three orders of magnitude. Moreover, although F decreases with increasing order of the perturbative solution for any ε, greater values of $\overline{B}$ have been calculated for the fourth-order perturbative solution than for the second-order one beyond $\varepsilon \approx 0.135$.

An explanation for the aforesaid trend can be found in Fig. 7, wherein plots of $|B(\theta)|$ vs θ are shown for perturbative solutions up to sixth order. A sinusoidal rough surface of order $j=5$ has been considered and results are given in the range $0 \leqslant \theta \leqslant 2\pi/j$, which represents an azimuthal period of surface roughness. If the corrugations of the scattering surface are shallow, as in Fig. 7(a), $|B(\theta)|$ decreases with increasing P everywhere on S. However, if the corrugations are deep, as in Fig. 7(b), higher-order perturbative solutions are more accurate than lower-order ones wherever $\rho_S(\theta) \geqslant \rho_0$, which occurs for $\theta \leqslant 36°$, but they are less accurate wherever $\rho_S(\theta) \leqslant \rho_0$, which occurs for $\theta \geqslant 36°$; the error of higher-order solutions is particularly large around the innermost point of the surface, which is at $\theta = 54°$. These results indicate that violations of the Rayleigh hypothesis may affect the behavior of the perturbation series on the scattering surface, even though they do not affect the behavior of the perturbation series in the far field, as discussed in Sec. V C. The mean boundary error is, therefore, not a reliable indicator of the far-field error.

Conformity of extinction-theorem perturbative solutions to the extinction theorem itself can be measured by use of the mean extinction error

$$\overline{X} = \frac{1}{2\pi} \int_0^{2\pi} |X(\theta)| d\theta, \quad X\theta = \left[\frac{\psi^{\text{inc}} - \tilde{\psi}^{\text{inc}}}{\psi^{\text{inc}}} \right]_{\rho=\rho_{\min}}, \quad (44)$$

where



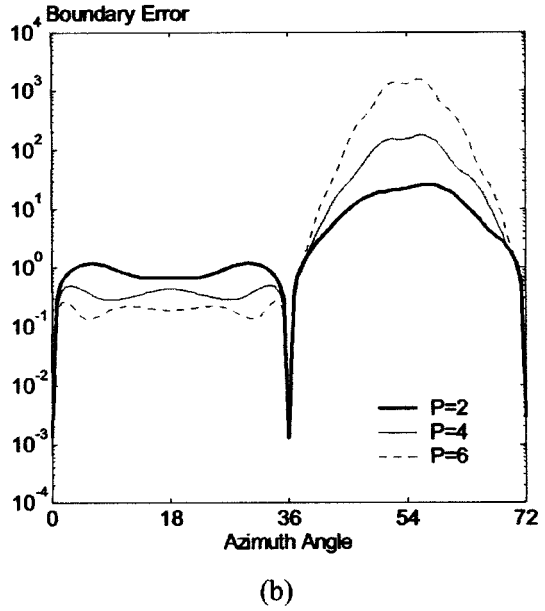
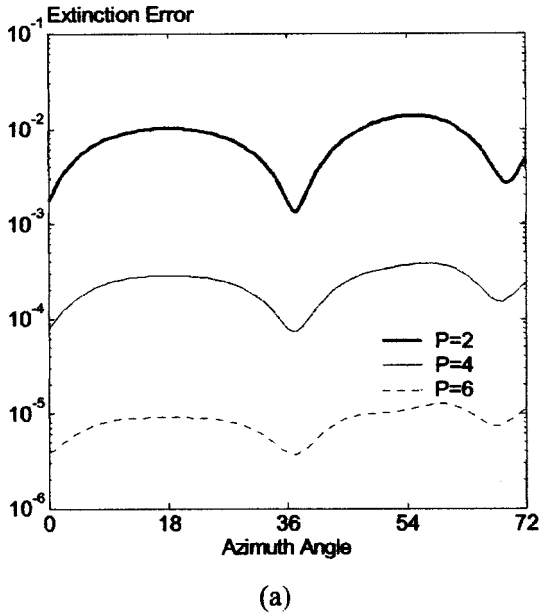


FIG. 7. Boundary error $|B(\theta)|$ vs azimuth angle θ (degrees) for Rayleigh-theory perturbative solutions ($a=1, j=5, \theta_{\text{inc}}=45°$) of various orders P: (a) shallow ($\varepsilon=0.05$) and (b) deep ($\varepsilon=0.2$) corrugations.

$$[\tilde{\psi}^{\text{inc}}]_{\rho=\rho_{\min}} = \sum_{p=0}^{P} \varepsilon^p \sum_{\mu=1}^{j} \sum_{n=-N}^{N} \sum_{\nu=-N}^{N} e^{i(\mu+jn)\theta} \times J_{\mu+jn}(\rho_{\min}) F^p_{\mu\nu} \mathbf{H}_{\mu+jn,\mu+j\nu} \quad (45)$$

is the extinction-theorem approximation to ψ^{inc} anywhere on the inscribed circle to S (i.e., at $\rho=\rho_{\min}$, $0 \leqslant \theta \leqslant 2\pi$); Eq. (45) has been obtained by use of Eqs. (5), (11), (12), (16), and (22). The truncation number N is determined numerically by checking the convergence of $\overline{X}$.

The merit of $\overline{X}$ and $|X(\theta)|$ as error measures for extinction-theorem perturbative solutions is probed by means



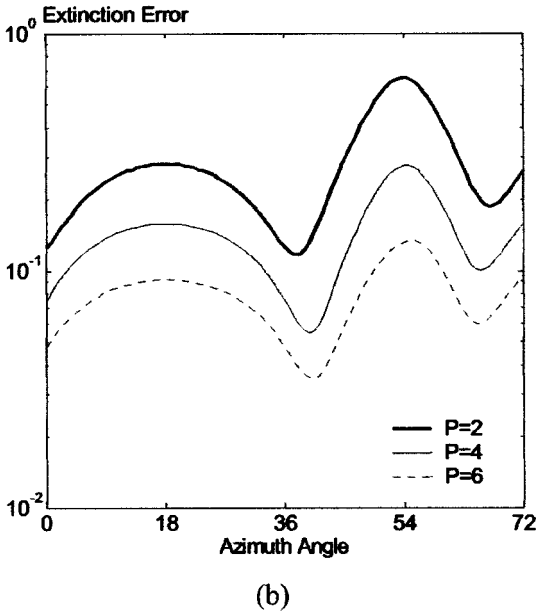


FIG. 8. Extinction error $|X(\theta)|$ (%) vs azimuth angle θ (degrees) of extinction-theorem perturbative solutions ($a=1, j=5, \theta_{\text{inc}}=45°$) of various orders P: (a) shallow ($\varepsilon=0.05$) and (b) deep ($\varepsilon=0.2$) corrugations.

of Figs. 8 and 9. Figure 8 includes plots of the extinction error $|X(\theta)|$ vs θ; the parameters of the scattering problem are those of Fig. 7. It may be seen that the extinction error is considerably smaller than the boundary error for any θ. Furthermore, higher-order extinction-theorem perturbative solutions exhibit less extinction error than lower-order ones for any θ, whereas such trend has not been observed for Rayleigh-theory perturbative solutions with regard to the boundary error. The dependence of $\overline{X}$ on the smallness parameter ε and the relation of $\overline{X}$ to the far-field error F are both shown in Fig. 9. Evidently, the mean extinction error is a reliable indicator of the error made by extinction-theorem solutions, but it underestimates the far-field error for large values of the smallness parameter.

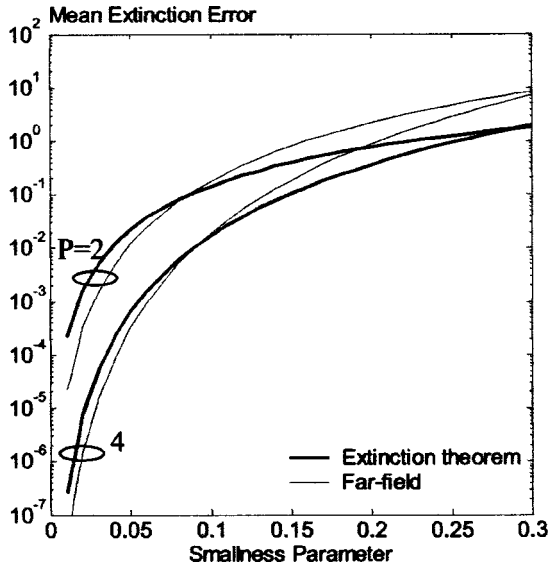


FIG. 9. Mean extinction error $\overline{X}$ (%) vs smallness parameter ε of second- and fourth-order extinction-theorem perturbative solutions ($a=1$, $j=5, \theta_{\text{inc}}=45°$); the far-field error F (%) serves as reference.

VI. CONCLUSIONS

The problem of wave scattering from a "soft" circular surface with sinusoidal roughness has been solved in three different ways. The solution of Sec. IV A is theoretically exact, whereas those of Secs. IV B and IV C make use of the small perturbation technique and they are based on the extinction theorem and the Rayleigh hypothesis, respectively.

The above solutions have been used in a numerical study of the accuracy of perturbative solutions to rough surface scattering. The results of Sec. V C indicate that the behavior of perturbative solutions depends on the height as well as on the slope of corrugations on the scattering surface. On the one hand, the error of the perturbative solutions increases monotonously with the height of surface corrugations. On the other hand, the convergence of the perturbative solutions depends critically on the slope of surface corrugations. Higher-order approximations are more accurate than lower-order ones when the corrugations are gently sloping, but they can diverge for steep corrugations. The usefulness of higher-order solutions is thus limited to surfaces with gentle slopes.

Furthermore, analytical as well as numerical comparisons of Rayleigh-theory to extinction-theorem perturbative solutions in Secs. V B and V C have shown that the two formulations yield identical results in the far field. Rayleigh-theory solutions have thus the advantage of being simpler in formulation. Nonetheless, the numerical investigation of Sec. V D has shown that Rayleigh-theory solutions cannot be associated with a reliable error estimate. On the contrary, extinction-theorem solutions can be validated by use of the extinction error, which correlates better with the far-field error.

The problem studied here is closely related to the one of wave scattering from rough planar surfaces, which has been studied extensively. It is, thus, worth noting that, whenever applicable, the conclusions drawn here are in qualitative agreement with those derived from the study of planar sur-

faces. It is also of interest that the shortcomings of Rayleigh-theory perturbative solutions when applied to points on the boundary surface, which were discussed in Refs. 18 and 19, are herein demonstrated quantitatively (see Fig. 7); such quantitative information is not, to the authors' knowledge, available elsewhere.

The results of the numerical investigation presented here can be of use to studies of wave scattering from randomly rough spheres or cylinders, in which case exact solutions are either unavailable or very demanding computationally. On the basis of the conclusions drawn above, it should be expected that higher-order solutions would be more appropriate for randomly rough surfaces with large correlation radii, that is, for surfaces with gently sloping roughness. Surfaces with highly localized roughness cannot be treated unless the root-mean-square height of surface roughness is very small; this limitation might be removed by reformulating the perturbative solution so that the slope of the roughness is also taken into consideration. Finally, an estimate of the error of the perturbative solution in the far field can be obtained by use of the mean extinction error.

[1] C. Yeh, ''Perturbation method in the diffraction of electromagnetic waves by arbitrarily shaped penetrable obstacles,'' J. Math. Phys. **6**, 2008–2013 (1965).

[2] V. Erma, ''An exact solution for the scattering of electromagnetic waves from conductors of arbitrary shape. I. Case of cylindrical symmetry,'' Phys. Rev. **173**, 1243–1257 (1968).

[3] C. Eftimiu, ''Electromagnetic scattering by rough conducting circular cylinders-I: Angular corrugation,'' IEEE Trans. Antennas Propag. **AP-36**, 651–658 (1988).

[4] C. Eftimiu, ''Electromagnetic scattering by rough conducting circular cylinders-II: Axial corrugation,'' IEEE Trans. Antennas Propag. **AP-36**, 659–663 (1988).

[5] C. Eftimiu, ''Electromagnetic scattering by rough conducting circular cylinders,'' Radio Sci. **23**, 1584–1594 (1989).

[6] N. Skaropoulos and D. Chrissoulidis, ''General perturbative solution to wave scattering from a soft random cylindrical surface,'' J. Acoust. Soc. Am. **106**, 596–604 (1999).

[7] V. Erma, ''Exact solution for the scattering of electromagnetic waves from conductors of arbitrary shape. II. General case,'' Phys. Rev. **176**, 1544–1553 (1968).

[8] V. Erma, ''Exact solution for the scattering of electromagnetic waves from bodies of arbitrary shape. III. Obstacles with arbitrary electromagnetic properties,'' Phys. Rev. **179**, 1238–1246 (1968).

[9] R. Schiffer, ''Light scattering by perfectly conducting statistically irregular particles,'' J. Opt. Soc. Am. A **6**, 385–402 (1989).

[10] R. Schiffer, ''Perturbation approach for light scattering by an ensemble of irregular particles of arbitrary material,'' Appl. Opt. **29**, 1536–1550 (1990).

[11] P. Jansson, ''Acoustic scattering from a rough sphere,'' J. Acoust. Soc. Am. **93**, 3032–3042 (1993).

[12] G. Farias, E. Vasconcelos, S. Cesar, and A. Maradudin, ''Mie scattering by a perfectly conducting sphere with a rough surface,'' Physica A **207**, 315–322 (1994).

[13] K. Muinonen, ''Light scattering by Gaussian random particles,'' Earth, Moon, Planets **72**, 339–342 (1996).

[14] K. Muinonen, ''Light scattering by stochastically shaped particles,'' in *Light Scattering by Nonspherical Particles*, edited by M. Mishchenko, J. Hovenier, and L. Travis (Academic, San Diego, 2000).

[15] T. Nousiainen, K. Muinonen, J. Avelin, and A. Sihvola, ''Microwave backscattering by nonspherical ice particles at 5.6 GHz using second-order perturbation series,'' J. Quant. Spectrosc. Radiat. Transf. **70**, 639–661 (2001).

[16] L. Cazandjian, ''Identity of the Rayleigh-Fourier and extinction theorem method solutions for scattering and transmission of sound at a rough fluid-solid interface,'' J. Acoust. Soc. Am. **90**, 416–425 (1991).

[17] J.-M. Chesneaux and A. Wirgin, ''Reflection from a corrugated surface revisited,'' J. Acoust. Soc. Am. **96**, 1116–1129 (1994); L. Cazandjian, ''Comments on 'Reflection from a corrugated surface revisited' [J. Acoust. Soc. Am. **96**, 1116–1129 (1994)],'' *ibid.* **98**, 1813–1814 (1995); J.-M. Chesneaux and A. Wirgin, ''Response to ''Comments on 'Reflection from a corrugated surface revisited' [J. Acoust. Soc. Am. **96**, 1116–1129 (1994)],'''' *ibid.* **98**, 1815–1816 (1995).

[18] J. W. Strutt (Lord Rayleigh), *The Theory of Sound* (Dover, New York, 1945).

[19] See for example, the papers: J. Uretsky, ''The scattering of plane waves from periodic surfaces,'' Ann. Phys. (N.Y.) **33**, 400–427 (1965); R. Petit and M. Cadilhac, ''Sur la diffraction d'une onde plane par un reseau infiniment conducteur,'' C. R. Acad. Sci. Paris, Ser. B **262**, 468–471 (1966); R. Millar, ''On the Rayleigh assumption in scattering by a periodic surface,'' Proc. Cambridge Philos. Soc. **65**, 773–791 (1969); A. Purcell, ''Reflection coefficients for scattering from a pressure-release sinusoidal surface,'' J. Acoust. Soc. Am. **100**, 2919–2936 (1996).

[20] D. Jackson, D. Winebrenner, and A. Ishimaru, ''Comparison of perturbation theories for rough-surface scattering,'' J. Acoust. Soc. Am. **83**, 961–969 (1988).

[21] J. Roginsky, ''Derivation of closed-form expressions for the T matrices of Rayleigh-Rice and extinction-theorem perturbation theories,'' J. Acoust. Soc. Am. **90**, 1130–1137 (1991).

[22] J. Soto-Crespo, M. Nieto-Vesperinas, and A. Friberg, ''Scattering from slightly rough random surfaces: a detailed study on the validity of the small perturbation method,'' J. Opt. Soc. Am. A **7**, 1185–1201 (1990).

[23] A. Wirgin, ''Reflection from a corrugated surface,'' J. Acoust. Soc. Am. **68**, 692–699 (1980).

[24] J. Hugonin, R. Petit, and M. Cadilhac, ''Plane-wave expansions used to describe the field diffracted by a grating,'' J. Opt. Soc. Am. **71**, 593–598 (1981).

[25] P. van den Berg, ''Reflection by a grating: Rayleigh methods,'' J. Opt. Soc. Am. **71**, 1224–1229 (1981).

[26] R. Masel, R. Merill, and W. Miller, ''Quantum scattering from a sinusoidal hard wall: Atomic diffraction from solid surfaces,'' Phys. Rev. B **12**, 5545–5551 (1975).

[27] W.-K. Tung, *Group Theory in Physics* (World Scientific, Philadelphia, 1985).

[28] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. Stegun (Dover, New York, 1970).

[29] R. Holford, ''Scattering of sound waves at a periodic, pressure-release surface: An exact solution,'' J. Acoust. Soc. Am. **70**, 1116–1128 (1981).

[30] D. McCammon and S. McDaniel, ''Application of a new theoretical treatment to an old problem: sinusoidal pressure release boundary scattering,'' J. Acoust. Soc. Am. **78**, 149–156 (1985).

[31] G. Bishop and J. Smith, ''A scattering model for nondifferentiable periodic surface roughness,'' J. Acoust. Soc. Am. **91**, 744–770 (1992).

[32] P. van den Berg and J. Fokkema, ''The Rayleigh hypothesis in the theory of diffraction by a cylindrical obstacle,'' IEEE Trans. Antennas Propag. **AP-27**, 577–583 (1979).

Estimation of broadband power levels radiated from turbulent boundary layer-driven ribbed plates having dissimilar sections

M. L. Rumerman[a)]
Naval Surface Warfare Center, Carderock Division, Signatures Directorate (Code 7204), 9500 MacArthur Boulevard, West Bethesda, Maryland 20817-5700

(Received 13 February 2003; revised 25 April 2003; accepted 12 May 2003)

Previous papers developed an approach for estimating the vibration-related broadband acoustic power radiated by a ribbed steel plate in water driven by turbulent boundary layer pressures. This was extended to apply to plates composed of identical ribbed sections joined at infinite impedance supports. In this paper the basic model is modified so that it can be applied to plates composed of dissimilar sections joined at clamped supports. The response is formulated directly in terms of power and the effects of the supports are accounted for by power scattering coefficients, which are calculated through a Wiener–Hopf analysis. Sample calculations are shown to agree with results of other formulations. Only frequencies below bending coincidence are considered.
[DOI: 10.1121/1.1587736]

PACS numbers: 43.30.Jx, 43.40.Rj, 43.20.Tb [ANN]

I. INTRODUCTION

Fluid-loaded flat plates subjected to turbulent flow are set into bending vibration by the imposed turbulent boundary layer (TBL) pressure excitation. Ffowcs Williams showed that if the plate is infinite and uniform, no enhancement of TBL-radiated acoustic power results from the vibration,[1] but the presence of a discontinuity, such as a support, changes the nature of the acoustic radiation.[2] Crighton considered two coplanar semi-infinite plates having different properties, and also demonstrated that the discontinuity at the junction influences the nature of the radiation.[3] Howe treated the problem of a TBL-driven infinite uniform plate supported along two parallel lines at which the plate could be either clamped or pinned, and derived expressions for radiated power and pressure.[4] Rumerman developed an approach to accounting for finite impedance reinforcing ribs in estimating power radiated by TBL-driven water-loaded steel plates in broad (e.g., one-third octave) frequency bands.[5,6] This formalism was extended to plates composed of identical ribbed plate sections that are joined at boundaries considered to be infinite impedance line supports.[7] The basis of the formalism is that the ribs (supports) can be considered to radiate independently of one another if the rib (support) spacing exceeds one-fourth the acoustic wavelength. The structural symmetry of this configuration determines that there is no net power flow across section boundaries, therefore, the only power loss at a boundary support is due to acoustic radiation. Although not characteristic of most real structures, this assumption allowed the basis of the approximate model to be tested by a comparison to exact results obtained by adapting the doubly periodic reinforced plate model developed by Mace.[8] The model was validated within these constraints.

In most real structures, adjacent panel sections are not generally identical, but may differ in material, plate thickness, applied damping, length between supports, and rib type and spacing. This has two effects on the analysis: (1) the net power flow across section boundaries does not vanish, which generates coupling loss or gain, and (2) boundary effects cannot easily be accounted for in terms of forces. An objective for this paper is to present a more general formulation that can account for different section properties. Consideration will be limited to adjacent plate sections joined at large impedance backing supports that may be assumed to form a clamped boundary to each section, a typical construction. The formulation depends upon the power transmission and radiation coefficients related to flexural waves incident upon the boundary between adjacent sections, which can be calculated using a Wiener–Hopf analysis. Only frequencies below bending coincidence will be considered.

II. FORMULATION

The formalism for the extended plate of *identical* sections is summarized by Eqs. (20)–(23) of Ref. 7. These equations give broadband estimates of the mean squared plate bending velocity and acoustic power radiated by a rib-reinforced section, when the section contains at least two resonances within the band. When this requirement is satisfied, interaction among the discontinuities can be ignored when estimating the broadband mean squared force due to each rib and boundary support. When the average rib spacing is greater than one-fourth the acoustic wavelength, rib interactions can be ignored when estimating the acoustic power radiated by each rib. Similarly, when the boundary support spacing is greater than one-fourth the acoustic wavelength, support interactions can be ignored when estimating the acoustic power radiated by each support force. When this criterion for the rib (boundary) spacing is not satisfied, the ribs (boundaries) are assumed to not radiate.

These considerations lead to the definition of spatially averaged radiation resistances due to the ribs and boundary

[a)]Electronic mail: rumermanml@nswccd.navy.mil

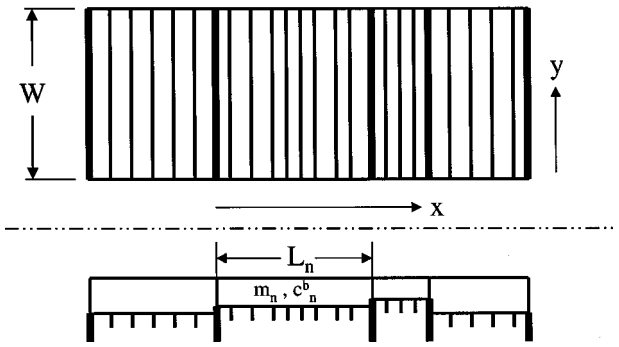


FIG. 1. Extended ribbed plate with dissimilar sections.

supports, Z^R and Z^B, respectively, given by Eqs. (1a) and (1b):

$$Z^R(\gamma;\omega)=\frac{NY_{\text{rad}}(\gamma;\omega)}{L|Y_\infty(\gamma;\omega)+Y_{\text{rib}}(\gamma;\omega)|^2};\quad ks>\pi/2$$
$$=0;\quad ks<\pi/2; \tag{1a}$$

$$Z^B(\gamma;\omega)=\frac{Y_{\text{rad}}(\gamma;\omega)}{L|Y_\infty(\gamma;\omega)|^2};\quad kL>\pi/2$$
$$=0;\quad kL<\pi/2. \tag{1b}$$

In these equations, ω is 2π times the frequency, k is the acoustic wave number in the ambient fluid, γ is the wave number corresponding to the direction along a rib or boundary, L is the distance between supports, N is the number of ribs within the section, and $s=L/(N+1)$ is the average rib spacing. Y_{rad} and Y_∞ are the radiation and input admittances, respectively, corresponding to a line force applied to the uniform plate in fluid, and Y_{rib} is the admittance of the rib.

Because all plate sections are identical in this model, and subject to the same excitation, there is no net coupling power transfer in or out of a section; therefore, the only effective power losses in a section are due to plate dissipation, rib radiation, and support radiation. When the sections are dissimilar, or the nature of the TBL drive varies among sections, net power can flow between adjacent sections, and must be accounted for as an additional coupling loss (or gain) mechanism. Because the sections do not have the same properties and/or responses, boundary effects cannot be easily represented in terms of forces and admittances. A power balance, using the power transmission and radiation coefficients at each section boundary, provides an easier approach.

Figure 1 shows an extended plate, of width W, consisting of sections joined at clamped boundary supports. Within each section, the ribs are identical and the plating properties invariant; however, these may change from section to section along with the average rib spacing. Because properties are invariant within a section, Eq. (1a) may be used to determine the rib radiation impedance for each section.

Consider the power balance for the nth section, as given by Eq. (2):

$$2\pi\sum_{j=-J}^{+J}[\Pi_n^{in,\text{TBL}}(j)+\Pi_n^{in,T}(j)]$$
$$=2\pi\sum_{j=-J}^{+J}[\Pi_n^{\text{diss}}(j)+\Pi_n^{\text{out},T}(j)+\Pi_n^{\text{rad},R}(j)+\Pi_n^{\text{rad},B}(j)]. \tag{2}$$

Each of the power terms corresponds to a mode in the y direction, whose wave number is $2\pi j/W$. Because the objective is to estimate the radiated power, the summation is limited to those modes for which $|j|\leqslant J$, where J is the largest integer that does not exceed $kW/2\pi$, (i.e., corner mode radiation is ignored compared to edge mode radiation). The factor 2π is consistent with the definition of terms given below, and $2\pi\Sigma_{j=-J}^{+J}\Pi_n^{in,\text{TBL}}(j)$ is the total power put into the edge modes of the nth section by the external TBL drive (exclusive of supersonic components that radiate independently of discontinuities). The other term on the left-hand side of Eq. (2) represents the power transmitted into the section across the boundaries. The terms on the right-hand side, respectively, represent the power dissipated within the section plating (e.g., via applied damping), power transmitted out of the section across the boundaries, power radiated due to the ribs and power radiated at the section boundaries. These power terms are given by Eqs. (3):

$$\Pi_n^{in,\text{TBL}}(j)$$
$$=S(\omega)L_n\,\text{Real}\int_{\substack{\text{residue}\\ \text{contrib of}\\ \text{subsonic pole}}}\frac{d\alpha}{Z_p\left(\alpha,\frac{2\pi j}{W}\right)+Z_f\left(\alpha,\frac{2\pi j}{W}\right)}$$
$$=2\pi S(\omega)\overline{Y_n^p}L_n=S(\omega)\frac{\pi Y_n^p L_n}{2m_n c_n^b}. \tag{3a}$$

In Eq. (3a), $\overline{Y_n^p}$ is the real contribution of the pole representing the subsonic structural wave to the line input admittance of the plate section, written as $Y_n^p/4m_nc_n^b$, where m_n is the mass per unit area of the section plating, c_n^b is the *in-vacuo* bending wave speed of the plating, and Y_n^p is dimensionless. The quantities Z_p and Z_f, respectively, are the wave number impedances of the plate and acoustic medium. The multiplier $S(\omega)$ is the frequency–wavenumber mean squared spectral density of the blocked TBL pressure, assumed wave number white. The factor 2π is distinct from that in Eq. (2). The index j will now be suppressed. Expressions for the wave number impedances are given in Appendix A:

$$\Pi_n^{\text{diss}}=\omega\varepsilon_n m_n L_n v_n^2. \tag{3b}$$

In Eq. (3b), $\varepsilon_n=\eta_n(\alpha_n^p/k_n^b)^4$ is the effective loss factor of the uniform fluid-loaded plating, defined relative to the mass of the plating alone, η_n is the plate's loss factor *in-vacuo*, and the wave numbers α_n^p and k_n^p are the in-water and *in-vacuo* wave numbers of the propagating bending waves, respectively. The effective loss factor is modified because the change in wave number due to fluid loading modifies the strain-displacement ratio:

$$\Pi_n^{\text{out},T} = (T_n^- + T_n^+)\frac{m_n c_n^b}{Y_n^p} v_n^2. \tag{3c}$$

In Eq. (3c), T_n^- is the power transmission coefficient for power flow from the nth section to the $(n-1)$th section, and T_n^+ is the power transmission coefficient for power flow from the nth section to the $(n+1)$th section, and it is assumed that the total mean squared velocity is composed equally of negative-going and positive-going components. Equation (3c) is derived in Appendix B. Similarly,

$$\Pi_n^{in,T} = \frac{m_{n-1} c_{n-1}^b}{Y_{n-1}^p} T_{n-1}^+ v_{n-1}^2 + \frac{m_{n+1} c_{n+1}^b}{Y_{n+1}^p} T_{n+1}^- v_{n+1}^2, \tag{3d}$$

$$\Pi_n^{\text{rad},B} = (A_n^- + A_n^+)\frac{m_n c_n^b}{Y_n^p} v_n^2. \tag{3e}$$

In Eq. (3e), A_n^- is the power radiation coefficient for power flow toward the boundary with the $(n-1)$th section, and A_n^+ is the power radiation coefficient for power flow toward the boundary with the $(n+1)$th section. Finally,

$$\Pi_n^{\text{rad},R} = L_n Z_n^R v_n^2. \tag{3f}$$

Because the power transmission coefficient is reciprocal with respect to direction,

$$T_{n-1}^+ = T_n^- \quad \text{and} \quad T_{n+1}^- = T_n^+ .$$

Insertion of Eqs. (3a)–(3f) into any of the cross-mode terms of Eq. (2) results in Eq. (4), in which the common factor 2π is removed:

$$\begin{aligned} S(\omega)\frac{\pi Y_n^p L_n}{2 m_n c_n^b} = &- \frac{m_{n-1} c_{n-1}^b}{Y_{n-1}^p} T_n^- v_{n-1}^2 - \frac{m_{n+1} c_{n+1}^b}{Y_{n+1}^p} T_n^+ v_{n+1}^2 \\ &+ \left(\omega \varepsilon_n m_n L_n + Z_n^R L_n + \frac{m_n c_n^b}{Y_n^p}(T_n^- + T_n^+) \right. \\ &\left. + \frac{m_n c_n^b}{Y_n^p}(A_n^- + A_n^+) \right) v_n^2. \end{aligned} \tag{4}$$

Such an equation can be written for each of the sections, and the system of equations can be solved for the mean squared velocity of each section. The power radiated by the jth mode of each section is found by summing the right-hand sides of Eqs. (3e) and (3f), and multiplying by 2π.

III. BOUNDARY COEFFICIENTS

A. Wiener–Hopf analysis

In order to use Eq. (4), the boundary power coefficients must be estimated. This can be done by representing two adjacent sections as two fluid-loaded semi-infinite plates joined at a clamped boundary, as shown in Fig. 2. A freely propagating bending wave $v(x) = V_0 \exp(i\alpha_L x)$, having amplitude V_0 and phase speed (with fluid loading) $c_{bL} = \omega/\alpha_L$, travels on the left-hand plate, and is normally incident upon the junction. Time dependence $\exp(-i\omega t)$ is assumed and suppressed. The bending wave carries with it an associated pressure wave on the plate surface given by $p(x) = (-i\omega\rho V_0/\sqrt{\alpha_L^2 - k^2})\exp(i\alpha_L x)$, where ρ is the fluid mass density and k is the acoustic wave number. The application of Wiener–Hopf analysis to structural acoustic problems is well established. Kouzov used an analogous method to obtain a general diffraction solution for a plane acoustic wave incident upon two coplanar dissimilar semi-infinite plates,[9] and then applied it to two particular situations.[10,11] Norris and Wickham applied it to scattering of an incident flexural wave from the junction of two plates,[12] the problem considered here, and Howe applied it to scattering of a flexural wave from the edge of a semi-infinite plate.[13] Howe's approach is adaptable to the two-plate problem and will guide the analysis here. Numerical results given by Norris and Wickham provide a check on the calculations.

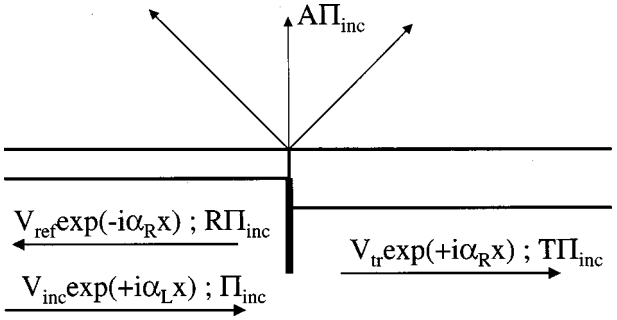


FIG. 2. Model for power scattering calculations.

The discontinuity at $x=0$ produces additional wave components, and the plate velocity and pressure for all x can be given by Eqs. (5a) and (5b):

$$v(x) = V_0 \exp(i\alpha_L x) + \frac{1}{2\pi}\int_{-\infty}^{+\infty} \overline{P}(\alpha)\frac{\sqrt{\alpha^2 - k^2}}{-i\omega\rho} \times \exp(+i\alpha x)\, d\alpha, \tag{5a}$$

$$P(x) = \frac{-i\omega\rho}{\sqrt{\alpha_L^2 - k^2}} V_0 \exp(+i\alpha_L x) + \frac{1}{2\pi}\int_{-\infty}^{+\infty} \overline{P}(\alpha) \times \exp(+i\alpha x)\, d\alpha. \tag{5b}$$

In these equations, the scattered field is given in terms of the scattered pressure, which has Fourier transform $\overline{P}$. The total pressure and total bending velocity must jointly satisfy the equation of motion for each plate given by Eq. (6), in which D is bending rigidity, m is the mass per unit area, and the subscript indicates the left or right plate:

$$\frac{D_{L,R}}{-i\omega}\frac{\partial^4 v(x)}{\partial x^4} - i\omega m_{L,R} v(x) = -p(x) \tag{6}$$

Because the incident wave is freely propagating on the left plate ($x<0$), it automatically satisfies Eq. (6) for the left plate. Inserting the fields into Eq. (6) yields Eq. (7a), in which $k_{bL}^4 = \omega^2 m_L / D_L$:

$$\int_{-\infty}^{+\infty}\left[\left(\frac{\alpha^4}{k_{bL}^4}-1\right)\sqrt{\alpha^2-k^2}-\frac{\rho}{m_L}\right]\overline{P}(\alpha)\exp(+i\alpha x)d\alpha=0;$$

$$x<0. \quad (7a)$$

The incident wave does not generally propagate freely on the right plate, and inserting the fields into Eq. (6) for the right plate ($x>0$), yields Eq. (7b), in which $k_{bR}^4=\omega^2 m_R/D_R$:

$$\int_{-\infty}^{+\infty}\left[\left(\frac{\alpha^4}{k_{bR}^4}-1\right)\sqrt{\alpha^2-k^2}-\frac{\rho}{m_R}\right]\times\left[\overline{P}(\alpha)-\frac{\omega\rho V_0}{\sqrt{\alpha_L^2-k^2}}\frac{1}{\alpha-\alpha_L(1+i0^+)}\right]\times\exp(+i\alpha x)d\alpha=0; \quad x>0. \quad (7b)$$

The notation $\alpha_L(1+i0^+)$ indicates that the pole is to be considered just above the real axis in the first quadrant of the α plane.

Because the integral in Eq. (7a) vanishes identically for arbitrary $x<0$, the integrand must be analytic in the lower half-α plane. Similarly, the integrand in Eq. (7b) must be analytic in the upper half-α plane. Let $\alpha=k\gamma$, $\alpha_L=k\gamma_L$, $\overline{P}(k\gamma)=J(\gamma)$, and $k/k_{bL,R}=\mu_{L,R}$; then Eqs. (8a) and (8b) apply:

$$W_L(\gamma)J(\gamma)=L(\gamma), \quad (8a)$$

$$W_R(\gamma)\left(J(\gamma)-\frac{\Psi_0}{\gamma-\gamma_L(1+i0^+)}\right)=U(\gamma). \quad (8b)$$

The functions $L(\gamma)$ and $U(\gamma)$ are analytic in the lower and upper γ half-planes, respectively, and

$$W_{L,R}(\gamma)=(\gamma^4-\mu_{L,R}^{-4})\sqrt{\gamma^2-1}-\frac{\rho\mu_{L,R}^{-4}}{km_{L,R}}, \quad (9a)$$

$$\Psi_0=\frac{\omega\rho V_0}{k^2\sqrt{\gamma_L^2-1}}. \quad (9b)$$

Equations (8a) and (8b) may be combined to produce Eq. (10):

$$J(\gamma)=\frac{L(\gamma)}{W_L(\gamma)}=\frac{U(\gamma)}{W_R(\gamma)}+\frac{\Psi_0}{\gamma-\gamma_L(1+i0^+)}. \quad (10)$$

Let $W_{L,R}(\gamma)$ each be written as the product of two functions; one analytic in the upper (+) half-plane and the other analytic in the lower (−) half-plane, i.e., $W_{L,R}(\gamma)=W_{L,R}^+(\gamma)W_{L,R}^-(\gamma)$. These have the property that $W_{L,R}^-(-\gamma)=W_{L,R}^+(+\gamma)$. Upon inserting this into Eq. (10), and after some manipulation, Eq. (11) results.

$$L(\gamma)\frac{W_R^-(\gamma)}{W_L^-(\gamma)}-\Psi_0\frac{W_L^+(\gamma_L)W_R^-(\gamma_L)}{\gamma-\gamma_L(1+i0^+)}=U(\gamma)\frac{W_L^+(\gamma)}{W_R^+(\gamma)}+\Psi_0\frac{W_L^+(\gamma)W_R^-(\gamma)-W_L^+(\gamma_L)W_R^-(\gamma_L)}{\gamma-\gamma_L(1+i0^+)}=\sum_{n=0}^{N}A_n\gamma^n. \quad (11)$$

The left-hand expression is analytic in a lower half-plane containing the real axis, and the middle expression is analytic in an upper half-plane containing the real axis; therefore, they are equal to one another and to the polynomial right-hand expression. Equations (10) and (11) may be combined to solve for $J(\gamma)$:

$$J(\gamma)=\frac{\Psi_0}{\gamma-\gamma_L(1+i0^+)}\frac{W_L^+(\gamma_L)W_R^-(\gamma_L)}{W_L^+(\gamma)W_R^-(\gamma)}+\frac{\Sigma_{n=0}^N A_n\gamma^n}{W_L^+(\gamma)W_R^-(\gamma)}. \quad (12)$$

The maximum exponent, N, is determined from the requirement that the pressure is continuous at $x=0$. Because $W_L^+(\gamma)W_R^-(\gamma)\to O(|\gamma^5|)$ for $|\gamma|\to\infty$, $N=3$. This provides four coefficients to satisfy the four boundary conditions at the junction of the two plates.

Using Eq. (5a), with $\alpha=k\gamma$, explicit expressions for the velocity fields in the left and right plates can be derived. The substitution $B_n=(k^2/\omega\rho)A_n$ has been made:

$$v(x)=V_0\exp(+ikx\gamma_L)-\frac{V_0}{2\pi i\sqrt{\gamma_L^2-1}}\times\int_{-\infty}^{+\infty}\frac{\sqrt{\gamma^2-1}}{\gamma-\gamma_L(1+i0^+)}\frac{W_R(\gamma_L)}{W_L(\gamma)}\times\frac{W_L^+(\gamma_L)}{W_R^+(\gamma_L)}\frac{W_L^-(\gamma)}{W_R^-(\gamma)}\exp(+ikx\gamma)d\gamma-\frac{1}{2\pi i}\int_{-\infty}^{+\infty}\frac{\sqrt{\gamma^2-1}}{W_L(\gamma)}\frac{W_L^-(\gamma)}{W_R^-(\gamma)}\times\sum_{n=0}^{3}B_n\gamma^n\exp(+ikx\gamma)d\gamma; \quad x<0; \quad (13a)$$

$$v(x)=V_0\exp(+ikx\gamma_L)-\frac{V_0}{2\pi i\sqrt{\gamma_L^2-1}}\times\int_{-\infty}^{+\infty}\frac{\sqrt{\gamma^2-1}}{\gamma-\gamma_L(1+i0^+)}\frac{W_R(\gamma_L)}{W_L(\gamma)}\frac{W_L^+(\gamma_L)}{W_R^+(\gamma_L)}\times\frac{W_R^+(\gamma)}{W_L^+(\gamma)}\exp(+ikx\gamma)d\gamma-\frac{1}{2\pi i}\int_{-\infty}^{+\infty}\frac{\sqrt{\gamma^2-1}}{W_R(\gamma)}\frac{W_R^+(\gamma)}{W_L^+(\gamma)}\times\sum_{n=0}^{3}B_n\gamma^n\exp(+ikx\gamma)d\gamma; \quad x>0. \quad (13b)$$

The residue of the pole at $\gamma_L(1+i0^+)$ cancels the incident wave for $x>0$. Due to the behavior of the integrands in the first integral of both Eqs. (13a) and (13b) at large $|\gamma|$, its contribution to the velocity and its first three derivatives is continuous across $x=0$. The contribution of the second integral in each equation to the velocity and its first derivative is continuous across $x=0$ for $n=0$ and $n=1$. Therefore, if particular values of B_0 and B_1 satisfy a clamped boundary

condition for the left plate, they will automatically do so for the right plate, and it can be shown that B_2 and B_3 both vanish.

To evaluate the integrals in Eq. (13a) for $x<0$, the contour is closed in the lower half-plane. Because the ratio of the split functions of γ is analytic in that region, it does not present a problem other than its evaluation, which can be done following Howe. The ratio of the split functions of γ_L is a constant, as is $W_R(\gamma_L)$. Therefore, the integrals present no more conceptual difficulty than analogous integrals for a uniform plate. A similar statement applies to the integrals in Eq. (13b) for $x>0$. After evaluating the required integrals, the coefficients B_0 and B_1, corresponding to the clamped boundary, are obtained by setting $v(x)=0$ and $\partial v(x)/\partial x = 0$ at either $x=0^+$ or $x=0^-$.

B. Power flow

The power associated with the incident bending wave is $\Pi_{\text{inc}}=(2m_L c_L^b/Y_L^p)V_0^2$. Its interaction with the junction at $x=0$ generates a bending wave reflected back into the left plate, a bending wave transmitted into the right plate, and an acoustic field. The amplitude of the reflected wave, V_{ref}, is associated with the residues of the integrands in Eq. (13a) corresponding to the pole $\gamma=-\gamma_L$, which is a zero of $W_L(\gamma)$. The corresponding reflected power is $\Pi_{\text{ref}}=(2m_L c_L^b/Y_L^p)V_{\text{ref}}^2$. The amplitude of the transmitted wave, V_{tr}, is associated with the residues of the integrands in Eq. (13b) corresponding to the pole $\gamma=+\gamma_R$, which is a zero of $W_R(\gamma)$. The corresponding transmitted power is $\Pi_{\text{tr}}=(2m_R c_R^b/Y_R^p)V_{\text{tr}}^2$. The radiated power associated with the acoustic field is given by substituting Eq. (12), with $A_n=(\omega\rho/k^2)B_n$, into Eq. (14):

$$\Pi_{\text{rad}}=\int_{-k}^{+k}\frac{\sqrt{k^2-\alpha^2}}{\omega\rho}|\overline{P}(\alpha)|^2 d\alpha$$

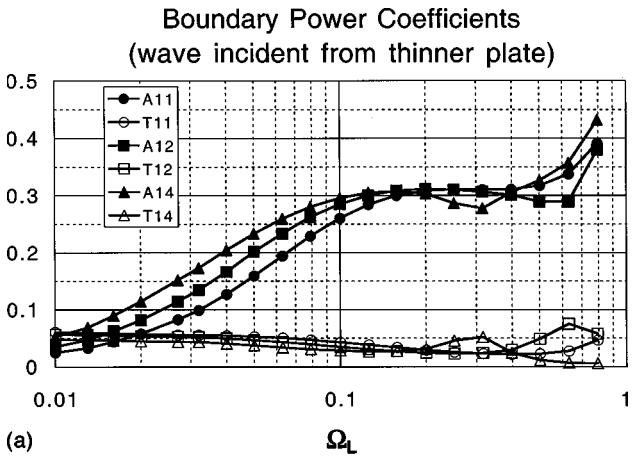
$$=\frac{k}{\rho c}\int_{-1}^{+1}|J^2(\gamma)|\sqrt{1-\gamma^2}\,d\gamma. \tag{14}$$

The three scattered power coefficients are found by dividing the associated scattered power by the incident power, with transmission coefficient $T=\Pi_{\text{tr}}/\Pi_{\text{inc}}$, reflection coefficient $R_L=\Pi_{\text{ref}}/\Pi_{\text{inc}}$, and acoustic radiation coefficient $A_L=\Pi_{\text{rad}}/\Pi_{\text{inc}}$.

C. Calculated results

Calculations were performed for steel plates with water on one side. The steel was assumed to have a bending modulus of 2.0×10^{12} dyn/cm^2 and a mass density of 7.8 g/cm^3. The water was assumed to have a mass density of 1.0 g/cm^3 and a sound speed of 1.5×10^5 cm/s. When the material properties are specified, universal curves can be derived in which the scattering coefficients depend only on the ratio of the thickness of the originating plate to that of the receiving plate and on the ratio of the frequency to the bending coincidence frequency of the originating plate, $\Omega_L=\mu_L^2$.

Figures 3(a) and (b) present transmission and radiation coefficients for originating plate to receiving plate thickness ratios of 0.25, 0.5, 1, 2, and 4. The reflection coefficients



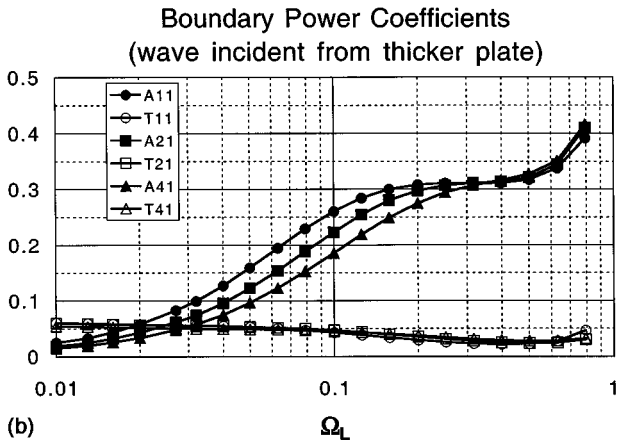


FIG. 3. (a) Power scattering coefficients for incidence onto thicker plate. Indices represent relative plate thicknesses. (b) Power scattering coefficients for incidence onto thinner plate. Indices represent relative plate thicknesses.

were also independently calculated, but are not shown. The three power coefficients should sum to unity, and this was achieved within about 0.1%. The results for $h_R=2h_L$ are in agreement with those given by Norris and Wickham.[12]

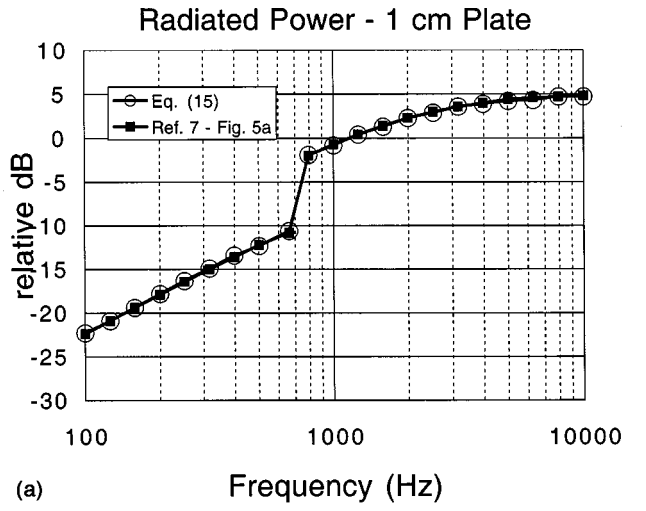
IV. SAMPLE CALCULATIONS OF RADIATED POWER

The results of boundary interaction calculations, similar to those displayed in Figs. 3(a) and (b), will be used to calculate the mean squared panel velocities and radiated power for two configurations of extended steel plates in water. The radiated power of the nth panel is given by Eq. (15):

$$\Pi_n^{\text{rad}}=2\pi\sum_{j=-J}^{+J}[\Pi_n^{\text{rad},R}(j)+\Pi_n^{\text{rad},B}(j)]. \tag{15}$$

A. Configuration 1

Configuration 1a corresponds to Fig. 5(a) of Ref. 7, so that the respective results can be compared. A steel plate with five identical sections is considered. The plate is 1 cm thick with an *in-vacuo* bending loss factor of 0.05, and has water on one side. Material properties are the same as given above. Each section has nine identical ribs with an average spacing of 50 cm, so that a section length is 500 cm. Only the contribution of the uniform cross-mode ($j=0$) will be considered, so the plate width, W, and the bending properties of the



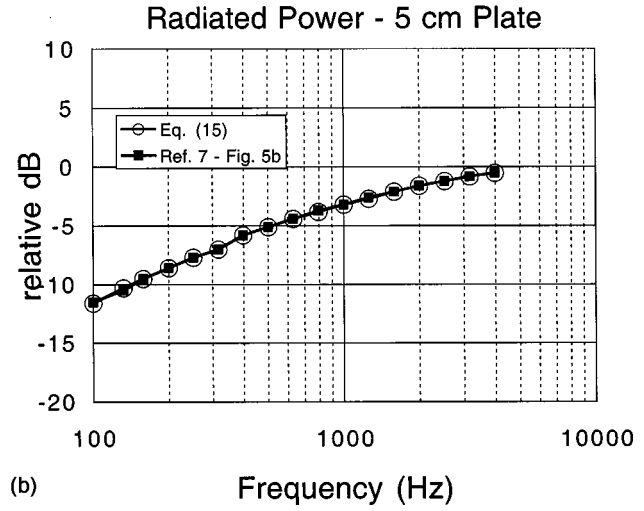


FIG. 4. (a) Radiation from 1 cm ribbed plate with identical sections—comparison with Ref. 7. (b) Radiation from 5 cm ribbed plate with identical sections—comparison with Ref. 7.

ribs are not needed. The power transmission coefficients at the outer edges of the first and fifth sections are assumed to be zero, so that no power leaves the system, except by radiation. Configuration 1b corresponds to Fig. 5(b) of Ref. 7. There are again five identical sections, but the plate thickness is 5 cm and the bending loss factor is 0.005. There are nine ribs with an average spacing of 100 cm, so that the section length is 1000 cm. For each configuration, the calculated results were multiplied by 0.23ω to obtain one-third-octave band levels presented in Figs. 4(a) and (b). The results overlay those of Ref. 7, as expected.

B. Configuration 2

Configuration 2 corresponds to Fig. 3 of Ref. 4. The section of interest is that portion of an infinite steel strip of width W sectioned off by two parallel clamped line supports that are a distance L apart, as illustrated in Fig. 5. There is water on each side of the plate system and there are no ribs. (The Wiener–Hopf analysis given earlier can be applied to this configuration by doubling the mass density of the fluid.) Excitation of the two semi-infinite extensions is not consid-

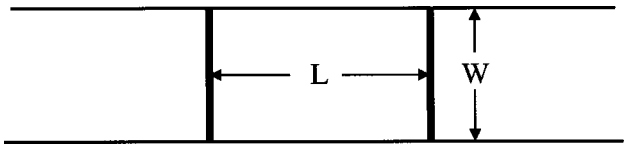


FIG. 5. Single section model used in Ref. 4.

ered, and they serve only as energy sinks into which power may be transmitted from the section. In this example, all radiating edge modes, $|j|<J$, are considered. In principle, this requires a Wiener–Hopf calculation for each mode, but the power coefficients are insensitive to mode number until $2\pi|j|\approx kW$. Assuming the coefficients to be independent of j, the power radiated by the panel is estimated by Eq. (16):

$$\Pi^{\mathrm{rad}}=2\pi\sum_{j=-J}^{+J}\Pi^{\mathrm{rad},B}(j)$$

$$\approx 2\pi\left(2\frac{kw}{2\pi}\right)\Pi^{\mathrm{rad},B}(0)$$

$$\approx 2kW\Pi^{\mathrm{rad},B}(0)=4kW\frac{mc^bA}{Y^p}v^2(0). \tag{16}$$

The mean squared velocity is obtained from Eq. (4), with one section and no ribs:

$$v^2(0)=\frac{S\dfrac{\pi Y^pL}{2mc^b}}{\varepsilon\omega mL+\dfrac{2mc^b(T+A)}{Y^p}}. \tag{17}$$

Reference 4 presents results in terms of the farfield squared pressure radiated to a point lying on the normal to the center of the section and attributable to one of the section edges [i.e., corresponding to half the power of Eq. (16)]. At frequencies sufficiently high so that $(\omega m/2\rho c)^2\gg 1$, but less than the bending coincidence frequency, the three-dimensional radiation approaches omnidirectionality. In the limit, the mean squared pressure on the normal at range R is estimated by Eq. (18):

$$P^2\approx\frac{\rho c\Pi^{\mathrm{rad}}}{4\pi R^2}. \tag{18}$$

At frequencies sufficiently low so that $(\omega m/2\rho c)^2\ll 1$, the three-dimensional radiation can be considered dipole-like, and the pressure squared on the normal is a factor of 3 higher than for omnidirectional radiation. Equation (19) was found to be a good directivity correction for frequencies below 80% of the coincidence frequency. Above the coincidence frequency, the radiation is directional:

$$D=\frac{3+\left(\dfrac{\omega m}{2\rho c}\right)^2}{1+\left(\dfrac{\omega m}{2\rho c}\right)^2}. \tag{19}$$

Upon combining Eqs. (16)–(19), the on-axis farfield mean squared pressure attributed to one support may be written in the form of Eq. (20):

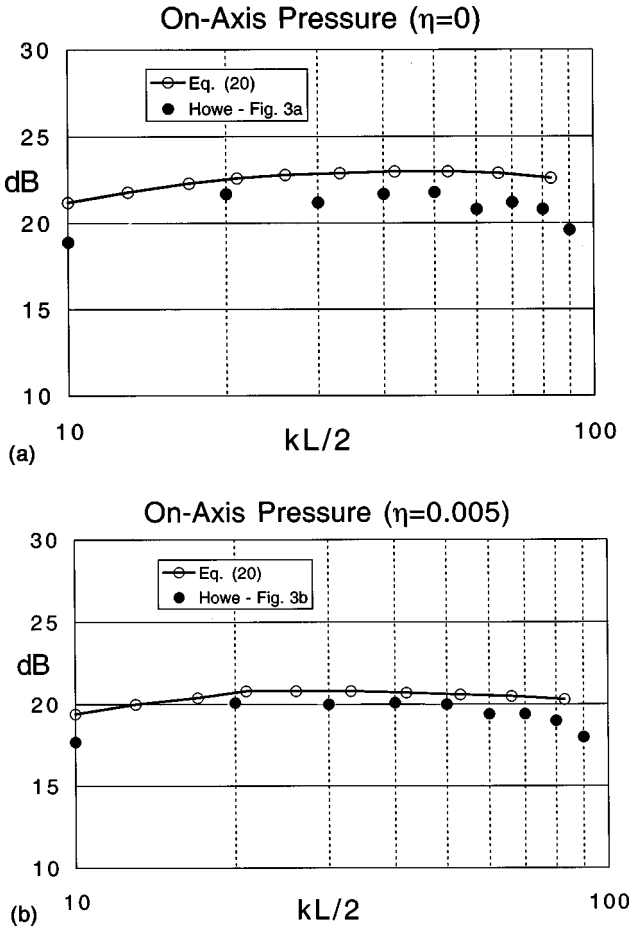


FIG. 6. (a) A comparison of results with Ref. 4 ($\eta=0$). (b) A comparison of results with Ref. 4 ($\eta=0.005$).

$$P^2 \approx \frac{S(\omega)}{R^2}\frac{W}{L/2}\left[\frac{\frac{1}{8}\omega\rho A D L^2}{\varepsilon\omega m L + \dfrac{2mc^b(T+A)}{Y^p}}\right]. \tag{20}$$

Figures 6(a) and (b) compare the quantity in brackets on the rhs of Eq. (20) with the corresponding quantity presented as the mean line in Figs. 3(a) and 3(b) of Ref. 4. [The mean was found by averaging the frequency-dependent squared pressure over a frequency interval corresponding to $\Delta(kL/2)=1$.] The method employed here inherently produces a mean squared pressure, and the results agree to within 2 dB.

V. CONCLUSION

Previous papers established a formalism for estimating broadband acoustic power levels radiated by TBL-driven extended steel plates (in water), which consisted of identical ribbed sections connected at infinite impedance supports. In this paper the formalism was generalized to apply to extended plates having nonidentical sections that are clamped at the supports. In order to do so, it was necessary to use a power balance approach, and to replace the force representation of the support with power scattering coefficients relating to flexural waves incident upon the boundary between adjacent sections. A Wiener-Hopf analysis was performed to show how these coefficients could be calculated. The formalism was applied to a previously considered extended plate of identical sections. The previous results were reproduced, providing a partial validation of the formalism and the calculations of power scattering coefficients. The formalism was also applied to the case of a single unribbed plate section, previously analyzed by a different method, and agreement to within 2 dB was achieved. Only frequencies below the bending coincidence frequency were considered because above the coincidence the subsonic response mode cannot receive much power from the TBL.[4] The dominant portion of the radiated power then arises from the supersonic portion of the wave number spectral density, and the plate vibration is usually of secondary influence.[1]

The burden of the method is carried by the calculation of the power scattering coefficients. For plate sections of the same material, they depend only on the plate thickness ratio and the fraction of bending coincidence frequency; therefore, they need to be calculated only once for a set of thickness ratios of practical interest and to as fine a resolution in dimensionless frequency as desired. A two-parameter table of each these coefficients can then be established, and values accessed by table lookup with or without interpolation.

Although the power estimation procedures given in this paper are in the context of turbulent boundary layer pressure excitation, no explicit use has been made of the existence of a flow. Two elements of the flow that are relevant are the influence of the convection ridge of the driving pressure spectral density on the plate response and the effects of the mean flow on the radiated acoustic power. Because, in underwater applications, a typical flow speed is a small fraction of the bending wave speed in steel plates and a smaller fraction of the sound speed in water, neither of these factors is important. Additional discussion is given in Appendix C.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research (ONR) Hydroacoustics Program: L. Patrick Purtell and Ki-han Kim, Program Managers.

APPENDIX A: GOVERNING EQUATIONS

The thin plate bending equation and the simple wave equation have been used to model the plate and acoustic fluid, respectively. When the analysis is performed in Fourier transform (i.e., wave number) space, their responses may be described in terms of their wave number impedances:

$$Z_p(\alpha,\beta) = -i\omega m\left(1-\frac{(\alpha^2+\beta^2)^2}{k_b^4}\right), \tag{A1}$$

$$Z_f(\alpha,\beta) = \frac{-i\omega\rho}{\sqrt{\alpha^2+\beta^{22}-k^2}}. \tag{A2}$$

Mechanical area damping of the plate vibration can be included by making the bending rigidity, D, complex so that $k_b^4 = m\omega^2/D(1-i\eta)$, where η is the bending loss factor *in-vacuo*.

The bending velocity of the fluid-loaded plate, due to a surface distribution of normal forces $f(x,y)$ having Fourier transform $\bar{f}(\alpha,\beta)$, is given by Eq. (A3):

$$v(x,y)=\frac{1}{4\pi^2}\int d\alpha\int d\beta\,\frac{\overline{f}(\alpha,\beta)}{Z_p(\alpha,\beta)+Z_f(\alpha,\beta)}\times\exp[i(\alpha x+\beta y)], \tag{A3a}$$

where

$$\overline{f}(\alpha,\beta)=\int dx\int dy\, f(x,y)\exp[-i(\alpha x+\beta y)]. \tag{A3b}$$

The power per unit area put into this system by a spatially homogeneous and temporally stationary distribution of normal forces, having mean squared wave number spectral density $S(\alpha,\beta;\omega)$, is given by Eq. (A4a):

$$\Pi_{in}^{\text{area}}=\text{Re}\int d\alpha\int d\beta\,\frac{S(\alpha,\beta;\omega)}{Z_p(\alpha,\beta)+Z_f(\alpha,\beta)} \tag{A4a}$$

Let the β integration be approximated by a summation:

$$\Pi_{in}^{\text{area}}\approx\frac{2\pi}{W}\text{Re}\sum_j\int\frac{S\left(\alpha,\frac{2\pi j}{W};\omega\right)}{Z_p\left(\alpha,\frac{2\pi j}{W}\right)+Z_f\left(\alpha,\frac{2\pi j}{W}\right)}d\alpha, \tag{A4b}$$

where W is a length associated with the y direction. The power input to a strip of width W in the y direction, and of length L in the x direction is then approximated by multiplying the rhs of Eq. (A4b) by WL. This is a heuristic demonstration of the form of the power balance in Eq. (2) in association with Eq. (3a) for the input power term, where S has been assumed to be wave number-white. The actual derivation was performed by developing the y dependence as a full range Fourier series, and finding mean squared statistics.

APPENDIX B: COUPLING POWER FLOW

Consider a line force F acting on the plate along the line $x=0$. The power it puts into the plate is $\Pi^{in}=F^2\overline{Y^p}$ and the propagating part of the induced bending velocity is $v^p=F\overline{Y^p}$. Upon replacing F in the power relation by its value in terms of the propagating velocity, and recognizing that half the power propagates to each side of the force, the power traveling in one direction, and incident upon a boundary, is $\Pi^{\text{inc}}=v_p^2/2\overline{Y^p}=2mc^b v_p^2/Y^p$. It is assumed that half of the mean squared velocity of the nth section, v_n^2, is associated with negative-traveling waves and half with positive-traveling waves. The power incident on each boundary is then given by $mc^b v_n^2/Y^p$, and multiplication by each of the boundary transmission coefficients results in Eq. (3c).

APPENDIX C: FLOW EFFECTS

In water, the maximum flow speed may be about 1500 cm/s and the convection speed approximately 70% of that or 1050 cm/s. The bending wave speed in a 1 cm steel plate at 100 Hz is about 9500 cm/s, and increases with frequency and for thicker plates. Therefore, the convection wave number is at least nine times the plate's bending wave number, and cannot drive the plate in bending. In fact, elasticity theory—not plate theory—should be used to compute the surface response of the plate to the convected pressures. Response levels exponentially attenuate from the outboard to the inboard face, and the effects on the ribs are insignificant. The convection ridge can also couple into thin plate modes through edge interactions at the section boundaries, but the resulting levels are also small in comparison to those due to the direct subconvective drive. These aspects were addressed by Chandiramani.[14,15]

In the presence of a mean flow, acoustic effects should be represented by the convected wave equation, not the simple wave equation. Use of this theory shows that the mean flow effects enter through the flow Mach number (the ratio of flow speed to sound speed in the fluid). For frequencies below the bending coincidence frequency, flow effects on radiated power are negligible if the Mach number is much smaller than unity. This is the case in underwater applications for which the sound speed is 150 000 cm/s and the maximum flow speed is about 1500 cm/s, giving a maximum Mach number of 0.01. Sound radiation from line force-driven plates in the presence of mean flow has been addressed by Wu and Zhu.[16–18]

[1] I. E. Efowcs Williams, "Sound radiation from turbulent boundary layers formed on compliant surfaces," J. Fluid Mech. **22**, 347–358 (1965).
[2] J. E. Ffowcs Williams, "The influence of simple supports on the radiation from turbulent flow near a plane compliant surface," J. Fluid Mech. **26**, 641–649 (1966).
[3] D. G. Crighton, "Radiation from turbulence near a composite flexible boundary," Proc. R. Soc. London, Ser. A **314**, 153–173 (1970).
[4] M. S. Howe, "Diffraction radiation produced by turbulent boundary layer excitation of a panel," J. Sound Vib. **121**, 47–65 (1988).
[5] M. L. Rumerman, "Estimation of broadband acoustic power due to rib forces on a reinforced panel under turbulent boundary layer-like pressure excitation. I. Derivations using string model," J. Acoust. Soc. Am. **109**, 563–575 (2001).
[6] M. L. Rumerman, "Estimation of broadband acoustic power due to rib forces on a reinforced panel under turbulent boundary layer-like pressure excitation. II. Applicability and validation," J. Acoust. Soc. Am. **109**, 576–582 (2001).
[7] M. L. Rumerman, "Estimation of broadband acoustic power radiated from a turbulent boundary layer-driven reinforced finite plate section due to rib and boundary forces," J. Acoust. Soc. Am. **111**, 1274–1279 (2002).
[8] B. R. Mace, "Sound radiation from a plate reinforced by two sets of parallel stiffeners," J. Sound Vib. **71**, 435–441 (1980).
[9] D. P. Kouzov, "Diffraction of a plane hydroacoustic wave on the boundary of two elastic plates," J. Appl. Math. Mech. **27**, 806–815 (1963).
[10] D. P. Kouzov, "Diffraction of a plane hydroacoustic wave at a crack in an elastic plate," J. Appl. Math. Mech. **27**, 1593–1601 (1963).
[11] D. P. Kouzov, "Diffraction of a cylindrical hydroacoustic wave at the joint of two semi-infinite elastic plates," J. Appl. Math. Mech. **33**, 225–234 (1969).
[12] A. N. Norris and G. R. Wickham, "Acoustic diffraction from the junction of two flat plates," Proc. R. Soc. London, Ser. A **451**, 631–655 (1995).
[13] M. S. Howe, *Acoustics of Fluid–Structure Interactions* (Cambridge University Press, New York, 1998), Sec. 4.4.
[14] K. L. Chandiramani, "Vibration response of fluid-loaded structures to low-speed flow noise," J. Acoust. Soc. Am. **61**, 1460–1470 (1977).
[15] K. L. Chandiramani, "Response of underwater structures to convective component of flow noise," J. Acoust. Soc. Am. **73**, 835–839 (1983).
[16] S. F. Wu and J. Zhu, "Sound radiation from two semi-infinite dissimilar plates subject to a harmonic line force excitation in mean flow. I. Theory," J. Acoust. Soc. Am. **97**, 2709–2723 (1995).
[17] J. Zhu and S. F. Wu, "Sound radiation from two semi-infinite dissimilar plates subject to a harmonic line force excitation in mean flow. II. Asymptotics and numerical results," J. Acoust. Soc. Am. **97**, 2724–2732 (1995).
[18] S. F. Wu and J. Zhu, "Effect of mean flow on responses of a fluid-loaded plate," J. Acoust. Soc. Am. **98**, 1786–1795 (1995).

Quantification of a multibeam sonar for fisheries assessment applications

N. A. Cochrane[a)]
Ocean Sciences Division, Department of Fisheries and Oceans, Bedford Institute of Oceanography, P.O. Box 1006, Dartmouth, Nova Scotia B2Y 4A2, Canada

Y. Li[b)]
Ocean Mapping Group, Department of Geodesy and Geomatics Engineering, University of New Brunswick, P.O. Box 4400, Station A, Fredericton, New Brunswick E3B 5A3, Canada

G. D. Melvin
Marine Fish Division, Department of Fisheries and Oceans, St. Andrews Biological Station, 531 Brandy Cove Road, St. Andrews, New Brunswick E5B 2L9, Canada

(Received 12 June 2001; revised 15 April 2003; accepted 1 May 2003)

The acoustic theory is developed for a multibeam fisheries-type sonar employing a circular arc of transducer elements. Specifically, numerical relations for transmit and receive beam patterns are derived and methodologies set forth for the derivation of appropriately scaled acoustic target strength and acoustic volume backscattering strength from an ideally performing multibeam device. Predicted and measured beam characteristics of a realizable multibeam sonar, a Kongsberg Simrad-Mesotech SM 2000, are compared. Practical techniques for the extraction of calibrated acoustic volume backscattering strength from real systems are advanced. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1587151]

PACS numbers: 43.30.Vh, 43.30.Yj, 43.30.Sf [DLB]

LIST OF SYMBOLS

a	Constant defining roll-off rate of beamformer window function
a_T	Radius of curvature of transducer array
c	Acoustic sound speed
C_{Cal}	Sonar calibration constant in decibels
d_{mid}	Lateral position of the midpoint of the span of the transducer summation arc when projected onto a line normal to the beamforming direction
d_n	Lateral position of array element n projected onto above normal line
$D_{e\theta}$, $D_{e\phi}$	Transmit or receive directivity functions of individual array elements: $D_{e\theta}(\Delta\theta)$—as function of off-axis angle $\Delta\theta$ in sonar equatorial plane $D_{e\phi}(\phi)$—as function of off-axis angle ϕ along sonar meridian
D_θ, D_ϕ	Directivity functions (sync function form) for rectangular acoustic radiator
D_S	Array source transmit pressure directivity function: $D_S(\theta_b)$—in sonar equatorial plane at longitudinal angle θ_b referenced to pressure radiated at 90° (array center) $D_S(\theta_b,\theta,\phi)$—in direction (θ,ϕ) referenced to pressure radiated at angle θ_b in equatorial plane $D_{S\theta}(\theta_b,\theta)$, $D_{S\phi}(\theta_b,\phi)$—special one-dimensional forms in directions θ and ϕ, respectively, referenced to pressure radiated in equatorial plane at angle θ_b
D_R	Array receive pressure directivity function: $D_R(\theta_b,\theta,\phi)$—beamformer voltage amplitude directivity function for pressure source at (θ,ϕ) while beamforming in direction θ_b. Voltage referenced to corresponding output with equatorial plane source at θ_b $D_{R\theta}(\theta_b,\theta)$, $D_{R\phi}(\theta_b,\phi)$—special one-dimensional forms in directions θ and ϕ, respectively, beamforming at angle θ_b referenced to voltage with equatorial plane pressure source at θ_b
I_i, I_r	Incident and reflected acoustic intensities, respectively
k	Acoustic wave number
$K_{v/p}$	Individual array element pressure sensitivity (volts/unit pressure) including effect of element-associated electronics
l	Array element physical dimension in longitudinal (fan) direction
n	Array element number index (1–80 for SM 2000)
N_1, N_2, etc.	Summation bounds on array element number index n
$p_{\mathrm{rec}}(\theta,t)$	Time-dependent incident pressure envelope at array surface for equatorial plane source at θ
$p_{\mathrm{xmit}}(\theta)$	Array equatorial plane transmit pressure amplitude in direction θ reduced to reference distance of 1 m from the transducer face. Second argument (ϕ) can be added in three-dimensional case.

[a)]Electronic mail: cochranen@mar.dfo-mpo.gc.ca
[b)]Currently at: IntelliCorp, Inc., 1975 El Camino Real West, Suite 201, Mt. View, CA 94040.

$p_e(\theta_i)$ Single-array element transmit pressure amplitude in direction θ_i reduced to reference distance of 1 m from the element. θ_i measured from normal to element surface in equatorial array plane.

R Range from sonar face

$\mathrm{Re}(x)$ Real component of complex x

S_{DR2}, $S_{DR\theta}$ Theoretical beamformer elemental voltage summations for angularly placed remote pressure source. Source produces unit elemental voltage output when located on given elemental axis

S_{DS}, S_{DS2}, $S_{DS\theta}$ Theoretical whole array transmit pressure summations at a remote angularly placed receiver. A unit pressure signal is received from a given element when receiver is located on same elemental axis

$S_v(\theta_b, t)$ Time-dependent volume backscattering strength in decibels while beamforming in direction θ_b

t Time from initiation of sonar pulse

TS, $\mathrm{TS_{Target}}$ Acoustic target strength in decibels, target strength of calibration target

$\underline{\mathbf{V}}_s$ Beamformer output voltage assuming nondemodulated elemental voltage inputs

V_s Beamformer output voltage (complex) assuming baseband-demodulated elemental voltage inputs

$V_N(\theta_b)$ Amplitude (real) of fully normalized beamformer summation in direction θ_b

$\underline{\mathbf{V}}_n$ Hypothetical output voltage, nondemodulated, from array element n including associated electronics

V_n Output voltage, baseband demodulated (complex), from array element n with associated electronics

$V_{\mathrm{BF}}(\theta_b, \theta, \phi)$ Normalized beamformer output voltage amplitude for pressure source at (θ,ϕ) while beamforming in direction θ_b

w Array element physical dimension in latitudinal (normal to fan) direction

$W(\theta_b, \theta_n)$ Beamformer window function multiplicatively applied to the voltage output of the array element at θ_n while beamforming in longitudinal direction θ_b

α Acoustic absorption coefficient

α_i Incidence angle of incoming pressure wave

Δl_n Path length differential from infinite range equatorial plane source at beamforming angle θ_b, to nearest point of array surface and to array element n

Δl_s Path length differential from infinite range equatorial plane source or receiver at arbitrary longitudinal angle, to nearest point on array surface and to array element n

Δt Sonar pulse length

Δt_n Acoustic propagation time differential corresponding to path differential Δl_n

$\Delta\theta_n$ Longitudinal absolute angular difference between the beamforming direction and the angular position of array element n

$\Delta\theta_s$ Longitudinal absolute angular difference between the direction of an arbitrary equatorial plane source or receiver and the angular position of array element n

θ,ϕ Longitudinal and latitudinal coordinates, respectively, in sonar spherical polar coordinate system where polar vector is normal to array elemental plane

θ_b, θ_n Longitudinal beamforming direction, longitudinal angle of array element n

$\Psi_D(\theta_b)$ Integrated beam pattern (2-way) for a receive beam synthesized at angle θ_b and a transmit pattern generated by exciting all array elements simultaneously in phase

$\Psi_{DS}(\theta_b)$ Modified integrated beam pattern, of similar form to $\Psi_D(\theta_b)$, but defined using receive directivities normalized to axial receive response of beam synthesized at $\theta_b = 90°$, and transmit directivities normalized to transmit response for $\theta = 90°$

ω Angular frequency of sonar carrier

Ω Solid angle

I. INTRODUCTION

Electronically synthesized multibeam sonar promises to become a valuable quantitative tool for fisheries survey and assessment. The simultaneous high-resolution ensonification of either a directed broad swath or a three-dimensional enclosing volume of water from or about a survey vessel affords areal coverages far exceeding those possible with conventional vertical beam echosounders typified by transducers of 5–15° conical beamwidth (Mayer *et al.*, 2002). Besides increased survey efficiency, other potential advantages include monitoring of the near-surface zone, the direct assessment of vessel avoidance, and wide-swath discrete target delineation and counting. Early studies of fish behavior and vessel avoidance using multibeam and its forerunner, sector scanning sonar, have been promising (Misund and Aglen, 1992; Gerlotto *et al.*, 1994; Misund *et al.*, 1995; Soria *et al.*, 1996; Cochrane and Melvin, 1997; Misund *et al.*, 1998; Gallaudet and de Moustier, 2002).

The most sophisticated multibeam systems are designed exclusively for bathymetric mapping (Heier, 1987) and normally gate-out water column signals (e.g., fish) as noise. The numerous fisheries-oriented multibeam and sector scanning sonars employed for real-time commercial harvesting of schooling pelagic fish species possess, at most, a rudimentary quantitative capability (Ona, 1994). Furthermore, exacting quantitative application has been slowed by: (a) Low precision and sometimes nonlinear signal processing oriented toward operational visual displays. (b) No provision

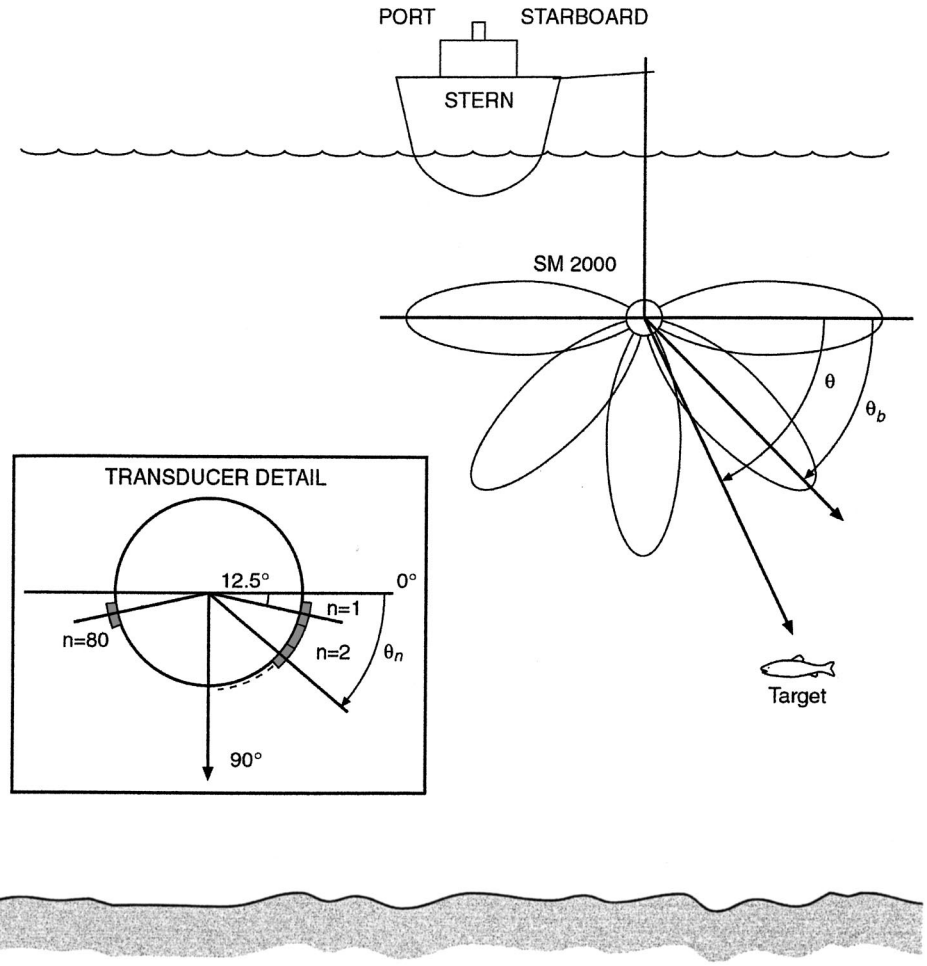


FIG. 1. Multibeam sonar deployment and beam geometry.

for extracting signals for external real-time or postprocessing. (c) Complex calibration requirements for multitransducer arrays compounded by numerous beam synthesis options. (d) Challenging ambiguities arising from highly variable fish target aspect in inverting standard acoustic measures to fish biomass. The recent Simrad-Mesotech SM 2000 multibeam sonar has the quantitative performance and data output capabilities to allow serious exploration of fisheries applications. Its properties, calibration, and utility for the extraction of quantitative acoustic backscatter measures as a preliminary step towards fisheries biomass estimation is the focus of this paper. The subsequent conversion of acoustic measures to biomass over a wide range of ensonification angles constitutes a separate complex problem beyond the scope of the present analysis.

Quantification of multibeam systems hinges on accurate description of their complex 3D beam patterns. A totally empirical methodology is excessively labor intensive, involving exhaustive far-field beam measurements over a surrounding 3D spherical surface. A purely theoretical approach neglects the nonideal performance of any realizable sonar. Our methodology combines the strengths of each approach while minimizing their specific weaknesses. Using a theoretical descriptive framework supplemented by the measured characteristics of a real multibeam sonar within a selected 2D plane, the extraction of calibrated target strength and volume backscattering strength to accuracies approaching those required for fisheries assessment is demonstrated. Theory and methodologies advanced are applicable to any multibeam system operating on the same principles as the illustrative unit.

II. HARDWARE

The Kongsberg Simrad-Mesotech SM 2000 multibeam sonar and its mode of utilization are first described. The 200-kHz sonar uses a virtually contiguous array of 80 rectangular 2.54-mm (equatorial or longitudinal fan direction)×21.1-mm (orthogonal latitudinal direction) transducer elements distributed over a 155° circular arc (bounded by centers of elements 1 and 80) of 0.1085-m nominal radius. On transmit, all 80 elements are excited in phase resulting in a broad and, to first-order, level radiation pattern within the angular span of the transducer arc. On receive, appropriately phase-adjusted elemental signals are real-time summed over multiple elements to yield an overlapping fan of 128 equispaced discrete beams of approximately 3° (longitudinal)×20° (latitudinal) width, spanning 180°. Baseband-demodulated elemental voltages can be alternatively recorded in complex form (Knight *et al.*, 1981) enabling flexible off-line beamforming, often better suited to system calibration, field data quantification, and exacting scientific analysis (Melvin *et al.*, 2003).

For fisheries applications the sonar unit is either hull deployed or towed with the array equatorial plane oriented

port–starboard (athwartships) and the array center nadir-directed (Fig. 1). At a typical 75-m sonar range setting, a 5/s maximum ping rate, and a 5-m/s survey vessel speed, complete port–starboard coverage (overlapping −3-dB points) is achieved for ranges exceeding 5 m with acceptable signal to noise for all beams lying within the active transducer arc. Any water column backscatter from ranges exceeding the transducer to bottom separation is often unacceptably contaminated by bottom-reflected energy. Contamination arises from a combination of receive beam sidelobe leakage, low-level preamp cross talk, and beamforming nonlinearities induced by bottom signal clipping. This is difficult to avoid when sonar directivity is primarily achieved in receive only. Typical SM 2000 fisheries imagery is displayed in Mayer *et al.* (2002) and Melvin *et al.* (2003). Preliminary investigations of SM 2000 target discrimination and localization performance have been reported by Chu *et al.* (2001a, 2001b).

III. THEORY

In this section, theoretical transmit and receive beam patterns are derived and the extraction of quantitative acoustic target strength and volume backscattering strength is addressed. Beginning with target strength (TS), defined for a single ensonified target, the theory is extended to synthesized sonar beams and then formulated for baseband-demodulated signals. A calibration parameter, $C_{\mathrm{Cal}}(\theta_b)$, is introduced to describe the deviations of a real sonar from ideal performance. Volume backscattering strength (S_v) is defined, first for a single beam, then formulated appropriately for a multibeam system. Evaluation of the integrated beam pattern Ψ_D using both theoretical and measured transmit and receive directivity functions is discussed. Finally, a redefined integrated beam-pattern measure, Ψ_{DS}, is introduced for simplified practical estimation of S_v.

A. Target strength

Consider a target at range R in the SM 2000 transducer array equatorial (beam fan) plane and at longitudinal angle θ from the horizontal plane, the center of the active transducer arc pointing vertically downward (Fig. 1).

Define $p_{\mathrm{xmit}}(90°)$ as the transmit pressure amplitude projected outward from the array's center of symmetry when all array elements are excited simultaneously in phase. Quantity p_{xmit} is reduced to a standard reference distance of 1 m. Define p_{rec} to be the returned target echo pressure amplitude at the sonar array face.

$p_{\mathrm{xmit}}(\theta)$ and p_{rec} can be related using the sonar equation (Medwin and Clay 1997)

$$\mathrm{TS}=20\log\frac{p_{\mathrm{rec}}}{p_{\mathrm{xmit}}(90°)D_S(\theta)}+40\log R+2\alpha R. \tag{1}$$

$D_S(\theta)=p_{\mathrm{xmit}}(\theta)/p_{\mathrm{xmit}}(90°)$ is the source directional response function and α the acoustic absorption coefficient.

A directionally sensitive beamformer estimator for p_{rec} enhances both signal-to-noise ratio and target angular discrimination. Implementation is by linear summation of array elemental voltages, $\underline{\mathbf{V}}_n$, individually time delayed to add in phase in the coincident beamforming and target direction, θ_b.

$$\underline{\mathbf{V}}_s(\theta_b,t)=\sum_{n=N_1}^{N_2}\underline{\mathbf{V}}_n(t+\Delta t_n)W(\theta_b,\theta_n). \tag{2}$$

For a specific θ_b, N_1 and N_2 denote the summation aperture of the array over elements of index number n, i.e., the range of discrete elements within which a predefined summation window weighting function, $W(\theta_b,\theta_n)$, is nonzero.

$\Delta t_n=\Delta l_n/c$ is the element-specific propagation path time delay. Δl_n is the propagation path length difference from an equatorial plane source at infinity at beamforming angle θ_b to, respectively, the array-surface center of element n, and the nearest point to the source on the (assumed) circular array surface.

$$\Delta l_n=(1-\cos(\Delta\theta_n))a_T, \tag{3}$$

where $\Delta\theta_n=|\theta_n-\theta_b|$, θ_n is the angular position of the nth transducer element, and a_T the array radius of curvature.

Down conversion of elemental voltages to baseband using a common phase reference (Knight *et al.*, 1981) can reduce data handling requirements in practical beamformer applications. If (complex) V_s and V_n denote down-converted $\underline{\mathbf{V}}_s$ and $\underline{\mathbf{V}}_n$, respectively, $\underline{\mathbf{V}}_n(t)=\sqrt{2}\,\mathrm{Re}[V_n(t)e^{i\omega t}]$ (Knight *et al.*, 1981) and similarly for $\underline{\mathbf{V}}_s(t)$, permitting Eq. (2) to be rewritten

$$V_s(\theta_b,t)=\sum_{n=N_1}^{N_2}V_n(t+\Delta t_n)W(\theta_b,\theta_n)e^{ik\Delta l_n}. \tag{4}$$

The V_n summation now involves the comparatively slowly varying, complex modulation envelopes of the elemental voltages at transmit frequency ω. The acoustic wave number $k=\omega/c$, c being the acoustic propagation velocity near the transducer. Further assuming a ''steady-state'' monochromatic approximation for a sonar pulse of sufficient time duration that the amplitude of the target echo envelope can be considered constant over a time comparable to the elemental arrival time differentials, Δt_n, Eq. (4) can be rewritten

$$V_s(\theta_b,t)=\sum_{n=N_1}^{N_2}V_n(t)W(\theta_b,\theta_n)e^{ik\Delta l_n}. \tag{5}$$

Heier (1987) recommends a minimum pulse duration of 10× the spread in arrival times for precise quantification.

Individual array elements are characterized by both a common directional response function, $D_{e\theta}(\alpha_i)=V(\alpha_i)/V(0°)$, where α_i is the incidence angle of the incoming pressure wave, and an on-axis pressure to voltage conversion constant (including electronics), $K_{v/p}$ (Medwin and Clay, 1997). Continuing with a pressure source in the beamforming direction θ_b, Eq. (5) may be expressed in terms of the received pressure envelope, $p_{\mathrm{rec}}(\theta_b,t)$

$$V_s(\theta_b,t)=K_{v/p}\sum_{n=N_1}^{N_2}(p_{\mathrm{rec}}(\theta_b,t)e^{-ik\Delta l_n})D_{e\theta}(\Delta\theta_n)\times W(\theta_b,\theta_n)e^{ik\Delta l_n}. \quad (6)$$

Equating Eqs. (5) and (6)

$$p_{\mathrm{rec}}(\theta_b,t)=\frac{\Sigma_{n=N_1}^{N_2}V_n(t)W(\theta_b,\theta_n)e^{ik\Delta l_n}}{K_{v/p}\Sigma_{n=N_1}^{N_2}D_{e\theta}(\Delta\theta_n)W(\theta_b,\theta_n)}. \quad (7)$$

Equation (7) serves as an estimator of incident source-field pressure amplitude since p_{rec}, unlike V_s in Eq. (5), is independent of chosen θ_b for an on-axis source. This is a consequence of the V_n sum normalization applied in the denominator.

Substituting p_{rec} from Eq. (7) in Eq. (1) with $\theta=\theta_b$ and removing the $40\log R+2\alpha R$ range dependency by applying an equivalent spreading and absorption time variable gain (TVG) function to the elemental voltages prior to summation

$$\mathrm{TS}(\theta_b,t)=20\log\frac{|\Sigma_{n=N_1}^{N_2}V_n(t)W(\theta_b,\theta_n)e^{ik\Delta l_n}|}{|p_{\mathrm{xmit}}(90°)D_S(\theta_b)K_{v/p}\Sigma_{n=N_1}^{N_2}D_{e\theta}(\Delta\theta_n)W(\theta_b,\theta_n)|}. \quad (8)$$

Note that TS inherits the θ_b and t dependencies of p_{rec}. Moving the product of constants $p_{\mathrm{xmit}}(90°)$ and $K_{v/p}$ outside the log term as an additive system calibration constant, C_{Cal}, for which a nominal θ_b dependency has been annotated for future consideration

$$\mathrm{TS}(\theta_b,t)=20\log\left(\frac{|\Sigma_{n=N_1}^{N_2}V_n(t)W(\theta_b,\theta_n)e^{ik\Delta l_n}|}{\Sigma_{n=N_1}^{N_2}W(\theta_b,\theta_n)D_{e\theta}(\Delta\theta_n)}D_S^{-1}(\theta_b)\right)+C_{\mathrm{Cal}}(\theta_b). \quad (9)$$

Signal paths are assumed sufficiently long to allow identical spreading and absorption corrections to be applied to each array element.

$D_{e\theta}(\Delta\theta_n)$ is the measured elemental transmit response, $p_e(\Delta\theta_n)/p_e(0°)$. Interelement baffles in the SM 2000 prevent its accurate derivation from simple planar radiator theory.

For simultaneous in-phase elemental excitation, the monochromatic, steady-state response is given by

$$D_S(\theta_b)=S_{DS}(\theta_b)/S_{DS}(90°), \quad (10)$$

where

$$S_{DS}(\theta_b)=\left|\sum_{n=N_3}^{N_4}D_{e\theta}(\Delta\theta_n)e^{ik\Delta l_n}\right|. \quad (11)$$

Summation limits, N_3 to N_4, encompass all array elements within $\pm 90°$ of θ_b, e.g., all elements visible from infinity in direction θ_b. Normalization is by the identical summation about the array center ($\theta_b=90°$).

For a theoretically "ideal" sonar, C_{Cal} in Eq. (9) is independent of θ_b. A single beamformer sum performed for an acoustic calibration target placed on-axis in *any* one beam suffices to define C_{Cal}, thereby, yielding a system (axially) calibrated for all beamforming angles, θ_b.

To accommodate target strength extraction from real as opposed to ideal sonars, a θ_b dependency has been introduced for C_{Cal} in Eq. (9). The angular dependency arises from elemental response mismatches and other minor system-specific perturbations from ideality in both the transmit source field and in the receive beamformer performance. $D_S(\theta_b)$ remains the theoretically computed directivity. The degree of independence of C_{Cal} on θ_b can be considered a quality measure of real sonar performance. While additional simplification of Eq. (9) is possible by the incorporation of both $D_S(\theta_b)$ and the divisive normalization terms into $C_{\mathrm{Cal}}(\theta_b)$ (one may legitimately proceed this way operationally), the θ_b dependence immediately becomes first order and the perturbation perspective is lost. Systematic observations of an acoustic calibration target angularly swept in the sonar equatorial plane allows straightforward determination of $C_{\mathrm{Cal}}(\theta_b)$, thereby achieving calibration of a real sonar.

B. Volume backscattering strength

1. General computation

Volume backscattering strength, denoted S_v (Medwin and Clay, 1997) is a fundamental quantitative measure in acoustic survey requiring neither on-axis target ensonification nor spatially resolved scatterers for effective utilization. Conceptually, S_v is the target strength of a unit volume of diffusely backscattering medium. Utilizing a short acoustic pulse

$$S_v(\theta_b,t)=\mathrm{TS}(\theta_b,t)-10\log\left((\Psi_D(\theta_b)/4\pi)4\pi R^2(t)\frac{c\Delta t}{2}\right), \quad (12)$$

where R is the instantaneous time dependent range, Δt the sonar pulse width and, $\Psi_D(\theta_b)$ is the combined transmit–receive integrated beam pattern applicable to a specific receive beam synthesized in direction θ_b.

The major bracketed term in (12) represents the "effective" instantaneously ensonified water volume at time-dependent range $R(t)$, for $c\Delta t \ll R(t)$, based on the total received energy flux. The latter is proportional to the beamformer output voltage squared. Ψ_D is the two-way "integrated beam pattern" (Medwin and Clay, 1997) or "integrated beam width factor" (Clay and Medwin, 1977). $\Psi_D/4\pi$ is the fraction of the diffusely backscattering spheri-

cal shell enclosing the transducer(s) ''effectively'' ensonified. It is defined by the ratio of total received acoustic power from the shell as measured by the receiving transducer (i.e., beamformed receiving array), to the power received using omnidirectional transducer(s) of similar axial transmit and receive sensitivities. Consequently

$$\Psi_D(\theta_b) = \int_{4\pi} (D_S(\theta_b,\theta,\phi) D_R(\theta_b,\theta,\phi))^2 d\Omega, \qquad (13)$$

where both D_S and D_R are defined in (θ,ϕ) unit spherical coordinates; ϕ denotes the latitudinal angle from the transducer equatorial plane of symmetry. $D_S(\theta_b,\theta,\phi)$ is the 3D transmit source acoustic directional response function defined as the ratio of projected pressure (measured at infinity) in direction (θ,ϕ) to the pressure projected along the θ_b *receive* beam axis, i.e., direction $(\theta_b,0°)$. The transmitted pressure amplitude is not θ_b dependent—only the reference pressure used to define D_S. Note that the triple argument form of D_S is defined differently than the single argument $D_S(\theta_b)$ of Eq. (9), the latter being equivalent to $D_S(90°,\theta_b,0°)$. $D_R(\theta_b,\theta,\phi)$ is the 3-D receive directional response function of the receive beamformer. It is defined as the ratio of the voltage output for a pressure source at infinity in direction (θ,ϕ), to the voltage output with the identical source placed on the receive beam axis i.e. direction $(\theta_b,0°)$. Integration is over solid angle, Ω, of the unit sphere enclosing the transducer.

TS in (12) is evaluated by Eq. (9)—except that $V_n(t)$ are precorrected to a $20\log R$ rather than $40\log R$ TVG response. The R^2 factor in (12) compensates the resultant TS term undercorrection, removing R from the S_v computation.

D_S in (13) can be evaluated by summing appropriately phase-delayed contributions from each planar element at infinity, all normalized by the identical sum for contributions directed along the θ_b beam axis

$$D_S(\theta_b,\theta,\phi) = S_{DS2}(\theta_b,\theta,\phi)/S_{DS2}(\theta_b,\theta_b,0°), \qquad (14)$$

where

$$S_{DS2}(\theta_b,\theta,\phi) = \left| \sum_{n=N_3}^{N_4} D_{e\theta}(\Delta\theta_s) D_{e\phi}(\phi) e^{ik\Delta l_s \cos\phi} \right|. \qquad (15)$$

Using the notation $\Delta\theta_s = |\theta_n - \theta|$, Δl_s's are computed in analogy to Δl_n previous, using $\Delta\theta_s$ in place of $\Delta\theta_n$. N_3 and N_4 define the range of transducer elements visible from direction θ. Expression (10) is recognized as the special case $D_S(90°,\theta_b,0°)$, i.e., the equatorial plane transmit directivity function in a given beamforming direction normalized to the central (90°) beam axial transmit response.

The $D_{e\theta}D_{e\phi}$ terms in (15) represent the planar element 2D directional response function expressed as the product of the corresponding one-dimensional line radiator responses (Medwin and Clay, 1997). One assumes the baffled elemental radiator response to be adequately approximated by the product of the line radiator response in the larger physical element (latitudinal) dimension and the empirically measured response in the lesser elemental (longitudinal) dimension. The latitudinal response is a sinc function

$$D_{e\phi}(\phi) = \frac{\sin\left(k\frac{w}{2}\sin\phi\right)}{k\frac{w}{2}\sin\phi}, \qquad (16)$$

where w is the common latitudinal (fore–aft) rectangular element dimension.

D_R in (13) is similarly evaluated by summing the elemental contributions from a point pressure source at infinity and normalizing by the corresponding sum for the source placed in the beamforming axial direction $(\theta_b,0°)$

$$D_R(\theta_b,\theta,\phi) = S_{DR2}(\theta_b,\theta,\phi)/S_{DR2}(\theta_b,\theta_b,0°), \qquad (17)$$

where

$$S_{DR2}(\theta_b,\theta,\phi) = \frac{\left|\sum_{n=N_1}^{N_2} W(\theta_b,\theta_n) D_{e\theta}(\Delta\theta_s) D_{e\phi}(\phi) e^{ik(\Delta l_n - \Delta l_s\cos\phi)}\right|}{\sum_{n=N_1}^{N_2} W(\theta_b,\theta_n) D_{e\theta}(\Delta\theta_n)}. \qquad (18)$$

The numerator in Eq. (17), defined by Eq. (18), is analogous to Eq. (7) with elemental voltages replaced by properly phased incident pressures and with a ϕ dependency added. The denominator of Eq. (17) references the directional response function to the θ_b beam axis in the equatorial plane.

Again, N_1, N_2, and the spatial path differentials, Δl's, must be chosen appropriate to each θ_b. N_1, N_2, $W(\theta_b,\theta_n)$, and Δl_b are functions of θ_b rather than θ.

Expression (13) can be numerically evaluated rigorously using (14) and (17) together with empirical $D_{e\theta}(\Delta\theta)$ and $D_{e\phi}(\phi)$ from (16). This computation can proceed in the (θ,ϕ) coordinate system by summation over appropriately chosen angular segments on incremental concentric rings about the sonar's polar axis (ϕ=90°). We have chosen, alternatively, to perform the numerical integration in a new spherical-polar coordinate system defined about the beamforming axial vector $(\theta_b,0°)$ such that the summation is performed on concentric rings about the beamforming axis.

2. Approximate method

A slightly less rigorous approach to integral evaluation, whereby the 3D integral (13) is reduced to a 2D approximation, yields significant computational savings while maintaining reasonable accuracy. This approach also furnishes

simplified mathematical forms for the incorporation of empirically measured directivities appropriate to a realizable sonar. One begins with two approximations

$$D_R(\theta_b,\theta,\phi)\cong D_{R\phi}(\theta_b,\phi)D_{R\theta}(\theta_b,\theta), \tag{19}$$

$$D_S(\theta_b,\theta,\phi)\cong D_{S\phi}(\theta_b,\phi)D_{S\theta}(\theta_b,\theta). \tag{20}$$

The above forms are equivalent to assuming $\cos(\phi)=1$ (small ϕ-phase approximation) in (15) and (18). The approximations decouple terms in θ and ϕ in (13), allowing

$$\Psi_D(\theta_b)\cong \int (D_{S\phi}(\theta_b,\phi)D_{R\phi}(\theta_b,\phi))^2 d\phi \times \int (D_{S\theta}(\theta_b,\theta)D_{R\theta}(\theta_b,\theta))^2 d\theta. \tag{21}$$

Also assumed is that incremental solid angle $d\Omega = \cos\phi\, d\phi\, d\theta$ can be approximated by $d\phi\, d\theta$ over a restricted range of ϕ. The first integral in (21) is considered to be the product of two identical sinc functions independent of θ_b, as implied by the theoretical forms (15) and (18), yielding

$$\Psi_D(\theta_b)\cong \int D_\phi^4 d\phi \int (D_{S\theta}(\theta_b,\theta)D_{R\theta}(\theta_b,\theta))^2 d\theta. \tag{22}$$

For $l\times w$-dimensioned rectangular radiators describable by sinc-function-product directivities, Clay and Medwin (1977) quote the evaluation formula

$$\Psi_D=\int_{4\pi}(D_\phi^2 D_\theta^2)^2 d\Omega\cong\frac{17.4}{k^2wl} \quad \text{for } kl \text{ and } kw\gg 1. \tag{23}$$

For small ranges of ϕ and θ (23) can be rewritten

$$\int_\phi D_\phi^4\, d\phi \int_\theta D_\theta^4\, d\theta\cong\frac{(17.4)^{1/2}}{kw}\times\frac{(17.4)^{1/2}}{kl}. \tag{24}$$

Since both terms in (24) refer to line elements stemming from planar radiator theory, the expression's symmetry allows one to insert the empirical line element approximation for the first term in (22)

$$\Psi_D(\theta_b)\cong\frac{(17.4)^{1/2}}{kw}\int (D_{S\theta}(\theta_b,\theta)D_{R\theta}(\theta_b,\theta))^2 d\theta. \tag{25}$$

The remaining one-dimensional integral in (25) can be readily numerically approximated and precomputed for discrete θ_b, ultimately enabling S_v evaluation from Eq. (12)

$$D_{S\theta}(\theta_b,\theta)=S_{DS\theta}(\theta_b,\theta)/S_{DS\theta}(\theta_b,\theta_b), \tag{26}$$

where

$$S_{DS\theta}(\theta_b,\theta)=\left|\sum_{n=N_3}^{N_4} D_{e\theta}(\Delta\theta_s)e^{ik\Delta l_s}\right|, \tag{27}$$

and

$$\Delta\theta_s=|\theta_n-\theta|. \tag{28}$$

Δl_s's are similarly computed in analogy to Δl_n's above using $\Delta\theta_s$ in place of $\Delta\theta_n$.

$D_{S\theta}$, as defined in Eq. (26), is θ_b dependent only by fact of its being referenced to the transmitted pressure in the θ_b receive beam axial direction. Summation limits N_3 and N_4 again define the index range of transducer elements visible from infinity in direction θ.

$D_{R\theta}$ is evaluated analogous to Eq. (17) with $\phi=0°$

$$D_{R\theta}(\theta_b,\theta)=S_{DR\theta}(\theta_b,\theta)/S_{DR\theta}(\theta_b,\theta_b), \tag{29}$$

where

$$S_{DR\theta}(\theta_b,\theta)=\frac{\left|\sum_{n=N_1}^{N_2} W(\theta_b,\theta_n)D_{e\theta}(\Delta\theta_s)D_{e\phi}(\phi)e^{ik(\Delta l_n-\Delta l_s)}\right|}{\sum_{n=N_1}^{N_2} W(\theta_b,\theta_n)D_{e\theta}(\Delta\theta_n)}. \tag{30}$$

N_1, N_2, and the Δl's must be chosen appropriate to each θ_b. Again N_1, N_2, $W(\Delta\theta_b)$, and Δl_b are functions of θ_b rather than of θ.

The one-dimensional integral in Eq. (25) is easily computed from theory, but, more importantly, its form invites empirical evaluation from experimentally measured directivity function products. Note that $D_{S\theta}(\theta_b,\theta)\cdot D_{R\theta}(\theta_b,\theta)$ is directly proportional to the resultant θ_b beamformer amplitude for a standard calibration target placed at angle θ.

3. Additional considerations

Simplified S_v computation using Eq. (12) in conjunction with (9) and (13) is achieved by referencing all directional response functions to the central (90°) beam axial response rather than to individual θ_b beam axial responses.

$D_R(\theta_b,\theta,\phi)$ in Eq. (13), as evaluated from Eq. (17), has been defined as $V_{BF}(\theta_b,\theta,\phi)/V_{BF}(\theta_b,\theta_b,0°)$, where $V_{BF}(\theta_b,\theta,\phi)$ denotes the normalized beamformer voltage amplitude while beamforming in direction θ_b for a source at (θ,ϕ). The fully normalized beamformer summation for an on-axis source in (17) is independent of θ_b, specifically

$$V_{\mathrm{BF}}(\theta_b,\theta_b,0°)=V_{\mathrm{BF}}(90°,90°,0°). \tag{31}$$

Therefore, $D_R(\theta_b)$ values referenced either way are, in theory, equivalent.

However, in transmit

(1) $D_S(\theta_b)$ in (9), as noted previously, is $p_{\mathrm{xmit}}(\theta_b,0°)/p_{\mathrm{xmit}}(90°,0°)$.
(2) $D_S(\theta_b,\theta,\phi)$ in (13) is $p_{\mathrm{xmit}}(\theta,\phi)/p_{\mathrm{xmit}}(\theta_b,0°)$.

If, in Eq. (12), $D_S(\theta_b)$ implicit in the TS(θ_b,t) term defined by Eq. (9) is moved inside the Ψ_D integral, as defined by Eq. (13), the resultant D_S product yields $D_S(90°,\theta,\phi)$, in our previously defined notation. This quantity is $p_{\mathrm{xmit}}(\theta,\phi)/p_{\mathrm{xmit}}(90°,0°)$, the transmit directional response function referenced to the central (90°) beam axis. The resultant modified Ψ_D integral-containing central beam referenced D_S is denoted Ψ_{DC}.

The R^2 term in (12) disappears on using a $20\log R$ system TVG, leaving

$$S_v(\theta_b,t)=20\log\left(\frac{|\Sigma_{n=N_1}^{N_2}V_n(t)W(\theta_b,\theta_n)e^{ik\Delta l_n}|}{\Sigma_{n=N_1}^{N_2}W(\theta_b,\theta_n)D_e(\Delta\theta_n)}\right)$$

$$+C_{\mathrm{Cal}}-10\log\left(\Psi_{DC}(\theta_b)\frac{c\Delta t}{2}\right). \quad (32)$$

The initial bracketed term is the beamformer output fully normalized. C_{Cal} is a constant, independent of θ_b, since an ''ideal'' sonar is being considered.

To compute S_v for ''real'' sonars, one again starts with definition (12) and TS(θ_b,t) defined by (9), but the θ_b dependence of C_{Cal} is retained. $\Psi_D(\theta_b)$ must be computed from *measured* rather than theoretical directivity functions. Again, it is convenient to state Ψ_D in terms of Ψ_{DC} characterized by constituent D_S's and D_R's normalized to the central (90°) beam axial responses. Measured beam-axial D_R is now θ_b dependent

$$\Psi_D(\theta_b)=\left(\frac{p_{\mathrm{xmit}}(90°,0°)}{p_{\mathrm{xmit}}(\theta_b,0°)}\bigg|_{\mathrm{meas}}\times\frac{V_{\mathrm{BF}}(90°,90°,0°)}{V_{\mathrm{BF}}(\theta_b,\theta_b,0°)}\bigg|_{\mathrm{meas}}\right)^2$$

$$\times\Psi_{DC}(\theta_b). \quad (33)$$

Proceeding from (33) and equating the outputs of Eq. (12) for identical targets placed successively at θ_b and at 90° in the equatorial plane

$$-10\log D_S^2(\theta_b)\Psi_D(\theta_b)$$

$$=-20\log\left(\frac{p_{\mathrm{xmit}}(\theta_b,0°)}{p_{\mathrm{xmit}}(90°,0°)}\bigg|_{\mathrm{theor}}\times\frac{p_{\mathrm{xmit}}(90°,0°)}{p_{\mathrm{xmit}}(\theta_b,0°)}\bigg|_{\mathrm{meas}}\right.$$

$$\left.\times\frac{V_{\mathrm{BF}}(90°,90°,0°)}{V_{\mathrm{BF}}(\theta_b,\theta_b,0°)}\bigg|_{\mathrm{meas}}\right)-10\log\Psi_{DC}(\theta_b)$$

$$=C_{\mathrm{Cal}}(90°)-C_{\mathrm{Cal}}(\theta_b)-10\log\Psi_{DC}(\theta_b). \quad (34)$$

If, in Eq. (12), the $D_S(\theta_b)$ factor in the TS(θ_b) term, defined by Eq. (9), is moved to the trailing Ψ_D containing term, and if Ψ_{DC} is substituted for Ψ_D using Eq. (34), the expression for real-sonar S_v reduces to a form analogous to Eq. (32)

$$S_v(\theta_b,t)=20\log\left(\frac{|\Sigma_{n=N_1}^{N_2}V_n(t)W(\theta_b,\theta_n)e^{ik\Delta l_n}|}{\Sigma_{n=N_1}^{N_2}W(\theta_b,\theta_n)D_e(\Delta\theta_n)}\right)$$

$$+C_{\mathrm{Cal}}(90°)-10\log\left(\Psi_{DC}(\theta_b)\frac{c\Delta t}{2}\right). \quad (35)$$

Introduction of central beam normalized Ψ_{DC} renders C_{Cal} a constant to be evaluated for $\theta_b=90°$, the central beam axis. The nonideal properties of a ''real'' sonar are now confined to the set of system-specific integrated beam-pattern coefficients, $\Psi_{DC}(\theta_b)$, which must be evaluated experimentally beam by beam.

In practice, $C_{\mathrm{Cal}}(90°)$ is readily obtained by use of a calibrated acoustic target. $\Psi_{DC}(\theta_b)$ can be estimated by several techniques of varying rigor. If irregularities in the latitudinal response are less than those in the equatorial response, Eq. (25) affords a practical approach. While latitudinal responses are derived from theory, the integral term can be empirically approximated for each θ_b as the finite sum

$$\sum\,[D_{S\theta}(\theta_b,\theta)D_{R\theta}(\theta_b,\theta)]^2\Delta\theta.$$

$D_{S\theta}$ and $D_{R\theta}$ must be renormalized to the directivities of the synthesized 90° beam axis to yield Ψ_{DC} rather than Ψ_D. Renormalization yields real-sonar $D_{S\theta}D_{R\theta}$ products directly proportional to the experimental θ_b beamformer amplitude for an equatorial plane calibration target placed at angle θ. In principle, a single continuous set of elemental amplitude observations, while a calibration target is moved slowly through the total range of θ, contains sufficient information to compute all relevant beamformer amplitudes and, in turn, the complete set of $\Psi_{DC}(\theta_b)$.

TABLE I. Empirical single element directional response, $D_{e\theta}(\Delta\theta)$, as measured by the manufacturer.

Angle $\Delta\theta$°	20 log D
0	0
2	−0.05
5	−0.14
10	−0.34
15	−0.65
20	−0.98
30	−2.44
40	−4.61
60	−9.58
80	−17.3
89	−21.3
>89	−∞

IV. THEORETICAL RESULTS

In this section, theoretical response patterns for a commercial SM 2000 multibeam sonar are presented in graphical form. Transmit and receive responses are illustrated in the longitudinal (equatorial) plane and in a select latitudinal plane. Integrated beam patterns are derived and compared over a range of beamforming angles using both rigorous theory and theory with simplifying approximations. We begin by first tabulating relevant sonar processing descriptors in subsection A prior to considering specific results in subsection B.

A. Modeling and processing parameters

For the theoretical simulations below, 1441 equiincremental beam directions over 180° are computed yielding visually continuous responses. Experimental measurements utilize either 128 or 256 beam directions over 180°. Sound velocity $c=1490$ m/s yielding $kw\cong 17.9$. The empirically measured longitudinal elemental response appears in Table I.

The window function for the current analysis (identical to that used in real-time firmware) is

$$W(\theta_b,\theta_n)=a+(1-a)\cos\left(\pi\left(\frac{|d_n-d_{\mathrm{mid}}|}{2d_{\mathrm{mid}}}\right)\right), \quad (36)$$

with $a=0.54$, yielding a Hamming response form. A rule-based 150° ''nominal'' (maximum) window summation aperture is utilized.

(a) Windowing is applied to a portion of the array arc projected onto a line perpendicular to the beamforming direction: Specifically, ''d_n'' is the lateral position of a perpendicular dropped from the array transducer at θ_n onto a line normal to the beamforming direction, θ_b. The normal line spans a designated array arc. ''Spanning'' means that vectors drawn from the ends of the normal line parallel to the beamforming direction just enclose a designated array arc. The ''designated array arc'' is the arc enclosing all transducers lying within a limiting angle, namely one half the nominal summation aperture, of the beamforming direction.

(b) If an ''end'' transducer is encountered within the limiting angle (150°/2), the end transducer defines one end of the designated arc. The quantity $|d_n - d_{\mathrm{mid}}|$ defines the displacement of d_n from the normal line midpoint, d_{mid}.

The window function is always symmetric about d_{mid}. However, if a transducer array end is encountered for a specific θ_b (the common case, since the nominal 150° summation aperture is just less than the 155° physical transducer arc), the effective summation aperture is reduced from the nominal value (150°), and the window function maximum also ceases to be coincident with the beamforming direction.

B. Specific results

Fundamental theoretical properties of the SM 2000 multibeam sonar are now explored. First, comparison of single-beam directivity responses in the equatorial plane reveals systematic beam widening on moving from the central beams to those oriented near the array angular extremes and beyond. The narrow symmetrical receive pattern computed by Eq. (17) for the 90° beam (total beamwidth 2.9° at the −3-dB points) is compared to the wider pattern (beamwidth 5.4°) for a beam at 12.5° (array edge) in Fig. 2. Widening of the receive pattern results from the decreased summation aperture employed for beams near the array edge. The Hamming window retains excellent sidelobe suppression (>40 dB) even in the presence of the sidelobe asymmetry arising from angular displacement of the beamforming direction from the summation aperture center.

The anticipated ''level'' beam-axial receive response for a moving pressure source in the sonar equatorial plane while beamforming in the target direction is shown in Fig. 3. This result constitutes a consistency check on the receive modeling algorithms employing full beamformer normalization. Simplified beamforming, by the omission of individual elemental directivities in the beamformer sum normalizations, which characterizes the manufacturer's real-time firmware in contrast to our off-line beamforming, results in minor deviations from the level response (Fig. 3).

The effect of employing more the rigorous Eq. (17) rather than the approximation of Eq. (16) in computation of the ϕ-dependent ''latitudinal'' receive response for $\theta_b=90°$ is shown in Fig. 4 while central lobe amplitude agreement is excellent, increasing divergence is noted for sidelobes beyond 20° from the equatorial plane. The angular placement of sidelobes is quite similar in both cases.

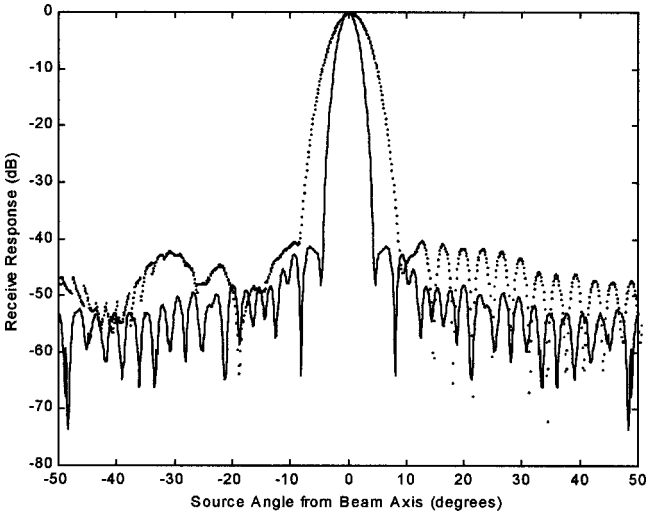


FIG. 2. Computed receive response ($20 \log D_R$) for moving source in equatorial plane ($\phi=0°$) and single fixed beams synthesized at $\theta_b=90°$ (solid line) and at $\theta_b=12.5°$ (points).

Transmit directivity functions rigorously computed from Eq. (14) for the sonar equatorial plane and for the 90° meridian (ϕ response) are shown in Figs. 5 and 6, respectively. Both directivities in decibel form are relative to the response at ($\theta_b=90°$, $\phi=0°$). The sinc function meridianal response approximation from Eq. (16) is shown in Fig. 6 for comparison. A comparatively high polar region response (Fig. 6) characterizes the rigorous computation due to radiated elemental pressures adding in-phase as $\phi \rightarrow 90°$, an effect not captured by the approximation. Lack of accurate modeling of the low-angle elemental radiation ''shadowing'' effects and diffractions off the physical ends of the unit probably pre-

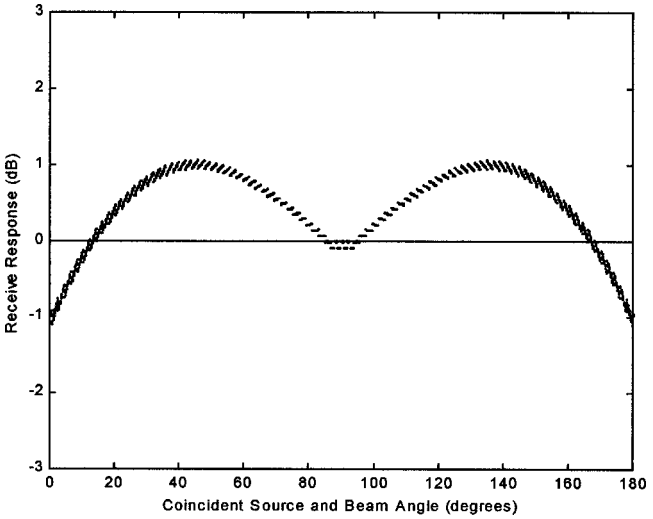


FIG. 3. Computed beam-axial receive responses in sonar equatorial plane ($\phi=0°$) referenced to response at $\theta_b=\theta=90°$. Constant amplitude pressure source moves with beamforming direction ($\theta_b=\theta$). The level, apparently continuous line (1440 angular points) is computed with full beamformer normalization. Discontinuous points show effect of omitting elemental directivities from beamformer normalization.

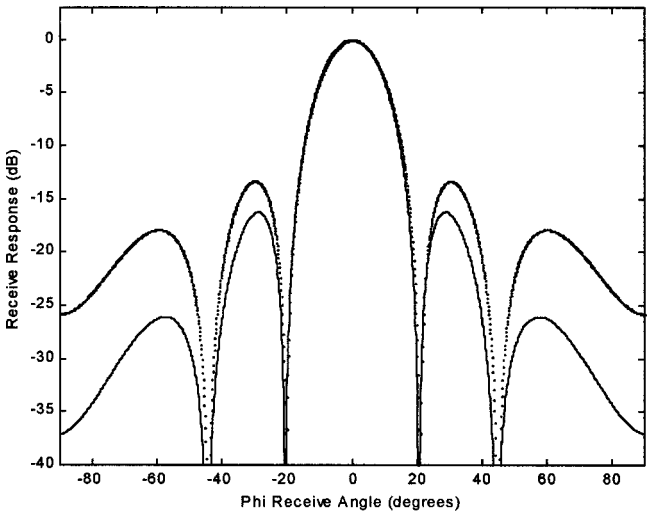


FIG. 4. ''Phi''-dependent receive responses at $\theta_b = \theta = 90°$. Rigorous computation—solid line. Line element, sinc function approximation—points.

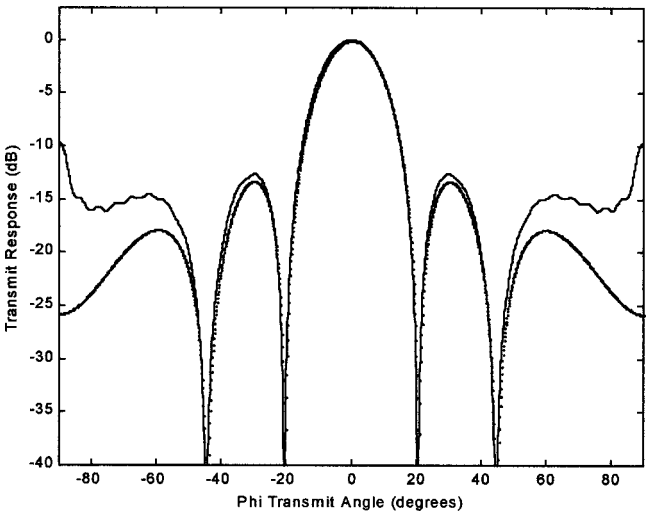


FIG. 6. ''Phi''-dependent transmit response at $\theta = 90°$ relative to response at $\theta_b = 90°$, $\phi = 0°$. Rigorous computation (solid line). Line radiator (sinc function) approximation (points).

clude accurate modeling very close to the poles.

Normalized theoretical $\Psi_{DC}(\theta_b)$, specifically, the decibel measure $10\log(\Psi_{DC}(\theta_b)/\Psi_{DC}(90°))$, is shown in Fig. 7. From Eq. (32), this quantity is seen to be equivalent to a plot of relative beamformer power (dB) versus beamforming angle θ_b for a uniform isotropic scattering medium. Three virtually coincident solid curves demonstrate the quantity is relatively insensitive to slight variants in estimational technique, providing elemental directivities are included in beamformer sum normalizations. The first curve is generated from Eq. (13), only using 90° beam referenced directivities, by rigorous numerical integration on the unit sphere to 45° off the axial vector of each discrete receive beam. Spherical coordinate integration increments of $\pi/800$ and $\pi/200$ radians in ''latitude'' and ''longitude,'' respectively, are utilized within the redefined coordinate system about each beam vector. Absolute $\Psi_{DC}(90°)$ is 1.233×10^{-2}. The second curve is generated by increasing the integration range to 90° off-axis. Absolute $\Psi_{DC}(90°)$ is 1.236×10^{-2}. The virtually unchanged summation demonstrates the summation integral is rapidly convergent. The third curve is computed using the efficient approximation given by Eq. (25), again with normalization to the central beam responses to yield Ψ_{DC}. Absolute $\Psi_{DC}(90°)$ is 1.217×10^{-2}. Finally, a fourth, distinctive, dotted curve shows the discrepancy induced in Ψ_{DC}, rigorously computed to 45° off-axis, on omitting individual elemental directivities from the sum normalizations. Absolute $\Psi_{DC}(90°)$ is 1.233×10^{-2}. The latter curve may be viewed as the angular response of the SM 2000 firmware real-time beamformer to a similar isotropic scattering medium. Good agreement with the fully normalized beamformers is largely restricted to the near-central beams.

In computing Ψ_{DC}, individual beam transmit directivities have been normalized to the 90° transmit response. Consequently, the modest oscillatory D_S dependency on angle (Fig. 5) is also reflected in Ψ_{DC} (Fig. 7). The systematic

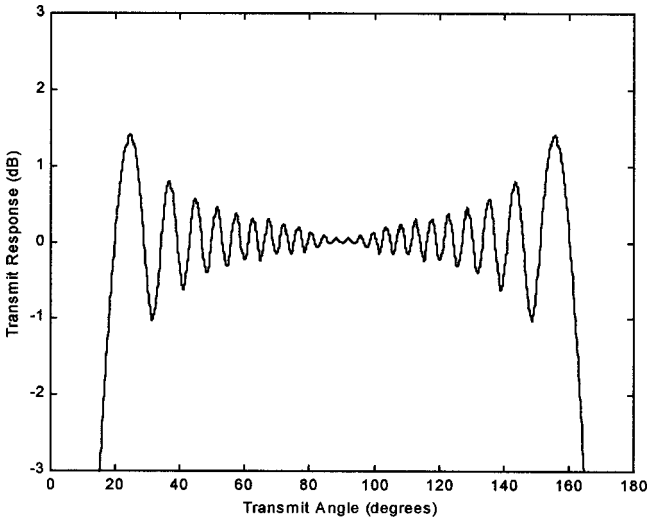


FIG. 5. Transmit response ($20\log D_S$) on equatorial plane ($\phi = 0°$) relative to response at $\theta_b = 90°$.

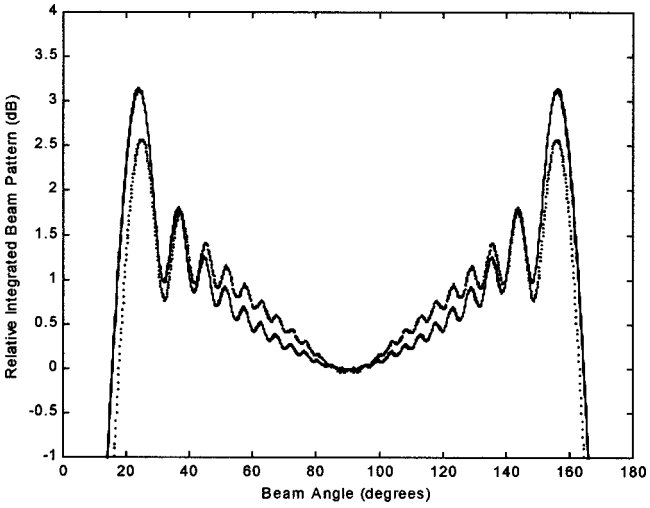


FIG. 7. Theoretical Ψ_{DC} versus beamforming angle. All curves plotted in dB (10 log ratio) relative to curve specific 90° beam value. Three nearly perfectly superimposed solid curves represent computational variants. A fourth curve (points) ignores the effect of elemental directivities in beamformer sum normalizations.

increase in Ψ_{DC} on moving outward from the central beam arises from receive beam progressive longitudinal widening, broader beams yielding larger surface integrals. As θ_b nears the physical array edges, the rapid drop-off in D_S dominates over beam widening, causing Ψ_{DC} to fall rapidly.

V. EXPERIMENTAL RESULTS

In this section both longitudinal and latitudinal experimentally measured sonar combined transmit–receive responses are compared to theoretical expectations. The measured combined transmit–receive responses are fundamental to the computation of empirical integrated beam patterns to follow.

Experimental measurements of the SM 2000 far-field combined transmit–receive response were conducted using the Defense R&D Canada—Atlantic's floating acoustic calibration barge facility. The sonar head array was submerged to between 10- and 20-m depth and temperature equilibrated. With the sonar pinging at a rate of 4–5/s the array was rotated at about 0.3°/s. Echoes received by each of the 80 elements from fixed 1.27- and 3.81-cm diameter tungsten carbide calibration spheres were continuously recorded for subsequent beamforming. Sphere target strengths computed by the method of Faran (1951), elaborated by Neubauer *et al.* (1974), were about −45.0 and −39.5 dB, respectively. Array rotation about the sonar head's polar axis allows generation of the combined transmit–receive equatorial response pattern. Similar rotation about the appropriate orthogonal axis provides the combined response along a fixed meridian.

For the above measurements, the transmit pulse length was increased to 350 μs, 10× the system default pulse length for a specified 10-m profiling range, while maintaining the original default receiver bandwidth. The maximum echo time delay between the central and outmost transducers is about 54 μs for a centered 150° array summation arc (Sec. IV A). Individual element directivities reduce the effective summation arc to, at most, 60° either side of the beamforming direction using the conservative −10-dB point of the elemental response. For this 120° ''effective'' aperture, the propagation delay falls to 37 μs or roughly 10% of the pulse width, consistent with the recommendations of Heier (1987) for phase rotation-type beamformers. To provide precise range compensation, a firmware TVG of $40\log R+2\alpha R$ form is applied prior to data logging using α precomputed from measured temperature and salinity by the methods of Francois and Garrison (1982a, b).

Numerical simulation of the SM 2000 near-field equatorial response (Cochrane, 2002) has established minimum target range criteria for accurate calibration. Computations (Cochrane, 2002) demonstrate that the central beam combined transmit–receive response using the Hamming ''low sidelobe'' window and a 150° maximum summation aperture is decreased by 1.5 dB at 2-m range, rapidly falling to 0.9 dB at 3 m, 0.5 dB at 5 m, and about 0.2 dB at 10-m range. For the present illustrations, only target ranges of 6.0 m or greater are considered where the near-field effects should be smaller than other sources of uncertainty. Synthesis of wider receive beams by use of smaller summation apertures reduces near-field effects at a given range.

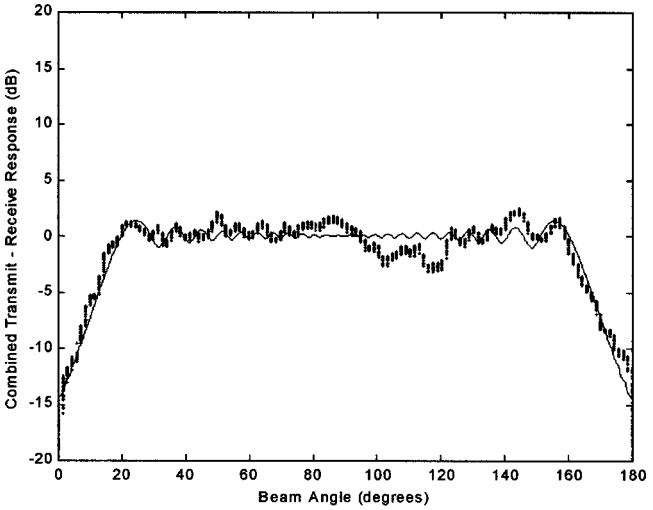


FIG. 8. Combined transmit–receive response inferred using a 3.81-cm diameter TC acoustic calibration sphere in equatorial plane at 6.0-m range (points—upper envelope) superimposed on computed theoretical response (solid line).

Baseband (complex) elemental voltages are beamformed using the normalized summation appearing in the first term of Eq. (35). To approximate the longitudinal combined transmit–receive response, the single overall target echo maximum from the set of all synthesized beams is plotted as a function of the corresponding beamforming direction on a ping-by-ping basis (Fig. 8). In the finished plot, a series of vertically aligned points with successively decreasing vertical spacing is observed for each beamforming direction. This pattern results from the progressive movement of the maximum amplitude echo from an immediately adjacent beam into the steeply rising skirt region of a given beam, through the rounded central region of the beam response, and then down the opposite skirt to eventually pass into the next beam. The upper envelope of the scatter plot constitutes a reasonable approximation to the combined transmit–receive angular response for beam-centered targets. During data reduction to scatter plot form, small excursions in target echo range are followed by tracking the time coordinates of the echo skirts ping by ping. Target amplitude is defined as the root-mean-square of the complex envelope amplitudes of, normally, 3 data points spanning the target echo center. Since the target echo angular reference is implicitly derived from the beamformer itself, the resultant pattern is largely immune to noise arising from (unwanted) random angular motions of the target relative to the sonar during measurement. Superimposed for comparison in Fig. 8 is the theoretical combined transmit–receive (same as transmit) response computed at infinity. Also for comparison in Fig. 9 is a direct hydrophone-measured transmit response at 12.5-m range in the sonar equatorial plane for 1° rotational increments, again superimposed on the theoretical transmit response at infinity.

The measured latitudinal (ϕ) combined transmit–receive response along the 90° meridian while beamforming in a fixed 90° longitudinal direction is shown in Fig. 10. The corresponding rigorously computed theoretical pattern is superimposed. Latitudinal target angles, in contrast to longitudinal angles, must be derived from an external angular ref-

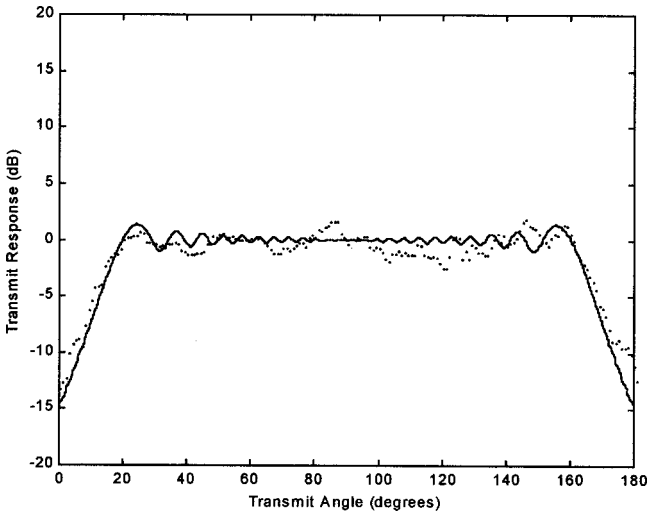


FIG. 9. Measured transmit response using reception hydrophone in equatorial plane at 12.5-m range (points) superimposed on computed theoretical response (solid line).

erence. Consequently, measured ϕ responses are more susceptible to errors induced by unwanted platform motion. The most stable calibration run (shown) was acquired while utilizing a comparatively small target, resulting in a relatively low signal-to-noise ratio.

VI. APPLICATIONS

To assist potential multibeam sonar users, systematic methodologies for extracting calibrated volume backscattering strength from field data are summarized. Extraction from both raw elemental data and real-time processed data is treated.

A. Calibrated S_v extraction from field data

Good quantitative performance is most easily assured by use of sufficiently long transmit pulses and sufficiently wide system bandwidths during both sonar calibration and subsequent field data collection such that

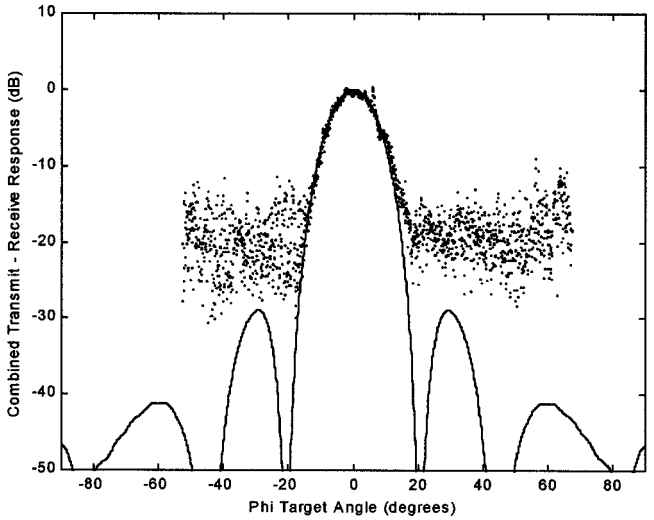


FIG. 10. Measured meridional (ϕ) combined transmit–receive response (points) for $\theta_b = \theta = 90°$ superimposed on rigorously computed theoretical response (solid line).

(a) A reasonable approximation to steady-state beamformer (single frequency, purely phase delay) operation is obtained;
(b) Negligible signal loss occurs due to the sonar's finite passband; and
(c) A single fixed frequency can be assumed in computation of calibration sphere target strengths.

When very short transmit pulses are utilized and high accuracy is required, more general approaches developed for conventional sounders should be applicable to multibeam (Vagle *et al.*, 1996).

The methodology is as follows:

(1) Evaluation of sonar calibration constant $C_{\mathrm{Cal}}(90°)$:

(a) With a suitable acoustic calibration sphere placed in the equatorial plane far field, record complex demodulated elemental voltages, corrected to $40\log R + 2\alpha R$ TVG, while the sonar is pinged and the array slowly rotated about its axis of symmetry.
(b) Beamform calibration data using the isolated normalized summation within Eq. (9) with beamwidths (defined by array summation aperture) used in field applications.
(c) Define $C_{\mathrm{Cal}}(90°)$ using Eq. (9) and the normalized beamformer summation amplitude at 90°, $V_{\mathrm{BF}}(90°,90°,0°)$ by our previous convention.

Since, by definition, $D_S(90°) = 1$

$$C_{\mathrm{Cal}}(90°) = \mathrm{TS}_{\mathrm{Target}} - 20\log V_{\mathrm{BF}}(90°,90°,0°). \tag{37}$$

(2) Numerically evaluate and tabulate $\Psi_{DC}(\theta_b)$ using beamformer outputs generated in step 1(b) above and Eq. (25). The directivities in Eq. (25) should be normalized to the axial transmit and receive responses of the 90° beam (or beam closest to 90°) to yield Ψ_{DC} as opposed to Ψ_D. The directivity products are experimentally given by $V_{\mathrm{BF}}(\theta_b,\theta,0°)/V_{\mathrm{BF}}(90°,90°,0°)$. The integration range over θ should considerably exceed the -3-dB to -3-dB θ_b beamwidth ($\pm 15°$ should be more than sufficient) with angular increments not exceeding 0.1° for the narrowest receive beams. Figure 11 shows an empirically computed Ψ_{DC} using a calibration-sphere-derived combined transmit–receive equatorial plane response measured at 10-m range. A rigorously computed theoretical curve assuming full elemental directivity normalization (Fig. 7) is superimposed.

(3) To extract S_v, beamform field data, corrected to $20\log \mathrm{R} + 2\alpha \mathrm{R}$ TVG, by expression (35) using predetermined $C_{\mathrm{Cal}}(90°)$, tabulated $\Psi_{DC}(\theta_b)$, and the field data pulse width. Alterations in receiver gain settings between calibration and field acquisition must be tracked and compensated in C_{Cal}.

B. S_v from real-time beamformed data

$S_v(\theta_b)$ can be readily evaluated from SM 2000 real-time beamformer outputs rather than the off-line computed beamformer summations. The manufacturer's real-time firmware beamformer omits the $D_e(\Delta\theta_n)$ term from the beam sum normalizations. Consequently, $D_R(\theta_b,\theta_b,\phi)$ referenced to

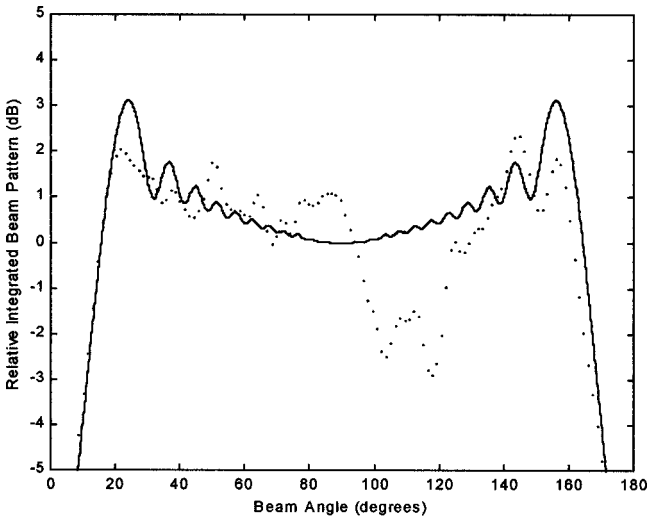


FIG. 11. Empirically computed Ψ_{DC} as function of beam angle (θ_b) derived from experimental combined transmit–receive response shown in Fig. 7 (points) superimposed on theoretical response (solid line). All values are plotted relative to the central theoretical curve value ($\theta_b = 90°$) having absolute magnitude 1.233×10^{-2} (dimensionless).

the 90° beam for an ideal sonar is, in theory, no longer independent of θ_b. Differences between the fully and partially normalized receive responses are, in theory, small but significant (Fig. 3). Consequently, the *theoretical* dependence of $\Psi_{DC}(\theta_b)$ on θ_b will differ between the fully and partially normalized beamformers. However, the normalization methodology is of no concern for S_v estimation using empirically computed Ψ_{DC}—providing that both the calibration data set defining $C_{\mathrm{Cal}}(90°)$ and $\Psi_{DC}(\theta_b)$, and the field data set for subsequent S_v extraction, are reduced with identical beamformers.

The preceding statement is quite general: For the data reduction in this presentation, the summation normalization was specifically chosen to yield, in the *ideal* case, constant beamformer amplitudes for constant amplitude *axial* pressure sources independent of θ_b. This summation form is advantageous in being simply related to acoustic target strength by knowledge of the theoretical transmit directivity function [Eq. (10)]. Differing summation normalization schemes—including the total elimination of any normalization, will yield different $\Psi_{DC}(\theta_b)$ function forms but equally valid estimates of S_v—provided the summation methodology is consistently followed.

VII. DISCUSSION AND CONCLUSIONS

Significant deviations of the measured sonar transmit–receive response from theory are revealed in Fig. 8. The marked broad response dip between angles of 95° and 125° appears real and is repeatable for various target ranges (data not shown). The isolated measured transmit response in Fig. 9 shows a corresponding dip, especially apparent when the curves of Figs. 8 and 9 are superimposed. This angular asymmetry in the transmit–receive response induces about a 3-dB dip in empirically computed Ψ_{DC} (Fig. 11). Clearly, accurate measurement of array responses are necessary to achieve a better than 1-dB accuracy in S_v, an accuracy generally expected for modern fisheries backscatter measures.

One challenging question arises in regard to the interpretation of S_v. In defining a beam-specific integrated beam pattern, Eq. (13), it is implicitly assumed that most of the integral contribution (backscattered power) arises from the beam's main lobe (2.9×18° width for 90° beam). This ensures a meaningful directional discrimination or volume localization of the backscattered signal. Rough estimates—the exact calculation is nontrivial—indicate reasonable directionality is attained for the ideal multibeam sonar, where receive far sidelobe rejection is of the order of −50 dB (Fig. 2). However, if the equatorial plane sidelobe rejection for real systems falls to −20 to −25 dB, as typically observed for the current generation hardware, it becomes increasingly questionable if clear volumetric localization for S_v is actually being achieved. For an isotropic backscattering medium, the ''apparent'' acoustic power returned from the far off-axis region—and for a real multibeam there is no guarantee that the observed ''sidelobe'' energy represents real signal as opposed to spurious intermodulation products—begins to approach that from the main lobe. In these circumstances it may be advantageous to widen the main receive lobe. For example, decreasing the maximum summation arc from 150° to 27° theoretically increases the −3-dB receive beamwidth from 2.9 to 11.0°. The wider beamwidth is still comparable to those used for traditional vertical beam fisheries assessment.

Optimal techniques for shading noncoplanar arrays recently explored by Gallaudet and de Moustier (2000) should provide a small theoretical advantage over the Hamming window function in terms of sidelobe rejection. For the current SM 2000 hardware, elemental mismatches and receiver front-end cross talk appear to be the critical factors determining real sidelobe characteristics. Consequently, further refinement of window functions, at present, seems nonadvantageous.

In regard to other primary acoustic measures, target strengths from Eq. (9) are valid for beam-centered targets only. For an isotropic scatterer randomly located in the water column, a valid target strength might be extracted by seeking the maximum from a 2D array of apparent TS values. The relevant TS array is computed across the fan, with highly overlapping beam patterns, then ping by ping, to allow the target to drift through the fan plane by virtue of the sonar's forward motion. The presence of multiple targets would require a target tracking facility, the operation of which would yield the number of targets transiting the fan by direct counting—an attractive possibility provided targets are sufficiently sparse to be uniquely resolved. For realistic nonisotropic biological targets, monitoring the transit through the fan plane would be more problematic, although a minimum target range criterion might be useful. Direct counting algorithms should be practical even for nonisotropic targets.

Finally, much work remains in both the modeling and the interpretation of off-vertical target strengths of fish. Excellent work has been done (Horne and Jech, 1999; McClatchie *et al.*, 1996; Foote, 1985) but generally, only dip angle dependence in dorsal aspect has been treated—not

both dip angle and orientational azimuth—as required for multibeam interpretation. Clearly, for schooling species like herring, orientation will be highly correlated over target aggregations. School movement relative to the water column can furnish orientational information, but parametrization is nontrivial during routine fisheries survey.

In conclusion, a general theoretical framework with practical methodologies has been advanced for the application of multibeam sonar to fisheries survey. The theoretical and empirical performance of a current generation multibeam sonar has been examined. While performance is sufficient to begin exploration of quantitative applications, significant departures from theoretical expectations, especially higher than anticipated sidelobe energy, indicate that hardware refinement is still possible—and probably necessary. In the long term, hardware will improve. For exacting quantitative applications, the efficacy in addressing the off-vertical-aspect interpretational problem will be critical. Progress in the interpretational problem will determine whether future wide-swath multibeam surveys will be limited to semiquantitative distributional mapping, or, alternatively, evolve into a precise technique for enhanced coverage biomass estimates, with accuracies approaching those achieved by current generation dual and split beam fisheries sounders.

ACKNOWLEDGMENTS

The authors thank many involved in this effort. First, the cooperation of Kongsberg Simrad Ltd., especially Bob Asplin, Steven Wright, and Alex Ivanov at their Mesotech, Port Coquitlam plant, and John Gillis at their Dartmouth office, is acknowledged. The Pelagics Research Council (now Herring Science Council) made available their SM 2000 hardware, while Defense Research & Development Canada—Atlantic (former DREA) provided time and technical assistance at their acoustic calibration facilities. Larry Mayer, formerly of the Dept. of Geodesy and Geomatics Engineering, the University of New Brunswick, arranged essential interinstitutional and external financial support. Substantial portions of this work were funded by the Canadian Department of Fisheries and Oceans' (DFO) National Hydroacoustic Program, by the DFO Strategic Science Fund, and by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Chu, D., Baldwin, K. C., Foote, K. G., Li, Y., Mayer, L. A., and Melvin, G. D. (**2001a**). ''Multibeam sonar calibration: Removal of static surface reverberation by coherent echo subtraction,'' in Proceedings: MTS/IEEE Oceans 2001, 2498–2502.

Chu, D., Baldwin, K. C., Foote, K. G., Li, Y., Mayer, L. A., and Melvin, G. D. (**2001b**). ''Multibeam sonar calibration: Target localization in azimuth,'' in Proceedings: MTS/IEEE Oceans 2001, pp. 2506–2510.

Clay, C. S., and Medwin, H. (**1977**). *Acoustical Oceanography* (Wiley, New York).

Cochrane, N. A. (**2002**). ''Near field considerations for Simrad-Mesotech SM 2000 multibeam sonar,'' Can. Tech. Rep. Fish. Aquat. Sci. 2417, iv+26 pp.

Cochrane, N. A., and Melvin, G. D. (**1997**). ''Scanning sonar as a supplement to conventional acoustic fish stock assessment in Atlantic Canada,'' in Proceedings: MTS/IEEE Oceans 97, pp. 551–556.

Faran, J. J. (**1951**). ''Sound scattering by solid cylinders and spheres,'' J. Acoust. Soc. Am. **23**(4), 405–418.

Foote, K. G. (**1985**). ''Rather-high-frequency sound scattering by swimbladdered fish,'' J. Acoust. Soc. Am. **78**(2), 688–700.

Francois, R. E., and Garrison, G. R. (**1982a**). ''Sound absorption based on ocean measurements. Part I. Pure water and magnesium sulfate contributions,'' J. Acoust. Soc. Am. **72**(3), 896–907.

Francois, R. E., and Garrison, G. R. (**1982b**). ''Sound absorption based on ocean measurements. Part II. Boric acid contribution and equation for total absorption,'' J. Acoust. Soc. Am. **72**(6), 1879–1890.

Gallaudet, T. C., and de Moustier, C. P. (**2000**). ''On optimal shading for arrays of irregularly spaced or noncoplanar elements,'' IEEE J. Ocean. Eng. **25**(4), 553–567.

Gallaudet, T. C., and de Moustier, C. P. (**2002**). ''Multibeam volume acoustic backscatter imagery and reverberation measurements in the northeastern Gulf of Mexico,'' J. Acoust. Soc. Am. **112**(2), 489–503.

Gerlotto, F., Fréon, P., Soria, M., Cottais, P. H., and Ronzier, L. (**1994**). ''Exhaustive observation of 3-D school structure using multibeam side scan sonar: Potential use for school classification, biomass estimation and behavior studies,'' ICES C.M. **B:26**, 12 pp.

Heier, O. (**1987**). ''Signal processing for precise ocean mapping,'' in *Progress in Underwater Acoustics*, edited by H. M. Merklinger (Plenum, New York), pp. 697–705.

Horne, J. K., and Jech, J. M. (**1999**). ''Multi-frequency estimates of fish abundance: Constraints of rather high frequencies,'' ICES J. Mar. Sci. **56**, 184–199.

Knight, W. C., Pridham, R. G., and Kay, S. M. (**1981**). ''Digital signal processing for sonar,'' Proc. IEEE **69**(11), 1451–1506.

Mayer, L., Li, Y., and Melvin, G. (**2002**). ''3D visualization for pelagic fisheries research and assessment,'' ICES J. Mar. Sci. **59**, 216–225.

McClatchie, S., Alsop, J., Ye, Z., and Coombs, R. F. (**1996**). ''Consequence of swimbladder model choice and fish orientation to target strength of three New Zealand fish species,'' ICES J. Mar. Sci. **53**, 847–862.

Medwin, H., and Clay, C. S. (**1997**). *Fundamentals of Acoustical Oceanography* (Academic, Boston).

Melvin, G. D., Cochrane, N. A., and Li, Y. (2003). ''Extraction and comparison of acoustic backscatter from a calibrated multi and single beam sonar,'' ICES J. Mar. Sci. **60**(3), 670–678.

Misund, O. A., and Aglen, A. (**1992**). ''Swimming behavior of fish schools in the North Sea during acoustic surveying and pelagic trawl sampling,'' ICES J. Mar. Sci. **49**, 325–334.

Misund, O. A., Aglen, A., and Frønæs, E. (**1995**). ''Mapping the shape, size, and density of fish schools by echo integration and a high-resolution sonar,'' ICES J. Mar. Sci. **52**, 11–20.

Misund, O. A., Fernö, A., Pitcher, T., and Totland, B. (**1998**). ''Tracking herring schools with a high resolution sonar. Variations in horizontal area and relative echo intensity,'' ICES J. Mar. Sci. **55**, 58–66.

Neubauer, W. G., Vogt, R. H., and Dragonette, L. R. (**1974**). ''Acoustic reflection from elastic spheres. I. Steady-state signals,'' J. Acoust. Soc. Am. **55**(6), 1123–1129.

Ona, E. (**1994**). ''Recent developments of acoustic instrumentation in connection with fish capture and abundance estimation,'' in *Marine Fish Behavior in Capture and Abundance Estimation*, edited by A. Fernö and S. Olsen (Blackwell Science, London), pp. 200–216.

Soria, M., Fréon, P., and Gerlotto, F. (**1996**). ''Analysis of vessel influence on spatial behavior of fish schools using a multi-beam sonar and consequences for biomass estimates by echosounder,'' ICES J. Mar. Sci. **53**, 453–458.

Vagle, S., Foote, K. G., Trevorrow, M. V., and Farmer, D. M. (**1996**). ''A technique for calibration of monostatic echosounder systems,'' IEEE J. Ocean. Eng. **21**(3), 298–305.

Simultaneous reconstruction of flow and temperature cross-sections in gases using acoustic tomography

Tat Hean Gan, David A. Hutchins,[a] and Peter W. Carpenter
School of Engineering, University of Warwick, Coventry CV4 7AL, United Kingdom

William M. D. Wright
Department of Electrical and Electronic Engineering, University College Cork, Ireland

(Received 15 January 2003; revised 2 May 2003; accepted 14 May 2003)

This paper describes the use of air-coupled ultrasonic tomography for the simultaneous measurement of flow and temperature variations in gases. Air-coupled ultrasonic transducers were used to collect through-transmission data from a heated gas jet. A transducer pair was scanned in two-dimensional sections at an angle to the jet, and travel time and amplitude data recorded along various paths in counter-propagating directions. Parallel-beam tomographic reconstruction techniques allowed images to be formed of variations in either temperature or flow velocity. Results have been obtained using heated jets, where it has been shown that it is possible to separate the two variables successfully. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1592523]

PACS numbers: 43.35.Ae, 43.38.Bs [DEC]

I. INTRODUCTION

Ultrasound has been used in many configurations to measure fluid flow, in both gas and liquid phases. The resulting instruments can be used in a wide range of applications in gases, such as industrial and domestic gas flow metering,[1] monitoring of gas flares on off-shore platforms,[2] and high-pressure gas flow measurement in gas distribution systems.[3] There are many measurements in liquids and in medical diagnosis that can also be performed. In liquids, ultrasonic flow meters have been used to measure very low liquid flow rates in the treatment of toxic liquids in semiconductor industries.[4] In medicine, ultrasound is now an established technique, such application being the measurement of blood flow in the nail bed of a human finger.[5]

The physical implementation of the technique usually relies on two main types of measurement—cross-propagation techniques[6] and Doppler measurements.[7] The cross-propagation technique is based on the transmission of ultrasound at a certain angle, and the change in time-of-flight of the transmitted ultrasound is related to the velocity of the flow. The Doppler technique is based on reflected or scattered ultrasonic signals from particles in the flow, which cause a shift in the received frequency content. While these two are the most common approaches to flow measurement using ultrasound, there are other methods such as cross-correlation,[8] vortex shedding[9] and open-channel measurements.[10]

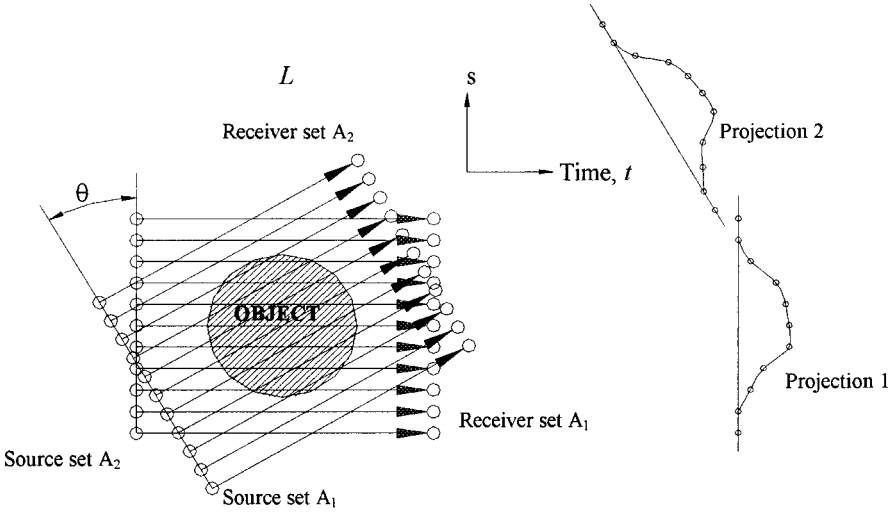
Although such ultrasonic techniques have been available for some years, further improvements to achieve modern demands such as temperature compensation and measurement accuracy are limited.[11] For instance, work by Willatzen[12] has shown that the presence of temperature gradients can cause errors in flow estimation, because of changes in the local sound velocity.

In the present work, we were interested in using the ultrasonic tomographic technique to measure complex flows, where both flow and temperature gradients are present. Previous authors reported that acoustic tomography could be used to estimate fluid flow, remembering that flow is a vector quantity. For instance, fluid flow estimation was demonstrated both theoretically and experimentally,[13,14] using travel time measurements, and this was later extended to stratified fluid flow.[15] Diffraction tomography was also investigate for ultrasonic measurement of fluid flow.[16]

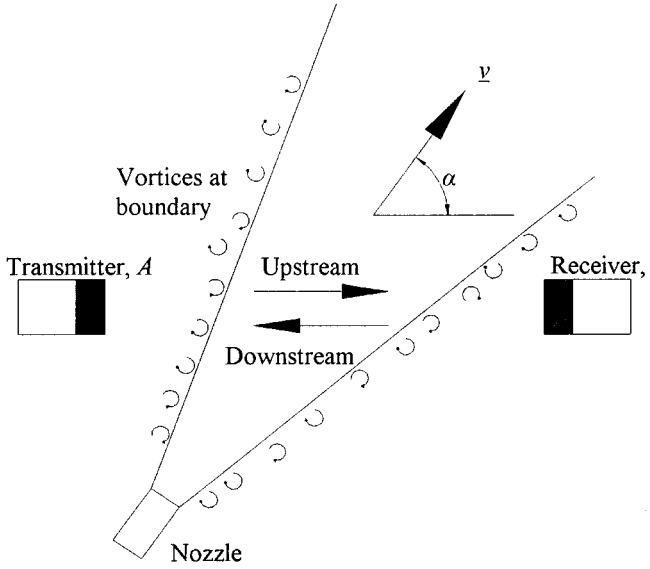
In the present case, we were particularly interested in the case of gas flow in turbulent jets, in the presence of temperature variations. This follows on from previous work,[17] where it was shown that either temperature or flow cross-sections could be measured in isolation using tomography. Here we present some experimental results, where it is shown that it is possible to reconstruct flow velocity and temperature variations simultaneously. This is illustrated for the particular case of a gas jet. Also included in this paper is a comparison between the ultrasonic technique and other scanned measurements systems, e.g., a hot-wire anemometer for flow and a thermocouple for temperature.

Tomographic reconstruction is a well-developed imaging technique that can be used to determine the cross-sectional values of a spatial by varying parameter from the transmission of acoustic signals through an object. This may be accomplished by transmitting an ultrasonic wave across the area of interest along a set of predetermined ray paths. There are many approaches to tomographic reconstruction. The first of these are iterative or series expansion methods, of which there are various approaches and methods of implementation.[18–22] This method is very flexible, in that it can correct for anisotropy and irregular geometries, although the result tends to be computationally intensive. Transform methods[23,24] are more efficient, but are less versatile and

[a]Author to whom correspondence should be addressed. Electronic mail: dah@eng.warwick.ac.uk



(a)



(b)

FIG. 1. (a) Schematic diagram of the parallel beam sampling geometry, defining the angle θ of each linear projection in the scan plane. (b) Configuration of the flow direction and the position of the transducers and hence scan plane at an angle α to the axis of a gas jet.

hence are not as popular. Probably the most widely used technique is filtered back-projection. This has been described in detail elsewhere.[25] It has the advantage in the present application that it is suitable for cylindrical scanning geometries of the type used here,[26] and is also fast and convenient to implement on a PC.

The present paper extends previous work, in that it shows that ultrasound can be used to reconstruct cross-sections in either flow or temperature, for fluid flow where both are varying spatially. In particular, it will be demonstrated that this can be done in gases, for a heated gas jet, using a combination of well-designed transducers and instrumentation, and appropriate measurement of the correct parameters for tomographic reconstruction. It also compares these results to those from point-to-point scanning of miniature flow and temperature sensors, demonstrating that the technique has some validity.

II. THEORETICAL BASIS OF THE TECHNIQUE

The technique uses a pair of ultrasonic capacitance transducers, which are highly efficient devices for generating ultrasonic signals in air. Their design is based on a micromachined silicon rigid backplate and a flexible Mylar membrane. The device construction is fully detailed elsewhere,[27] but the devices work well over the approximate frequency range of 100 kHz to 1 MHz. In the present measurements, a pair of such transducers, one acting as a source and the other as a receiver, was scanned in unison around the plane of interest using a simple ''translate-rotate'' sheme. This is shown schematically in Fig. 1, where the two transducers were fixed at points A and B, at a distance L apart. The transducer pair was scanned linearly across the region of interest, with a sampling interval of Δs [Fig. 1(a)]. When a projection was completed, the transducer pair was then ro-

tated through an angle θ in the plane of measurement, as again illustrated in Fig. 1(a). The process of translation and rotation was then repeated up to 360°, so that signals could be propagated along the same paths in opposite directions, as required for the reconstruction process. Due to the discrete sampling of projections, the spatial sampling interval was limited by the Nyquist sampling theorem, and this is given by $B=1/2\Delta s$, where B is the spatial bandwidth and Δs is the linear spacing of each ray.

For each ray-path, the transit time of the transmitted signal t_{AB} was recorded, and used to estimate the change in local sound velocity across the area of interest [as shown in Fig. 1(b) for a gas jet, with the ray-path at some angle α to the jet axis]. However, as stated above, t_{AB} will depend on two main factors. The transit time will increase and decrease with the speed depending on the flow direction. In addition, the local temperature (T) also affects the speed of sound (c) according to the following dependence, quoted from previously published work:[28]

$$c=331.31\sqrt{\frac{T}{273.16}}\ \mathrm{ms}^{-1}. \tag{1}$$

Hence, both temperature T and flow velocity $\underline{v}$ can cause a change in the apparent value of sound speed. The former is a scalar effect, whereas flow is a vector quantity, and it is this information that can be used to distinguish their contribution to t_{AB}.

Assume that, in still air at some temperature T, the local sound speed is c_o. In the absence of flow, the acoustic travel time, t, across a distance L is given by $t=L/c_o$. When flow is present and properties, such as temperature, vary across the flow region, the acoustic travel time is altered by the effect of both U_θ (the component of the flow velocity in direction AB), and by any changes δc in the acoustic speed due to changes in T, so that the acoustic travel time becomes

$$t_{AB}=\frac{L}{c_0+(1/L)\int_0^L(\delta c+U_\theta)dx'}, \tag{2}$$

where the integral takes account of variations in both U_θ and δc across the flow region. For low-speed flow it can be assumed that

$$|\delta c+U_\theta|\ll c_0, \tag{3}$$

so that, to a good approximation,

$$t_{AB}-\frac{L}{c_0}=-\frac{1}{c_0^2}\int_0^L(\delta c+U_\theta)dx'. \tag{4}$$

For the ray going from B to A, U_θ changes sign, so that when the two opposing acoustic travel times are added, the contribution from the flow speed of the gas cancels, leaving the total contribution S_θ from the temperature variations along the selected path between the transducers only:

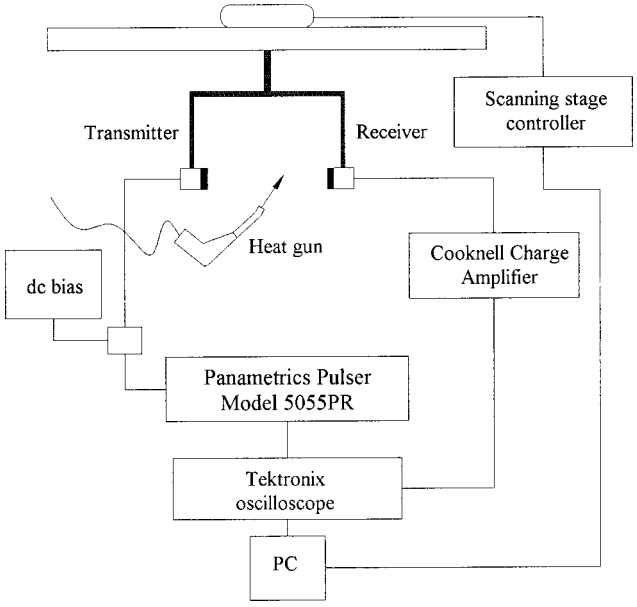


FIG. 2. A schematic diagram of the experimental arrangement for tomographic imaging.

$$S_\theta=t_{AB}+t_{BA}-\frac{2L}{c_0}=-\frac{2}{c_0^2}\int_0^L\delta c\,dx. \tag{5}$$

Similarly, subtracting the travel times gives

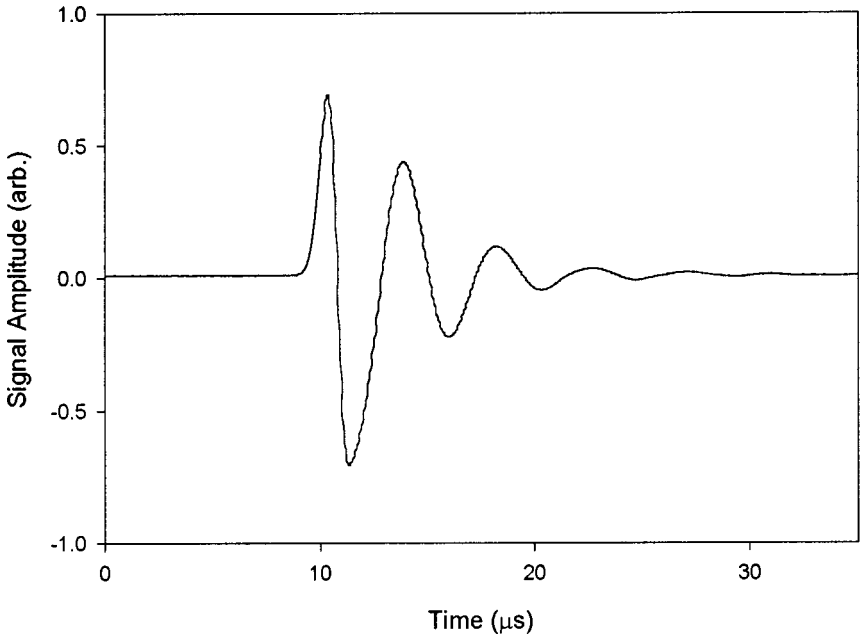
$$D_\theta=t_{AB}-t_{BA}=-\frac{2}{c_0^2}\int_0^L U_\theta dx', \tag{6}$$

and the integrated contribution (D_θ) is due only to gas flow. Hence, if ultrasonic signals are sent in counter-propagating directions along each ray path, addition of time of arrival would allow variations in δc [and hence temperature, via Eq. (1)] to be reconstructed, whereas subtraction leads to the measurement of variations in U_α (i.e., variations in the flow velocity) at the angle α. Flow and temperature can thus be measured independently in the same experiment.

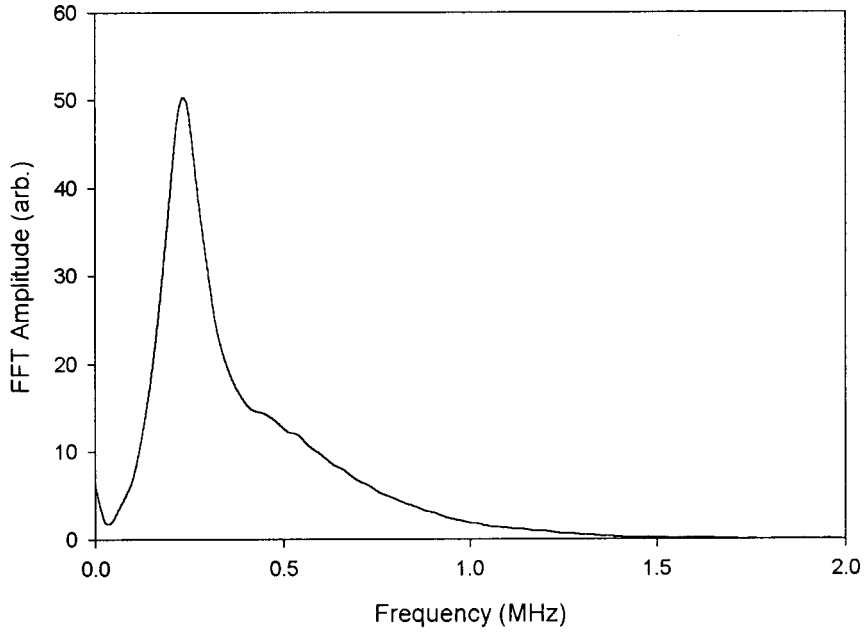
Below, it is demonstrated that the use of well-designed transducers, operating at frequencies of up to 1 MHz, can be used in air to collect tomographic data from flowing gases at a range of temperatures. This information can then be used to reconstruct images of both temperature and flow variations independently. The technique is demonstrated using gas jets at both room temperature and at elevated temperatures.

III. APPARATUS AND EXPERIMENT

Figure 2 shows the apparatus used to perform ultrasonic measurements of flow and temperature in gas jets. One form of jet was created using a Steinel Model HL1605s heat gun, shown in the figure, with a 22-mm-diam nozzle. This provided an approximate volumetric flow rate of 250 lmin^{-1} at temperatures of up to 50 °C at the plane of measurement. The ultrasonic capacitance transducers were separated by a distance of 142 mm, and were mounted horizontally as shown in the figure on a computer-controlled X-Y scanning



(a)



(b)

FIG. 3. A typical signal transmitted from source to receiver in the absence of gas flow, showing (a) a time waveform and (b) the corresponding spectrum.

stage. Each of the transducers could act as either a transmitter or receiver, so that signals could be transmitted in either direction through the selected region of flow. Time waveforms were recorded in each location on the digital oscilloscope, and transferred to the PC for automated software picking of arrival time t_{AB}.

The transducers, described more fully elsewhere,[26] were designed to have an excellent bandwidth and sensitivity for operation in air. A Panametrics 5055PR pulser provided voltage transients for excitation of the source, this being superimposed on a +100-V dc bias to attract the polymer film onto the backplate, to improve bandwidths. The receiver was connected to a Cooknell CA6/C charge amplifier, which again supplied the receiver with a 100-V dc bias. The results were then digitized using a Tektronix TDS540 oscilloscope before storing the arrival time for further data processing. Figure 3 shows a typical waveform transmitted from one device to the other in the absence of gas flow. The time response is well-damped, and the frequency response demonstrates that a reasonable bandwidth was available for the measurement. This improved the accuracy of time-of-flight measurements, which could then be fed into Eqs. (5) and (6).

Each full tomographic scan consisted of 61 linear projections ($\theta = 3°$), each of which had 101 rays for which an

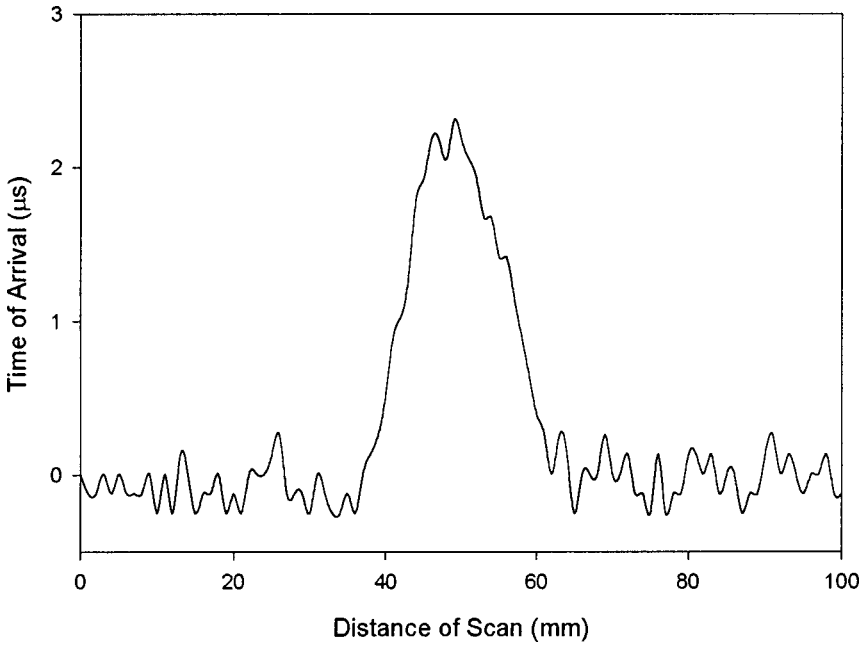


FIG. 4. Variation of time of flight (in μs) at a distance of 15 mm above a heat gun, measured using a through-transmission ultrasonic technique. The gas jet was heated and of approximately 28 mm diameter.

ultrasonic waveform was recorded at every 1-mm interval. Each projection involved scanning the transducers at right angles to their propagation axis using the linear stage. This was also attached to a rotational stage, so that projections could be taken at various angles in the plane of interest. Note that for all scans reported, a difference technique was employed to remove variations in room temperature and any unknown delays that might affect the ultrasonic time of flight. This was carried out by normalizing the data from each projection to the first waveforms taken in the ambient still air, outside the gas jet volume. Thus, systematic variations in room temperature or other factors were removed from the data prior to reconstruction. Data was collected in the form of time waveforms for each path. Note that at any one time there was only one signal to be recorded, using a single source–receiver pair. At no stage was there mixing of signals from multiple sources. In the case of counter-propagating signals along the same path, the source and receiver were interchanged in space during the 360° scan.

To provide independent measurements with which to compare the tomographic reconstructions from ultrasonic data, two point measurement devices were also scanned across the measurement plane. The first was a hot-wire anemometer (Model TA5 by Airflow Developments Ltd.), which measured flow velocity in a particular direction. The second was a K-type thermocouple, attached to a 9-V voltage supply unit, which measured the local temperature at that point. The same X-Y stage and control electronics was used for these scans, so that direct comparisons could be made.

Flow and temperature measurements, using both the ultrasonic tomography approach and the point detectors, were performed for various values of α, the angle between the gas jet axis and the plane containing the tomographic scan. Initial experiments were performed with the angle α between the gas jet axis and the tomographic scan plane [see Fig. 1(b)] at 90°, i.e., with the measurement plane perpendicular to the axis of flow (to enhance sensitivity to temperature

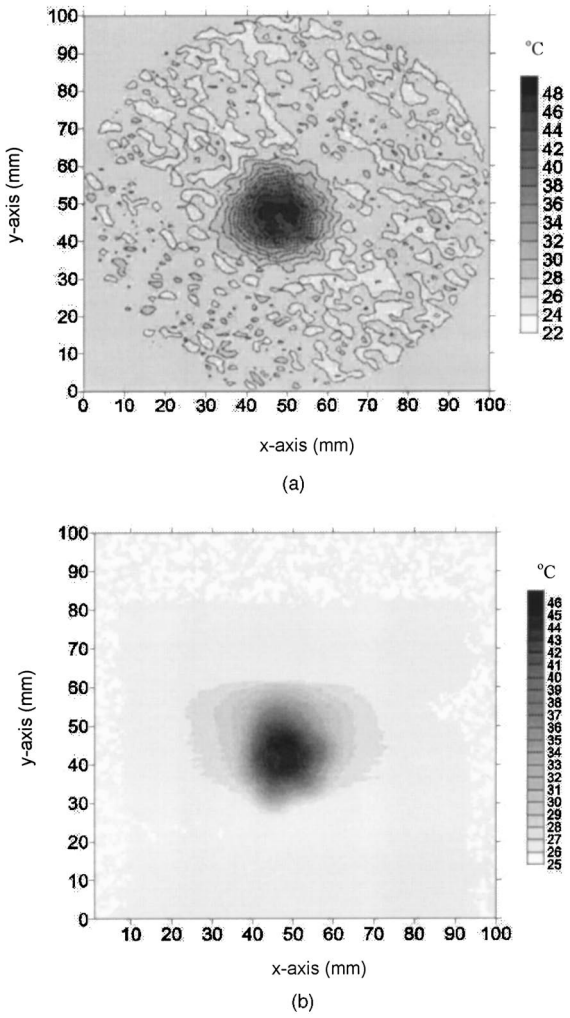


FIG. 5. Temperature measurement at 15 mm above the heated air jet from the heat gun, using (a) ultrasonic tomographic reconstruction and (b) a scanned thermocouple. $\alpha = 90°$ for this measurement so that the effects of temperature were maximized.

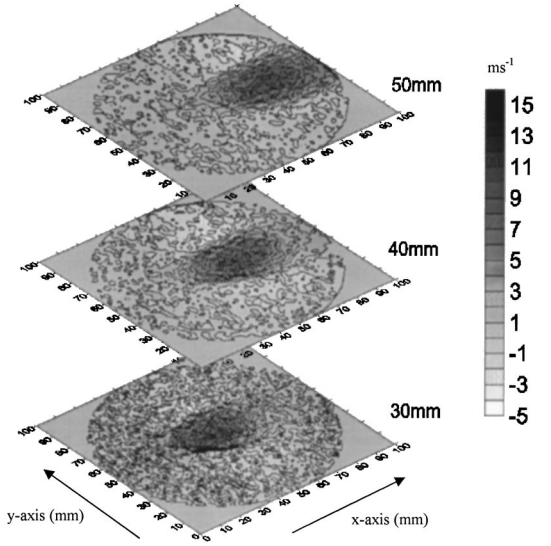


FIG. 6. Tomographic reconstruction of air flow velocity within the jet from the heat gun at different heights. The scan was performed at $\alpha = 45°$.

effects with respect to flow). Other measurements were then performed at $\alpha = 45°$, at various vertical heights above the exit nozzle, so that both temperature and flow effects would contribute to the measurements. A final measurement was also made with the heat gun at $\alpha = 40°$, so that the effect of flow would be further enhanced.

IV. RESULTS AND DISCUSSION

A. Temperature measurements at $\alpha = 90°$

Experiments were performed initially to establish whether temperature variations could be reconstructed, using Eq. (5), without the effects of flow being a major factor. Data were thus collected with the gas jet vertical, and with the tomographic X-Y scan plane in the horizontal plane, i.e., with $\alpha = 90°$, where the effect of flow would be minimized. Note that the choice of $\alpha = 90°$ optimized the sensitivity of the measurement to temperature effects in relation to the effect of flow (which increases as α decreases). The scans were performed at a height of 15 mm above the heat gun. The results of ultrasonic measurements, leading to cross-sections of temperature via Eq. (1), were then compared to temperature cross-sections measured across the same plane using the thermocouple. The time of arrival of the ultrasonic signal was found to change by about 2.3 μs across the heated gas jet region of approximately 28 mm diameter, as shown in Fig. 4.

Figure 5(a) shows a tomographic reconstruction of the temperature profile 15 mm above the nozzle, obtained using ultrasonic data. This image can be compared to that obtained using the thermocouple [Fig. 5(b)]. The darker areas in both images represent the heated area and the size is well correlated. Both have a diameter of approximately 25 mm, with the temperature range being very similar. This demonstrates that the ultrasonic approach shows promise for completely noninvasive measurement of temperature variations in gases.

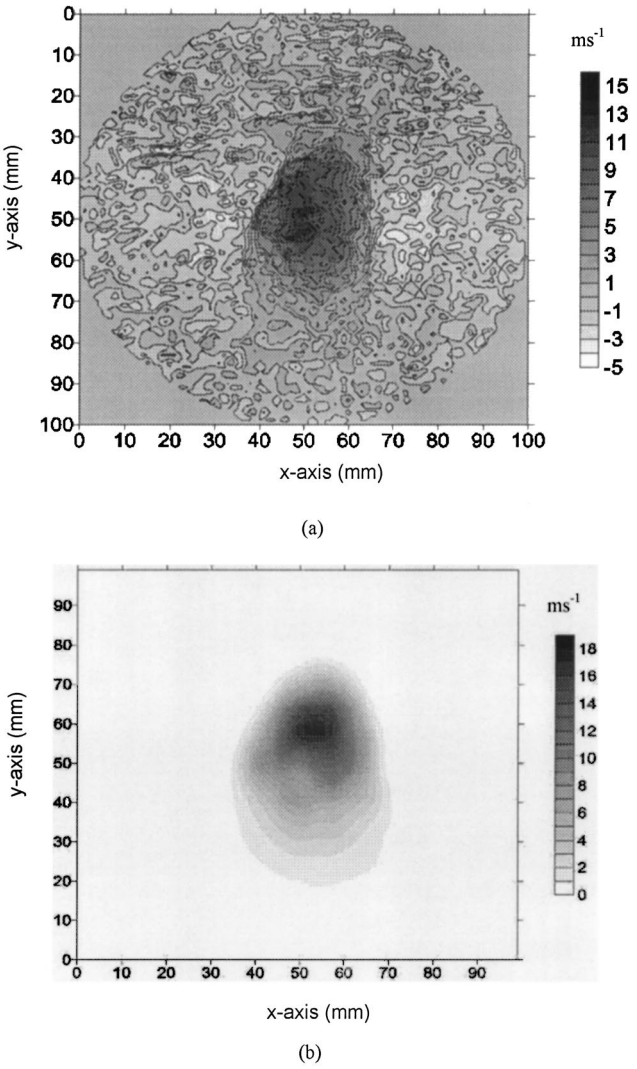


FIG. 7. Details of flow velocity cross-sections at a distance of 30 mm above the heat gun. Shown are the results of (a) tomographic reconstruction from ultrasonic data and (b) a scanned hot-wire anemometer.

B. Simultaneous measurements of temperature and flow at $\alpha = 45°$

The direction of the heat gun was now set to $\alpha = 45°$ to the horizontal scanning plane. This made the measurement sensitive to components arising from both temperature and flow effects, the latter increasing in dominance as α increased in value. Equations (5) and (6) were used to obtain variations in δc [and hence temperature, via Eq. (1)] and U_θ across the same area of flow, using addition and subtraction of the counter-propagating ultrasonic signals along each ray path. The result was the simultaneous reconstruction of temperature and flow cross-sections.

The tomographic reconstruction of flow velocity images was performed using ultrasonic data at axial distances of 30, 40, and 50 mm from the exit nozzle, as shown in Fig. 6, and the maximum value of flow velocity at each the vertical height was found to be approximately 15.5, 12.5 and 11 ms^{-1}, respectively. As distance from the nozzle increased, the gas jet is seen to diverge, and the axial flow velocities reduce, as would be expected. At a distance of 30

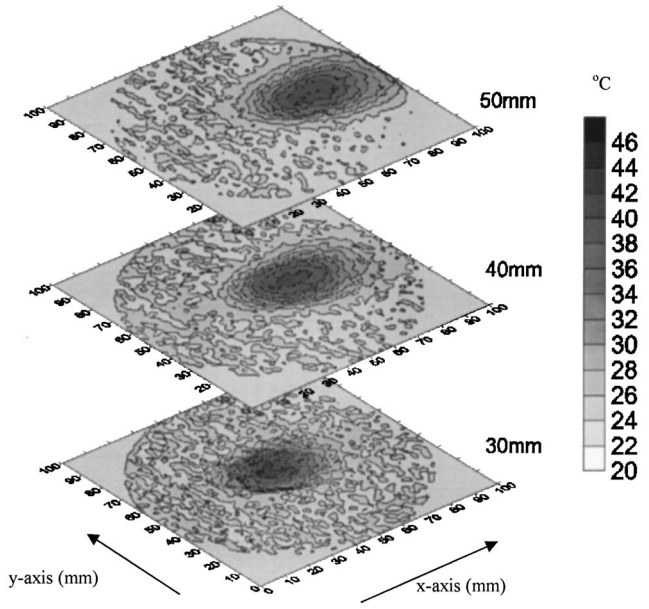


FIG. 8. Tomographic reconstruction of temperature variations, resulting from a scan at $\alpha = 45°$ of the heat gun at different heights.

mm, the hot wire anemometer was also scanned across the same area, to compare variations in axial flow velocity with those obtained from tomographic reconstruction. The maximum velocity detected by the point detector was found to be approximately $20\ \mathrm{ms}^{-1}$, and this agreed reasonably well with the value obtained from ultrasonic measurements, as shown in Fig. 7, where the two flow velocity cross-sections are compared.

Figure 8 shows the cross-sections in temperature that were reconstructed from ultrasonic data, at the three same heights above the heat gun nozzle as those shown earlier in Fig. 6. It will be evident from the figure that, as the axial distance from the exit nozzle increased, the average temperature reduced, as would be expected (these are displayed as lighter shades on the images). Note that, as before, the divergence of the hot air jet is evident as the increased diameter with axial distance. Note also that temperatures at 10 mm axial distances from the nozzle were predicted to approach 50 °C.

Figure 9(a) shows the reconstructed image at a height of 30 mm. This is compared to the measurements from the thermocouple, scanned as appoint detector over the same area, and presented in Fig. 9(b). The images show that the two results are similar to within a few degrees.

C. Measurements at $\alpha = 40°$

As a final experiment, the flow angle was reduced to 40°. This was performed so that the sound propagation is affected more by the flow and less by variations in temperature. The reconstructed image at the vertical height of 30 mm is as shown in Fig. 10(a). The hot-wire anemometer result is presented in Fig. 10(b). The flow imaging seems to be more successful as the velocity is now closer to the value imaged using the anemometer. It can thus be concluded that $\alpha = 40°$ can be used to optimize flow measurements for the same heated gas jet. Note that changing the angle would

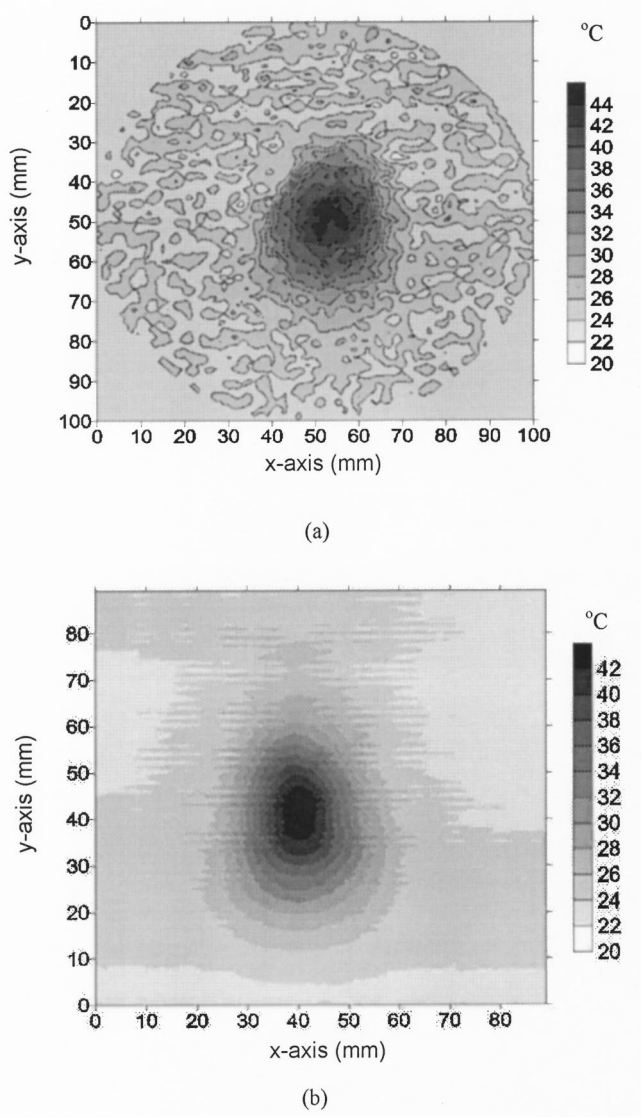


FIG. 9. Comparison of temperature cross-sections at a distance of 30 mm above the heat gun. (a) Temperature map derived from ultrasonic data and (b) measured via a scanned thermocouple.

have little effect on the reconstruction of temperature profiles, as temperature is a scalar component of the equations. This has been observed in practice.

V. CONCLUSIONS

Ultrasonic tomographic imaging has been performed using a pair of broadband capacitance transducers to measure flow and temperature. A difference technique was employed to remove variations in room temperature and any unknown delays. This was carried out by normalizing the data from each projection to the first waveform in a particular linear scan, at some angle θ.

This experiment investigated the simultaneous reconstruction of temperature and flow velocity cross sections within a heated gas jet. The experiments were first conducted at $\alpha = 90°$ and it was shown that temperature measurement can be corrected by performing a 360° tomographic scan. When $\alpha = 45°$ it was shown that good results could be obtained if counter-propagating signals were used to separate

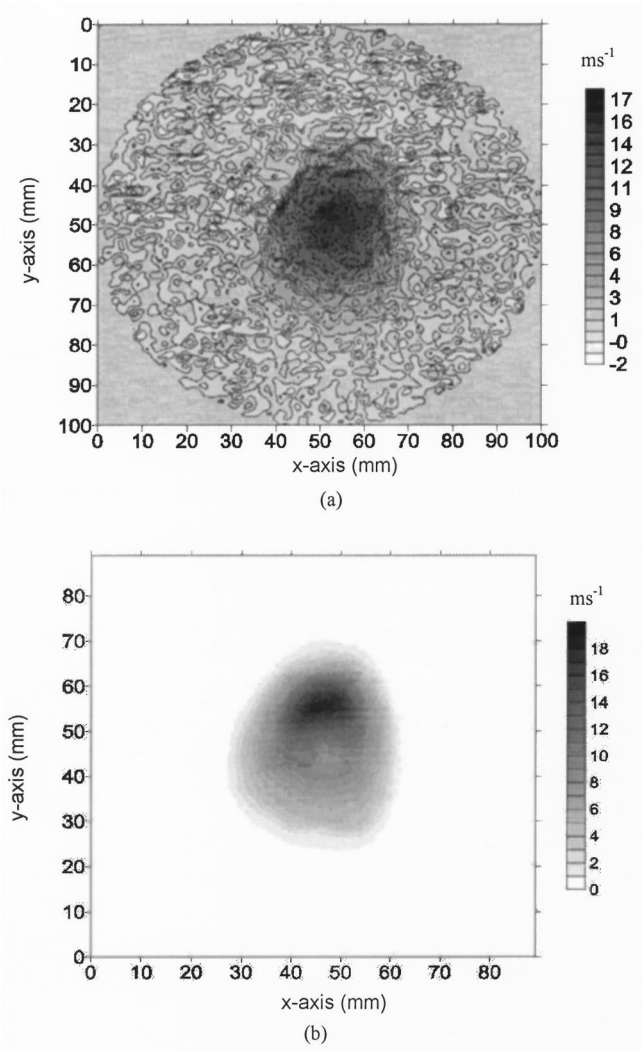


FIG. 10. Flow cross-sections taken at $\alpha = 40°$ and at a height of 30 mm above the heat gun. (a) Flow velocity variations derived from acoustic data and (b) measured via a scanned hot wire anemometer.

the two components. It was found, however, that either flow or temperature measurement could be optimized by changing the angle α. Temperature would be best measured at $\alpha = 90°$, so that flow effects were minimized. Conversely, temperature effects could be minimized by decreasing α to 45° or 40° and enhancing the effect of flow.

[1] G. Buonanno, "On field characterization of static domestic gas flowmeters," Measurement **27**, 277–285 (2000).
[2] E. A. Hammer, "Flare gas metering and high pressure gas metering using ultrasonic sensors," IEE Colloq. Ultra. Process Indus. **8**, 1–4 (1993).
[3] M. Dell'Isola, M. Cannizzo, and M. Diritti, "Measurement of high-pressure natural gas flow using ultrasonic flowmeters," Measurement **20**, 75–89 (1997).
[4] M. Takamoto, H. Ishikawa, K. Shimizu, H. Monji, and G. Matsui, "New measurement method for very low liquid flow rates using ultrasound," Flow Meas. Instrum. **12**, 267–273 (2001).
[5] D. E. Kruse, J. J. Mai, R. H. Silverman, M. F. Insana, D. J. Coleman, and K. W. Ferrara, "Optimization of real-time high frequency ultrasound for blood flow imaging in the micro-circulation," Proc. SPIE **4325**, 284–292 (2001).
[6] R. C. Asher, "Ultrasonic sensors for the process industry," Meas. Control **30**, 138–140 (1997).
[7] L. C. Lynnworth, *Ultrasonic Measurements for Process Control—Theory, Techniques, Applications* (Academic, London, 1989).
[8] V. Skwarek, H. Windorfer, and V. Hans, "Measuring pulsating flow with ultrasound," Measurement **29**, 225–236 (2001).
[9] V. Hans, G. Poppen, E. von Lavante, and S. Perpeet, "Vortex shedding flowmeters and ultrasound detection: signal processing and influence of bluff body geometry," Flow Meas. Instrum. **9**, 79–82 (1998).
[10] B. L. Thurley, "The fundamentals of open channel flow measurement," in *Proc. Symp. Appl. of Flow Measuring Tech. (Flow-Con 77)* (1977), pp. 171–186.
[11] H. Köchnert, A. Melling, and M. Baumgärtner, "Optical flow field investigations for design improvements of an ultrasonic gas meter," Flow Meas. Instrum. **7**, 133–140 (1996).
[12] M. Willatzen, "Temperature gradients and flow-meter performance," Ultrasonics **39**, 383–389 (2001).
[13] S. J. Norton, "Tomographic reconstruction of 2-D vector fields: application to flow imaging," Geophys. Res. **97**, 161–168 (1988).
[14] H. Braun and A. Hauck, "Tomographic reconstruction of vector fields," IEEE Trans. Signal Process. **39**, 464–471 (1991).
[15] K. B. Winters and D. Rouseff, "Tomographic reconstruction of stratified fluid flow," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **40**, 26–33 (1993).
[16] M. N. Rychagov and H. Ermert, "Reconstruction of fluid motion in acoustic diffraction tomography," J. Acoust. Soc. Am. **99**, 3029–3035 (1996).
[17] W. M. D. Wright, D. W. Schindel, D. A. Hutchins, P. W. Carpenter, and D. P. Jansen, "Ultrasonic tomographic imaging of temperature and flow fields in gases using air-coupled capacitance transducers," J. Acoust. Soc. Am. **104**, 3446–3455 (1998).
[18] J. M. S. Prewitt, "Algorithms in computerized tomography," in *Medical Imaging Techniques—A Comparison* (Plenum, New York, 1979), pp. 287–312.
[19] K. T. Ladas and A. J. Devaney, "Application of an ART algorithm in an experimental study of ultrasonic diffraction tomography," Ultrason. Imaging **15**, 48–58 (1993).
[20] D. P. Jansen, D. A. Hutchins, P. J. Ungar, and R. P. Young, "Acoustic tomography in solids using a bent ray SIRT algorithm," Nondestr. Test. Eval. **6**, 131–148 (1991).
[21] A. H. Delaney and Y. Bresler, "A fast iterative tomographic reconstruction algorithm," Int. Conf. Acoust., Speech, & Signal Process. (1995), pp. 2295–2298.
[22] L. Wang and R. S. Granetz, "Series expansion method in three-dimensional tomography," J. Opt. Soc. Am. A Opt. Image Sci. Vis. **10**, 2292–2295 (1993).
[23] H. H. Stark, J. W. Woods, I. Paul, and R. Hingorani, "Direct Fourier reconstruction in computer tomography," IEEE Trans. Acoust., Speech, Signal Process. **29**, 237–245 (1981).
[24] A. J. Devaney, "A fast filtered backpropagation algorithm for ultrasound tomography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **34**, 330–340 (1987).
[25] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging* (IEEE, New York, 1988).
[26] T. H. Gan, D. A. Hutchins, D. R. Billson, and F. C. Wong, "Ultrasonic Tomographic Imaging of Encased Highly-Attenuating Solid Media," J. Quant. Non-Destr. Eval. **13**, 131–152 (2001).
[27] D. W. Schindel, D. A. Hutchins, L. Zou, and M. Sayer, "The design and characterization of micromachined air-coupled capacitive transducers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 42–50 (1995).
[28] R. Hickling and S. P. Marin, "The use of Ultrasonics for gauging and proximity sensing in air," J. Acoust. Soc. Am. **79**, 1151–1160 (1986).

Guided wave helical ultrasonic tomography of pipes

Kevin R. Leonard and Mark K. Hinders[a)]
Department of Applied Science, College of William and Mary, Williamsburg, Virginia 23187-8795

(Received 5 August 2002; revised 20 May 2003; accepted 23 May 2003)

Ultrasonic guided waves have been used for a wide variety of ultrasonic inspection techniques. We describe here a new variation called helical ultrasound tomography (HUT) that uses guided ultrasonic waves along with tomographic reconstruction algorithms that have been developed by seismologists for what they call ''cross borehole'' tomography. In HUT, the Lamb-like guided waves travel the various helical criss-cross paths between two parallel circumferential transducer arrays instead of the planar criss-cross seismic paths between two boreholes. Although the measurement itself is fairly complicated, the output of the tomographic reconstruction is a readily interpretable map of a quantity of interest such as pipe wall thickness. In this paper we demonstrate HUT via laboratory scans on steel pipe segments into which controlled thinnings have been introduced.
 [DOI: 10.1121/1.1593068]

PACS numbers: 43.35.Pt [DEC]

I. INTRODUCTION

Guided ultrasonic waves have many useful properties that can be exploited for nondestructive testing (NDT) of large-area structures such as piping systems.[1,2] First and foremost, they can be made to propagate significant distances in either pitch–catch or pulse-echo configurations. This is an inherent advantage over typical point-by-point bulk wave ultrasonic measurements because guided waves can accumulate information about the structure along the entire propagation path. Furthermore, they follow the curvature of the structure and can be directed along any desired helical path from circumferential to meridional. As the guided waves propagate, they involve the entire thickness of the pipe wall in a complex elastic deformation and recovery so that they are sensitive to flaws on the inner diameter, outer diameter, and in between. They can be launched and received by external transducer belts, internal pigs, and by a wide variety of contact and noncontact transducer configurations. The primary problem has always been that their inherent complexity often defeats those who try to implement guided wave inspection techniques outside of controlled laboratory environments. Uncontrolled and uncontrollable variations in seemingly insignificant details of the measurement setup drive one toward simpler and more straightforward schemes. A secondary problem is that even perfectly controlled guided wave measurements are *quite* difficult to interpret. In the laboratory, scientists are comfortable sorting out subtle variations in complex waveforms by drawing on their extensive background and intuitive understanding of the underlying physics, but in the field, technicians responsible for interpreting the NDT measurements are not.

Early work on guided waves in hollow cylinders was done by Gazis[3,4] and others.[5–8] Many authors have recently used guided ultrasonic waves to detect flaws in pipes and tubes[9–27] by generating selected modes in the pipe and using pulse-echo measurements to locate flaws. Alleyne *et al.*[13–15] have shown that a ring of transducers around the circumference of the pipe can excite axially symmetric modes that propagate long distances down the pipe. By choosing a nondispersive mode they are able to retain the signal's shape and amplitude as it propagates. Similarly, Rose *et al.* have explored several methods for selecting different modes for pipe inspection. These methods include using comb transducers[17,18] to generate longitudinal guided waves and partial circumferential loading[19] to focus flexural modes to the area of interest in the pipe. Other authors have also explored the use of EMATs[20–22] and PVDF[23] transducers.

Ultimately, any real-world guided wave measurement scheme is complicated by mode conversion at flaw interfaces, bends in the pipe, and loading on both the inner and outer diameters.[24–32] This inherent complexity makes guided waves very informative, but at the same time very difficult to utilize. An infinite number of modes are theoretically available for use, each with characteristic dispersion and through-thickness displacement properties. At low frequencies, longitudinal, torsional, and flexural pipe modes dominate, but at higher frequencies these structural modes are less important and the waves see the pipe more like a curved plate. It is these helical guided wave modes that are of interest in our measurements. This is in contrast to previous work done in this area that has mostly concentrated on generating the longitudinal and flexural pipe modes. Of course it is important to point out that there is no clear demarcation between these two perspectives once the guided waves are nonaxially symmetric, as in the work by Rose.[33] There are a wide variety of models appropriate for studying this type of vibration cylindrical shells,[34–38] many of which come from the underwater sound community. One can even formally connect plate waves to the corresponding Lamb-like waves in cylindrical shells by replacing the source by a periodic array of excitations, and then replacing the circular cylinder by an equivalent ''unwrapped'' two-dimensional plate.[39] A recent review

[a)]Author to whom correspondence should be addressed: Professor Mark Hinders, College of William and Mary, Department of Applied Science, P.O. Box 8795, Williamsburg, VA 23187-8795. Phone: (757) 221-1519; fax: (757) 221-2050; electronic mail: hinders@as.wm.edu

of guided ultrasonic waves with an extensive bibliography is quite useful.[40]

That the helical modes can be considered similarly to Lamb wave modes in plates, allows us to extend our previous work on Lamb wave tomography[41–46] to pipe inspection systems. The Lamb-like guided waves form a series of helical criss-cross paths that are a useful tomographic geometry equivalent to what seismologists use in cross-borehole tomography.[47] The tomographic reconstruction is key because it allows the complexity of the guided wave modes to be harnessed, resulting in a readily interpretable ''map'' of the structural flaws of interest. For example, if one selects a mode with high sensitivity to thickness variations—i.e., it speeds up or slows down when the pipe wall thickness changes—then by noting deviations from the expected arrival times of the many criss-cross helical paths, the tomographic reconstruction is easily rendered as a wall-thickness map. For many piping applications, accurately mapping out wall thickness changes due to corrosion or erosion is a primary concern. With appropriate software expert systems to identify and sort out the features of interest from the received waveforms, the entire measurement process can be automated so that the human inspector needs to have mastered none of the complexity of guided wave ultrasonics.

In this paper we first cover some necessary background on guided waves and tomography. We then explain in detail our particular implementation for guided wave helical ultrasonic tomography of pipes. A series of measurements with our laboratory scanner for flawed pipe segments with accompanying reconstructions are then shown. Finally, we discuss the results and indicate future directions for this research.

II. DESCRIPTION OF GUIDED WAVES AND TOMOGRAPHY

Elastic waves in extended solids are either longitudinal or transverse, characterized by compressional or shear vibrations, respectively. Boundaries cause mode coupling and often intermixing of these waves to the extent that it is no longer useful to try to distinguish them. Locally, plate-like structures have two boundaries and we refer to the intermixed propagating compressional and shear vibrations as Lamb waves. Lamb waves come in two families of modes: symmetric and antisymmetric. The lowest-order symmetric mode is a thickness stretching and contraction while the lowest-order antisymmetric mode is a constant-thickness flexing. Higher-order modes have increasingly complex through-thickness displacements. Each mode has its own characteristic dispersion properties.[1,2]

Curved plates behave similarly, except that for pipe-like objects—where the plate curves around upon itself—some additional complexities arise. The first one is that the pipe as a whole can exhibit three families of propagating modes: torsional, flexural, and extensional. These can be distinct from, or intermingled with, Lamb wave (plate) modes in the pipe. Generally speaking, we can hope to distinguish pipe modes from plate modes in that the former are lower in frequency and longer in range while the latter are higher-frequency more localized phenomena. Of course there is no clear demarcation between the two regimes, and in practice it is necessary to be able to sort them out or the unwanted modes will corrupt the datasets of interest. A second complexity that arises for pipe-like objects is that more than one helical mode can travel between any two transmit and receive locations. In Fig. 1 we show schematically a pair of transducers on a pipe, and several helical rays propagating away from the transmitting transducer. One of these rays takes the most direct path partway around the pipe to the receiving transducer while others, with steeper helical paths, will travel further around the pipe circumferentially and will miss the receiver. Others will make one or more complete loops around the pipe and end up at the receiver. Although one could envision ''aiming'' the waves in a narrow enough beam to avoid this confusion, tomographic considerations require that both the transmitting and receiving transducers be as omnidirectional as possible. This adds yet a third complexity; the helical waves are generated in both clockwise and counterclockwise directions. Depending on the relative angular positions of the transmitting and receiving transducers these pairs of modes may interfere either constructively or destructively.

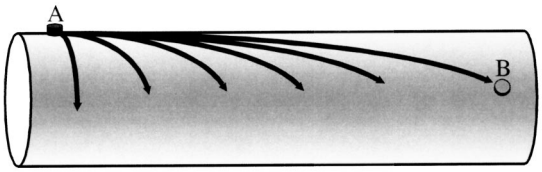


FIG. 1. A pipe segment is shown with a transmitting transducer at A and a receiving transducer at B. Six helical Lamb wave paths are shown, with the shallowest only just beginning to wrap around the pipe before it reaches B. The steepest possible helical path is the circumferential guided wave that will ring around the pipe and be received at A. In between these two extremes, the guided wave modes are launched at A for all helicities (four are shown) some of which will travel around the pipe one, two, and more times before being recorded at B. Note also by symmetry that all of these waves go in the opposite directions as well, and these mirror modes will, in general, overlap with the guided wave modes of interest at the receiver B. Sophisticated software to almost instantaneously sort out and identify all of these various modes in thousands of jumbled waveforms is key to making the HUT technique work.

Of course, all of this complexity is what one faces in the absence of any flaws. Flaws scatter and refract the guided waves, and even cause energy to be converted from one mode to another. These effects can be severe for strongly scattering flaws such as cracks. Thickness changes due to corrosion, erosion, or gouging can cause some modes to be cut off, resulting in a fairly dramatic reflection of those same modes or a dumping of the wave energy into other modes. In our previous work on Lamb wave tomography in plates we have developed some techniques to handle much of this complexity,[44,46] and have now found that we are able to deal satisfactorily with the added difficulties of pipe-like geometries. If we envision a series of pitch–catch helical guided wave measurements on a pipe with a large number of transducers along two separated circumferential belts, the set of helical crisscross ray paths is a tomographic geometry. In Fig. 2 we illustrate this by showing a pipe that is ''unwrapped'' and laid flat. The transducers now lie along parallel lines in what seismologists call a crosshole geometry. Because the helical waves can wrap around the pipe more

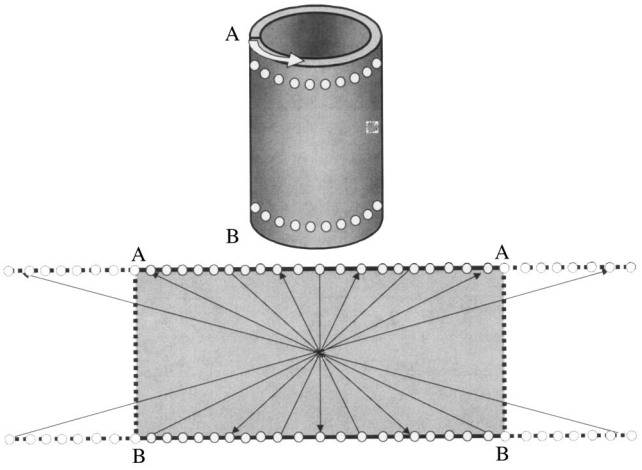


FIG. 2. Parallel circumferential arrays of transducers can be seen to give a cross-hole tomographic geometry when the pipe is mentally ''unwrapped.'' The two identical parallel circumferential array belts of transducers wrap around the pipe. Each transducer in one belt transmits helical Lamb waves, which are received by all of the transducers in the other belt. Mentally break the pipe longitudinally along the line AB and then unwrap the pipe to lie flat. The circumferential belts of transducers now lie along the lines AA and BB in the ''unwrapped'' pipe. Note that the Lamb waves travel along the criss-cross rays shown passing through the flaw. Because the helical waves can wrap around the pipe more than once we can consider the ''extra'' regions to the left and right of AB. These longer ray paths give better tomographic reconstruction because they pass through the flaw from additional angles.

than once, we can consider the ''extra'' regions and the longer ray paths to give better tomographic reconstruction because these rays pass through the region of interest from additional angles. One of the fundamental limitations of HUT is that the wave vector coverage is incomplete because ''rays'' do not go through the region of interest from all angles. To improve reconstructions in cross-borehole tomography, seismologists often place a line of receivers on the surface of the ground between the boreholes, and for tomographic plate inspection with Lamb waves we mimic a four-sided square perimeter array so that the rays pass through from all angles. At this early stage of development we are able to employ only the first arriving modes, but we recognize that the desire for better reconstructions will drive us toward using the ''twice around'' steep helical modes as well because it is not practical to surround the region of interest from all sides to achieve complete wave vector coverage.

III. TOMOGRAPHY EQUATIONS

We have found in our previous work that the iterative families of algorithms developed in the seismological literature are better suited to our purposes than are the convolution-backprojection algorithms developed for medical imaging and other applications.[43–46] In particular, we find that the simultaneous iterative reconstruction tomography (SIRT) algorithms[47] are relatively robust, computationally efficient, and insensitive to experimental noise. They also have the inherent advantage of being widely applicable to a variety of geometries and incomplete datasets. Moreover, we have been able to adapt and extend these algorithms somewhat to account for material anisotropy and ''ray bending'' due to scattering from flaws.[44]

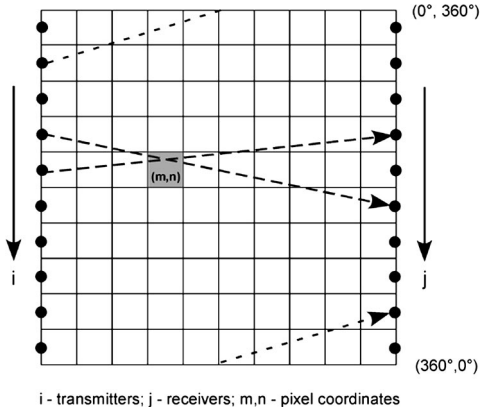


FIG. 3. Helical ultrasound tomographic reconstruction geometry unwrapped. Transmitter and receiver locations are shown by the filled circles along the left and right columns. Three different ray paths are shown via the dashed-line arrows.

In this work we have assumed, for simplicity, that the waves travel in straight paths and have ignored scattering. Figure 3 shows the particular tomographic reconstruction geometry for our scanner once it is unwrapped from the pipe. The reconstruction algorithm attempts to solve for the Lamb wave velocity within each grid cell, $v[m,n]$. In general this can be done by solving for the slowness $1/v[m,n]$ from the following system of linear equations:

$$T[i,j] = \sum_{m,n \in \mathrm{ray}[i,j]} t[i,j,m,n] = \sum_{m,n \in \mathrm{ray}[i,j]} \frac{d[i,j,m,n]}{v[m,n]}. \quad (1)$$

In this system of equations, $T[i,j]$ is the total time it takes the wave to travel from the transmitter to the receiver and $t[i,j,m,n]$ is the amount of time that ray$[i,j]$ travels within

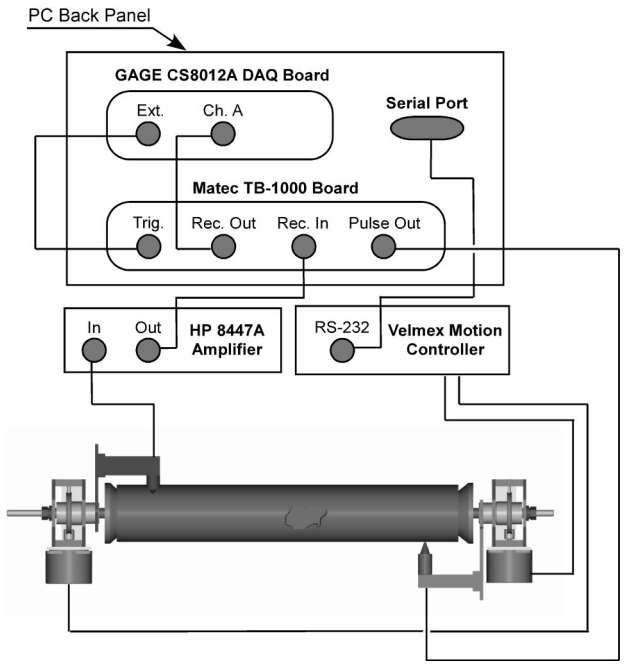


FIG. 4. Data acquisition block diagram for the computer-controlled HUT scanner.

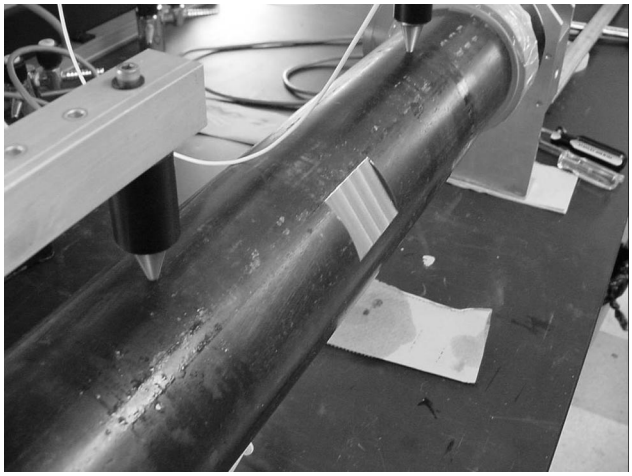


FIG. 5. A thick-walled steel pipe sample is shown in the HUT scanner. Conical delay lines are used on the spring-loaded contact transducers in order to minimize their footprint.

the cell$[m,n]$. $t[i,j,m,n]$ is equivalent to the length of ray$[i,j]$ in cell$[m,n]$ (denoted by $d[i,j,m,n]$) divided by the cell velocity $\nu[m,n]$. The segment lengths, $d[i,j,m,n]$, are calculated theoretically and $T[i,j]$ is measured experimentally. The solution of this system of equations yields a velocity map over the entire region. Given the operating frequency, we can then convert this velocity map into a thickness map used to detect flaws within the region of interest. For our current scanning system (1) provides a system of N^2 equations, where $N=180$ typically. The limitation on the number of transducer elements (N) is a combination of the footprint of the transducer, the radius of the inspected pipe, the allowable scanning time, and the desired reconstruction quality. To avoid the computationally expensive inversion of such a large matrix, we use the SIRT algorithm outlined in Kak and Slaney[47] to solve this inverse problem. The adaptation of this algorithm to our problem has four main steps.

(i) Step 1: First we determine the segment lengths $d[i,j,m,n]$ theoretically. Then, using an initial guess $\nu^0[m,n]$ for the cell velocities, we calculate the initial estimated arrival times for each ray:

$$T^0[i,j]=\sum_{m,n\in \mathrm{ray}[i,j]}\frac{d[i,j,m,n]}{\nu^0[m,n]}. \tag{2}$$

In subsequent iterations, the estimated arrival times are cal-

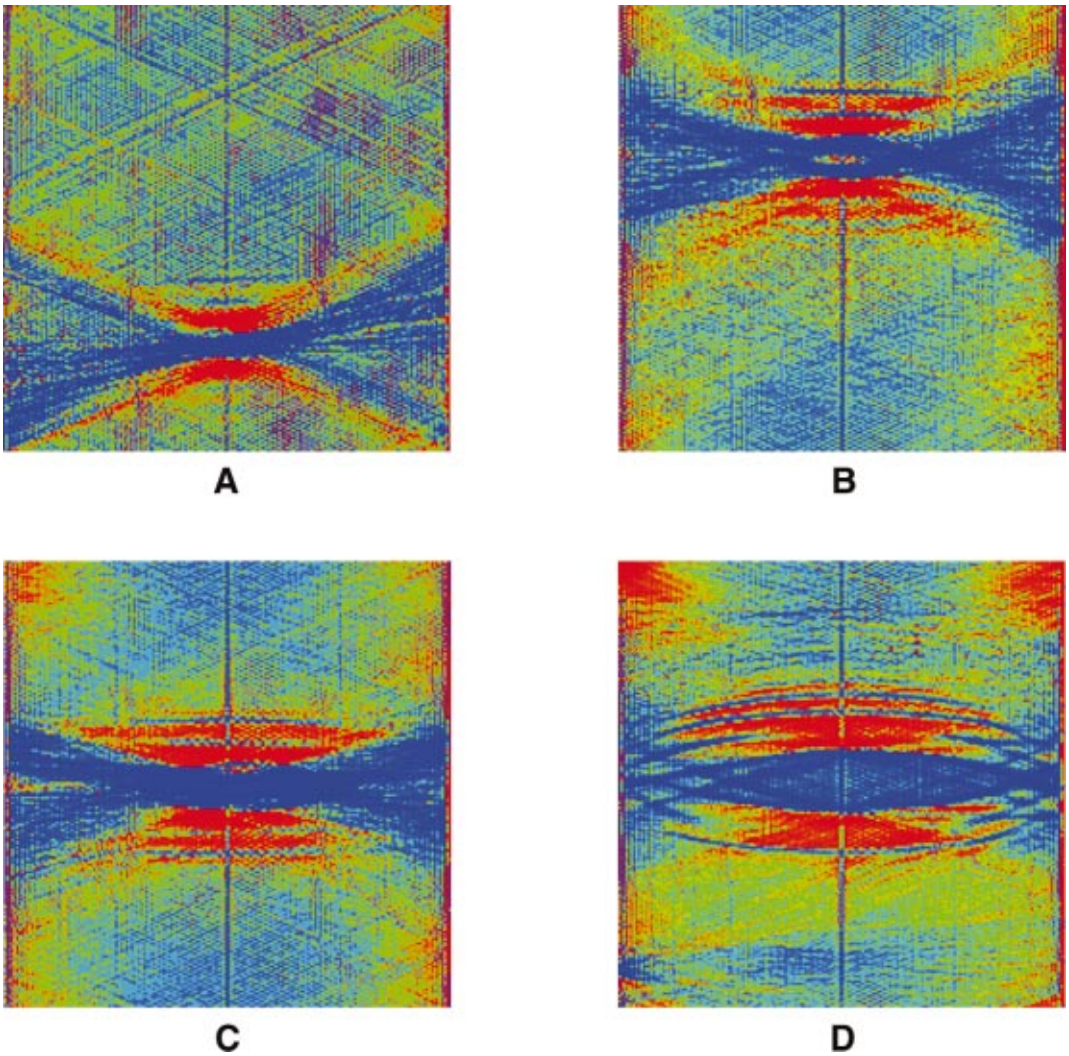


FIG. 6. Four reconstructions are shown for increasing flaw size in the same thick-walled steel sample shown in Fig. 5. The red horizontal striations and the blue criss-cross artifact indicate the location of the flaws, which can be seen to increase in size, as expected.

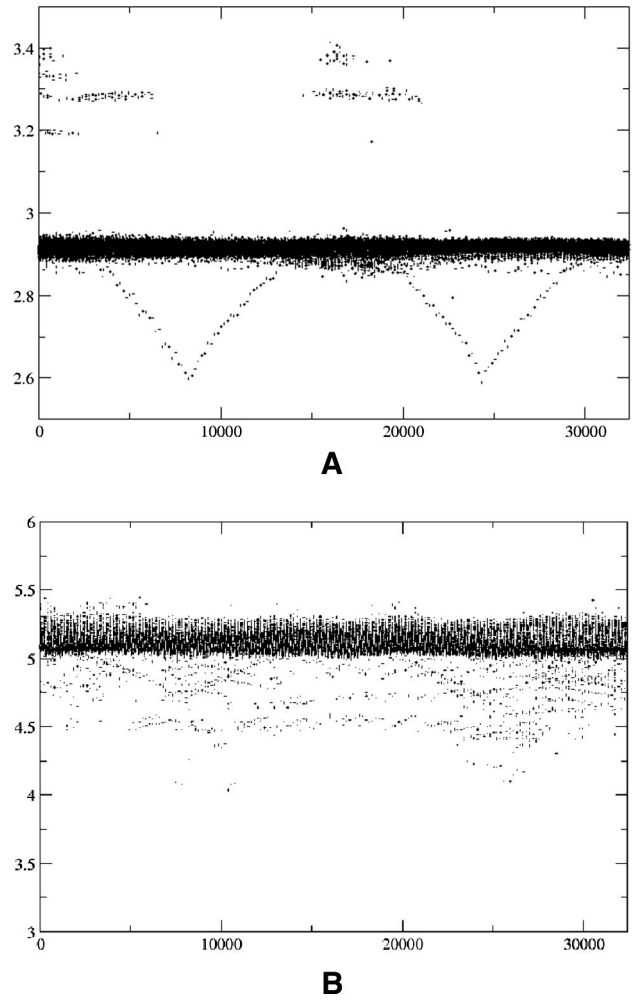


FIG. 7. Velocity scatter plots are shown for unflawed pipe samples at 0.5 MHz (a) and 1.35 MHz (b). If our arrival-time extraction algorithms were perfect, all of the data points would be clustered about the dark horizontal bands.

culated with the updated cell velocities calculated below in step (3):

$$T^k[i,j] = \sum_{m,n \in \text{ray}[i,j]} \frac{d[i,j,m,n]}{\nu^k[m,n]}, \tag{3}$$

where k is the iteration number.

(ii) Step 2: For every ray, calculate the difference between the velocities in the current iteration from those in the previous iteration for each cell that the ray passes through:

$$\Delta \frac{1}{\nu_{m,n \in \text{ray}[i,j]}[m,n]} = \frac{T^k[i,j] - T^{k-1}[i,j]}{L[i,j]}, \tag{4}$$

where $L[i,j]$ is the length of ray$[i,j]$ and $T[i,j]$ is the experimentally measured arrival time for the ray$[i,j]$. During each iteration, cycle through each ray and record how many times each individual cell has a change in velocity and what that change is.

(iii) Step 3: Finally, update each cell's velocity by taking the average of the differences recorded for that cell in step 2 and adding it to the cell's current velocity:

$$\frac{1}{\nu^{k+1}[m,n]} = \frac{1}{\nu^k[m,n]} + \Delta_{\text{AVG}} \frac{1}{\nu[m,n]}. \tag{5}$$

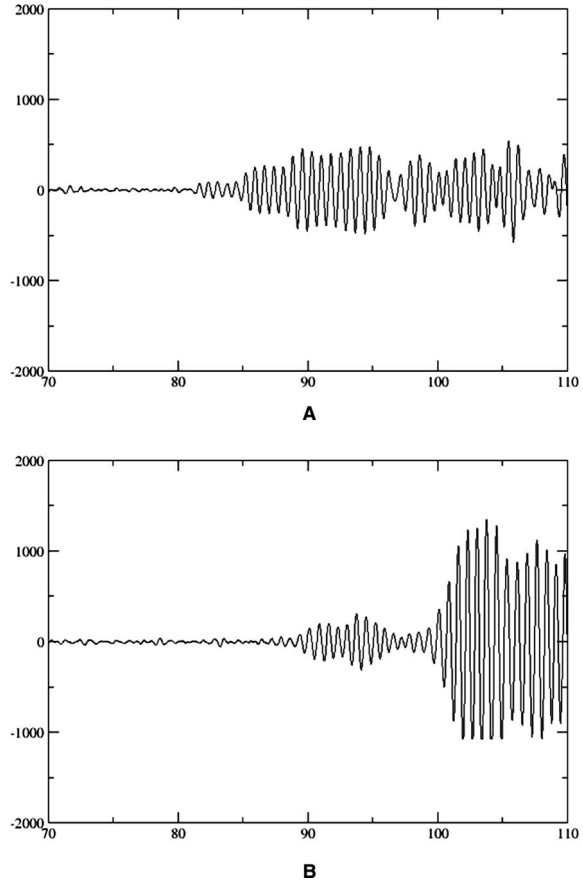


FIG. 8. Typical waveforms for meridional (a) and helical (b) guided wave modes.

(iv) Step 4: Steps (1)–(3) are repeated until the required accuracy is reached. For our results below 50 global iterations were used.

IV. DESCRIPTION OF SETUP AND EXPERIMENTAL RESULTS

We have constructed the apparatus shown in Fig. 4 to mimic two circumferential belts of transducers via a pair of transducers incremented by stepper motors under computer control. For each transmitter location a guided wave is launched by exciting the contact transducer with a toneburst while the receiver is paused briefly in turn at all of the circumferential positions to catch the various helical Lamb waves. The frequency of the tone burst is chosen to select the guided wave modes of interest, typically those that will be first arriving in order to minimize complications from overlapping modes. The recorded waveforms at each pitch–catch pair location are digitized and processed on the computer to extract the arrival times and/or amplitudes of the mode(s) of interest. Additional projections are taken by stepping the transmitter in turn to all of the circumferential positions and repeating the process. The complete set of arrival times or amplitudes are then passed to the tomographic reconstruction codes.

A block diagram of the data acquisition equipment for the HUT scanner is also shown in Fig. 4. The Matec™ TB-1000 PC ISA board is used to form a tone burst of typically

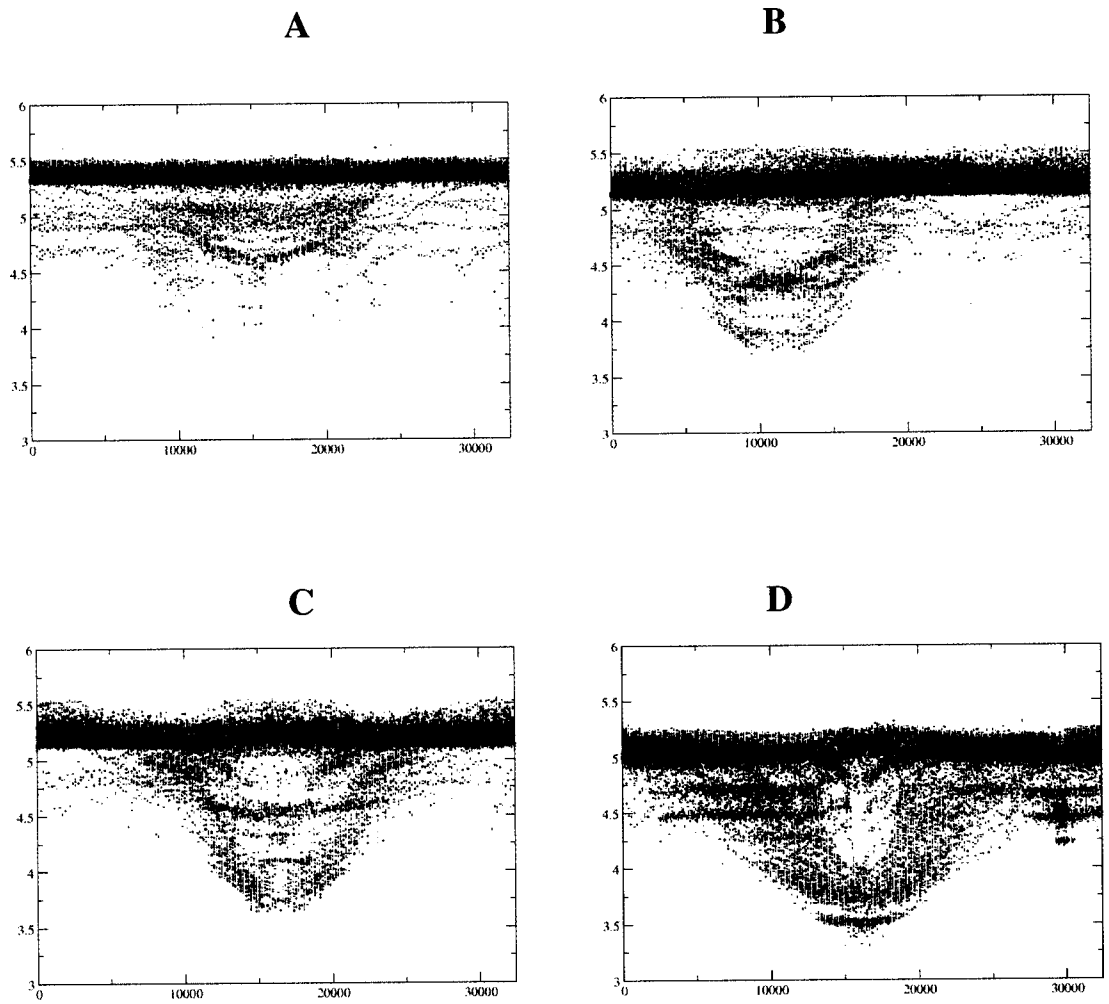


FIG. 9. Velocity scatter plots are shown for flawed samples from Fig. 6, as extracted by our algorithms versus ray number. The rays that pass through the flaws indicate a different velocity because the modes we select for our measurements are dispersive: *instantaneous speed varies with local pipe wall thickness*. A second effect also occurs at the strongly scattering edges, which is a diffraction of the guided wave modes. This also often shows up as a slowing, because the waves tend to skirt around the flaw and hence take an effectively longer path from transmitter to receiver.

3–15 cycles. In the reconstructions shown here, a 15-cycle 1.35 MHz tone burst drives the transmitting transducer, a 2.25 MHz center frequency broadband longitudinal transducer, which excites Lamb-like guided wave modes in the pipe. The received signal is amplified, filtered, and then digitized by a GAGE™ CS8012A DAQ board with 12-bit resolution and 100 MHz sampling rate. Initial testing was performed on a steel pipe with an ID of 75 mm and an OD of 102 mm. The distance between the transmitter and receiver at the same circumferential angle was 320 mm. Glycerin was used as a coupling agent because it provided good transmission of the signal and had a high enough surface tension that the coupling was consistent, even on the bottom side of the pipe. Each recorded waveform was gated around the first arriving mode, which in our measurements was the S6 mode.

For this pipe sample that mode has a group velocity of a little over 5 mm/μs and is reasonably nondispersive. The next fastest modes we record are less than 4 mm/μs with several slower modes present, but in the 3–4 mm/μs range, so they separate out from the first arriving S6 mode. After the data had been recorded, the arrival time for the fastest mode was extracted for each waveform and then sent to the tomographic algorithm described above, resulting in the reconstructions shown below. All of the signal processing algorithms are fully automated since the large number of individual pitch–catch measurements necessary precludes any manual analysis of the waveforms. The waveform processing, mode-extraction, and tomographic reconstructions take only a total of a few seconds on a modern PC running Linux.

Figure 5 is a photo of the thick-walled steel pipe sample with a rectangular flat thinned region. This sample, with four different lengths of flaws, was used for the tests. The thinning is about 25% at the center of the rectangular region, with only a surface slope change (but no thickness change) at the circumferential edges and a varying-depth step discontinuity at the axial edges. The photo also shows the conical delay lines that are used with the contact transducers in order to minimize their footprint.

Figure 6 shows four reconstructions for increasing length of the rectangular thinning. The red horizontal striations and the blue criss-cross artifact indicate the location of the flaws, which can be seen to increase in size, as expected. Note that the scans are shown for different angular positions of the flaws, but all other scanning, reconstruction, and rendering parameters are remained unchanged. No additional image processing or enhancement has been done to improve

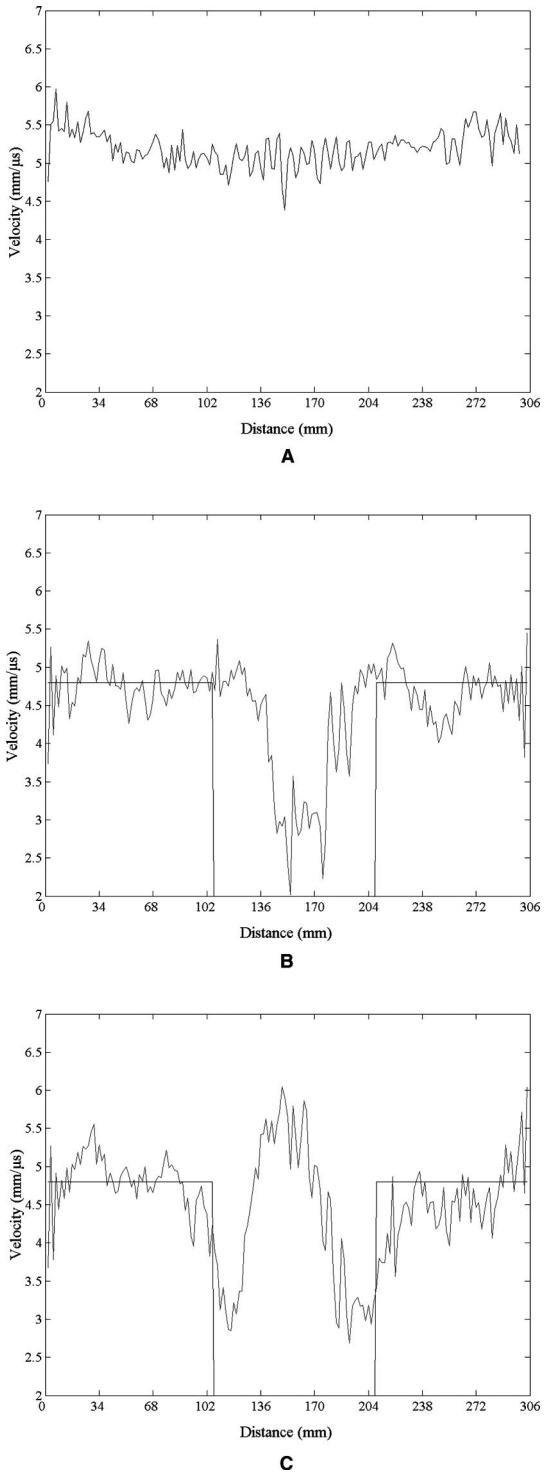


FIG. 10. Each graph represents a different horizontal cross section of the tomographic reconstruction shown in Fig. 6(b). These velocity profiles can be converted to thickness profiles. (a) A horizontal cross section outside the flawed region. The background level is around 5.25 mm/μs, which is to be expected from the unflawed portion of the sample. (b) This horizontal cross section is taken from the red region that can be seen on the top edge of the flaw in Fig. 6(b). (c) The velocity profile for the horizontal cross section directly through the center of the flaw.

the appearance of these ''raw'' reconstructions. Note also that we have not yet made other than a few minor changes to our Lamb wave tomography codes. Significant enhancement is expected as this software is optimized.

V. DISCUSSION

Figure 7 shows two scatter plots of the velocities extracted by our algorithms versus the ray number for the pipe in its unflawed state, scanned at two different frequencies: 0.5 and 1.35 MHz. A ''perfect'' result would show all of the velocities in the tight band at about 2.9 mm/μs (7 A) and 5.2 mm/μs [Fig. 7(b)] which are the velocities for the mode of interest at the two frequencies. Note several things from these plots. In Fig. 7(a), two clusters of rays gave points between 3.2 and 3.4 mm/μs, which likely means that our thresholding to ignore the early pipe modes needs to be adjusted. Also apparent is the weak regular structure that has two dips. We speculate that this is when the two counterhelically propagating mode pairs destructively interfere (pitch and catch directly opposite each other) and thus the mode of interest zeros out. Consequently, our algorithms then pick out one of the nonhelical pipe modes. In Fig. 7(b) the pipe modes were not apparent because of the higher frequency, and we see a fair number of points where our algorithms appear to have missed the mode of interest and instead picked out slower (smaller velocity) modes below the dark band at about 5.2 mm/μs.

Figure 8 shows two typical waveforms recorded by our system on the unflawed pipe sample at 1.35 MHz. The top one is for a meridional wave, where both transducers are at the same angular position. The bottom one is when they are 90° rotated from one another. Both are single shot without averaging and are shown on an arbitrary amplitude scale that is consistent between them. Note that the signals we are dealing with do not have distinct, isolated modes, and that the essential character of the waveforms changes dramatically throughout the measurements, even in the absence of flaws. We tend to use the subtle arrival-time changes in the first-arriving modes because coupling variations in our laboratory scanning system introduce uncertainty into amplitude measurements and because our mode identification and extraction algorithms are not quite robust enough yet to deal with the later mutually overlapping modes.

The four plots in Fig. 9 show the same representation of velocities for the four flawed samples as Fig. 7. For those rays that do not pass through the flaw, the velocity is constant, and our algorithm reliably returns a velocity in the band just above 5 mm/μs, which is the velocity of the S6 mode of interest for these scans. Because the modes we select for our measurements are dispersive, their instantaneous speed varies with local pipe wall thickness and the rays that pass through the flaws indicate a different velocity.

Additionally, diffraction of the guided wave modes occurs at the strongly scattering edges of the flaws. This often shows up as a slowing of the guided wave, because the waves tend to skirt around the flaw and hence take an effectively longer path from transmitter to receiver. Figure 10 shows the tomographic velocity profiles for three different horizontal slices through Fig. 6(b). It can be seen that outside the flaw the velocity remains close to the background level of about 5.25 mm/μs. The other two profiles are taken from different locations within the flaw. Figure 10(c) shows how the waves have an apparent decrease in velocity as they encounter the edges of the flaw and an increase in velocity at

the center of the flaw, where they do not interact with strongly scattering edges. Accounting for the complicated physics involved and developing optimized algorithms to extract the different mode arrivals is a key part of our future work, the goal of which is clear and unambiguous reconstructions that localize, size, and quantify flaws in real time.

ACKNOWLEDGMENTS

This work was supported in part by the U.S. Army under Contract No. DAAE30-02-C-1061 and by the Virginia Space Grant Consortium. The authors would also like to thank Jonathan Stevens for constructing much of the instrumentation.

[1] I. A. Viktorov, *Rayleigh and Lamb Waves—Physical Theory and Applications* (Plenum, New York, 1967).
[2] J. L. Rose, *Ultrasonic Waves in Solid Media* (Cambridge University Press, Cambridge, 1999).
[3] D. C. Gazis, ''Three dimensional investigation of the propagation of waves in hollow circular cylinders I. Analytical foundation,'' J. Acoust. Soc. Am. **31**, 568–573 (1959).
[4] D. C. Gazis, ''Three dimensional investigation of the propagation of waves in hollow circular cylinders II. Numerical results,'' J. Acoust. Soc. Am. **31**, 573–5738 (1959).
[5] T. R. Meeker and A. H. Meitzler, ''Guided wave propagation elongated cylinders and plates,'' Phys. Acoust. **1A**, 111–167 (1964).
[6] J. J. R. Zemmanek, ''An experimental and theoretical investigation of elastic wave propagation in a cylinder,'' J. Acoust. Soc. Am. **52**, 265–283 (1972).
[7] W. Mohr and P. Holler, ''On inspection of thin walled tubes for transverse and longitudinal flaws by guided ultrasonic waves,'' IEEE Trans. Sonics Ultrason. **SU-23**, 369–374 (1976).
[8] M. G. Silk and K. F. Bainton, ''The propagation metal tubing of ultrasonic wave mode equivalent to Lamb waves,'' Ultrasonics **17**, 11–19 (1979).
[9] J. J. Ditri, ''Utilization of guided elastic waves for the characterization of circumferential cracks in hollow cylinders,'' J. Acoust. Soc. Am. **96**, 3769–3775 (1994).
[10] J. L. Rose, J. J. Ditri, A. Pilarski, K. Rajana, and F. T. Carr, ''A guided wave inspection technique for nuclear steam generator tubing,'' NDT & E Int. **27**, 307–330 (1994).
[11] J. L. Rose, K. M. Rajana, and F. T. Carr, ''Ultrasonic guided wave inspection concepts for steam generator tubing,'' Mater. Eval. **52**, 307–311 (1994).
[12] D. N. Alleyne and P. Cawley, ''The excitation of Lamb waves, in pipes using dry coupled piezoelectric transducers,'' J. Nondestruct. Eval. **15**, 11–20 (1996).
[13] D. N. Alleyne and P. Cawley, ''Long range propagation of Lamb waves in chemical plant pipework,'' Mater. Eval. **55**, 504–508 (1997).
[14] D. N. Alleyne, P. Cawley, A. M. Lank, and P. J. Mudge, ''The Lamb wave inspection of chemical plant pipework,'' *Review of Progress in QNDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1997), Vol. 16, pp. 1269–1276.
[15] M. J. S. Lowe, D. N. Alleyne, and P. Cawley, ''Defect detection in pipes using guided waves,'' Ultrasonics **36**, 147–154 (1998).
[16] D. Guo and T. Kundu, ''A new transducer holder mechanism for pipe inspection,'' J. Acoust. Soc. Am. **110**, 303–309 (2001).
[17] J. L. Rose, D. Jiao, and J. Spanner, Jr., ''Ultrasonic guided wave NDE for piping,'' Mater. Eval. **54**, 1310–1313 (1996).
[18] M. J. Quarry and J. L. Rose, ''Multimode guided wave inspection of piping using Comb transducers,'' Mater. Eval. **57**, 1089–1090 (1999).
[19] Z. Sun, J. L. Rose, M. Quarry, and D. Chin, ''Flexural mode tuning in pipe inspection,'' *Review of Progress in QNDE*, edited by D. O. Thompson and D. E. Chimenti (American Institute of Physics, Melville, NY, 2002), Vol. 21, pp. 262–269.
[20] W. Böttger, H. Schneider, and W. Weingarten, ''Prototype EMAT system for tube inspection with guided ultrasonic waves,'' Nucl. Eng. Des. **102**, 356–376 (1987).
[21] M. Hirao and H. Ogi, ''An SH-wave EMAT technique for gas pipeline inspection,'' NDT & E Int. **31**, 127–132 (1999).
[22] G. A. Alers and J. D. McColskey, ''Measurement of residual stress in bent pipelines,'' in Ref. 19, pp. 1681–1686.
[23] P. Wilcox, M. Castaings, R. Monkhouse, P. Cawley, and M. J. S. Lowe, ''An example of the use of interdigital PVDF transducers to generate and receive a high order Lamb wave mode in a pipe,'' in Ref. 19, Vol. 16, pp. 919–926.
[24] D. N. Alleyne, M. J. S. Lowe, and P. Cawley, ''The reflection of guided waves from circumferential notches in pipes,'' ASME J. Appl. Mech. **65**, 635–641 (1998).
[25] M. J. S. Lowe, D. N. Alleyne, and P. Cawley, ''The mode conversion of a guided wave by a part-circumferential notch in a pipe,'' ASME J. Appl. Mech. **65**, 649–656 (1998).
[26] Z. Sun, Y. Mao, W. Jian, and D. Zhang, ''Investigation on interaction of Lamb waves and circumferential notch in pipe by means of wavelet transform,'' IEEE Ultrason. Symp. 827–830 (2000).
[27] Demma P. Cawley and M. J. S. Lowe, ''Guided waves in curved pipes,'' in Ref. 19, Vol. 21, pp. 157–164.
[28] J. D. N. Cheeke, X. Li, and Z. Wang, ''Observation of flexural Lamb waves (A0 mode) on water-filled cylindrical shells,'' J. Acoust. Soc. Am. **104**, 3678–3680 (1998).
[29] K. Shannon, X. Li, Z. Wang, and J. D. N. Cheeke, ''Mode conversion and the path of acoustic energy in a partially water-filled aluminum tube,'' Ultrasonics **37**, 303–307 (1999).
[30] C. Aristegui, M. J. S. Lowe, and P. Cawley, ''Guided waves in fluid-filled pipes surrounded by different fluids,'' Ultrasonics **39**, 367–375 (2001).
[31] J. N. Barshinger, and J. L. Rose, ''Ultrasonic guided wave propagation in pipes with viscoelastic coatings,'' in Ref. 19, Vol. 21, pp. 239–246.
[32] R. Long, K. Vine, M. J. S. Lowe and P. Cawley, ''The effect of soil properties on acoustic wave propagation in buried iron water pipes,'' in Ref. 19, Vol. 21, pp. 1310–1317.
[33] J. L. Rose, Z. Sun, S. J. Mudge, and M. J. Avioli, ''Guided wave flexural mode tuning and focusing for pipe testing,'' Mater. Eval. **61**, 162–167 (2003).
[34] M. C. Junger and D. Feit, *Sound, Structures, and Their Interaction* (The MIT Press, Cambridge, MA, 1972).
[35] J. Callahan and H. Baruh, ''A closed-form solution procedure for circular cylindrical shell vibrations,'' Int. J. Solids Struct. **36**, 2973–3013 (1999).
[36] V. B. Poruchikov, ''Response of a cylindrical elastic shell to an applied impulse,'' Mech. Solids **35**, 147–152 (2000).
[37] V. A. Kovalev, L. Yu. Kossovich, and A. V. Nikonov, ''Transient waves in a cylindrical shell subjected to sudden harmonic loads,'' Mech. Solids **35**, 143–152 (2000).
[38] 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**, 1764–1760 (2001).
[39] A. D. Pierce and H. G. Kil, ''Elastic wave propagation from point excitations on thin-walled cylindrical shells,'' J. Sound Vib. **112**, 399–406 (1990).
[40] J. L. Rose, ''A baseline and vision of ultrasonic guided wave inspection potential,'' J. Pressure Vessel Technol. **124**, 273–282 (2002).
[41] J. C. P. McKeon and M. K. Hinders, ''Lamb wave contact scanning tomography,'' *Review of Progress in QNDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1999), Vol. 18, p. 951.
[42] J. C. P. McKeon and M. K. Hinders, ''Parallel projection and crosshole Lamb wave contact scanning tomography,'' J. Acoust. Soc. Am. **106**, 2568–2577 (1999).
[43] E. V. Malyarenko and M. K. Hinders, ''Fan beam and double crosshole Lamb wave tomography for mapping flaws in Aging Aircraft Structures,'' J. Acoust. Soc. Am. **108**(10), 1631 (2000).
[44] E. V. Malyarenko and M. K. Hinders, ''Ultrasonic Lamb wave diffraction tomography,'' Ultrasonics **39**, 269–281 (2001).
[45] M. K. Hinders, K. R. Leonard, and E. V. Malyarenko, ''Blind test of Lamb wave diffraction tomography,'' in Ref. 19, Vol. 21, pp. 278–283.
[46] K. R. Leonard, E. V. Malyarenko, and M. K. Hinders, ''Ultrasonic Lamb wave tomography,'' Inverse Probl. **18**, 1795–1808 (2002).
[47] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging* (IEEE Press, New York, 1988).

Bubble levitation and translation under single-bubble sonoluminescence conditions

Thomas J. Matula[a)]
Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105

(Received 28 January 2003; revised 8 May 2003; accepted 9 May 2003)

Bubble levitation in an acoustic standing wave is re-examined for conditions relevant to single-bubble sonoluminescence. Unlike a previous examination [Matula *et al.*, J. Acoust. Soc. Am. **102**, 1522–1527 (1997)], the stable parameter space $[P_a, R_0]$ is accounted for in this realization. Forces such as the added mass force and drag are included, and the results are compared with a simple force balance that equates the Bjerknes force to the buoyancy force. Under normal sonoluminescence conditions, the comparison is quite favorable. A more complete accounting of the forces shows that a stably levitated bubble does undergo periodic translational motion. The asymmetries associated with translational motion are hypothesized to generate instabilities in the spherical shape of the bubble. A reduction in gravity results in reduced translational motion. It is hypothesized that such conditions may lead to increased light output from sonoluminescing bubbles.
 [DOI: 10.1121/1.1589753]

PACS numbers: 43.35.Hl, 43.25.Yw [AJS]

I. INTRODUCTION

Single-bubble sonoluminescence (SBSL) refers to the periodic[1] energetic collapse of a single bubble, leading to the emission of light.[2] SBSL is obtained by levitating a bubble near the pressure antinode of an acoustic standing wave. Bubble levitation occurs because of a balance between the time-averaged acoustic radiation (Bjerknes) force and the time-averaged buoyancy force.[3,4]

Besides the obvious mechanistic rationale for studying SBSL, there are important issues related to nonlinear bubble oscillations coupled to bubble translation that have yet to be explored in detail. Applications include fluid processing[5] and medical ultrasound.[6] In these and other cases, bubbles can form and grow, and be manipulated by acoustic waves. Although much is known with linear coupled equations,[7] it is important to understand the coupled behavior of nonlinear pulsation and translation. Bubble levitation systems are an ideal starting point because they are principally understood. Bubble levitation and translation in these well-defined systems can provide clues as to how ultrasound interacts with highly nonlinear bubble oscillations. For example, in SBSL bubbles are usually seeded far from their levitation position. It is the action of the relatively large acoustic radiation force that pulls them toward the antinode. The translation of these bubbles to their levitation position has not yet been studied. In this paper, we will utilize a coupled oscillation/translation set of equations to examine bubble translation under SBSL conditions. Recently, a similar formulation was used to show that maximum translation occurs during bubble collapse.[8]

In previous work, we showed that the levitation position of an SBSL bubble (relative to the pressure antinode) moved upwards with drive pressure (the opposite direction to small-amplitude forcing).[9] However, the calculations assumed that the ambient bubble size remained the same for all drive pressures. We now know that the bubble size changes with drive pressure.[10,11] Our paper will take into account this new information. For simplicity, we will limit ourselves to just a single dissolved gas concentration.

Several applications will be examined. We will compare the levitation position within a rectangular and spherical resonator driven at the same frequency. We will look at the initial bubble formation, and the translation of the bubble from initial formation to its equilibrium levitation position. We will also examine bubble levitation under various gravitational conditions, and show a comparison with an experiment. Finally, we will show that the translational velocity of an SBSL bubble increases with drive pressure, and may be the mechanism for the extinction threshold, the upper limit of energy concentration for SBSL bubbles.

II. THEORY

The model we use consists of a radial equation of motion coupled to a translational equation of motion. The calculations are limited in that we do not account for the vapor pressure (and corresponding vapor diffusion), or that a translating bubble may depart from its spherical shape. However, because we are interested in the instantaneous motion of the bubble, we do not decouple the equations by utilizing an ''average'' bubble radius.

A. Radial model

The radial equation of motion is a modified Rayleigh-Plesset equation that is common in SBSL modeling. The equation is[2]

$$\rho\left(RR+\frac{3}{2}R^2\right)=[p_g-P(t)-P_0]+\frac{R}{c}\frac{d}{dt}(p_g)$$
$$-4\eta\frac{R}{R}-\frac{2\sigma}{R}. \tag{1}$$

a)Electronic mail: matula@apl.washington.edu

In the calculations, we used an ambient pressure $P_0=1.103\times10^5$ Pa, frequency $f=25$ kHz, speed of sound in water $c=1500$ m/s, water density $\rho=1000$ kg/m^3, shear viscosity $\eta=10^{-3}$ kg/(m·s), and surface tension $\sigma=0.072$ N/m. The temporal component of the driving pressure was $P(t)=-P_a\sin(\omega t)$. The internal gas pressure was calculated from

$$p_g(t)=\left(P_0+\frac{2\sigma}{R_0}\right)\left[\frac{R_0^3-h^3}{R(t)^3-h^3}\right]^{\gamma}, \tag{2}$$

with $\gamma=1$, to keep the calculations simple.

B. Diffusional stability

In order to solve the radial equation of motion, knowledge of the ambient radius R_0 and drive pressure P_a must be known. These parameters are obtained by our current understanding that stable SBSL bubbles are diffusionally stable; that is, there is no net mass flux during a given acoustic cycle (we are neglecting the shape-stability criteria, since most of the results are applied to bubbles known to be shape stable). A common equation to describe the diffusion stability condition is[2]

$$\frac{c_\infty}{c_0}=\frac{\int R^4 p_g\,dt}{P_0\int R^4\,dt}, \tag{3}$$

where c_∞ refers to the gas concentration far from the bubble, and c_0 refers to the saturated gas concentration. In this paper, we assume that the concentration ratio is 0.001. This condition applies to air bubbles in water, which are known to be mostly filled with argon (because the diatomic gases N_2 and O_2 are dissociated and their products condense in the surrounding water).[12]

Solving the radial equation of motion for many different drive pressures and ambient radii, subject to the stability condition generates a particular set of $[P_a,R_0]$ pairs which are stable against diffusion. The $[P_a,R_0]$ pairs make up what is commonly referred to as the parameter space for SBSL.

Therefore, the first order of business is to determine the set of $[P_a,R_0]$ pairs that provides stable SBSL at 25 kHz. In order to be complete, we included pairs that are outside the expected SBSL range (smaller bubbles are below the luminescence threshold, while larger bubbles are above the extinction threshold, where shape instabilities that are not discussed here occur). Figure 1 shows the $[P_a,R_0]$ pairs from $R_0=1\,\mu$m up to $R_0=6.8\,\mu$m. In some instances we will perform a sample calculation with a specific value of $R_0=4.0\,\mu$m, within the predicted stability region for SBSL; the corresponding pressure amplitude is $P_a=1.44$ atm.

C. Resonators

Two geometries for levitation cells (or resonators) are considered; spherical, and rectangular with a square cross section. The distinction is important because the spatial pressure profile is used in the translation calculations. To be consistent, we will assume that both the rectangular and spherical resonators are water-filled, with an air boundary such that the pressure goes to zero at the boundary, and that the resonance frequency of both is 25 kHz. The center of the resonators is taken to be $x,y,z=0$. We also assume that the horizontal forces are symmetrical, and thus cancel (the bubble is levitated at $x,y=0$). In the vertical direction, the body force of gravity produces an asymmetry, giving rise to vertical translation of the bubble.

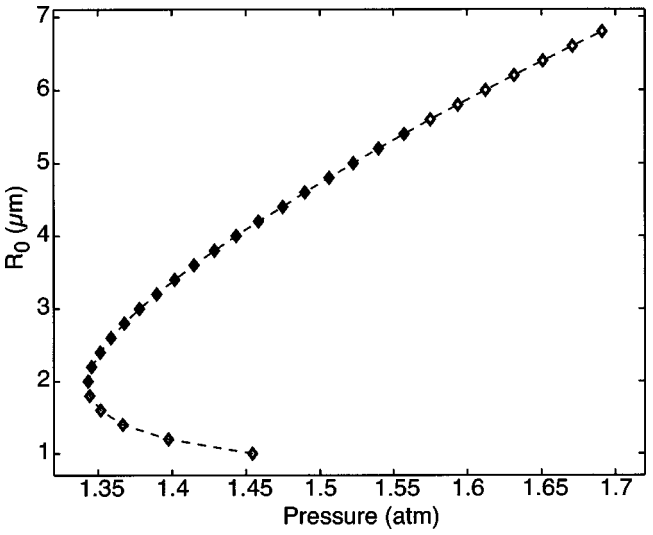


FIG. 1. Stability parameter space with $C_\infty/C_0=0.001$. The filled markers represent the SBSL region. The calculated luminescence threshold is near $P_a\approx1.35$ atm and the extinction threshold is near $P_a\approx1.55$ atm. These points serve only as a guide, and are not meant to be absolute.

The spatial pressure profile in a spherical resonator goes as $P(r)=P_a j_0(kr)=P_a\sin(kr)/(kr)$, where $k=\omega/c$. This zeroth-order spherical Bessel function has its first zero (resonator wall) at π. Thus, the resonator's radius is found from $ka=2\pi fa/c=\pi$, or $a=c/(2f)=3$ cm.

The vertical spatial pressure profile in the rectangular resonator goes as $P(z)=P_a\cos(k_z z)$, where $k_z=2\pi/\lambda_z$. The idea is to find a geometry to compare with the spherical resonator ($f=25$ kHz, "radius"$=3$ cm). We therefore take the radius as half its height. This implies that the rectangular resonator is 6 cm tall, and has a wavelength $\lambda_z=12$ cm. The cross-section dimensions are found by noting that the wave number $k=2\pi f/c=\sqrt{2k_x^2+k_z^2}$ (with a square cross section, $k_x=k_y$). With this information, we find that the rectangular geometry is 4.9 cm square by 6 cm tall.

D. Levitation position (simple relation)

In this section we look for an analytical expression that describes the equilibrium bubble position z for a given set of $[P_a,R_0]$ pairs. Following the work of Eller[3] and Matula *et al.*,[9] we consider that at equilibrium there exists a balance between the time-averaged acoustic radiation (Bjerknes) force and the buoyancy force. These forces are given by

$$F_{\text{rad}}=-V(t)\nabla P, \tag{4}$$

and

$$F_{\text{buoy}}=\rho g z, \tag{5}$$

where V(t) is the instantaneous volume of the bubble, g ($=9.8$ m/s^2 in earth-based laboratories) is the acceleration due to gravity and $z=z(t)$ is the position of the bubble,

measured relative to the pressure antinode. Therefore: $\langle F_{\text{rad}}\rangle = \langle F_{\text{buoy}}\rangle$. Note that the Bjerknes force depends on the geometry of the system, because of the ∇P term.

1. Rectangular resonator

The pressure here is given by $P = -P_a \cos(k_z z)\sin(\omega t)$, and the gradient of the pressure at the equilibrium position is $\nabla P = P_a \sin(k_z z_0)\sin(\omega t) \approx P_a k_z z_0 \sin(\omega t)$ for small z_0. The magnitude of the Bjerknes force is thus $\langle F_{\text{rad}}\rangle \approx P_a k_z z_0 \langle V(t)\sin(\omega t)\rangle$. By equating the two time-averaged forces, and solving for z_0, one obtains[9]

$$z_{0R} \approx \frac{\rho g}{k_z^2 P_a}\frac{\Lambda_1}{\Lambda_2}, \tag{6}$$

where

$$\Lambda_1 = \frac{1}{T}\int_0^T V(t)dt \quad \text{and} \quad \Lambda_2 = \frac{1}{T}\int_0^T V(t)\sin(\omega t)dt. \tag{7}$$

We added the subscript "*R*" to remind us that the levitation position is for a rectangular resonator.

2. Spherical resonator

Here, the pressure is given by $P = -P_a j_0(kr)\sin(\omega t) = -P_a \sin(kr)/(kr)\sin(\omega t)$, and the pressure gradient at equilibrium $r_0 = z_0$ (since the motion is vertical) is

$$\begin{aligned}\nabla P &= -kP_a j_1(kz_0)\sin(\omega t)\\ &= \frac{-P_a}{z_0}\left[\frac{\sin(kz_0)}{kz_0} - \cos(kz_0)\right]\sin(\omega t)\\ &\approx \frac{-P_a k z_0}{3}\sin(\omega t),\end{aligned} \tag{8}$$

for small z_0. Therefore, the equilibrium bubble position in a spherical resonator is

$$z_{0S} \approx \frac{3\rho g}{k^2 P_a}\frac{\Lambda_1}{\Lambda_2}, \tag{9}$$

where Λ_1 and Λ_2 are given by Eq. (7) and the "*S*" reminds us that the levitation position is for a spherical resonator. Equations (6) and (9) are thus the levitation positions for a rectangular and spherical cell, respectively.

In Fig. 2 we show how these two expressions compare. In all cases, the bubble position in a rectangular resonator is slightly higher than in the corresponding spherical resonator. Later, we will compare the spherical levitation expression with a numerical solution that includes other forces to show that there is little difference when the other forces are neglected. These other forces are, however, important when considering the translation of bubbles.

E. *Z*(*t*) model

For simplicity, we assume that the bubble remains spherical at all times. The translational equation of motion is obtained by summing up all the forces acting on the bubble. In addition to the acoustic radiation and buoyancy forces described above, there is also an added mass force which accounts for changes in momentum when a translating bubble changes size

$$F_{\text{mass}} = -\frac{1}{2}\rho\frac{d}{dt}[V(t)u_r], \tag{10}$$

and a drag force

$$\frac{-12\pi\rho u_r^2 R^2(t)}{\text{Re}}. \tag{11}$$

Here, $u_r = u_b - u_l$ is the relative velocity of the bubble, where u_l is the liquid velocity. u_l is obtained by solving Euler's equation; the Reynolds number is given by $\text{Re} = 2\rho u_r R/\mu$. There are several formulations for the drag force.[13] We examined both a Stokes drag and a "Crum" drag,[4] and found little difference. The addition of a "history" force is neglected here.[8]

The total net force gives rise to an acceleration of the bubble; thus[14]

$$m_b\dot{u}_b = -V(t)\nabla P + \rho g z - \frac{1}{2}\rho\frac{d}{dt}[V(t)u_r] - \frac{12\rho u_r^2 R^2}{\text{Re}}, \tag{12}$$

where the bubble mass $m_b = \rho_b V(t)$. In the following sections we examine the levitation position of a bubble by considering all the forces. We will also examine the instantaneous motion of a bubble during a given acoustic cycle.

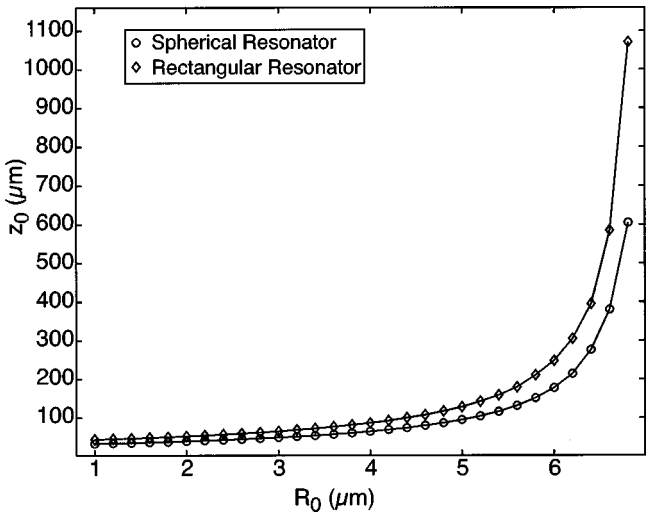
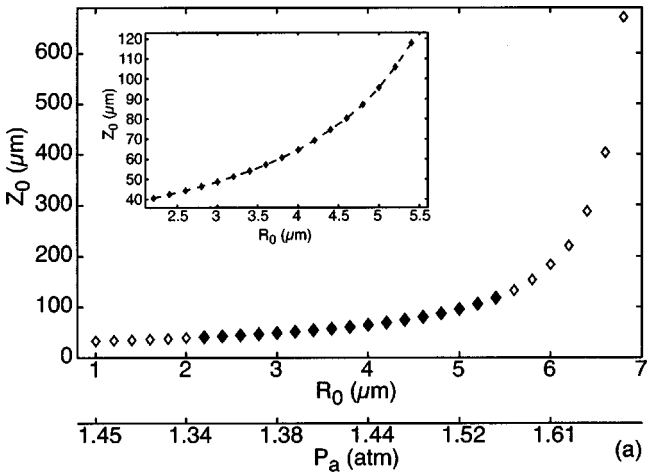


FIG. 2. The equilibrium levitation position of a bubble for a given $[P_a, R_0]$ pair. The squares are for a rectangular resonator and the circles are for a spherical resonator.

F. Levitation position (all forces considered)

In order to be brief, we show only the case of a 6-cm-diameter spherical resonator operating at 25 kHz. The rectangular resonator results are not significantly different. We calculated the final equilibrium position of the bubble by solving the translational equation of motion for a given set of $[P_a, R_0]$ pairs over many acoustic cycles until the average bubble position converged to an equilibrium position. The equilibrium position for the various $[P_a, R_0]$ pairs is shown in Fig. 3. Both a two-dimensional and three-dimensional rep-



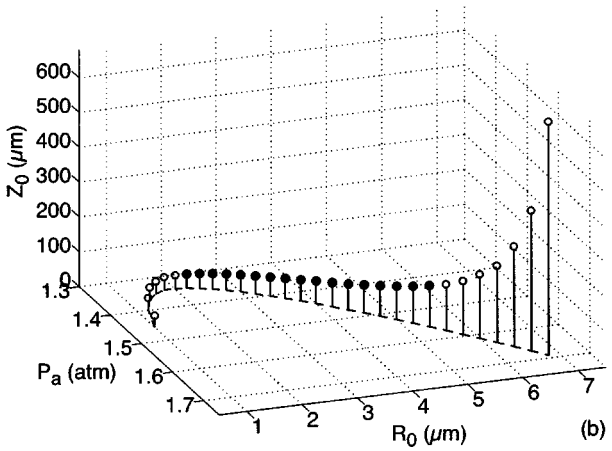


FIG. 3. (a) Steady-state stability position of a bubble above the pressure antinode for various ambient radii, or equivalently, drive pressures. The inset shows an expanded region in the SBSL parameter space, and corresponds to the darker markers. (b) A representation of the steady-state stability position of a bubble in the diffusive equilibrium parameter space. Stable SBSL bubbles are shown as darker, filled circles. Open circle positions below $\approx$2.5 μm are unstable, while open circle positions above $\approx$6.5 μm are above the extinction threshold.

resentation of the levitation position are shown. The three-dimensional view also includes the parameter space curve. Figure 4 compares the analytical expression for the levitation position, Eq. (9), with the more complicated steady-state solution of Eq. (12). There is little difference in the SBSL parameter space.

G. Bubble motion at equilibrium

Although the bubble is stationary in the averaged sense, there does exist instantaneous translation motion. The motion occurs because the bubble is levitated slightly above the pressure antinode, and the periodic growth and collapse results in periodic translational forces on the bubble.[9] In Fig. 5 we show the steady-state motion of a 4.0-μm bubble for several acoustic cycles. The radial motion (a) is repetitive, cycle to cycle. The translational motion (b) varies almost sinusoidally during an acoustic cycle, and appears to drift a little from cycle to cycle. The drift is probably due to numerical round-off errors in the calculations. The magnitude of the translational velocity (c) is maximized during the final moments of collapse.

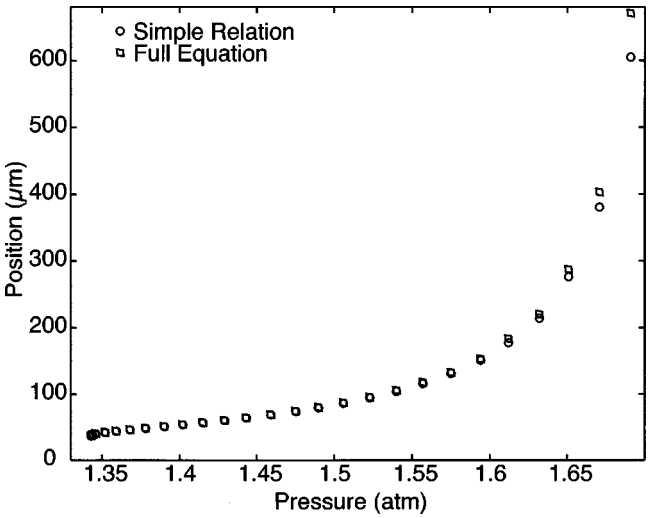


FIG. 4. A comparison between the numerical solution obtained by solving the full translational equation and the analytical solution obtained by equating the time-averaged acoustic radiation and buoyancy forces. In the stable SBSL parameter space, there is little difference between the two solutions.

A closer examination of the bubble motion during a single acoustic cycle is shown in Fig. 6(a). The sinusoidal translational motion in the z direction is interrupted during the final moments of collapse, when the added mass force dominates the motion, causing the bubble to accelerate.

An expanded view during the final moments of collapse is shown in Fig. 6(b). Note that the details shown here are critically dependent on the details of the gas within the bubble. Vapor trapping would undoubtedly have an influence over the details. The important point here is that the translational acceleration of the bubble dramatically increases during the final moments of collapse. We will examine this behavior a little more closely in the gravitational subsection.

The calculations above are meant to illustrate the physics of the coupled oscillation/translation of a sonoluminescence bubble. It should be noted that real systems have other

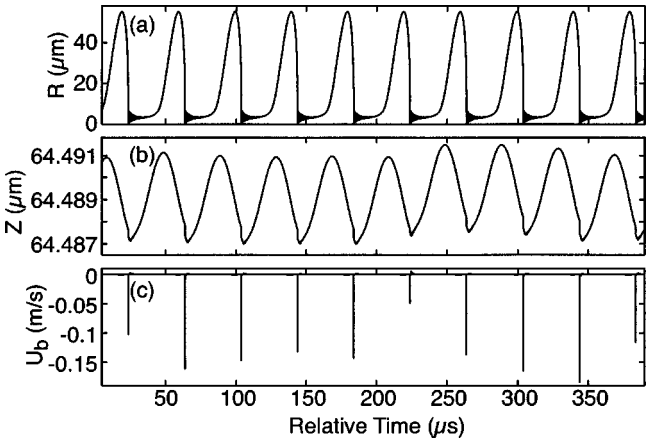


FIG. 5. The steady-state radius (a), position (b), and translational velocity (c) are shown here for ten consecutive acoustic cycles for a 4.0-μm radius bubble driven at 1.44 atm. From (c) we can see that the added mass force dominates during collapse.

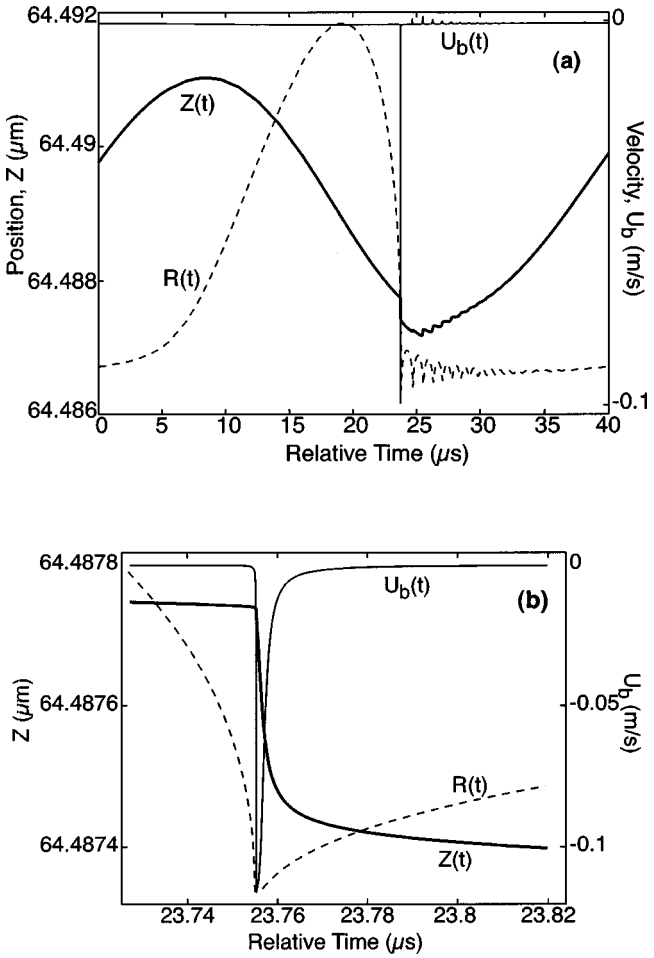


FIG. 6. (a) One acoustic cycle of a 4.0-μm radius bubble (corresponding pressure amplitude $P_a = 1.44$ atm). Shown are the vertical position $z(t)$ and translational velocity $U_b(t)$. The radius $R(t)$ plot is overlaid only for reference purposes. (b) A close-up view of the position $z(t)$ and translational velocity $U_b(t)$ during the final moments of bubble collapse and initial rebound. The radius $R(t)$ plot is overlaid only for reference purposes.

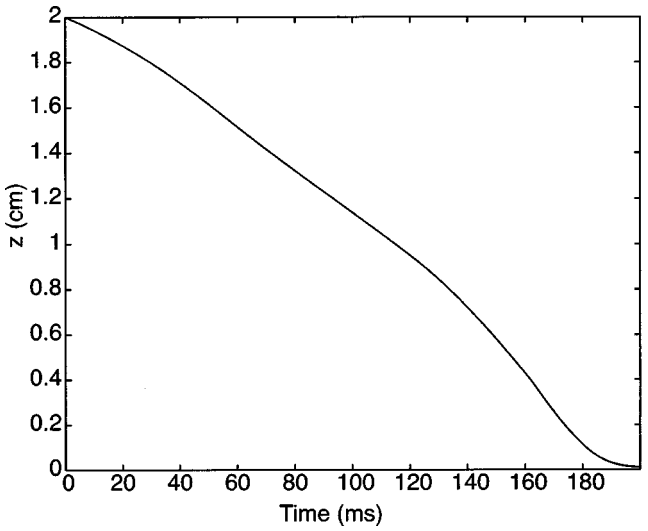


FIG. 7. Simulated experiment of a dissolving bubble being drawn to its levitation position. The initial bubble size is 100 μm, and decreased linearly to 10 μm after 5000 acoustic cycles. At each acoustic cycle, the pressure amplitude is adjusted to account for spatial variations in the field.

influences that we have not considered. For example, we have assumed that the system is driven by a single frequency. Holzfuss *et al.*[15] showed that inclusion of higher harmonics (present in any real system) can have dramatic influences on bubble stability, and in some cases results in bubble levitation *below* the calculated pressure antinode for a single-frequency drive. In another study, Holzfuss *et al.*[16] showed that the shock wave emitted by the collapsing bubble can be reflected and refocused back towards the bubble. The refocused wave may also influence bubble stability. The degree to which these influences affect bubble stability would of course depend on the specific levitation system.

III. APPLICATIONS

A. SBSL startup

Sonoluminescence experiments are often carried out by creating one or more bubbles near the air/water interface, with the drive pressure already set at the correct level for stable SBSL. For example, in our spherical resonators, a current pulse through a nichrome wire a couple of centimeters above the antinode creates bubbles that are then drawn to their levitation position very quickly. Coalescence and dissolution to reach a stable steady state is an automatic response of the system. The time it takes for a bubble to reach its levitation position has not been studied, but it appears to be on the order of milliseconds. We attempted to simulate this SBSL startup phase in order to determine if the calculated time scales seemed reasonable.

In this numerical experiment, we ''created'' a bubble 2 cm above the antinode and followed its motion for 5000 consecutive acoustic cycles. At the end of each acoustic cycle, we used the final value of its position as an initial condition for the next acoustic cycle. Because realistic bubbles are usually bigger than ambient SBSL bubbles, our initial bubble size was set to 100 μm. We assumed that the bubble would dissolve during its transition to its levitation position, since the water is highly degassed (we also assumed the bubble remained spherical at all times, although by translating through the fluid, it may deform). We thus numerically decreased the bubble size of each acoustic cycle linearly from its initial value to 10 μm at the end of the calculation (5000 acoustic cycles later). The final R_0 of 10 μm is about the size of an air bubble, before chemical diffusion removes the diatomic gases.[17]

We furthermore needed to account for a pressure amplitude that depends on the position of the bubble at the end of each acoustic cycle. That is, we chose a pressure amplitude that would generate a 4.0-μm stable SBSL bubble ($P_a = 1.44$ atm). However, a bubble located 2 cm from the antinode will feel a reduced pressure [by the factor $\sin(kr)/kr$]. We accounted for this by calculating a new driving pressure amplitude at each location of the bubble, over the entire 5000 acoustic cycles.

To recap, we began with a 100-μm bubble at a distance 2 cm above the antinode. We determined how far the bubble traveled at this local pressure amplitude in a single acoustic cycle. At the end of the acoustic cycle, we updated the initial conditions (R_0, P_a, and z_0,) and recalculated its motion. In

TABLE I. Levitation position, translational velocity, and translational acceleration for a 4.0-μm radius bubble ($P_a = 1.44$ atm).

Acceleration* g (m/s^2)	z (μm)	u_b (μm/μs)	$\dot{u}_b$ (μm/μs^2)
2	130	3×10^{-1}	2×10^{4}
1	65	2×10^{-1}	1×10^{4}
10^{-2}	65×10^{-2}	2×10^{-3}	200
10^{-4}	65×10^{-4}	5×10^{-5}	3

Fig. 7 we show how such a changing bubble is pulled to its levitation position. In this numerical experiment, the bubble nears its levitation position in about 200 ms (5000 acoustic cycles). For comparison purposes, if the initial bubble size was 50 μm, the translation time was about 100 ms; if the final bubble size was 50 μm, the translation time was on the order of 150 ms; if the initial and final bubble sizes were 20 and 10 μm, the translation time was close to 60 ms. Although our assumptions for initial, final, and rate of change in size are arbitrary, these time scales seem reasonable.

B. Gravitational effects

According to Eqs. (6) and (9), the levitation position of a bubble is directly proportional to the gravitational acceleration. Perhaps even more important is that the translational velocity and acceleration are also functions of g. A bubble moving through a fluid fast enough may lose its ability to focus enough energy to generate light, and thus this translational motion may be the dominant mechanism for the extinction threshold.[18]

We show in Table I the calculated levitation position of a 4.0-μm bubble ($P_a = 1.44$ atm) in a spherical resonator. The position, translational velocity, and acceleration are given for hypergravity ($\approx 2g$, $g = 9.8$ m/s^2), normal gravity ($1\,g$), in NASA's parabolic research aircraft, a KC135A ($g' \approx 10^{-2} g$), and in the International Space Station (ISS), where $g'' \approx 10^{-4} g$.

Comparison with experiment is difficult due to the lack of available data. We have preliminary levitation data from an experiment performed several years ago that shows the bubble position as it transitions from a hypergravity state ($\approx 2g$'s) to a microgravity state in the KC135A aircraft (see Fig. 8). The bubble moves about 400 μm during the transition. We calculate that a 4.0-μm bubble should move about 130 μm (see Table I). It should be emphasized that the calculation is performed under specific conditions that do not match the experimental conditions (which are not precisely known). For example, the equilibrium position of a 5.4-μm bubble (with a gas concentration ratio of 0.001) changes by about 240 μm, not 130 μm. Changing the gas concentration ratio would also change the results.

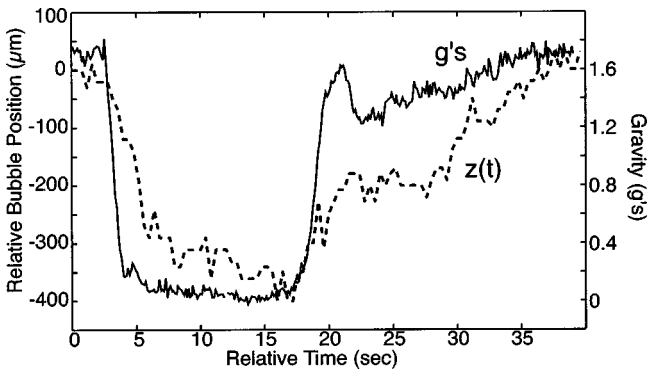


FIG. 8. The gravitational acceleration in NASA's parabolic research aircraft over a single parabola. The bubble translated about 400 μm during this parabola. As expected, in microgravity, the bubble moved closer to the antinode.

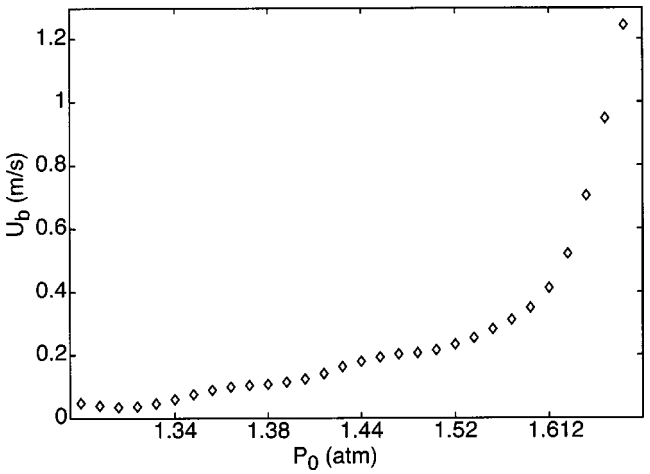


FIG. 9. The maximum translational velocity of the bubble is plotted as a function of drive pressure amplitude.

Measurements of the translational velocity or acceleration are even more difficult because the extent of the motion is very small (see Figs. 5 and 6). We hypothesize that these translational movements can lead to bubble instabilities. Specifically, the extinction threshold (the maximum energy concentration achievable in SBSL) may be a measure of instabilities brought about by bubble translation. It has been shown that instabilities limit the collapse strength,[11] and we argue that the mechanism for instabilities is bubble translation. Not enough information is known at this time to resolve the issue, but we show in Fig. 9 that the maximum translational velocity increases with drive pressure. If the dominant extinction mechanism is buoyancy, then one should expect that under microgravity conditions where the bubble velocity is reduced (see Table I), the extinction threshold should increase.

IV. CONCLUSIONS

Bubble levitation and translation was considered for conditions relevant to single-bubble sonoluminescence. Analytical expressions for bubble levitation were derived for a spherical resonator, and compared with a rectangular resonator. Comparable results were found using a more complete equation of motion for a bubble. The maximum bubble velocities and accelerations in microgravity are orders of magnitude lower than under terrestrial conditions, and may result in an increased extinction threshold.

ACKNOWLEDGMENTS

The author wishes to acknowledge S. Kargl for many helpful discussions. This work is supported by NASA NAG3-2379.

[1] D. F. Gaitan and L. A. Crum, ''Observation of sonoluminescence from a single, stable cavitation bubble in a water/glycerine mixture,'' in *Frontiers of Nonlinear Acoustics*, 12th ed. (Elsevier Applied Science, London, UK, 1990); B. P. Barber and S. J. Putterman, ''Observation of synchronous picosecond sonoluminescence,'' Nature (London) **352**, 318–320 (1991).
[2] M. P. Brenner, S. Hilgenfeldt, and D. Lohse, ''Single-bubble sonoluminescence,'' Rev. Mod. Phys. **74**, 425–484 (2002).
[3] A. I. Eller, ''Force on a bubble in a standing acoustic wave,'' J. Acoust. Soc. Am. **43**, 170–171 (1968).
[4] L. A. Crum, ''Bjerknes forces on bubbles in a stationary sound field,'' J. Acoust. Soc. Am. **57**, 1363–1370 (1975).
[5] Y. Hao, H. N. Oguz, and A. Prosperetti, ''The action of pressure-radiation forces on pulsating vapor bubbles,'' Phys. Fluids **13**, 1167–1177 (2001).
[6] P. A. Dayton, K. E. Morgan, A. L. S. Klibanov, G. Brandenburger, K. R. Nightingale, and K. W. Ferrara, ''A preliminary evaluation of the effects of primary and secondary radiation forces on acoustic contrast agents,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 1264–1277 (1997).
[7] L. A. Crum and A. I. Eller, ''The motion of air bubbles in stationary sound field,'' J. Acoust. Soc. Am. **48**, 181–189 (1970).
[8] A. J. Reddy and A. J. Szeri, ''Coupled dynamics of translation and collapse of acoustically driven microbubbles,'' J. Acoust. Soc. Am. **112**, 1346–1352 (2002).
[9] T. J. Matula, S. M. Cordry, R. A. Roy, and L. A. Crum, ''Bjerknes force and bubble levitation under single-bubble sonoluminescence conditions,'' J. Acoust. Soc. Am. **102**, 1522–1527 (1997).
[10] R. G. Holt and D. F. Gaitan, ''Observation of stability boundaries in the parameter space of single bubble sonoluminescence,'' Phys. Rev. Lett. **77**, 3791–3794 (1996).
[11] S. Hilgenfeldt, D. Lohse, and M. P. Brenner, ''Phase diagrams for sonoluminescing bubbles,'' Phys. Fluids **8**, 2808–2826 (1996).
[12] D. Lohse, M. P. Brenner, T. F. Dupont, S. Hilgenfeldt, and B. Johnston, ''Sonoluminescing air bubbles rectify argon,'' Phys. Rev. Lett. **78**, 1359–1362 (1997); T. J. Matula and L. A. Crum, ''Evidence of gas exchange in single-bubble sonoluminescence,'' *ibid.* **80**, 865–868 (1998); J. A. Ketterling and R. E. Apfel, ''Experimental validation of the dissociation hypothesis for single bubble sonoluminescence,'' *ibid.* **81**, 4991–4994 (1998).
[13] J. Magnaudet and D. Legendre, ''The viscous drag force on a spherical bubble with a time-dependent radius,'' Phys. Fluids **10**, 550–554 (1998).
[14] T. Watanabe and Y. Kukita, ''Translational and radial motions of a bubble in an acoustic standing wave field,'' Phys. Fluids A **5**, 2682–2688 (1993).
[15] J. Holzfuss, M. Ruggeberg, and R. G. Holt, ''Acoustical stability of a sonoluminescing bubble,'' Phys. Rev. E **66**, 046630 (2002).
[16] J. Holzfuss, M. Ruggeberg, and A. Billo, ''Shock wave emissions of a sonoluminescing bubble,'' Phys. Rev. Lett. **81**, 5434–5437 (1998).
[17] B. P. Barber and S. J. Putterman, ''Light scattering measurements of the repetitive supersonic implosion of a sonoluminescing bubble,'' Phys. Rev. Lett. **69**, 3839–3842 (1992); D. Lohse and S. Hilgenfeldt, ''Inert gas accumulation in sonoluminescing bubbles,'' J. Chem. Phys. **107**, 6986–6997 (1997).
[18] T. J. Matula, J. E. Swalwell, V. Bezzerides, P. Hilmo, M. Chittick, L. A. Crum, D. K. Kuhns, and R. A. Roy, ''Single-bubble sonoluminescence in microgravity,'' J. Acoust. Soc. Am. **102**, 3185 (1997); T. J. Matula, R. A. Roy, L. A. Crum, and D. W. Kuhns, ''Preliminary experimental observations of the effects of buoyancy on single-bubble sonoluminescence in microgravity and hypergravity,'' *ibid.* **100**, 2717 (1996).

Thermal modeling and performance analysis of a thermoacoustic refrigerator

David G. Holmberg and G. S. Chen
Institute of Applied Mechanics, National Taiwan University, Taipei 106, Taiwan

H. T. Lin
Vehicle Engineering Department, Chung Cheng Institute of Technology, National Defense University, Tai-Hsi, Tao-Yuan 33509, Taiwan

Andrew M. Wo
Institute of Applied Mechanics, National Taiwan University, Taipei 106, Taiwan

(Received 9 January 2001; revised 20 March 2003; accepted 15 May 2003)

A heat-driven thermoacoustic refrigerator has been designed and tested. A detailed thermal model of the device is presented. Energy balances within the system are discussed using external, heat exchanger, and stack control volumes in order to clarify the relationships of work and heat fluxes below and above onset. Thermal modeling is discussed as a tool for performance analysis as well as for determining system heat losses and finding input heat flows required by a thermoacoustic code. A method of using the control volume balance equations to find stack work and device efficiencies is presented. Experimental measurements are compared to DELTAE thermoacoustic modeling predictions. Modeling results show that viscous losses within the system have a significant impact on the device performance as well as on the ability of DELTAE to accurately predict performance. Modeling has led to an understanding of system performance and highlighted loss sources that are areas for improvement in a redesign. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1590971]

PACS numbers: 43.35.Ud [SGK]

I. INTRODUCTION

The basic science of using a temperature gradient along a surface to generate sound (or vice versa) has been presented well by Swift in his review article (Swift, 1988). Since that time, the development of numerical thermoacoustic (TA) modeling codes (especially DELTAE, Ward and Swift, 1996) has aided in a rapid increase in the number of research programs working on thermoacoustic systems. More recently, Swift has combined additional material into book format (Swift, 2002), giving a review of the current state of TA knowledge and research while pointing out the challenges ahead.

Thermal modeling of thermoacoustic devices (i.e., modeling heat movement throughout a system and not just in the stack) is not often discussed in the literature. In Swift's introductory thermoacoustics book (2002), methods for measuring insulation heat loss and heat removal in water-cooling flows are discussed. Backhaus and Swift (2000) give a more detailed discussion of pertinent heat losses in their Stirling-TA engine and how they were calculated. However, the relationships of these losses to thermoacoustic heat flows, and the use of thermal modeling for performance analysis, is not discussed explicitly. Adeff and Hofler (2000) discuss a transient method used to find the load on their heat pump stack. The arrangement of their device, with the heat pump stack located on the opposite end of their half-wave resonator, allowed them to use lumped capacitance analysis to find the loading on the heat pump stack from transient solar tests. Their approach allows finding actual cooling power, but is not useful in the case of the present apparatus (heat-driven thermoacoustic apparatus, HDTA), nor of many other devices.

This paper seeks to address a proper understanding of the place of a thermal system model relative to numerical thermoacoustic modeling of a TA device, and to document the use of thermal modeling as a tool for obtaining stack work and thermal efficiencies from experimental temperature and heat-flow measurements. After a discussion of thermal modeling, the design of the HDTA setup is presented. Then, control volume analysis is used to show interaction of thermal and acoustic fluxes and demonstrate how work terms can be found from the thermal model. Finally, HDTA measurements are compared to DELTAE predictions with discussion of thermal and viscous loss estimation.

II. THERMAL MODELING

An accurate thermal model not only translates input powers and measured temperatures into stack fluxes versus heat losses (insulation and others), but by itself it can be used to find stack work fluxes (acoustic power) even without pressure measurement, at least for some device configurations.

Thermal modeling of any system has its foundation in first-law energy balances of different system control volumes (e.g., a heat exchanger or stack). The heat fluxes across the control volume boundaries include solid conduction, convection, and radiation. In TA devices above onset, these same control volume balances have additional acoustic work fluxes in them as well as changes in some of the thermal conduction terms. Any heat-transfer textbook (e.g., Incropera and DeWitt, 1985) will discuss conduction (Fourier's law),

convection, and radiation flux estimation, as well as the implications of nonconstant material properties, inhomogeneous materials, and application of finite-element methods for more complex control volumes. More detail on contact resistance estimation can be found in Madhusudana (1996). Property data sources include references such as the TPRC (Touloukian and DeWitt, 1970) for older materials. The website at *matweb.com* is an excellent database of recent data on a variety of materials.

An accurate thermal model must be detailed: accurate device dimensions, contact resistances, thermal conductivity variation with temperature, internal and external radiation and convection, and consideration of nonuniform (i.e., 3D) temperature profiles within system elements. The temperature distribution within a part should be estimated, and if necessary measured or computationally modeled. Some estimate of the temperature variation is useful at least for placement of thermocouples, or interpretation and error analysis of thermocouple signals. Generally, considering all of these elements in the design phase can lead to a better design with more accurate measurements.

In its typical form, a thermal model is a system of equations with the first-law balances of several control volumes within a device. Below onset, an accurate thermal model will balance. That is, supplying measured power and temperature data as inputs to the model will result in all control volumes showing a balanced condition. The external balance will show that the combined input and load power is all seen exiting the device, either in the cooling water stream or in insulation/convection loss. An internal part balance will show that the sum of all fluxes crossing the control volume boundary equals zero.

Below-onset data are useful for tuning the model parameters necessary to find heat losses above onset. Ideally, a device is operated at the same thermal conditions above and below onset so that thermal losses are the same and can be separated from TA fluxes. In the case of HDTA, five below-onset data sets (using lower mean gas pressure to avoid onset at higher temperatures) were used to tune values of some uncertain thermal properties: insulation thermal resistance, thermal conductivity of some materials at elevated temperatures, contact resistance between flanges, etc. Additional measurements (temperature surveys at different points on individual parts, heat flux probe measurements) were made to help understand temperature profiles and thermal losses. Tuning involved finding a set of property values that resulted in all control volumes balancing for a given data set, and then repeating this for all below-onset data sets. Mean values of the below-onset property values were used for above-onset data sets.

The thermal model was implemented in a spreadsheet format with measured power levels and temperatures and thermal properties as inputs, and the values of each control volume balance equation and subequations as output. The model includes TA fluxes only as a by-product, appearing as a result of the model for above-onset data sets. See further discussion below.

III. DELTAE THERMOACOUSTICS DESIGN SOFTWARE

Design of the HDTA apparatus and performance analysis were aided by DELTAE (Design Environment for Low-amplitude ThermoAcoustic Engines, Ward and Swift, 1996), an executable code that takes device parameters (physical dimensions and material properties) of basic acoustic and thermo-acoustic elements and then integrates along the connected elements to predict ideal device performance based on low-amplitude (linear) thermo-acoustic equations. Great flexibility is provided for applying boundary conditions such that any relevant variables can be fixed. Both the user manual and Swift's recent book (Swift, 2002) provide guidance in using the code.

Besides the linear acoustics limitation, DELTAE has two other limitations relevant to this paper. First, while DELTAE accounts for conduction losses in the solid stack material and gas within the stack, it does not account for thermal conduction losses through the device body to ambient or shunt losses around the stacks through the solid body of the device shell. These thermal losses effectively decrease performance and must be accounted for separately. Second, DELTAE does not automatically account for flow effects such as turbulent flow disturbance present due to discontinuities between device elements, or bulk fluid motion streaming effects which will tend to reduce performance.

IV. THERMOACOUSTIC APPARATUS DESIGN

The thermoacoustically driven and cooled heat-driven thermoacoustic apparatus (HDTA) is a nominal quarter-length resonator device with two stacks sandwiched by three heat exchangers, and a compliance on the bottom, Fig. 1. Heat (Q_{IN}) at the hot end (HXF) generates 264-Hz sound in the prime mover stack (stack 1). This acoustic energy drives the heat pump stack 2 which pumps energy from the cold heat exchanger (HXL) to the midheat exchanger (HXM), where heat from the hot end and cold end is removed via circulating room-temperature water. The device was initially conceived with solar application in mind which led to a hot-end design with the hot heat exchanger copper fins (HXF) extending from the stack to the copper end block. This arrangement would allow sunlight, gas-fired, or electrically generated heat to be absorbed in the copper end cap and directly conducted to the stack.

DELTAE optimization of HDTA involved the sizing of various subcomponents of the device: heat exchangers, stacks, and ducts. Various device parameters are presented in Table I. The gas (He 60%–Ar 40%) was selected to minimize Prandtl number, thus decreasing viscous losses (Belcher *et al.*, 1999). The mean pressure (2 bar) was chosen to fit an overall desired length of the device to the stack ceramic structure and working gas. The stack material (Corning Celcor® 400-cpsi square cell extruded ceramic) was chosen based on its strength, low thermal conductivity, high-temperature capability, and availability.

The heat exchangers are all standard copper fin design. A 0.7-mm-thick stainless-steel shell surrounding HXF and part of stack 1 minimizes thermal structure losses. The Teflon rings serve the dual purpose of insulation between metal

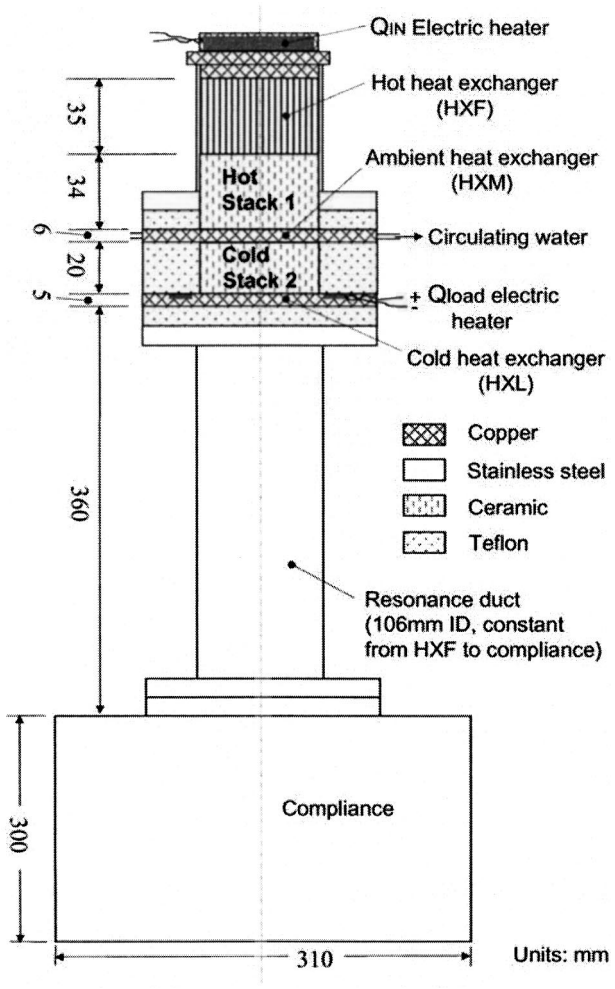


FIG. 1. HDTA thermoacoustics refrigerator schematic (to scale). HXF copper fins extend from endcap to stack 1. HXM heat exchanger is cooled with circulating room temperature water (in six 3.2-mm brass tubes). HXL heat exchanger has an electric heater for loading purposes.

flanges and heat exchangers as well as gripping the stacks. Water flowing through brass tubes passing through the HXM fins (three tubes in either direction) circulates in a counterflow arrangement to minimize temperature nonuniformity. Differential temperature measurement across the water inlet and outlet and flow rate measurement give the heat removed. The cold heat exchanger has a flat ring-shaped electric heater attached around the flange to provide loading (Q_{load}) capability.

Bolts pass through the stack/heat-exchanger pile, Fig. 2, holding the pile together. This pile and bolt design has the undesirable side effect of a complicated thermal model which must account for conduction through the bolts (Q_{bolt}) from hot end to cold end with heat loss along the length of the bolt with various contact resistances. The resonator duct and compliance were designed for simplicity; the abrupt resonator-duct/compliance (RC) transition has a significant flow pressure loss penalty.

The HDTA setup was instrumented with thermocouples (1-mm diameter, SS sheathed) in each heat exchanger as well as the stainless-steel flanges above HXM (SShot) and below HXL (SScold). Thermocouple accuracy at room temperature is approximately 0.2 °C, which is less than the temperature

TABLE I. Physical parameters of the HDTA device and gas.

Parameter	Value	Units
Mean pressure	2.00	bar
Operating frequency	264	Hz
Gas mixture	He0.6 Ar0.4	
Typical ΔT across hot stack 1	500	K
Gas therm. pen. depth (300 K, 800 K)	0.312, 0.511	mm
Gas visc. pen. depth (300 K, 800 K)	0.137, 0.317	mm
Stack thermal conductivity (k_s)	1.46	W/mK
Stack specific heat (c_p)	1000	J/kgK
Stack material density (ρ)	2510	Kg/m^3
Stack square cell web thickness ($2l$)	0.18	mm
Stack cell web separation ($2a$,$2b$)	1.10	mm
Stack thermal penetration depth (δ_s)	0.0265	mm
Hot stack 1 length	33.7	Mm
Cold stack 2 length	20.0	mm
HXF length	35.0	mm
HXM length	6.0	mm
HXL length	5.0	mm
HXF, HXM fin separation	3.0	mm
HXL fin separation	4.0	mm
HX (all) fin thickness	1.0	mm

variation seen in the parts at operating conditions: temperature nonuniformity in HXM was observed to be large radially due to the bolt influence, and circumferentially varied by approximately 10 °C on the external surface (coldest at the water inlets and outlets), but thermocouple temperatures near the interior wall agreed within 1 °C at operating conditions. Pressure measurements were made in a port just below HXL in the resonator wall. HDTA power inputs have an accuracy (based on voltage, resistance, and current measurements) of 1%. HDTA sensors and power levels (current and voltage) are monitored using LABVIEW PC-based software and data-acquisition hardware.

When tested at elevated temperatures, approximately 3 h were required to reach thermal equilibrium at the first power-set point, with approximately 1-h settling time between set points. The below-onset data series, run for thermal modeling purposes, had no external HXL heat loading ($Q_{\text{load}} = 0$ W). Above-onset data were collected with Q_{load} set to 0 W and 10 W.

V. DEVICE THERMAL MODEL

The thermal model recognizes below-onset heat fluxes in seven control volumes: the external balance, three heat exchangers, two stainless-steel flanges (above HXM and below HXL), and the bolts. The external control volume surrounds the device (although cutting through the resonator duct), giving the summation of all fluxes entering or exiting the device as shown in Fig. 2, left side. The heat flux, ''*Q*,'' and work flux, ''*W*,'' terms seen in Fig. 2, will be discussed below. *Solid arrows* are **below-onset** fluxes that appear in the thermal model and *open arrows* are **TA** fluxes that appear **above onset** (and do not appear explicitly in thermal model, as discussed below). Control volumes for the three heat exchangers and two stacks are also shown in Fig. 2.

The stainless-steel flange control volumes are not shown in detail in Fig. 2 (and will not be discussed below) because no TA terms enter the control volume. The bolt control vol-

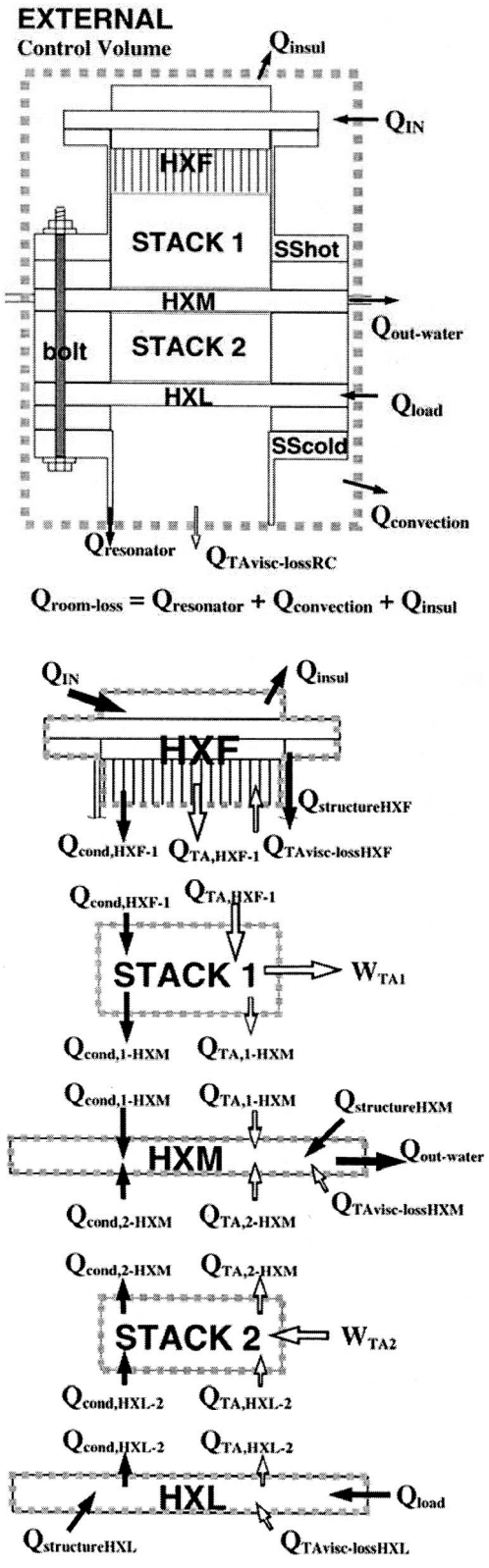


FIG. 2. HDTA control volumes for Eqs. (1)–(6). Arrows show heat and work flows as given in Eqs. (1)–(6), with direction giving sign. Solid arrows represent below-onset heat flows (explicit in thermal model). Open arrows are above-onset acoustic work or viscous loss terms. The bolt schematic is representative of the 16 bolts around the device circumference. Fluxes related to the SShot, SScold, and bolt control volumes are not shown.

ume includes 16 bolts in contact with each steel flange, heat exchanger, and Teflon ring. This is modeled using 1D finite elements starting at the head at SShot down and ending with the nut at SScold. The model includes conduction fluxes across contact resistances between the bolt and each layer at different temperatures along the length of the bolt. Because the Teflon internal temperature distributions are unknown, the Teflon was treated as a thermal resistor between a bolt segment and either a heat exchanger or a steel flange. The bolt equations matrix was solved iteratively as a first step in balancing the thermal model. The equations are not presented here as they are numerous and no TA terms are involved.

Each control volume in the thermal model produces a balance equation stating that fluxes crossing the boundary (dotted line of Fig. 2) must sum to zero. Below onset, only the fluxes shown as solid arrows in Fig. 2 exist and the thermal model (which only recognizes these fluxes) will balance. Above onset, thermoacoustic terms enter the control volume balances (open arrows in Fig. 2), including: work fluxes, changing stack conduction fluxes, surface viscous losses, and flow losses. The two stack control volumes, also seen in Fig. 2, have nontrivial balances only above onset; below onset they are merely thermal conduction paths between heat exchangers and are only included in the thermal model as such.

Following are the first-law energy balance equations for each control volume that includes thermoacoustic (TA) fluxes. Arrow directions in Fig. 2 indicate positive sign for a given flux. All viscous work losses are treated as heat inputs at the location of the loss. Prime mover stack 1 work (W_{TA1}) is defined as positive leaving the control volume (produced), while stack 2 work (W_{TA2}) is positive entering the control volume (consumed).

A. External balance

$$0 = Q_{IN} + Q_{load} - Q_{out\text{-}water} - Q_{room\text{-}loss} - Q_{TAvisc\text{-}lossRC}, \quad (1)$$

$Q_{IN} \equiv$ electric power input at HXF

(V_{rms} (corrected for lead resistance) $* I_{rms}$),

$Q_{load} \equiv$ electrical load at HXL,

$Q_{out\text{-}water} \equiv$ energy leaving in water cooling at HXM,

$Q_{room\text{-}loss} \equiv$ sum of losses to room air via convection, including insulation loss,

$Q_{TAvisc\text{-}lossRC} \equiv$ TA work dissipated in resonator and compliance due to: (a) surface viscous losses, and (b) abrupt resonator-compliance transition.

B. HXF balance

$$0 = Q_{IN} - Q_{insul} - Q_{structureHXF} - Q_{cond,HXF\text{-}1} - Q_{TA,HXF\text{-}1} + Q_{TAvisc\text{-}lossHXF}, \quad (2)$$

$Q_{IN} \equiv$ electric power input at HXF,

$Q_{insul} \equiv$ energy loss through hot-end insulation,

$Q_{structureHXF} \equiv$ conduction loss through stainless-steel shell surrounding stack 1 plus a small radiation loss to cooler regions in view,

$Q_{cond,HXF\text{-}1} \equiv$ conduction from HXF into stack 1, equal to $Q_{cond,1avg}$ below onset,

$Q_{TA,HXF\text{-}1} \equiv$ TA heat flow from HXF to stack 1,

$Q_{TAvisc\text{-}lossHXF} \equiv$ TA work that is converted to heat due to viscous losses in HXF.

C. Stack 1 balance

$$0 = Q_{cond,HXF\text{-}1} - Q_{cond,1\text{-}HXM} + Q_{TA,HXF\text{-}1} - Q_{TA,1\text{-}HXM} - W_{TA1}, \quad (3)$$

$Q_{cond,HXF\text{-}1} \equiv$ conduction from HXF into stack 1, equal to $Q_{cond,1avg}$ below onset,

$Q_{cond,1\text{-}HXM} \equiv$ conduction from stack 1 into HXM, equal to $Q_{cond,1avg}$ below onset,

$Q_{TA,HXF\text{-}1} \equiv$ TA heat flow from HXF to stack1,

$Q_{TA,1\text{-}HXM} \equiv$ TA heat flow from prime mover stack 1 to HXM,

$W_{TA1} \equiv$ stack work produced in stack 1.

D. HXM balance

$$0 = Q_{cond,1\text{-}HXM} + Q_{cond,2\text{-}HXM} + Q_{structureHXM} - Q_{out\text{-}water} + Q_{TA,1\text{-}HXM} + Q_{TA,2\text{-}HXM} + Q_{TAvisc\text{-}lossHXM}, \quad (4)$$

$Q_{cond,1\text{-}HXM} \equiv$ conduction loss from stack 1 to HXM,

$Q_{cond,2\text{-}HXM} \equiv$ conduction loss from stack 2 to HXM,

$Q_{structureHXM} \equiv$ sum of internal structural (Teflon casing and bolts) fluxes into HXM,

$Q_{out\text{-}water} \equiv$ energy leaving in the cooling water flow at HXM,

$Q_{TA,1\text{-}HXM} \equiv$ TA heat flow from prime mover stack 1 to HXM,

$Q_{TA,2\text{-}HXM} \equiv$ TA heat flow from heat pump stack 2 to HXM,

$Q_{TAvisc\text{-}lossHXM} \equiv$ TA work that is converted to heat due to surface viscous losses in HXM.

E. Stack 2 balance

$$0 = Q_{cond,HXL\text{-}2} - Q_{cond,2\text{-}HXM} + Q_{TA,HXL\text{-}2} - Q_{TA,2\text{-}HXM} + W_{TA2}, \quad (5)$$

$Q_{cond,HXL\text{-}2} \equiv$ conduction from HXL into stack 2, equal to $Q_{cond,2avg}$ below onset,

$Q_{cond,2\text{-}HXM} \equiv$ conduction from stack 2 into HXM, equal to $Q_{cond,2avg}$ below onset,

$Q_{TA,HXL\text{-}2} \equiv$ TA heat flow from HXL into stack 2,

$Q_{TA,2\text{-}HXM} \equiv$ TA heat flow from heat pump stack 2 to HXM,

$W_{TA2} \equiv$ work consumed in stack 2, used to pump heat from HXL to HXM.

F. HXL balance

$$0 = Q_{load} + Q_{structureHXL} - Q_{cond,HXL\text{-}2} - Q_{TA,HXL\text{-}2} + Q_{TAvisc\text{-}lossHXL}, \quad (6)$$

$Q_{load} \equiv$ electrical power external load to heat pump stack 2,

$Q_{structureHXL} \equiv$ sum of internal structural (Teflon casing and bolts) fluxes into HXL,

$Q_{cond,HXL\text{-}2} \equiv$ conduction from HXL into stack 2,

$Q_{TA,HXL\text{-}2} \equiv$ TA heat flow from HXL into stack 2,

$Q_{TAvisc\text{-}lossHXL} \equiv$ TA work that is converted to heat due to surface viscous losses in HXL plus minor flow losses at HXL fin ends.

Each of these control volume equations can be rearranged by bringing the TA terms to the other side of the equal sign, thus separating TA terms from the below-onset terms recognized by the thermal model. However, care must be taken in dealing with stack conduction terms. Below onset, the conduction flux (loss) in a stack is constant along the length of the stack, producing a linear temperature profile if thermal conductivity is constant. This flux (stack 1) is defined as

$$Q_{cond,1avg} \equiv (T_{HXF} - T_{HXM})/R_{stack1}, \quad Q_{cond,2avg} \equiv (T_{HXL} - T_{HXM})/R_{stack2}, \quad (7)$$

where R_{stack1} and R_{stack2} are the total thermal resistances of the stacks in units of (K/W) and include conduction through the stack solid material and gas volume (with thermal conductivity taken at the mean stack temperature), and across

gaps at either end of the stacks (where gas properties are taken at the heat exchanger temperatures). Therefore, $Q_{\text{cond,avg}}$ can be calculated based on thermal properties and heat exchanger temperatures below and above onset. Above onset, however, there is some deviation from a linear profile for a long stack (i.e., for a stack that does not meet the mathematical ''short'' stack condition) due to a continual flow of energy from the stack surface to the oscillating flow (or vice versa) along the stack's length. This means that some of the energy that enters stack 1 by conduction in the solid and gas ($Q_{\text{cond,HXF-1}}$) is later converted to work, or that enters in the TA heat flow ($Q_{\text{TA,HXF-1}}$) later becomes pure conduction loss. In other words, while below onset $Q_{\text{cond,1avg}}=Q_{\text{cond,HXF-1}}=Q_{\text{cond,1-HXM}}$, above onset these equalities do not hold.

For analysis purposes, it is desired to separate heat used for work production from the conduction loss. For this purpose the average conduction flow based on the thermal model, $Q_{\text{cond,1avg}}$, is subtracted from the total heat flow entering stack 1 at HXF to give the heat used in work production. That is

$$Q_{\text{TAin,1}} \equiv Q_{\text{TA,HXF-1}} + Q_{\text{cond,HXF-1}} - Q_{\text{cond,1avg}}. \tag{8}$$

If the prime mover stack is viewed as a heat engine, then $Q_{\text{TAin,1}}$ is equivalent to the heat flow entering from the hot reservoir at HXF and used to produce work, W_{TA1}, with rejected heat, $Q_{\text{TAout,1}}$, to the cold reservoir at HXM. $Q_{\text{TAin,1}}$ is essentially Q_{IN} (power in at HXF) with all conduction losses (to room, through structure and stack to HXM) removed. $Q_{\text{TAout,1}}$ and the stack 2 equivalents are defined as

$$Q_{\text{TAout,1}} \equiv Q_{\text{TA,1-HXM}} + Q_{\text{cond,1-HXM}} - Q_{\text{cond,1avg}}, \tag{9}$$

$$Q_{\text{TAin,2}} \equiv Q_{\text{TA,HXL-2}} + Q_{\text{cond,HXL-2}} - Q_{\text{cond,2avg}}, \tag{10}$$

$$Q_{\text{TAout,2}} \equiv Q_{\text{TA,2-HXM}} + Q_{\text{cond,2-HXM}} - Q_{\text{cond,2avg}}. \tag{11}$$

Then, substituting Eqs. (8)–(11) into Eqs. (1)–(6) and rearranging gives

$$[\text{EXT balance}] = Q_{\text{IN}} + Q_{\text{load}} - Q_{\text{out-water}} - Q_{\text{room-loss}} = Q_{\text{TAvisc-lossRC}}, \tag{12}$$

$$[\text{HXF balance}] = Q_{\text{IN}} - Q_{\text{insul}} - Q_{\text{structureHXF}} - Q_{\text{cond,1avg}} = Q_{\text{TAin,1}} - Q_{\text{TAvisc-lossHXF}}, \tag{13}$$

$$[\text{STK1 balance}] = Q_{\text{cond,1avg}} - Q_{\text{cond,1avg}} = 0 = Q_{\text{TAin,1}} - W_{\text{TA1}} - Q_{\text{TAout,1}}, \tag{14}$$

$$[\text{HXM balance}] = Q_{\text{cond,1avg}} + Q_{\text{cond,2avg}} + Q_{\text{structureHXM}} - Q_{\text{out-water}} = -(Q_{\text{TAout,1}} + Q_{\text{TAout,2}} + Q_{\text{TAvisc-lossHXM}}), \tag{15}$$

$$[\text{STK2 balance}] = Q_{\text{cond,2avg}} - Q_{\text{cond,2avg}} = 0 = Q_{\text{TAin,2}} + W_{\text{TA2}} - Q_{\text{TAout,2}}, \tag{16}$$

$$[\text{HXL balance}] = Q_{\text{load}} + Q_{\text{structureHXL}} - Q_{\text{cond,2avg}} = Q_{\text{TAin,2}} - Q_{\text{TAvisc-lossHXL}}, \tag{17}$$

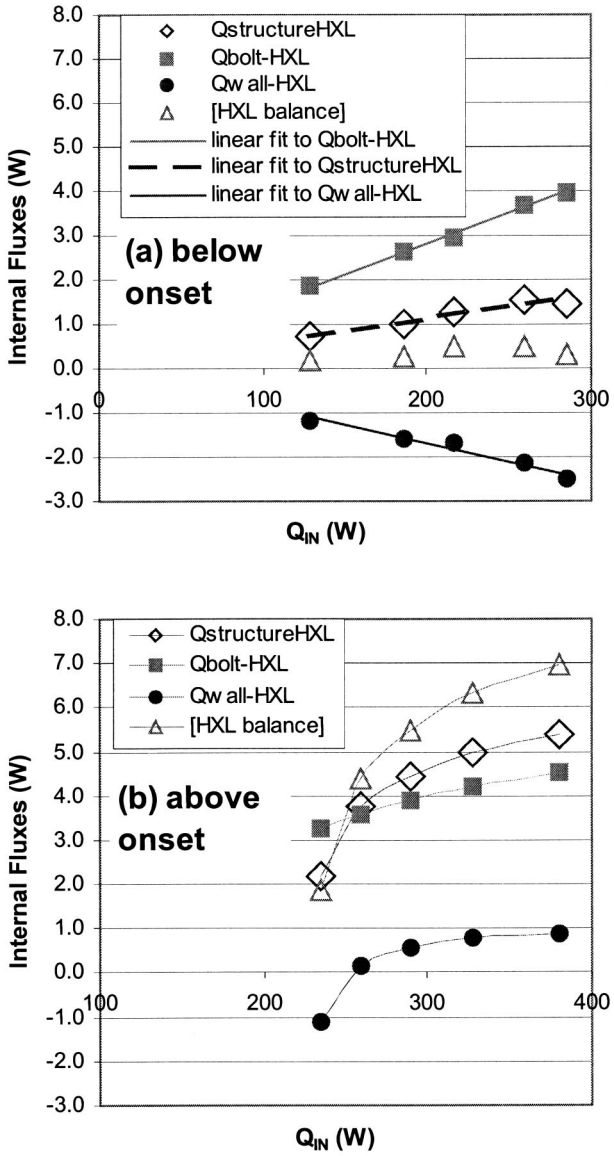


FIG. 3. Heat flows from HXL control volume as measured for (a) below-onset data and (b) above-onset low load ($Q_{\text{load}}=0$) data. Here, $Q_{\text{bolt-HXL}}+Q_{\text{wall-HXL}}=Q_{\text{structure-HXL}}$ of Fig. 2, and [HXL balance] equals the first-law energy imbalance for below-onset data (a) and equals $Q_{\text{TAin,2}}-Q_{\text{TAvisc-lossHXL}}$ for above-onset data (b). Lines in (a) are straight line fits to data through origin. Lines in (b) are for guiding the eye only.

where each balance refers to the summation of fluxes in the thermal model. Below onset, the TA terms (second line of each of the above equations) will all equal zero so that the balances will equal zero (to the accuracy of the model). The thermal model only recognizes the below-onset terms, and thus sees the balance as written in the first line of each equation above. Above onset the TA terms will not equal zero, and thus the thermal model energy balance will not equal zero. This imbalance of the thermal model at a particular control volume is therefore seen to be equal to the sum of the TA terms at that particular control volume.

For the HDTA device, an accurate thermal model was needed for finding the real heat load at HXL. While this is true for any device, HDTA's bolt design called for careful analysis. Figure 3 shows the various HXL heat flows at different operating temperatures, here for below-onset [Fig.

3(a)] and above-onset [Fig. 3(b)] data. $Q_{\text{bolt-HXL}}$ is the flux from the bolts into HXL either directly across the bolt thread to copper contact resistance, or indirectly through the Teflon resistor. $Q_{\text{wall-HXL}}$ is the sum of the fluxes from HXM and the lower steel flange (SScold, Fig. 2) entering HXL through the Teflon. The sum of $Q_{\text{bolt-HXL}}$ and $Q_{\text{wall-HXL}}$ equals $Q_{\text{structureHXL}}$ seen here. In Fig. 3(a), below-onset data shows that as Q_{IN} is increased at HXF, T_{HXL} increases, which leads to a competing effect where the hot bolt is conducting heat to HXL while heat is simultaneously transferred from HXL through Teflon to HXM and SScold.

Figure 3(b) gives the same fluxes for above-onset ($Q_{\text{load}}=0$) data. In this case HXL is cooled below room temperature and the previously negative wall flux is now seen to reverse, leading to a largely increased $Q_{\text{structureHXL}}$. The [HXL balance] is seen to increase and is now equal to $Q_{\text{TAin,2}}-Q_{\text{TAvisc-lossHXL}}$ [Eq. (17)] for this above-onset data. The difference between the [HXL balance] and $Q_{\text{structureHXL}}$ values is equal to $Q_{\text{cond,2avg}}(Q_{\text{load}}=0)$. The total load on HXL ($Q_{\text{TAin,2}}$) reaches 7.4 W at the $Q_{\text{IN}}=380$-W setting, with $Q_{\text{bolt-HXL}}$ the largest component. Increasing Q_{load} (for the 10-W series, not shown) raises T_{HXL} and therefore reduces $Q_{\text{structureHXL}}$.

Concerning accuracy of thermal model estimated fluxes, the plot of [HXL balance] in Fig. 3(a) shows representative error levels for below onset data, i.e., all control volumes generally balanced within half a watt. However, half a watt is a significant percentage of the low fluxes seen at HXL. Unfortunately, uncertainties for above-onset fluxes are higher. The control volumes for SShot and SScold (which have no TA terms present) should balance above onset as well as below. However, an imbalance at SScold was seen to grow with decreasing T_{HXL} (increasing Q_{IN}) for above-onset data, with a maximum imbalance of 3.2 W at the minimum T_{HXL} [where $Q_{\text{IN}}=380$ W, Fig. 3(b)]. SShot showed some smaller imbalance also. This forces the conclusion that TA effects within the device are changing thermal conditions significantly away from the below-onset case: for example, cooling of HXL at the center (and heating of HXM) could cause a significant change in the 3D temperature distribution within the heat exchangers that did not exist below onset, and which would likely not be seen by the limited number of thermocouples. The changed temperature profiles will affect structural fluxes and thus balances as seen at SShot and SScold. The conclusion is that uncertainties on $Q_{\text{structureHXL}}$ must be assumed to be of similar magnitude, i.e., approximately 1 W near onset up to 3 W at lowest T_{HXL}, or an uncertainty of about 50% on $Q_{\text{structureHXL}}$ and thus on $Q_{\text{load}\delta\text{E}}$.

The complicating effects of the bolts can also be seen by considering the effect of a changing T_{HXL} on the HXF heat balance. The bolt heat loss, and thus the structure loss at HXF, is influenced by T_{HXL}, an effect that was observed experimentally. The presence of the bolt, therefore, can change the heat input to DELTAE and the total heat loss even while T_{HXF}, T_{room}, and T_{HXM} are unchanged. This indicates the need for a loss correlation that includes T_{HXL} rather than a more simple function of $(T_{\text{HXF}}-T_{\text{HXM}})$. In any case, a loss correlation should account for loss components relative to appropriate temperature differences, and a careful thermal model accomplishes this.

VI. USING THERMAL MODEL FOR PERFORMANCE ANALYSIS

A thermal model provides the ''room-loss'' (insulation and convection) terms and structural losses (heat that bypasses the stacks through device shell), and based on these gives the required inputs to a thermoacoustics code such as DELTAE, where for DELTAE (that does not account for room loss or structure loss)

$$Q_{\text{IN}\delta\text{E}}\equiv Q_{\text{IN}}-Q_{\text{insul}}-Q_{\text{structureHXF}},$$
$$Q_{\text{load}\delta\text{E}}\equiv Q_{\text{load}}+Q_{\text{structureHXL}}. \tag{18}$$

The thermal model also provides a tool to evaluate system performance. Temperature and power measurements are the inputs to the thermal model. The results of the thermal model, combined with pressure measurements and numerical modeling, give device performance. The thermal model provides inputs to a thermoacoustics code, but may be able by itself to provide stack work measurements.

To see the power of the thermal model in performance analysis, consider the following use of the balance equations [Eqs. (12)–(17)]. First, the [EXT balance] could give an experimental measure of the viscous loss in the resonator and compliance which is equal to the imbalance in the external control volume. This is true because the external control volume is drawn cutting through the resonator below HXL (Fig. 2), with heat loss to the resonator based on the temperature difference between HXL and the compliance (at room temperature), and also because the resonator and compliance are uninsulated with sufficient room convection to remove this energy. However, in practice, uncertainty for the insulation loss is greater than for the viscous loss component, and so the external balance is used to find insulation loss estimates.

Second, stack work terms can be obtained from Eqs. (13)–(17). The viscous loss terms must be either estimated or neglected. For this example, viscous losses are assumed negligible, with the resulting equations

$$[\text{HXF balance}]=Q_{\text{TAin,1}}, \tag{13′}$$
$$[\text{STK1 balance}]=0=Q_{\text{TAin,1}}-W_{\text{TA1}}-Q_{\text{TAout,1}}, \tag{14′}$$
$$[\text{HXM balance}]=-Q_{\text{TAout,1}}-Q_{\text{TAout,2}}, \tag{15′}$$
$$[\text{STK2 balance}]=0=Q_{\text{TAin,2}}+W_{\text{TA2}}-Q_{\text{TAout,2}}, \tag{16′}$$
$$[\text{HXL balance}]=Q_{\text{TAin,2}}. \tag{17′}$$

For each of these equations, the left-hand-side balance is known from the thermal model when above-onset data are entered as inputs. Unfortunately (for the HDTA thermal model), there are only five equations but six unknowns. With the current HDTA hardware configuration, both stacks are cooled by one heat exchanger (HXM). Separating the stacks and using two room-temperature heat exchangers (e.g., by placing stacks on opposite ends of a half-wavelength resonator as done by Adeff and Hofler, 2000) would allow separation of $Q_{\text{TAout,1}}$ and $Q_{\text{TAout,2}}$, and solution of the stack work terms. While $Q_{\text{TAin,1}}$ is the actual heat used to produce work,

$Q_{\mathrm{TAin},2}$ is the actual heat removed from HXL solely by the efforts of W_{TA2} acting in the heat pump stack, i.e., the true load on HXL.

In essence, the work estimates found in this way are valid even at high pressure fluctuation levels because they are based on measurements, not computations. The accuracy of these work estimates depends on the accuracy of the thermal model as well as estimates of viscous losses.

One method of estimating viscous losses requires both the aid of a thermoacoustic code as well as pressure measurement. Thermoacoutic viscous surface losses are calculated in a code such as DELTAE. Acoustic power dissipated by viscosity acting at interior surfaces is proportional to the fluctuating velocity squared, which (assuming acoustic frequency, mean pressure, and temperature match) is proportional to the fluctuating pressure squared (Swift, 1988). Thus, surface viscous losses can be taken from DELTAE multiplied by the ratio $(|p|_{\mathrm{measured}}/|p|_{\mathrm{DELTAE}})^2$. Flow viscous losses can also be estimated using nonoscillating pipe flow ''minor-loss'' coefficients according to the method of Swift (2002), which at least gives some estimate of flow losses. The acoustic power dissipated due to these minor losses is shown to be proportional to fluctuating velocity cubed, $|U|^3$, which by the same argument allows DELTAE results to be used with correction by $(|p|_{\mathrm{measured}}/|p|_{\mathrm{DELTAE}})^3$.

Having $Q_{\mathrm{TAin},1}$ and $Q_{\mathrm{TAin},2}$ allows for ''lossless'' (no conduction loss) estimates of stack performance. For stack 1

$$\eta_{1\text{-lossless}}=\frac{W_{TA1}}{Q_{TA\mathrm{in},1}}, \tag{19}$$

and for stack 2

$$COP_{\mathrm{lossless}}=\frac{Q_{TA\mathrm{in},2}}{W_{TA2}}. \tag{20}$$

Even for the HDTA device, where the work terms cannot be isolated, several useful thermal efficiencies can be considered. First, there is the gross performance criterion

$$\eta_{\text{therm-gross}}=Q_{\mathrm{load}}/Q_{\mathrm{IN}}, \tag{21}$$

which gives the external power load moved for a given external power input, including all losses. A second thermal efficiency is

$$\eta_{\text{therm-}\delta\mathrm{E}}=Q_{\mathrm{load}\delta\mathrm{E}}/Q_{\mathrm{IN}\delta\mathrm{E}}, \tag{22}$$

which is the amount of energy removed at HXL for a given heat input to stack 1 at HXF subtracting out room and structure losses, but not removing viscous and stack conduction losses. $Q_{\mathrm{load}\delta\mathrm{E}}$ and $Q_{\mathrm{IN}\delta\mathrm{E}}$ are also the inputs required by DELTAE. Finally, based on the HDTA HXF and HXL balances, and with minor terms estimated with the aid of DELTAE, a lossless thermal efficiency is

$$\eta_{\text{therm-lossless}}=Q_{\mathrm{TAin},2}/Q_{\mathrm{TAin},1} \tag{23}$$

which gives an indication of the efficiency that could be achieved with the current physical geometry if perfect materials (nonconducting except as needed for thermoacoustic performance in the stack) were used such that no conduction losses were present.

Therefore, depending on hardware configuration and magnitude of viscous losses, useful measures of device performance can be found without any acoustic modeling or measurements. A thermoacoustics code such as DELTAE serves best in a design capacity, rather than in performance analysis. However, it is also a powerful tool for separating out thermoacoustic terms that the thermal model cannot differentiate, as noted above for aiding in estimating viscous losses.

Work terms presented in the Results section are taken directly from the results of a DELTAE model, which is limited by both the accuracy of the thermal model (supplying inputs) as well as of the thermoacoustics model of the device. The HDTA DELTAE model includes ''minor'' flow loss terms at HXF, HXM, HXL, and at the resonator/compliance (RC) junction. Of these, only the RC loss is large. Minor-loss coefficients were taken (with some engineering judgment required) from Idelchik's handbook (1994), and applied according to the recommendations of Swift (2002). DELTAE calculations were run with Q_{IN} fixed and Q_{load} as a target allowing T_{HXF}, T_{HXL}, and pressure fluctuation level to vary.

VII. RESULTS AND DISCUSSION

The goal of system thermal modeling as well as of thermoacoustic modeling is to understand the system and analyze performance. In the case of the HDTA device, modeling has led to an understanding of system performance and highlighted loss sources that are areas for improvement in a redesign. A good example of this was given in Sec. V concerning the heat flows into HXL (Fig. 3). The thermal model clearly shows the influence of the bolt design on the actual loading on the HXL heat exchanger, even showing the connection between HXL temperature and the heat input at HXF. The thermal model provided insulation and stack conduction losses as well as the required heat inputs to DELTAE.

Results of DELTAE modeling, using thermal model provided heat inputs, show that DELTAE does not accurately predict device performance, although results are reasonable and helpful for understanding the system. Figure 4 shows predicted versus measured prime mover stack temperature differentials, showing that DELTAE predictions are about 20% above measurements for the $Q_{\mathrm{load}}=0$ W data. DELTAE predicted frequency, Fig. 5, is correspondingly high. Pressure fluctuation levels show a crossing trend, Fig. 6.

These differences in DELTAE and measurements are most likely due to several factors. First, it is suspected that the abrupt transition between resonator duct and compliance has a strong influence on actual device performance that is not modeled by DELTAE. The abrupt transition may produce a resonator effective duct length different from that which DELTAE calculates, which would affect predicted frequency and hot-end temperatures. The drop in predicted pressure fluctuation levels below measurements in Fig. 6 (max pressure fluctuation level in Figs. 4–6 was $|p|/P_m=5.3\%$) suggests that the calculated resonator/compliance (RC) viscous losses are overpredicted at higher pressure fluctuation levels, likely due to an overprediction of the minor flow loss pressure drop estimate in DELTAE which is based on nonoscillating flow

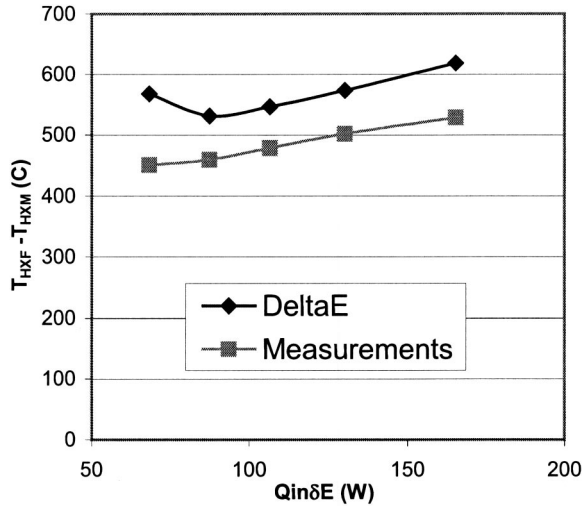


FIG. 4. Measured prime mover stack temperature difference and DELTAE predicted temperature difference versus $Q_{IN\delta E}$ [DELTAE input power, Eq. (18)] for low load ($Q_{load}=0$ W) data. Data with $Q_{load}=10$ W showed greater disagreement.

correlations, and increases as $|U|^3$. The viscous surface loss (normally calculated by DELTAE) in the compliance may also be overly large since DELTAE applies simple conservation of momentum at the RC interface, while the actual fluid motion in the compliance is certainly not simple uniform oscillating flow. Finally, it is possible that the thermal model is overpredicting $Q_{\text{structureHXL}}$, as mentioned in Sec. V. This increased load at HXL would drive the DELTAE model closer to onset with higher resulting T_{HXF}.

The results above demonstrate that there are elements of the HDTA device that are not accurately modeled in DELTAE. However, understanding these elements (e.g., the fluid dynamics of the compliance) is not required. Instead, DELTAE results give work and viscous loss estimates that can be used together with thermal modeling results. This in turn is helpful for pinpointing loss sources leading to device improvement.

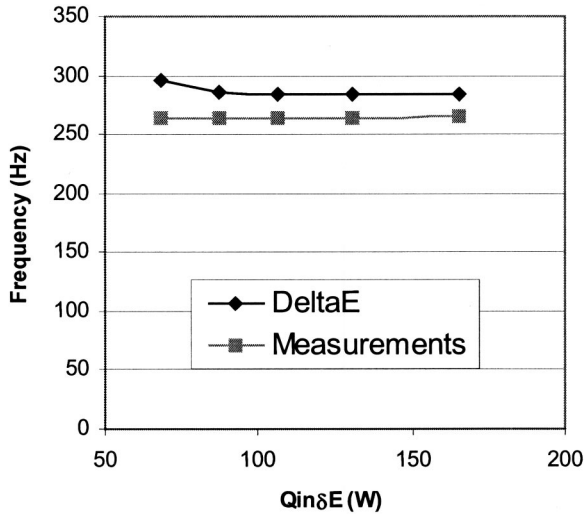


FIG. 5. Measured TAFA resonance frequency and DELTAE predicted frequency versus DELTAE input power [Eq. (18)] for low load ($Q_{load}=0$ W) data.

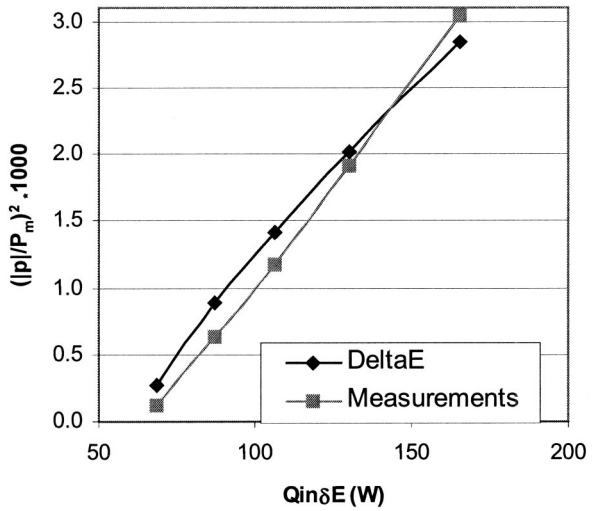


FIG. 6. Measured TAFA normalized pressure fluctuation levels and DELTAE predicted values for low load ($Q_{load}=0$ W) data.

Some device performance information is presented in Figs. 7 and 8. Figure 7 shows the heat terms from the HXF control volume for the different low load ($Q_{load}=0$) data sets showing relative magnitudes of heat flows. The insulation loss is seen to be relatively high and an easy target for improvement. Stack and structure losses also may have room for improvement. Utilizing a more efficient stack (e.g., parallel plate) could make better use of the available heat ($Q_{TAin,1}$) for work production.

Figure 8 shows the work balance in HDTA, where the work balance states that all work produced in stack 1 (W_{TA1}) is consumed within the device either usefully as W_{TA2} or wasted on viscous losses. Clearly, $Q_{\text{TAvisc-lossHXF}}$, while the smallest band of Fig. 7, is actually the largest viscous loss in the system, and therefore HXF redesign is a serious candidate for improvement. The other obvious drain on work is the RC viscous flow loss that could be eliminated with a

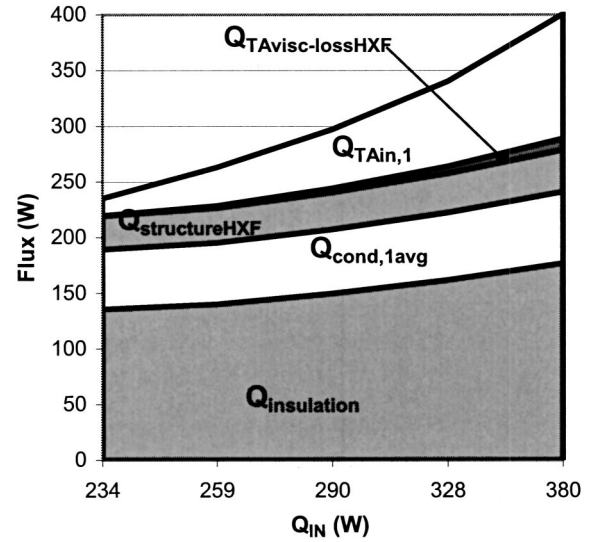


FIG. 7. Stacked area plot giving heat terms from the HXF control volume for the low load ($Q_{load}=0$) data. The heat available for work production in stack 1 ($Q_{TAin,1}$) and viscous loss increase more rapidly with Q_{IN} than other terms.

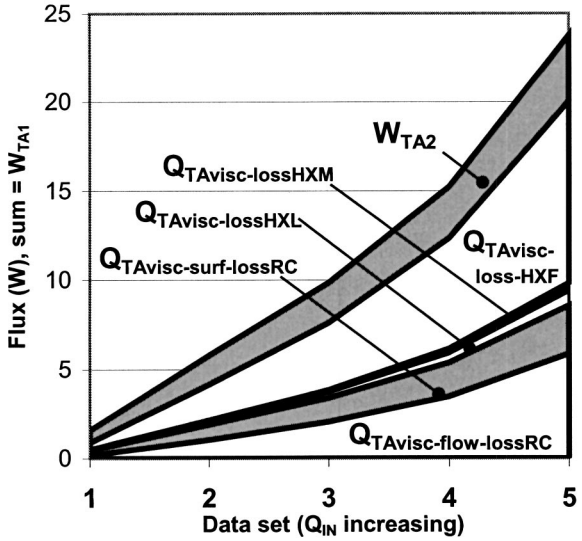


FIG. 8. Stacked area plot giving terms in work balance: $W_{TA1}=W_{TA2}+$viscous losses. Viscous losses in HXF ($Q_{TAvisc\text{-}lossHXF}$) and in the resonator-duct and compliance ($Q_{TAvisc\text{-}lossRC}$) are seen to be larger than the stack 2 work (W_{TA2}).

smooth transition from the resonator duct to the compliance. Note that all TA viscous loss and work fluxes in Figs. 7 and 8 were estimated in DELTAE.

Figure 9 shows the results of the thermal efficiencies discussed earlier [Eqs. (23)–(25)]. Gross efficiency is about 3% for the 10-W load case, but could be higher at higher loads. The DELTAE efficiency (with structural and insulation heat losses removed) reaches above 10%. The high efficiencies for the $Q_{load}=10$ W and $Q_{IN}=260$ W setting occur when T_{HXL} is 8 °C above T_{HXM}. The conduction-lossless efficiency is highest at the point where useful input power (Q_{TAin1}) is at its lowest while HXL load is still high (and therefore $T_{HXL}>T_{HXM}$). However, the main reason for $\eta_{therm\text{-}lossless}$ being

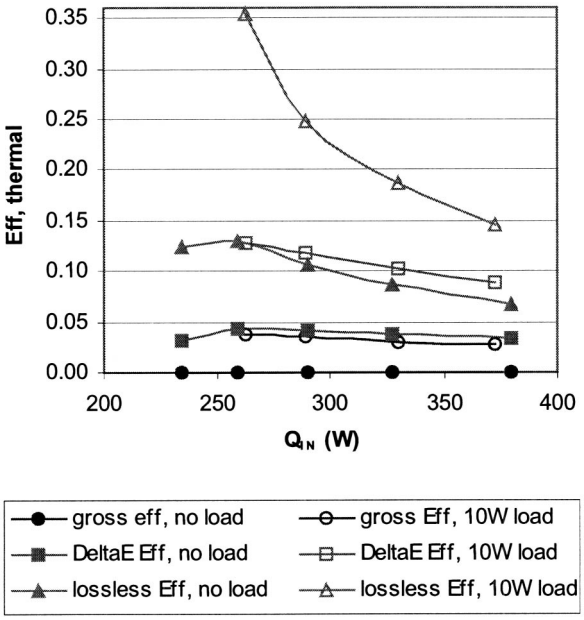


FIG. 9. Thermal efficiency results corresponding to Eqs. (23)–(25).

significantly above $\eta_{therm\text{-}\delta E}$ is the removal of stack 1 conduction losses, a nonrealistic possibility. Nonetheless, these thermal model results give targets to shoot for and clearly show where losses exist.

VIII. CONCLUSION

The design, modeling, and performance of the National Taiwan University's Institute of Applied Mechanics and Chung Cheng Institute of Technology thermoacoustic group's heat-driven thermoacoustic refrigerator have been presented. Thermal modeling, based on control volumes of various components in the device, has been discussed as a tool for determining system losses, for finding input heat flows to a thermoacoustic code, and for performance analysis. The relationships of external input powers to the input heats required by DELTAE have been discussed relative to the actual heat flows in the prime mover and heat pump stacks that are used for work production and heat pumping, respectively. The relationships of heat losses and acoustic work flows have been derived along with a method for finding stack work from the thermal model. The present modeling has proven useful in identifying and understanding thermal and viscous losses in the system, specifically in the bolted stack/heat-exchanger pile design, in the hot heat exchanger design, and in the abrupt transition between resonator duct and compliance. These specific loss sources are believed to have contributed to DELTAE's poor prediction of device performance, and elimination of these losses in a redesign could lead to significant improvements in the device.

ACKNOWLEDGMENTS

The authors would like to acknowledge the help of Mr. Joseph F.S. Lee in the construction of the thermoacoustic device, and to thank G. Swift and A. Atchley for their advice and encouragement. Partial funding of this work was supported by the National Science Council of the Republic of China under Grants NSC 89-2212-E-014-021 and NSC 89-2811-E-002-004.

Adeff, J. A., and Hofler, T. J. (**2000**). "Design and construction of a solar-powered, thermo-acoustically driven, thermoacoustic refrigerator," presented at ASA Atlanta meeting, June 2000, and available online at: http://www.physics.nps.navy.mil/hofler/stadtar.htm

Backhaus, S., and Swift, G. W. (**2000**). "A thermoacoustic-Stirling heat engine: Detailed study," J. Acoust. Soc. Am. **107**(5), 3148.

Belcher, J. R., Slaton, W. V., Raspet, R., Bass, H. E., and Lightfoot, J. (**1999**). "Working gases in thermoacoustic engines," J. Acoust. Soc. Am. **105**(5), 2677–2684.

Idelchik, I. E. (**1994**). *Handbook of Hydraulic Resistance*, 3rd ed. (Begell House, New York).

Incropera, F. P., and DeWitt, D. P. (**1985**). *Fundamentals of Heat and Mass Transfer*, 2nd ed. (Wiley, New York).

Madhusudana, C. V. (**1996**). *Thermal Contact Conductance* (Springer, New York).

Swift, G. W. (**1988**). "Thermoacoustic engines," J. Acoust. Soc. Am. **84**(4), 1145–1180.

Swift, G. W. (**2002**). *Thermoacoustics* (Acoustical Society of America, New York).

Touloukian, Y. S., and DeWitt, D. P. (**1970**). *Thermophysical Properties of Matter*, TPRC Data Series (IFI/Plenum, New York).

Ward, B., and Swift, G. W. (**1996**). "Design Environment for Low-amplitude Thermoacoustic Engines," Los Alamos National Laboratory, LA-CC-93-8.

On the optimization of the effective electromechanical coupling coefficients of a piezoelectric body[a)]

Boris Aronov
Acoustic Research Laboratory, Department of Electrical and Computer Engineering, The University of Massachusetts, Dartmouth, 285 Old Westport Road, North Dartmouth, Massachusetts 02747-2300 and BTECH Acoustics, 1445 Wampanoag Trail, Suite 115, East Providence, Rhode Island 02915

(Received 12 March 2003; revised 12 May 2003; accepted 19 May 2003)

Effective electromechanical coupling coefficients are defined based on the expression for the internal energy of a piezoelectric body [B. S. Aronov, ''Energy analysis of a piezoelectric body under nonuniform deformation,'' J. Acoust. Soc. Am. **113**, 2638–2646 (2003)]. The condition is considered under which the effective coupling coefficient can be maximized up to the value of the material coupling coefficient for any given distribution of deformation. A simple practical way to optimize the effective coupling coefficient of a transducer by changing its electrode shape is illustrated with examples of transducers vibrating in longitudinal and flexural modes. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1592163]

PACS numbers: 43.38.Ar, 43.38.Fx, 43.38.Pf [AJZ]

I. INTRODUCTION

Electromechanical coupling coefficients (further referred to as coupling coefficients k_c) are important parameters of piezoelectric materials and piezoelectric bodies subjected to deformation. The basic energy expression for k_c, as introduced in Ref. 1, is

$$k_c^2=\frac{\text{energy stored in the mechanical form}}{\text{total input energy}}. \tag{1}$$

The calculations of the piezoelectric material coupling coefficients, k_m, related to a uniformly deformed small volume of the material, were made using this expression. In the case of a deformed piezoelectric body, k_c is called an effective coupling coefficient, k_{eff}. The effective coupling coefficient, k_{eff}, is related to the static strain distribution in a body corresponding to a particular mode of vibration. In general, $k_{\text{eff}}\leqslant k_m$, and particularly $k_{\text{eff}}=k_m$ for a uniformly deformed body. Using the expression (1), it is difficult to calculate and especially to analyze possible ways to optimize the effective coupling coefficients unless the energies involved are represented analytically. More suitable in this sense is the formula for the coupling coefficients introduced in Ref. 2

$$k_m=\frac{U_m}{\sqrt{U_e U_d}}. \tag{2}$$

The extension of Eq. (2) to the case of the effective coupling coefficient is given in Ref. 3 as

$$k_{\text{eff}}^2=\frac{(\int_{\tilde{V}} U_m\, d\tilde{V})^2}{(\int_{\tilde{V}} U_e\, d\tilde{V})(\int_{\tilde{V}} U_d\, d\tilde{V})}. \tag{3}$$

In Eqs. (2) and (3), U_m, U_e, and U_d are the densities of the mutual, elastic, and dielectric energies, respectively, as they are defined in Ref. 2 and $\tilde{V}$ is the volume of a piezoelectric body. No further general analysis of the concept of the effective coupling coefficients has been published to the best of my knowledge.

The expression (2), which differs from the physically clear definition of the coupling coefficient in the expression (1), is introduced in a formal way in Ref. 2 that has shortcomings to be discussed. Nonetheless, correct results can be obtained using Eq. (2), and it will be shown here that Eqs. (1) and (2) are equivalent in a certain sense. It is hard to use Eqs. (2) and (3) for the general analysis of k_{eff} unless the procedure of integration of the energy densities over the volume of a piezoelectric body is clearly specified. Therefore, before considering the optimization of k_{eff}, which is the main objective of this paper, the concepts of k_m and k_{eff} are discussed in order to introduce a clear analytical expression suitable for the calculation and general analysis of k_{eff}. This is done in Sec. II, where expression (1) is represented in an analytical form and the validity of Eq. (2) is considered. In Sec. III the improved definition for k_{eff} is used in order to reach a general conclusion regarding the ways to optimize k_{eff}. The underlying physics is illustrated with an example of a length expander bar. Examples of various transducer types are considered from the standpoint of maximizing k_{eff} in Sec. IV.

II. ON THE DEFINITION OF THE ELECTROMECHANICAL COUPLING COEFFICIENT

The analytical definition of the coupling coefficient k_c based on the expression (1) can be introduced using the analytical results of the internal energy in a piezoelectric body presented in Ref. 4. In the case of a small volume of piezoelectric material under uniform deformation, $k_c=k_m$ and k_m can be represented in the following equivalent forms:

$$k_m^2=\frac{w_m^E}{w_{el}^S+w_m^E}=\frac{|w_{em}|}{w_{el}^S+w_m^E}=\frac{w_m^E}{w_{el}^T}, \tag{4}$$

[a)]The results of this work were reported in part at the 136th meeting of the Acoustical Society of America [J. Acoust. Soc. Am. **104**, 1845(A) (1998)].

 0001-4966/2003/114(2)/792/9/$19.00

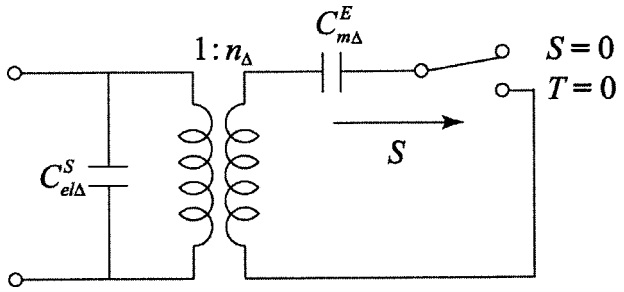


FIG. 1. The equivalent electromechanical circuit for a unit volume of a piezoelectric body.

where w^S_{el}, w^T_{el}, w^E_m, and w_{em} are the components of the internal energy density defined in Ref. 4. They represent the electrical energies under constant S and T, the mechanical energy under constant E, and the electromechanical energy, respectively. Note that $|w_{em}|=w^E_m$ according to Eq. (11) in Ref. 4. Energy densities may be obtained for any particular mechanical boundary conditions using the piezoelectric equations of state

$$T_i=c^E_{ik}S_k-e_{3i}E_3, \quad D_3=e_{3i}S_i-\varepsilon^S_{33}E_3, \tag{5}$$

or

$$S_i=s^E_{ik}T_k+d_{3i}E_3, \quad D_3=d_{3i}T_i+\varepsilon^T_{33}E_3. \tag{6}$$

The standard piezoelectric notations used in Ref. 2 are adopted and the common convention of the summation with respect to repeated indices is applied in Eqs. (5) and (6). In particular, for an element of a thin length expander bar, in which the stresses are zero in all the directions except for in the longitudinal direction, the energy densities are

$$w^{S_i}_{el}=\frac{1}{2}\varepsilon^{S_i}_{33}E^2_3, \quad w^E_m=\frac{1}{2}\frac{S^2_i}{s^E_{ii}}, \quad w_{em}=\frac{1}{2}\frac{d_{ii}S_iE_3}{s^E_{ii}}, \tag{7}$$

where $\varepsilon^{S_i}_{33}=\varepsilon^T_{33}(1-d^2_{3i}/s^E_{ii}\varepsilon^T_{33})$ and $i=1$ and 3 for the transverse and axial electric field, respectively. The substitution of Eq. (7) into Eq. (4) gives

$$k^2_{3i}=\frac{d^2_{3i}}{s^E_{ii}\varepsilon^T_{33}}.$$

In general, when the energy densities depend on a single mechanical coordinate S, their components can be represented as follows:

$$w^S_{el}=\frac{1}{2}C^S_{el\cdot\Delta}E^2, \quad w^E_m=\frac{1}{2}\frac{S^2}{C^E_{m\cdot\Delta}}, \quad w_{em}=\frac{1}{2}n_\Delta ES. \tag{8}$$

The energy relations in Eq. (8) can be derived from the electrical circuit with parameters $C^S_{el\cdot\Delta}$, $C^E_{m\cdot\Delta}$, and n_Δ as shown in Fig. 1, which can thus be considered as the equivalent electrical representation of a unit volume of an electromechanical transducer. Parameters such as $C^E_{el\cdot\Delta}$, $C^E_{m\cdot\Delta}$, and n_Δ have to be specified for particular boundary conditions. For instance, for an elemental volume inside the length expander bar these parameters are expressed as

$$C^S_{el\cdot\Delta}=\varepsilon^{S_i}_{33}, \quad C^E_{m\Delta}=\frac{1}{s^E_{ii}}, \quad n_\Delta=\frac{d_{3i}}{s^E_{ii}}. \tag{9}$$

Taking into account Eq. (9), the definition of k_m in Eq. (4) can be rewritten in the form

$$k^2_m=\frac{1}{1+\dfrac{C^S_{el\cdot\Delta}}{n^2_\Delta C^E_{m\cdot\Delta}}}. \tag{10}$$

Equation (4) for k_m is obtained by the operation in the transmit mode. Although it could be concluded by reciprocity that k_m defined in the "receive mode" (i.e., with T or S as independent variables) should be the same, it is useful to derive this relation independently. And, one of the reasons is that the expressions for the internal energy components in the case of the independent mechanical input are needed to validate Eq. (2) for k_m.

In the receive mode an alteration of the internal energy may occur only due to the independent mechanical energy supply. If it is supposed that electrical boundary conditions are ideal, namely, the electrical output is open and $D_3=0$, then

$$\tilde{w}_{\rm int}=\tilde{w}_m=\tilde{w}^D_m=\tfrac{1}{2}\tilde{S}_i\tilde{T}_i. \tag{11}$$

The sign "~" is used with the internal energy components and related quantities in the receive mode (with independent T or S) in order to distinguish them from the analogous components in the transmit mode (with independent E). Thus, the piezoelectric equations of state in Eq. (5) become

$$\tilde{T}_i=c^E_{ik}\tilde{S}_k-e_{3i}\tilde{E}_{3{\rm oc}}, \quad 0=e_{3i}\tilde{S}_i+\varepsilon^S_{3i}\tilde{E}_{3{\rm oc}}, \tag{12}$$

where $\tilde{E}_{3{\rm oc}}$ is the electric field due to the piezoelectric effect in the open-circuited element. Thus, by substituting Eq. (12) into Eq. (11) the internal energy is expressed as

$$\tilde{w}_{\rm int}=\tfrac{1}{2}\tilde{S}_i\tilde{T}_i=\tfrac{1}{2}c^E_{ik}\tilde{S}_i\tilde{S}_k-\tfrac{1}{2}e_{3i}\tilde{S}_i\tilde{E}_3. \tag{13}$$

Obviously, the first term in Eq. (13) represents the mechanical energy of the short-circuited element

$$\tilde{w}^E_m=\tfrac{1}{2}c^E_{ik}\tilde{S}_i\tilde{S}_k. \tag{14}$$

If one uses the relation

$$\tilde{E}_{3{\rm oc}}=-\frac{e_{3i}\tilde{S}_i}{\varepsilon^S_{33}}, \tag{15}$$

obtained from Eq. (12), the second term in Eq. (13) can be transformed into

$$\tilde{w}^S_{el}=-\tfrac{1}{2}e_{3i}\tilde{S}_i\tilde{E}_3=\tfrac{1}{2}\tilde{E}^2_3\varepsilon^S_{33}. \tag{16}$$

Let us denote

$$\tilde{w}_{me}=-\tfrac{1}{2}e_{3i}\tilde{S}_i\tilde{E}_{3{\rm oc}}, \tag{17}$$

as the mechanoelectrical energy. As it follows from Eq. (16), $\tilde{w}_{me}$ is always positive. This means that in the case of $i=1$ ($e_{3i}<0$) the expansion as the positive strain ($\tilde{S}_1>0$) leads to $\tilde{E}_{3{\rm oc}}>0$ (the direction of $\mathbf{E}_{3{\rm oc}}$ coincides with the direction of the polar axis $\mathbf{P}$). In the case of $i=3$ ($e_{33}>0$)

the expansion ($\tilde{S}_3>0$) leads to $\tilde{E}_{3\mathrm{oc}}<0$ ($\mathbf{E}_{3\mathrm{oc}}$ is opposite to $\mathbf{P}$). In order to reduce the complications regarding the signs of e_{3i} and $\tilde{E}_{3\mathrm{oc}}$ we can substitute $\tilde{w}_{me}$ by its absolute value $|\tilde{w}_{me}|$ in the expressions using the mechanoelectrical energy. The internal energy in the receive mode in Eq. (13) can then be represented as

$$\tilde{w}_{\mathrm{int}}=\tilde{w}_m^D=\tilde{w}_m^E+|\tilde{w}_{me}|=\tilde{w}_m^E+\tilde{w}_{el}^S. \tag{18}$$

The expression (1) for k_c has to be rephrased for the receive mode as

$$k_c^2=\frac{\text{energy stored in the electrical form}}{\text{total input mechanical energy}}. \tag{19}$$

Therefore, k_m can be represented in one of the forms

$$k_m^2=\frac{\tilde{w}_{el}^S}{\tilde{w}_m^E+\tilde{w}_{el}^S}=\frac{|\tilde{w}_{me}|}{\tilde{w}_m^E+|\tilde{w}_{me}|}=\frac{\tilde{w}_{el}^S}{\tilde{w}_m^D}. \tag{20}$$

The expressions for w_m^E, w_{el}^S, and $|w_{em}|$ in Ref. 4 will coincide with Eqs. (14), (15), and (17) for $\tilde{w}_m^E$, $\tilde{w}_{el}^S$, and $|\tilde{w}_{me}|$, when substituting $\tilde{S}_i$ and $\tilde{E}_{3\mathrm{oc}}$ for S_i and E_3. Likewise, the equivalent circuit in Fig. 1 and the expression for k_m in Eq. (10) are valid for both receive and transmit modes.

Let us now turn to the expression for k_m in Eq. (2) and consider its derivation in Ref. 2. Starting from the general expression for the internal energy

$$w_{\mathrm{int}}=\tfrac{1}{2}S_iT_i+\tfrac{1}{2}D_mE_m, \tag{21}$$

and using the piezoelectric equations (6), the authors obtained

$$\begin{aligned}w_{\mathrm{int}}&=\tfrac{1}{2}T_is_{ik}^ET_k+\tfrac{1}{2}T_id_{im}E_m+\tfrac{1}{2}E_md_{im}T_i+\tfrac{1}{2}E_3^2\varepsilon_{33}^T\\&=U_e+2U_m+U_d,\end{aligned} \tag{22}$$

where $U_e=T_is_{ik}^ET_k/2$, $U_m=E_md_{im}T_i/2$, and $U_d=\varepsilon_{33}^TE_3^2/2$ are the elastic, mutual, and dielectric energy, respectively. Afterwards, the definition for k_m in Eq. (2) is introduced by analogy with the correlation coefficient between two actions. However, it has to be remembered that $E_mD_m/2$ and $T_iS_i/2$ in Eq. (21) are independent electrical and independent mechanical energies, that is, Eqs. (21) and (22) should be represented in our designations in the form of

$$\begin{aligned}w_{\mathrm{int}}&=\tfrac{1}{2}\tilde{S}_i\tilde{T}_i+\tfrac{1}{2}E_mD_m\\&=\tfrac{1}{2}\tilde{T}_is_{ik}^E\tilde{T}_k+\tfrac{1}{2}\tilde{T}_id_{im}\tilde{E}_m+\tfrac{1}{2}T_id_{im}E_m+\tfrac{1}{2}\varepsilon_{33}^TE_3^2\\&=\tilde{w}_m^E+\tilde{w}_{me}+w_{em}+w_{el}^T.\end{aligned} \tag{23}$$

Now, it is clear that the middle terms in Eq. (22) being outwardly similar are not equal in general and could not be doubled. They may be equal in only a particular case, when a certain relation exists between otherwise independent electrical and mechanical actions. This relation being obtained from the equality $w_{em}=\tilde{w}_{me}$ is

$$\frac{E_3}{\tilde{T}_i}=-\frac{d_{3i}}{\varepsilon_{33}^T}. \tag{24}$$

Because of the above-mentioned inaccuracy, the piezoelectric equations (5) having strains as independent variables cannot be used to derive the expression for k_m in Eq. (2). In fact, as described by Eqs. (21) and (22), the same procedure applied in this case leads to

$$\begin{aligned}w_{\mathrm{int}}&=\tfrac{1}{2}S_ic_{ik}^ES_k+\tfrac{1}{2}S_ie_{im}E_m-\tfrac{1}{2}E_me_{im}S_i+\tfrac{1}{2}\varepsilon_{33}^SE_3^2\\&=U_e+U_d,\end{aligned} \tag{25}$$

and the mutual term disappears.

Despite the inaccurate derivation of Eq. (2), the expression of this kind can be used to determine k_m because it can be obtained from the original definition of Eq. (1) in the form of Eq. (4). Indeed, if it is taken into account that $|w_{em}|=w_m^E$ in the case that the stresses are independent variables (see Ref. 4), Eq. (4) for k_m^2 can be represented as

$$k_m^2=\frac{|w_{em}|}{w_{el}^T}=\frac{|w_{em}|}{w_{el}^T}\cdot\frac{|w_{em}|}{w_m^E},\quad \text{i.e.,}\quad k_m=\frac{|w_{em}|}{\sqrt{w_{el}^Tw_m^E}}. \tag{26}$$

In the case that the strains are independent variables, it can be found from Eq. (4) that

$$\frac{|w_{em}|}{w_{el}^S}=\frac{k_m^2}{1-k_m^2}=\frac{|w_{em}|}{w_{el}^S}\cdot\frac{|w_{em}|}{w_m^E},$$

i.e.,

$$\frac{k_m}{\sqrt{1-k_m^2}}=\frac{|w_{em}|}{\sqrt{w_{el}^Sw_m^E}}. \tag{27}$$

Equations (26) and (27) can be used to determine k_m as well as Eqs. (4) and (19).

Consider now the effective coupling coefficient, k_{eff}, for the arbitrary deformed piezoelectric body. Using the expression for the internal energy of a piezoelectric body under nonuniform deformation analyzed in Ref. 4, Eq. (1) for k_{eff} can be represented as

$$k_{\mathrm{eff}}^2=\frac{W_{em}}{W_{el}^S+W_{em}}=\frac{W_m^E+\Delta W}{W_{el}^S+W_m^E+\Delta W}, \tag{28}$$

where W_{el}^S, W_{em}, W_m^E, and ΔW are the internal energy components of a piezoelectric body defined in Ref. 4.

It has to be noted that the concept of k_{eff} makes sense only in connection with a certain distribution of strain in a body, and is valid so long as this distribution remains invariable. In other words, each single mechanical degree of freedom of a piezoelectric body is characterized by its coupling coefficient k_{eff}. Therefore, prior to calculating k_{eff} the particular displacement distribution $\xi(\mathbf{r})$ in the body should be known. This displacement distribution can be represented as

$$\xi(\mathbf{r})=\xi_0\theta_\xi(\mathbf{r}), \tag{29}$$

where ξ_0 is the displacement of the reference point with coordinate $\mathbf{r}_0$ and $\theta_\xi(\mathbf{r})$ is the mode shape, i.e., the displacement distribution normalized in such a way that $\theta_\xi(\mathbf{r}_0)=1$. Let the corresponding strain distribution be

$$S(\mathbf{r})=S_0\theta_S(\mathbf{r}). \tag{30}$$

In the case that the body has one mechanical degree of freedom, the energies involved in Eq. (28) may be expressed by

means of corresponding equivalent parameters defined in Ref. 4, i.e.,

$$W_{el}^S=\frac{1}{2}C_{el}^S V^2, \quad W_m^E=\frac{1}{2}\frac{\xi_0^2}{C_m^E},$$
$$W_{em}=\frac{1}{2}nV\xi_0, \quad \Delta W=\frac{1}{2}\frac{\xi_0^2}{\Delta C}, \tag{31}$$

and Eq. (28) for k_{eff}^2 can be modified as follows:

$$k_{\text{eff}}^2=\frac{1}{1+\dfrac{C_{el}^S}{n^2 C_m}}, \tag{32}$$

where $C_m=C_m^E\cdot\Delta C/(C_m^E+\Delta C)$. Equation (32) is especially convenient to calculate k_{eff} for particular transducer types.

III. ON THE OPTIMIZATION OF THE EFFECTIVE COUPLING COEFFICIENT

In general, under nonuniform deformation the effective coupling coefficient is smaller than the material coupling coefficient, i.e., $k_{\text{eff}}<k_m$, which means that the ability of a piezoelectric material to perform electromechanical conversion is not fully utilized. The question arises whether the electromechanical conversion in a body under nonuniform deformation can be improved. To answer this question the analytical formulation for k_{eff} in Eq. (28) may be considered. First, as stated in Ref. 4, in the case that initially $\Delta W\neq 0$, k_{eff} can be increased by segmenting the mechanical system of the transducer in the direction of the electric field, which leads to $\Delta W\doteq 0$. Therefore, for further analysis, it is assumed that $\Delta W=0$ in Eq. (28) and as a result $C_m=C_m^E$ in Eq. (32). The expressions for the energies of a body with one mechanical degree of freedom may be represented in Eq. (28) as follows (see Ref. 4):

$$W_{el}^S=\frac{1}{2}C_{el}^S V^2=\int_{\tilde{V}} w_{el}^S(\mathbf{r})\,d\tilde{V},$$
$$W_m^E=\frac{1}{2}\frac{\xi_0^2}{C_m^E}=\int_{\tilde{V}} w_m^E(\mathbf{r})\,d\tilde{V}, \tag{33}$$
$$W_{em}=\frac{1}{2}nV\xi_0=\int_{\tilde{V}} w_{em}(\mathbf{r})\,d\tilde{V}.$$

Suppose that the electric field is a function of coordinates

$$E_3(\mathbf{r})=E_3(\mathbf{r}_0)\,\theta_E(\mathbf{r}), \tag{34}$$

where $\theta_E(\mathbf{r})$ is the normalized electric field distribution. Taking into account Eqs. (8), (30), and (34), the energy densities w_{el}^S, w_m^E, and w_{em} as the functions of the distribution of strain or electric field can be represented as

$$w_{el}^S(\mathbf{r})=w_{el}^S(\mathbf{r}_0)\,\theta_E^2(\mathbf{r}),$$
$$w_m^E(\mathbf{r})=w_m^E(\mathbf{r}_0)\,\theta_S^2(\mathbf{r}), \tag{35}$$
$$w_{em}=w_{em}(\mathbf{r}_0)\,\theta_E(\mathbf{r})\,\theta_S(\mathbf{r}).$$

Upon substituting Eq. (35) into Eq. (33), and Eq. (33) in Eq. (28), we arrive at the expression

$$k_{\text{eff}}^2=\frac{w_{em}(\mathbf{r}_0)\int_{\tilde{V}}\theta_E(\mathbf{r})\,\theta_S(\mathbf{r})\,d\tilde{V}}{w_{el}^S(\mathbf{r}_0)\int_{\tilde{V}}\theta_E^2(\mathbf{r})\,d\tilde{V}+w_m^E(\mathbf{r}_0)\int_{\tilde{V}}\theta_S^2(\mathbf{r})\,d\tilde{V}}, \tag{36}$$

from which k_{eff} may be found in a particular case as soon as the distribution of strain and electric field are known. The following particular cases are of interest:

(1) Uniform strain and uniform electric field

$$\theta_S(\mathbf{r})=1, \quad \theta_E(\mathbf{r})=1.$$

This is, for example, the case that thin rings and spherical shells with fully electroded inner and outer surfaces vibrate in the fundamental (breathing) mode. Equation (36) in this case becomes

$$k_{\text{eff}}^2=\frac{w_{em}(\mathbf{r}_0)}{w_{el}^S(\mathbf{r}_0)+w_m^E(\mathbf{r}_0)}, \tag{37}$$

and the comparison with Eq. (4) results in $k_{\text{eff}}=k_m$.

(2) Arbitrary strain distribution and uniform electric field

$$\theta_S(\mathbf{r})<1 \quad \text{at } \mathbf{r}\neq\mathbf{r}_0, \quad \theta_E(\mathbf{r})=1.$$

From Eq. (36), it can be concluded that $k_{\text{eff}}<k_m$.

(3) Arbitrary strain distribution and electric field, that matches the strain distribution

$$\theta_E(\mathbf{r})=\theta_S(\mathbf{r}).$$

Substitution of $\theta_E(\mathbf{r})=\theta_S(\mathbf{r})$ into Eq. (36) results in Eq. (37), and hence $k_{\text{eff}}=k_m$.

Thus, theoretically, the effective coupling coefficient k_{eff} for a piezoelectric body under nonuniform deformation can be increased up to the corresponding k_m, if to match the electric field distribution with the strain distribution.

To illustrate how this condition can be fulfilled, and the physics behind this condition, at first we present the examples of the length expander bars, namely, side-electroded, end-electroded, and segmented axially polarized bars vibrating in the fundamental mode. The peculiarities of the energy conversion in these piezoelectric elements were considered in Ref. 4, where the geometry of the bars can be found in Fig. 5 and the expressions for equivalent parameters are given in Eqs. (44)–(46). The substitution of the equivalent parameters of the bars into Eq. (32) gives the following values of k_{eff}:

$$k_{\text{eff}}^2=\frac{1}{1+\dfrac{\pi^2}{8}\dfrac{k_{3i}^2}{1-k_{3i}^2}}, \tag{38}$$

where $i=1$ for the side-electroded bar and $i=3$ for the segmented bar with the number of segments greater than six $(N>6)$, and

$$k_{\text{eff}}^2=\frac{8k_{33}^2}{\pi^2} \tag{39}$$

for the solid end-electroded bar. Thus, k_{eff} of a solid end-electroded bar is smaller than that of a segmented bar. For

example, if PZT-4 is considered, $k^2_{\text{eff}}=0.40$ for the solid bar and $k^2_{\text{eff}}=0.45$ for the segmented bar, while the material coupling coefficient is $k^2_m \doteq 0.5$.

Qualitatively, this fact is explained in Ref. 4 by considering the density of the electrical energy $w_{el}(x)$ consumed in an elemental volume of a bar transducer

$$w_{el}(x)=[w^S_{el}(x)+w_{el\cdot m}(x)]\sim[i_S(x)+i_m(x)], \tag{40}$$

where $w^S_{el}(x)$ and $i_S(x)$ are the electrical energy density and the current through a clamped element, $w_{el\cdot m}(x)$ and $i_m(x)$ are the motional part of the electrical energy density and the motional current due to the electromechanical conversion. It is shown in Ref. 4 that the distribution of the motional part of the electrical energy between the elements of the side-electroded and segmented bars is proportional to their contribution to the electromechanical conversion, while in a solid end-electroded bar the motional energy is distributed uniformly regardless of the contribution of elements to the electromechanical conversion. Thus, the consumption of the electrical energy is more effective and the value of k_{eff} is larger in a segmented bar than in a solid bar. However, at $E_3(x)=\text{constant}$ [or $\theta_E(x)=1$] the consumption of the electrical energy in the "clamped" elements $w^S_{el}(x)$ still remains uniform in both the side-electroded and the segmented bars, although the elements experiencing small strain contribute very little to the electromechanical energy, $w_{em}(x)$. In fact, according to Eq. (8) $w_{em}(x)\sim E_3(x)\cdot S(x)$ and $w^S_{el}\sim E^2_3(x)$, i.e.,

$$\frac{w_{em}(x)}{w^S_{el}(x)}\sim\frac{S(x)}{E_3(x)}\sim\frac{\theta_S(x)}{\theta_E(x)}. \tag{41}$$

In the fundamental mode of a length expander bar, the strain distribution is $\theta_S(x)=\sin(\pi x/l)$, and in the case that electrodes are full as illustrated in Fig. 2(a), the electric field is $E_3(x)=\text{constant}$, i.e., $\theta_E(x)=1$. Therefore, $w_{em}(x)/w^S_{el}(x)\sim\sin(\pi x/l)$.

The goal is to achieve $E_3(x)\sim S(x)$, in other words, distributions $\theta_E(x)=\theta_S(x)$, that would result in $w_{em}(x)/w^S_{el}(x)=\text{constant}$ in Eq. (41), which is the case that the consumption of energy $w^S_{el}(x)$ exactly matches the conversion of the energy into the mechanical form.

The electric field and strain distribution is seldom exactly matched in practical devices. Even the staircase distribution of electric field shown in Fig. 2(b) is very complicated and scarcely worthwhile to pursue matching. Fortunately, an almost optimum value of k_{eff} can be achieved by means of the simple approximation of the electric field to the strain distribution shown in Fig. 2(c), namely, just by removing the electrodes from the parts of the mechanical system that experience relatively small deformation.

To make an appropriate quantitative analysis, consider Eq. (32) for k_{eff}. In this expression the term $n^2C^E_m/C^S_{el}$, which we denote by α

$$\alpha=\frac{n^2C^E_m}{C^S_{el}}, \tag{42}$$

is linked to k_{eff} by the relation

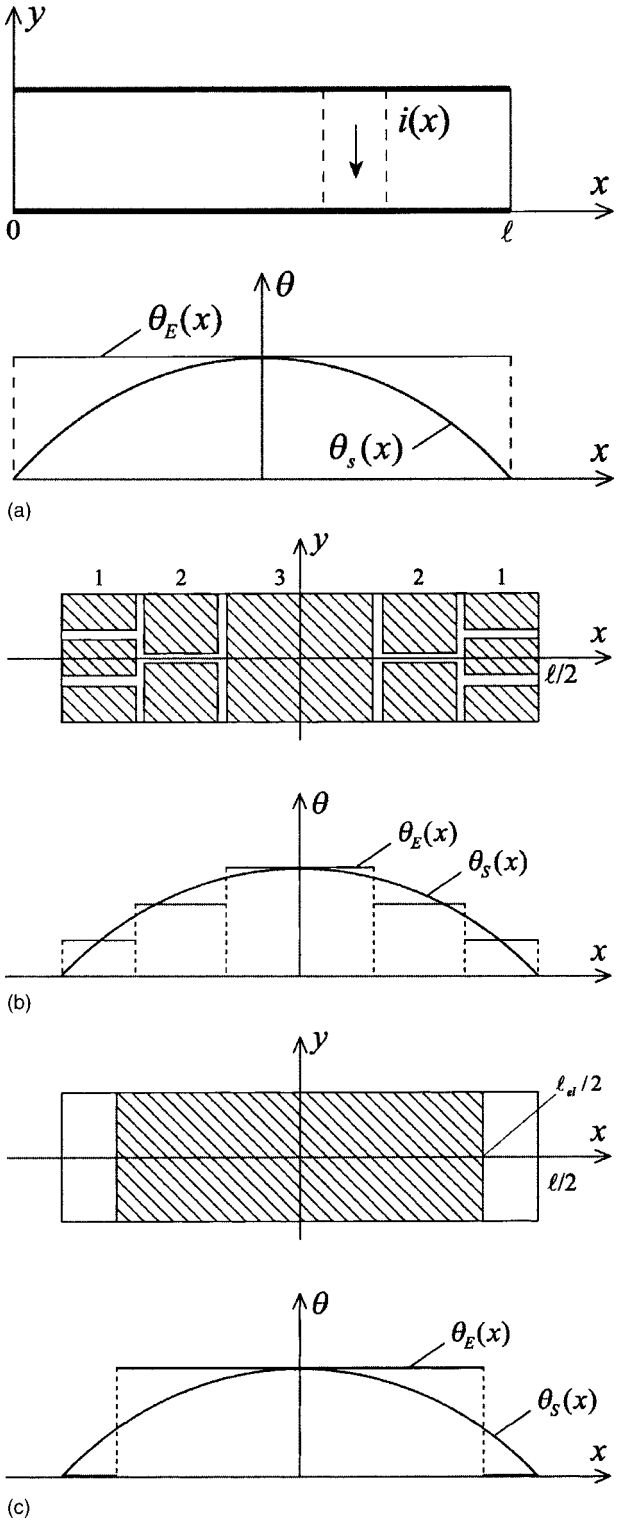


FIG. 2. The length expander bars with different electrode shapes: (a) full electrode; (b) segmented electrode (two halves of part 2 and three thirds of part 1 are connected electrically in series and then all the parts 1, 2, and 3 are connected electrically in parallel); and (c) partial electrode.

$$\alpha=\frac{k^2_{\text{eff}}}{1-k^2_{\text{eff}}}. \tag{43}$$

Let α_0 and k_{eff_0} be the coefficients corresponding to a certain electrode configuration of comparison (the reference configuration), that is

$$\alpha_0 = \left(\frac{n^2 C_m^E}{C_{el}^S} \right)_0 = \frac{k_{\text{eff}_0}^2}{1 - k_{\text{eff}_0}^2}. \tag{44}$$

The coefficient

$$A = \frac{\alpha}{\alpha_0} = \frac{k_{\text{eff}}^2}{k_{\text{eff}_0}^2} \frac{1 - k_{\text{eff}_0}^2}{1 - k_{\text{eff}}^2}, \tag{45}$$

characterizes the relative change of k_{eff} caused by the change of the electrode shape with respect to the reference configuration. In particular, when the reference configuration is optimal in the sense that k_{eff} reaches its maximum value, $\alpha_0 = \alpha_{3i} = k_{3i}^2/(1 - k_{3i}^2)$. If $k_{\text{eff}_0}^2 \ll 1$, as it is in the case of the transverse piezoelectric effect in PZT ceramics [$k_{31}^2 < (0.1\text{–}0.15)$ practically for all PZT compositions]

$$A \doteq \frac{k_{\text{eff}}^2}{k_{\text{eff}_0}^2}, \tag{46}$$

within 5%. In the case that k_{33} is used, and especially for the materials with large values of the coupling coefficients, for instance, such as single-crystal ceramics ($k_{33} \doteq 0.9$), the exact relation (45) has to be employed and

$$\frac{k_{\text{eff}}^2}{k_{\text{eff}_0}^2} = A \frac{1}{1 - (1 - A) k_{\text{eff}_0}^2} = A \frac{1 + \alpha_0}{1 + A \alpha_0}. \tag{47}$$

When the electrode shape changes, the equivalent compliance C_m^E remains the same, if the inactive parts of the mechanical system are electroded and short-circuited. If these parts are not electroded, then strictly speaking one has to use the Young's modulus of the unpolarized ceramic upon evaluating the contribution of the inactive parts to the total potential energy of the mechanical system. However, the potential energy density, w_{pot}, of these parts is relatively small because of small strain (they are not used for the electromechanical conversion exactly for this reason). Thus, the values of the total W_m^E, and accordingly C_m^E, should not change significantly in comparison with those for a fully active mechanical system. Therefore

$$A \doteq \frac{n^2 / C_{el}^S}{(n^2 / C_{el}^S)_0},$$

and we can judge the change of k_{eff} based on the behavior of the term n^2/C_{el}^S.

Note that it is assumed throughout this article that the mechanical system of the transducer remains uniformly made up of ceramic material. In general, the unelectroded parts of the mechanical system can be replaced by an inactive material with different elastic properties. In this case the same analysis is applicable, but the changes of C_m^E also have to be taken into account.

Substitution of $w_{el}^{S_1}$ and w_{em} from Eq. (7) into Eq. (35), and carrying out the integrals in Eq. (33) over the volume of the bar produces the following general expressions for n and $C_{el}^{S_1}$ in the case of a side-electroded length expander bar:

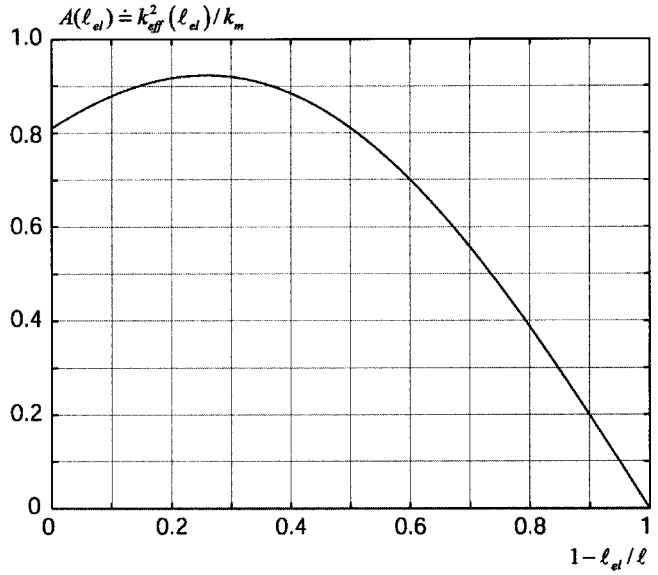


FIG. 3. Illustration of the effective coupling coefficient dependence on the electrode length.

$$n = \frac{d_{31} b}{s_{11}^E} \int_{-\ell/2}^{\ell/2} \theta_E(x)\, \theta_S(x)\, dx, \tag{48}$$

$$C_{el}^{S_1} = \frac{\varepsilon_{33}^T (1 - k_{31}^2)}{t} \int_{-\ell/2}^{\ell/2} \theta_E^2(x)\, dx. \tag{49}$$

With these expressions in use, it appears that for the electric field distribution corresponding to the electrode connection shown in Fig. 2(b) in the case that $\alpha_0 = \alpha_{31}$, we obtain A=0.95, i.e., $k_{\text{eff}} = 0.97 k_{31}$ (which is almost the maximum possible value) compared with $k_{\text{eff}} = 0.90 k_{31}$ in the case of the uniform electric field.

In the case of the electrode shape shown in Fig. 2(c)

$$\theta_E(x) = 1 \quad \text{at} \quad |x| \leqslant \frac{\ell_{el}}{2} \quad \text{and} \quad \theta_E(x) = 0 \quad \text{at} \quad |x| > \frac{\ell_{el}}{2}, \tag{50}$$

$$n = \frac{2 b d_{31}}{S_{11}^E} \sin \frac{\pi \ell_{el}}{2\ell}, \quad C_{el}^{S_1} = \frac{\varepsilon_{33}^T (1 - k_{31}^2) b \ell_{el}}{t}, \tag{51}$$

$$A(\ell_{el}) = \frac{8\ell}{\pi^2 \ell_{el}} \cdot \sin^2 \frac{\pi \ell_{el}}{2\ell}. \tag{52}$$

The plot for Eq. (52) is depicted in Fig. 3, from which it can be concluded that the maximum value for k_{eff} can be achieved by removing 0.125 of electrode length from each end of the bar. In this case $A = 0.94$, i.e., practically the same value as for the rather complicated electrode configuration presented in Fig. 2(b). Another interesting conclusion can be made that in the case that $\ell_{el} = 0.5\ell$ the value of k_{eff} is the same as with the electrodes of the full size. Both of these conclusions are qualitatively valid in general, namely, the effective coupling coefficient k_{eff} of a piezoelectric body can be increased or remain as high as the case of replacing its less active parts by an inactive material with approximately the same elastic properties.

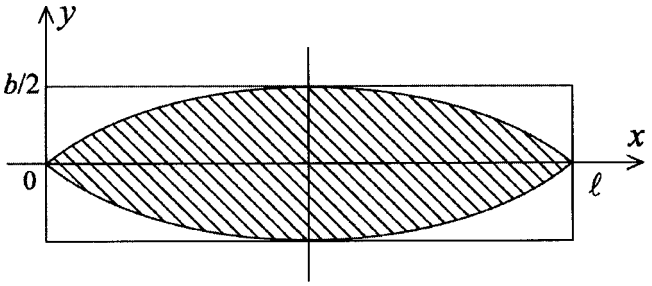


FIG. 4. The electrode shape, leading to the isolated fundamental mode $\theta_{\xi}(x)=\cos(\pi x/l)$.

If one considers a piezoelectric transducer as having the mechanical system with multiple degrees of freedom each corresponding to a normal mode of vibration, then maximizing k_{eff} for a particular mode makes this mode isolated, i.e., the only electromechanically coupled one. It is worth mentioning that the opposite statement is not valid, i.e., an isolated mode does not always have maximum k_{eff}. For example, the electrode shape shown in Fig. 4, in which case

$$\theta_E(x,y)=1 \quad \text{at } |y|\leqslant \frac{b}{2}\sin\frac{\pi x}{\ell}$$

and

$$\theta_E(x)=0 \quad \text{at } |y|>\frac{b}{2}\sin\frac{\pi x}{\ell}, \tag{53}$$

leads to the isolated fundamental mode $\theta_{\xi}(x)=\cos(\pi x/\ell)$. In this case, however, evaluating k_{eff} by applying the above-considered procedure gives $k_{\text{eff}}=0.78k_{31}$, i.e., an even smaller value than in the case of the uniform electrical field.

Very often it is more reasonable to consider the optimization of k_{eff} in terms of reducing the amount of piezoelectric material in the mechanical system of a transducer without loss in k_{eff} rather than obtaining the maximum possible k_{eff}. Thus, as we have already seen, up to 0.50 of the volume of a length expander bar can be substituted by an inactive material without loss in k_{eff}. Even more significant gain can be produced in the case that the mechanical system experiences flexural deformation.

IV. ON THE EFFECTIVE COUPLING COEFFICIENTS OF THE BEAMS AND PLATES UNDER FLEXURE

The peculiarity of mechanical systems experiencing flexural deformation is that nonuniform strain distribution takes place in several dimensions. In the case of a thin beam and round plate with axisymmetric electrodes shown in Fig. 5, for example, there are distributions in the direction of the thickness and the length or the radius. Therefore, the values of k_{eff} for regular designs with uniform electrical field have to be smaller than in the case of a one-dimensional strain distribution, and the gains due to optimization have to be more significant. For a thin beam with simply supported ends vibrating in the first flexural normal mode

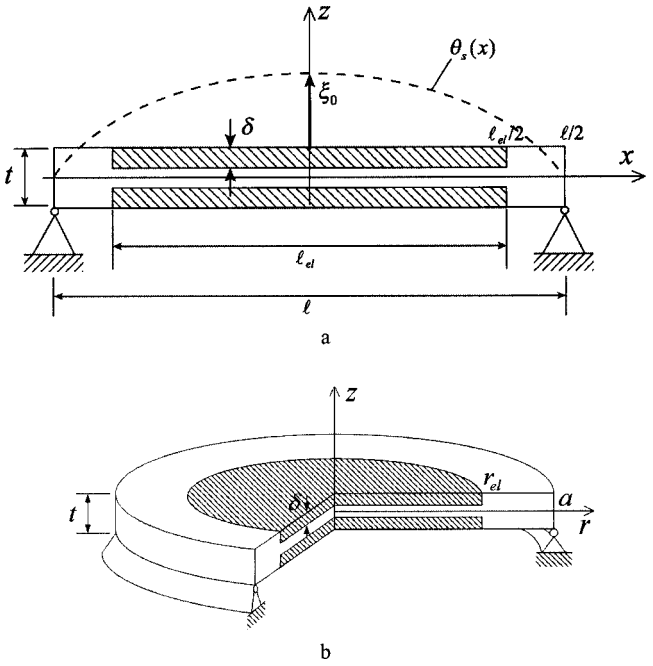


FIG. 5. Mechanical systems under flexure: (a) rectangular bar with simply supported ends, and (b) round plate with simply supported edge.

$$\xi_z(x)=\xi_0\cos\frac{\pi x}{\ell},$$
$$S_x(x,z)=-z\frac{\partial^2\xi_z}{\partial^2 x}=z\xi_0\frac{\pi^2}{\ell^2}\cos\frac{\pi x}{\ell}. \tag{54}$$

In the case that electrodes are inserted in mechanical system as shown in Fig. 5(a), we represent the electrical field in the body of the beam as follows:

$$E_3(x,z)=\theta_{Ex}(x)\,\theta_{Ez}(z)\,\frac{V}{\delta}, \tag{55}$$

where

$$\theta_{Ex}(x)=1 \quad \text{at } |x|\leqslant\frac{\ell_{el}}{2} \quad \text{and } \theta_{Ex}=0 \quad \text{at } |x|>\frac{\ell_{el}}{2}, \tag{56}$$

$$\theta_{Ez}(z)=1 \quad \text{at } |z|\geqslant\frac{t}{2}-\delta \quad \text{and } \theta_{Ez}(z)=0 \quad \text{at } |z|\leqslant\frac{t}{2}-\delta. \tag{57}$$

Here, ℓ_{el} and δ are the length and thickness of the active piezoelectric layers, which are assumed to be connected electrically in parallel, and $\ell_{el}=\ell$ and $\delta=t/2$ correspond to the beam fully made of active piezoelectric material. Using Eqs. (8) and (35) and carrying out the integrals in Eq. (33) over the volume of the beam produces

$$n(\delta,\ell_{el})=\frac{bd_{31}}{s_{11}^E}\frac{(t-\delta)\pi}{\ell}\sin\frac{\pi\ell_{el}}{2\ell},$$
$$C_{el}^{S_1}=\frac{2\varepsilon_{33}^T(1-k_{31}^2)b\ell_{el}}{\delta}. \tag{58}$$

The term $n^2(\delta,\ell_{el})/C_{el}^{S_1}(\delta,\ell_{el})$, which is normalized to its value at $\delta=t/2$ and $\ell_{el}=\ell$, may be represented in the form

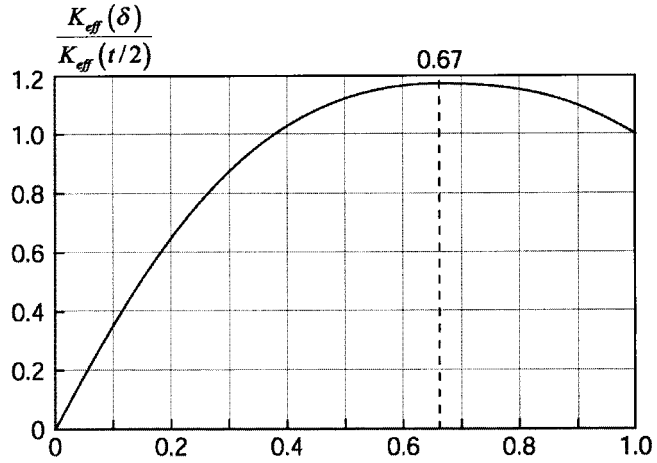


FIG. 6. Illustration of the k_{eff} dependence on the thickness of a piezoelectric layer.

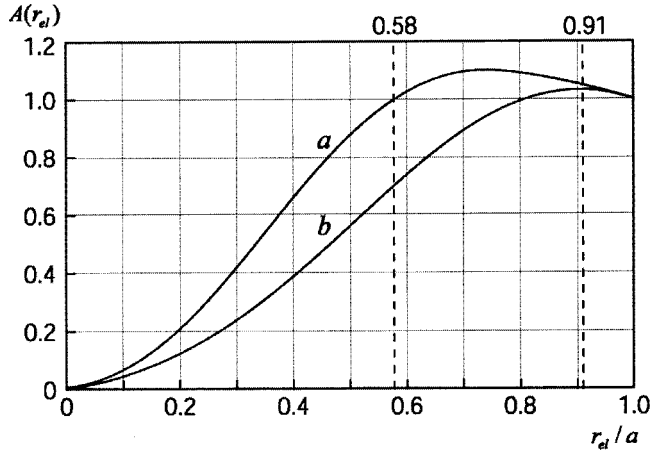


FIG. 7. Illustration of the k_{eff} of circular plates: (a) with supported boundary; (b) with free boundary.

$$A \doteq \frac{\alpha(\delta,\ell_{el})}{\alpha(t/2,\ell)} = A(\ell_{el})A(\delta), \tag{59}$$

where $A(\ell_{el})$ is given by Eq. (52) and depicted in Fig. 3, and

$$A(\delta) = 8\frac{(1-\delta/t)^2\delta}{t}. \tag{60}$$

The coefficient $A(\delta/t)$ is given in Fig. 6 as a function of δ/t. This function has the maximum at $\delta = t/3$, and it has the same value at $\delta = t/6$ as at $\delta = t/2$. Summarizing the results illustrated in Figs. 3 and 6, one may conclude that the maximum value of k_{eff} can be achieved at $\delta = t/3$ and $\ell_{el} = 0.75\ell$. In this case $k_{\text{eff}} = 0.91k_{31}$ and the volume of an active material is half of the total volume of the beam. Note that in the case of $\delta = t/6$ and $\ell_{el} = 0.5\ell$ ($\tilde{V}_{\text{active}} = 0.17\tilde{V}_{\text{total}}$), the effective coupling coefficient has the same value as in the case that the beam is fully made of piezoelectric material. This value is $k_{\text{eff}} = 0.78k_{31}$.

The electromechanical conversion in the circular axially symmetric plates [Fig. 5(b)] vibrating in flexure is considered in Ref. 5. For a circular plate the expression analogous to Eq. (59) for a beam vibrating in flexure may be represented as

$$A \doteq \frac{\alpha(\delta,r_{el})}{\alpha(t/2,a)} = A(\delta/t)A(r_{el}/a). \tag{61}$$

The factor $A(\delta/t)$ is the same as given by Eq. (60) and as illustrated in Fig. 6. The factor $A(r_{el}/a)$ is different for different boundary conditions. For the simply supported and free plates, $A(r_{el}/a)$ as the function of r_{el}/a is given in Fig. 7. In the case that the plates are fully made of piezoelectric material $k_{\text{eff}} = 0.9k_{31}$ for the simply supported boundary, and $k_{\text{eff}} = 0.8k_{31}$ for the free boundary. The same value of k_{eff} can be obtained at $r_{el} = 0.8a$ and $\delta = t/6$ ($\tilde{V}_{\text{active}} = 0.2\tilde{V}_{\text{total}}$) for the simply supported plate, and at $r_{el} = 0.58a$ and $\delta = t/6$ ($\tilde{V}_{\text{active}} = 0.1\tilde{V}_{\text{total}}$) for the free plate. The maximum values of k_{eff} are: $k_{\text{eff}} = 0.98k_{31}$ ($r_{el} = 0.9a$, $\delta = t/3$ and $\tilde{V}_{\text{active}} = 0.55\tilde{V}_{\text{total}}$) for the simply supported plate, and $k_{\text{eff}} = 0.92k_{31}$ ($r_{el} = 0.75a$, $\delta = t/3$ and $\tilde{V}_{\text{active}} = 0.37\tilde{V}_{\text{total}}$) for the free plate.

It can be noted that, in the case of a fully electroded circular plate with a clamped outer boundary, $k_{\text{eff}} = 0$. This is due to the fact that the total lateral strain distribution in the plate under flexure and the distribution of the electrical charge density, respectively, change their signs at $r \doteq a/\sqrt{2}$, as shown in Fig. 8. Thus, the total charge of the electrode appears to be zero. In order to ''revive'' the electromechanical conversion the electrodes have to be split on the line $r = a/\sqrt{2}$, and the two parts have to be connected in opposite phase. This will make the first step to match electric field to the strain distribution. Then, the above-described procedure of k_{eff} optimization can be applied.

It has to be noticed that all the numerical results related to k_{eff} optimization were obtained under the assumption that the elastic properties of the inactive parts of a mechanical system are approximately the same as for the active piezoelectric parts. In the case that the elastic properties of the active and inactive parts differ significantly, the mechanical system has to be treated as nonuniform and the numerical results may change, although the qualitative conclusions remain valid.

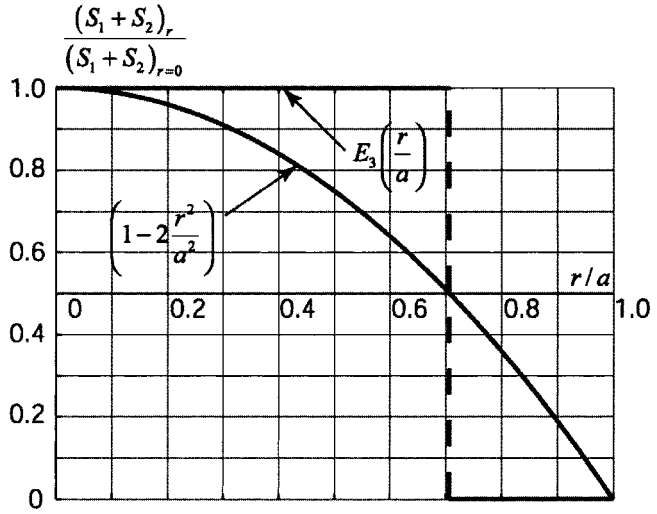


FIG. 8. The distribution of the lateral strain and the charge density over the electrodes of the circular plate with clamped boundary under flexure.

V. CONCLUSIONS

The presented treatment shows that the analytical expression in Eq. (28) for the effective coupling coefficient, k_{eff}, and its modification given in Eq. (32), which are based on the concept of the internal energy of a piezoelectric body considered in Ref. 4, may be successfully used for optimizing the electromechanical conversion of energy under nonuniform deformation. It was shown that the absolute maximum of k_{eff}, which is equal to the corresponding coupling coefficient of the piezoelectric material, can theoretically be achieved for any particular mode of the strain distribution by special electrode design, leading to the distribution of electric field matching the strain distribution. It is illustrated with typical examples of bars, beams, and plates vibrating in the longitudinal and flexural modes that almost optimum results can be obtained just by removing electrodes from the parts of a piezoelectric body, which experience relatively small strains. In practice, these parts of piezoelectric material may be replaced by an inactive material having approximately similar elastic properties. Another option illustrated with the same examples is to significantly reduce the amount of the piezoelectric material in an electromechanical transducer without much reduction of the coupling coefficient k_{eff}.

ACKNOWLEDGMENTS

The author wishes to thank Dr. David A. Brown and Tetsuro Oishi for their cooperation and assistance in preparing and revising the paper for publication. This work was supported in part by BTECH Acoustics, ONR 321SS Elswick and Lindberg, and SBIR N02-066.

[1] W. G. Cady, *Piezoelectricity* (McGraw-Hill, New York, 1946).

[2] D. A. Berlincourt, D. R. Curren, and H. Jaffee, ''Piezoelectric and piezomagnetic materials and their function in transducers'' in *Physical Acoustics*, edited by W. P. Mason (Academic, New York, 1964), Vol. IA.

[3] R. Holland and EeR Nisse E. P., *Design of Resonant Piezoelectric Devices* (MIT Press, Cambridge, 1969).

[4] B. S. Aronov, ''Energy analysis of a piezoelectric body under nonuniform deformation,'' J. Acoust. Soc. Am. **113**, 2638–2646 (2003).

[5] B. S. Aronov, ''On electromechanical transformation of energy under bending strains in thin piezoelectric plates,'' Dopov. Akad. Nauk Ukr. RSR, Ser. A, No. 3 (1980) (Ukrainian), translated in Phys. Abstr. 71320 (1 Aug. 1980).

A new microphone system for near whispering

Sungjoon Choi and Wonkyu Moon
Department of Mechanical Engineering, Pohang University of Science and Technology (POSTECH), San 31, Hyojadong Namgu, Pohang, Kyungbuk 790-784, South Korea

Jeong Hyun Lee
iCurie Lab., 5th Daewon Building, 946-18, Daechi-dong, Kangnam-gu, Seoul 135-846, South Korea

(Received 20 June 2002; revised 25 April 2003; accepted 10 May 2003)

A new microphone system was developed to monitor the human voice near the microphone in a noisy environment. The system is equipped with two special functions in addition to the usual microphone functions: reduction of air-blow effects by the mouth and focused reception to a sound source. A wind filter was developed to reduce the air-blow effects from the mouth during speaking. This filter is a plate perforated by an array of small holes; the method used to design the filter is also presented. To achieve focused reception, four microphones were used in conjunction with a new signal-processing method. The proposed signal-processing method effectively increases the directivity in the desired direction. Additionally, it provides the system with focusing on the source since the source is located adjacent to the system. A prototype of the proposed system was fabricated and subjected to performance tests. The results showed that air-blow effects can be reduced by up to 20 dB and the directional gain is more than 4 dB. The proposed microphone system shows such good performance that it can be used in mobile phones for whispering communication. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1589752]

PACS numbers: 43.38.Hz, 43.38.Kb, 43.38.Si [AJZ]

I. INTRODUCTION

The widespread use of mobile phones in public places has led to an increasing need for a microphone system that can selectively monitor sounds from a particular source. Phone calls in public places cause problems such as unwanted noise impinging on others and leakage of private information. Despite these problems, many people feel it convenient to make phone calls in any place and at any time they desire. The problems associated with mobile phone use in public places would be largely solved by the development of a microphone system that could selectively monitor sounds by focusing on a sound source at a particular location. Such a microphone system would allow the user to whisper into the phone, provided that it is suitable for use near the mouth of the speaker.

Two special functions would be required if the microphone system is to monitor the near whispering of a human: (1) elimination of the effects of the air blows from the mouth, and (2) the ability to focus on a sound source close to the microphones. The former is very important when the microphone system is operated close to the human mouth because air is naturally blown from the mouth during speaking. These air blows cannot be eliminated by the care of the speaker because they accompany the pronunciation of consonants such as ''p'' and ''t''. The second special function is very useful when one makes a phone call in a very noisy place.

In the present study we develop a novel microphone system that includes the two special functions described above. A plate perforated by an array of small holes, referred to as the wind filter, is introduced to eliminate or at least reduce the effects of air blows from the mouth on the acoustic signal received by the microphones. Since the microphone system operates close to and in front of the mouth to maximize the whispering sounds reception, the wind filter, which is located very close to the front of the mouth, can reduce the effects of the air blows. The wind filter is designed to let the air blows from the mouth by-pass through it freely to the opposite side of the microphone array such that their effects on the microphones could be minimized. In addition, the arrangement of the microphones in the system should also be properly designed so as to reduce the effects of the air blows from the mouth.

To achieve the second special function, source-focused reception, four microphones are used. This approach is similar to that used in existing microphone array systems, except that the proposed system becomes more directive at near field and at low frequencies by the signal-processing method. However, since the source of the target sounds (i.e., the speaker's mouth) is located very close to the plane of the microphones and wind filter, the source-focused reception function does not need to be capable of focusing on a point location; it is sufficient to increase the directivity in the direction where the mouth will be located. Existing methods for microphone arrangement and signal processing that have been used previously to design source-focused microphone systems cannot be directly applied to the microphone system developed here, because in the present application the distance between the microphones is too small to generate a satisfactory focusing effect.

Since the 1950s, various technologies have been developed that eliminate background noise by using multiple microphones with appropriate signal-processing techniques.[1–6] Mitchell developed a nonlinear signal-processing technique (MAXMIN and MINMAX types) that can extract noise-reduced signals from the signals received by four micro-

phones. This method works properly provided that the signals are nonoverlapping and have no reverberation effect.[2] Flanagan *et al.* used two or more microphones to reduce the reverberation-tail effects and devised a signal-processing method that reduces the signal distortion more effectively with two or more microphones than when perceiving sounds using a single microphone.[3] In addition, they designed a two-dimensional microphone array that can receive sounds focused on a desired source in a large noisy room without reverberation effects after passing the signals from the multiple microphones through the signal-processing procedures.[4] In another study, they designed a system for increasing the directivity of a single microphone. This method controlled the directivity by placing multiple circular tubes of different lengths on the path to the single microphone, as does the end-fire microphone array.[5] Zelinski *et al.* adopted an approach that uses four microphones to increase the directivity, and a signal-filtering technique called the adaptive Wiener filter (WFS).[6] They demonstrated that their approach considerably reduces background noise. It is worth noting that most of the methods that have been developed to increase directivity require large separations between the microphones to achieve their effects.

There is a technique that can increase the directivity of a microphone system composed of several microphones with their separation distance much smaller than the wavelength of signals: the signal-processing techniques based on the directivity of the multipoles.[7–10] However, in general the multipole microphone system has lower performance in directivity than the focused microphone systems described above, and their directivity decreases at low frequencies. Furthermore, they can be used only for receiving sounds from a faraway source instead of a near source.

As mentioned above, small devices such as cellular phones cannot accommodate microphone systems that require large separations between microphones. Moreover, the existing methods for increasing the directivity by use of an array of microphones and signal processing are effective only if the sources of signals and noise are at a sufficient distance from the microphone system, which is not the case for near whispering. Therefore, a new method needs to be developed to improve the directivity of the microphone system for close whispering.

In this paper, we present a new microphone system and describe the methods used to design the system on the basis of acoustics. A new wind filter is introduced for reducing the air blows from the mouth, and a new signal-processing technique is developed to improve the directivity of the four-microphone array when the sound source is very close to the system in a given direction. A prototype of the new microphone system was fabricated and subjected to performance tests.

II. SYSTEM OPERATION PRINCIPLES AND CONCEPTUAL DESIGN

A. Operation principles

The microphone system is to be installed into mobile phones as an additional flip-style plate such as are found in the flip-type and folder-type cellular phones illustrated in Fig. 1. Hence, it should have a thin profile and be narrower than the phone. The proposed configuration places the microphone system in front of and close to the mouth of the user. The close proximity of the microphone system to the user's mouth means that it will be severely affected by the air blows caused by popping and sibilance during speaking. So, the system should be designed to minimize the effects of the air blows on the microphones. This may be the first functional requirement of the system.

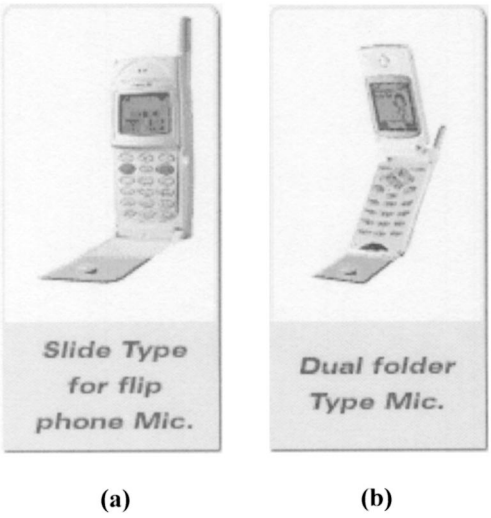


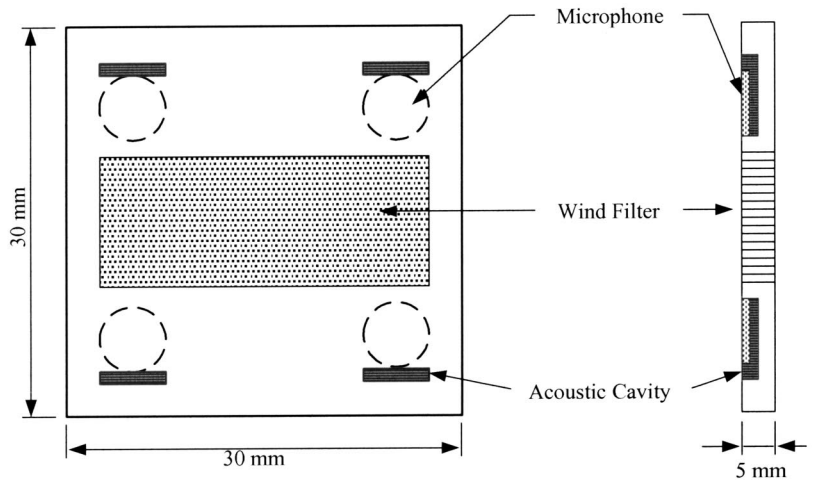
FIG. 1. Installation examples of the system.

Another function is required to receive whispering sounds effectively in various environments: source-focused reception characteristics. Since the system may be used in various environments, it should be very sensitive to the voice signals from the user but as insensitive as possible to the sounds from other sources. Therefore, the function is required to focus the sound reception of the system on a specific source.

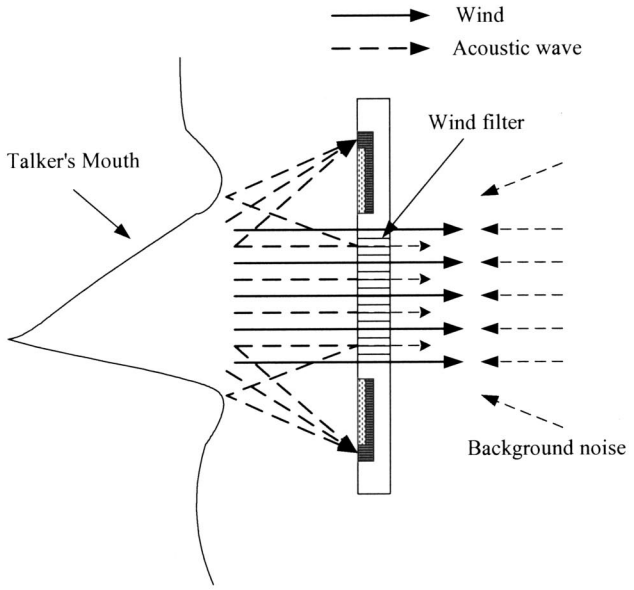
B. Functional requirements

1. Eliminating air-blow effects from the mouth

As described above, the mouth generates strong air blows during speaking, and the noise caused by these air blows must be eliminated if the microphone system is to be effective. In most cases, the effects of air blows become significant when they directly touch the microphone membrane because they generate pressure variations at the membrane. The ''windscreen'' is a widely used device for reducing the effects of wind on microphone reception. This device damps out the air flows by passing them through porous materials, thereby reducing the wind effects. The windscreen requires relatively large volume in order to reduce the wind effects considerably; hence, it cannot be properly installed on the cellular phone without increasing its total volume. Moreover, the acoustic pressure obtained at the microphones is reduced by the windscreen. Due to these reasons, the windscreen is not adequate to apply to the microphone system for near whispering.



(a)



(b)

FIG. 2. Schematics and operating scheme of the microphone system; (a) schematics of the microphone system; (b) operating scheme of the microphone system.

Therefore, we take a different approach to reduce the air-blow effects. In our design the air flow is passed through the system without increasing the pressure near the microphones. This approach can be adopted because the air blows from the mouth influence only a limited range around the mouth.

2. Source-focused reception

The microphone system needs to effectively receive sounds from a given direction (toward the talker's mouth). In addition, it is desirable that the sensitivity of the system decreases rapidly with increasing distance from the source, because the sound source (the mouth) is always located close to the microphone system during operation. As in existing microphone systems for focused reception, we used multiple (four) microphones. And, signal-processing techniques were used to improve the performance of the multimicrophone system.

C. Components of the system and their functions

A schematic diagram of the proposed microphone system is presented in Fig. 2(a). The size of the system is limited by the size of the mobile phone into which it is installed. In fact, given that the system will be installed as a flip-type plate, the size of the system is further restricted to that of the flip component in a flip-type phone [Fig. 1(a)] or the lower part of a folder-type phone [Fig. 1(b)]. Therefore, the maximum size will be approximately a square of dimensions 30×30 mm. In addition, the system should be thinner than the latest thin mobile phones.

As shown in Fig. 2(a), the system consists of a wind filter, four microphones, and four acoustic cavities. The power amplifier circuits are installed inside the plate-like whole system shown in Fig. 2. The wind filter is the area at the center of the square plane containing an array of microchannels [the shaded area in Fig. 2(a)]. These microchannels are small holes with diameters of several hundred micrometers, which penetrate through the plate. Since air blows from

the mouth must pass through these small holes, the pressure variations produced by air blows should be greatly diminished at the microphones. Thus, the pressure signals measured by the microphones should not include the effects of air blows. There is a reason why an array of microchannels is used to reduce the air-blow effects instead of one large hole on the area where the mouth is located. Since sound waves as well as air blows from the mouth pass freely through a large hole, a relatively low proportion of the sounds generated from the mouth is transmitted to the microphones, and a relatively large proportion of them is delivered to people around the speaker. Therefore, the array of microchannels is designed so that only a small portion of the sound waves is transmitted to the outside of the system while the air flows easily pass through the channels.

The purpose of the wind filter is identical to that of the windscreen: to reduce wind effects. However, wind filters and windscreens interact with air flows in completely different ways. The wind filter passes air flows with as small a pressure drop as possible, whereas the windscreen damps out the air flows as fast as possible. Since the dimensions of the windscreen must be relatively large to effectively reduce air-blow effects, it is not suitable for use in microphone systems for detecting near whispering. In contrast, the wind filter is very effective for such systems because the flow cross-sectional area is small for air blows from the mouth.

To further reduce the effects of air blows, the microphones are installed facing away from the sound source at locations to the side of the wind filter. Since the microphones face away from the source, acoustic cavities are necessary to guide the sound waves from the source to the microphones. The proposed system uses four microphones placed symmetrically at locations outside the wind filter in order to increase the directional reception to the sound source. The signals received by the four microphones can be processed to improve the directivity. Since the sound source (the mouth) is located near the center of the four microphones, simple addition of the signals from the microphones can achieve a considerable increase in the directivity if the distances between them are large enough. However, due to the small size of the proposed system, simple addition of signals provides little improvement in the directional property of reception. Thus, a new signal-processing method needs to be developed to overcome this problem. In Sec. III C a new signal-processing method is introduced that considerably improves the directional characteristics of reception when the sound source is located near the wind filter in the direction perpendicular to the plane of the microphone system.

III. DESIGN OF EACH COMPONENT

A. Wind filter

The wind filter has two main purposes: (1) to reduce the wind effect to the utmost, and (2) to prevent others near the speaker from overhearing his/her voice by using attenuation and reflection when a sound wave passes through the microchannel. To achieve these purposes, the wind filter is manufactured as an array of microchannels. Microchannel arrays allow wind to pass through the system because of its inertial tendency to move straight forward, whereas they restrict the passage of sound waves out of the system. This allows the sound waves to be delivered into the microphone through the acoustic cavity without interference from wind effects. Therefore, the optimum size of the microchannels in the wind filter must be determined by taking into account the air (fluid) flow and the acoustic wave propagation.[11–14]

The optimal diameter for the microchannels was determined using existing theories for fluids and the propagation of acoustic waves. First, fluid flow was modeled using Hagen–Poiseuille theory.[11]

$$\Delta p = 32\rho V^2 \frac{L}{D\cdot \mathrm{Re}} = 32\rho V \nu \frac{L}{D^2}, \tag{1}$$

where Δp is the pressure difference across the length of the channel, L is the length of the channel, ρ is the air density, V is the air velocity, ν is the kinematic viscosity, Re is the Reynolds number, and D is the diameter of the channel. We set the channel length L to 5 mm, which would be close to the maximum thickness that could be used in a flip-type cellular phone. The air velocity V is assumed to be 1 m/s.

Figure 3(a) shows the change in air pressure as a function of the microchannel diameter. The pressure difference is observed to increase rapidly for diameters less than 100 μm, indicating that the diameter of the microchannel should be over 100 μm to allow the smooth passage of air.

Next, the propagation of acoustic waves in the microchannel was modeled by considering the absorption coefficient that expresses the loss that occurs when sound waves pass through a microchannel, as follows:[12]

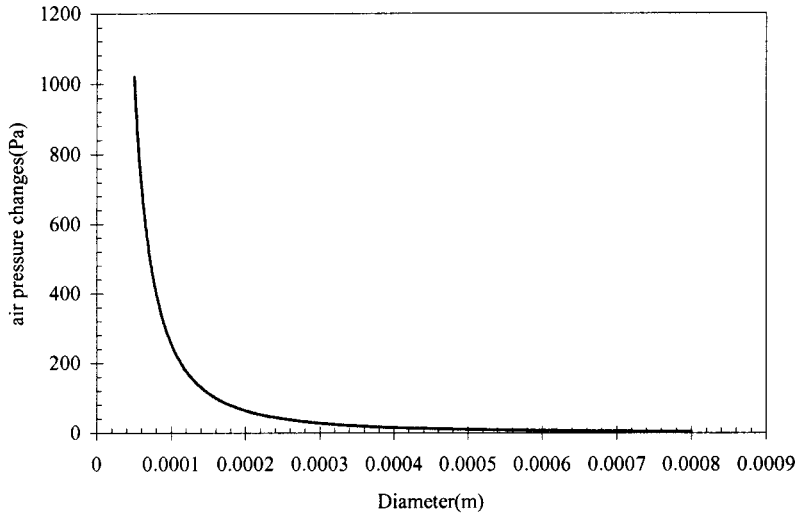
$$\alpha = \frac{2}{Dc}\left(\frac{\eta\omega}{2\rho_0}\right)^{1/2}\left(1+\frac{\gamma-1}{\sqrt{\mathrm{Pr}}}\right), \tag{2}$$

where α is the absorption coefficient, c is the speed of sound, η is the coefficient of shear viscosity, ρ_0 is the equilibrium density, γ is the specific heat ratio, and Pr is a nondimensional parameter, the Prandtl number of the air. When acoustic waves are passing through the microchannel, since the change of sound pressure is dependent on the exponential function of length variables (L) and absorption coefficient (α), it may be expressed as follows:

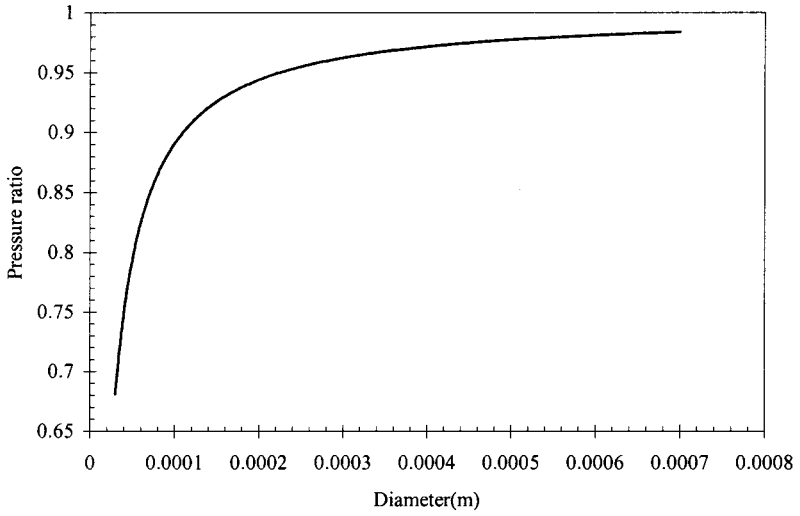
$$p = P_0 \exp(-\alpha L), \tag{3}$$

where P_0 is the equilibrium pressure. The result of this equation has a similar value to ''the low reduced frequency solution'' at the frequency of interest. The low reduced frequency solution refers to the solution of the differential equations for acoustic propagation in a circular channel when the reduced frequency is much less than 1. Zwikker derived the solution by approximating in the low reduced frequency region the full Kirchhoff's equations for acoustic wave propagation in a circular channel.[13] In that frequency region, the result from Eq. (3) is also similar to the numerical solution obtained by Tijdeman using a Newton–Raphson procedure.[14]

Since the microphone system is to be used primarily for receiving human voices, the frequency of 1500 Hz is selected to show the changes in sound-pressure amplitude as a function of microchannel diameter. The ratios of the sound-



(a)



(b)

FIG. 3. Simulations for the design of the microchannels; (a) pressure difference at inlet and outlet of channels of various diameters; (b) ratios of acoustic pressures at inlet and outlet of channels of various diameters.

pressure amplitude at the outlet to that at the inlet are plotted in Fig. 3(b) as a function of microchannel diameter. Similar to the behavior of the air pressure shown in Fig. 3(a), the ratio of the acoustic pressures at the channel inlet and outlet changes rapidly at a microchannel diameter of 100 μm. That is to say, the pressure changes rapidly with changing microchannel diameter for diameters less than 100 μm in both the cases of airflow and sound wave propagation in the microchannel. Optimally, air flow should be smooth through the microchannel, which means the optimum diameter will be in a region showing small pressure change in Fig. 3(a). At the same time, the diameter should ideally be in a region of rapidly changing sound pressure in Fig. 3(b) in order to optimize sound wave propagation through the microchannel. The optimal choice for the microchannel diameter will therefore lie before the rapid rise in Fig. 3(a) and before the rapid descent in Fig. 3(b). Thus, the ideal diameter for the microchannels is about 100 μm. However, there are many technological difficulties that preclude the construction of an array of microchannels of diameter 100 μm and length 5 mm. Therefore, for the functional test we employed an array of microchannels whose diameter is about 500 μm and length is 5 mm. These are the smallest channels that can be fabricated using existing mechanical manufacturing technology.

We anticipate that future innovations in micromachining technology will enable the construction of arrays of microchannels of diameter 100 μm and length 5 mm.

B. Acoustic cavity

The acoustic cavity acts as the waveguide that propagates the sound wave into the microphone. An important consideration in the design of the microphone system is the extent of the transmission loss when an acoustic cavity is used. If the size of the acoustic cavity is too small, the transmission loss becomes too high and as a result the acoustic power to be propagated to the microphone is reduced, which decreases the output of the microphone. Conversely, if the size of the acoustic cavity is large, the probability of air inflow through the cavity is high even though the transmission loss is reduced considerably. Therefore, acoustic theory must be used to decide on an appropriate cavity size.

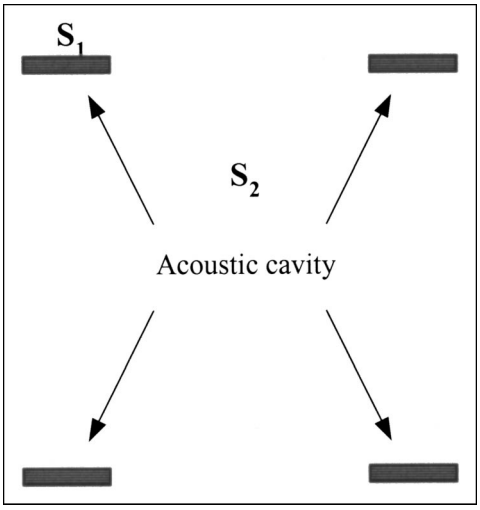


FIG. 4. Locations and areas of acoustic cavities in the system.

Figure 4 is a front view of the acoustic cavity. Here, we employ a method for deciding the optimum cavity size that considers only the effect of the cavity and ignores the effect of the wind filter. The transmission loss, TL, can be expressed as

$$\begin{aligned}\mathrm{TL} &= 10\log\left(\frac{\Pi_i}{\Pi_t}\right)\\ &= 10\log\left(\frac{I_i \Sigma S_i}{\Sigma S_i I_t}\right)\\ &= 10\log\left(\frac{\Sigma S_i}{\Sigma S_i T_{I_i}}\right) = 10\log\left(\frac{S_1+S_2}{S_1 T_{I_1}+S_2 T_{I_2}}\right),\end{aligned} \tag{4}$$

where Π_i, Π_t, S_1, S_2, and I_i are the total power incident on the source side of the partition, the total power transmitted through the partition, the area of the acoustic cavity, the area of the entire system excluding the cavity, and the acoustic intensity, respectively. And, T_{I_1} and T_{I_2} are the intensity transmission coefficient through the acoustic cavities of total area S_1 and the intensity transmission coefficient of the area S_2, respectively. In the calculations for design, the value of 1 is assigned to T_{I_1} and, to obtain the value of T_{I_2}, the transmission loss of a single-leaf partition is used, which is expressed as follows:

$$T_{I_2} = 10^{-2\log(f\cdot\rho_s)+4.7}, \tag{5}$$

where f is the frequency of the incoming acoustic waves and ρ_s is the surface density. After a value is assigned to ρ_s, which reflects the physical properties of the material used to construct the system, the transmission loss in the system can then be expressed as a function of S_1 and f by substituting Eq. (5) into Eq. (4). When the material was assumed to have the qualities of lucite ($\rho_s = \rho_{\mathrm{lucite}} t$, $\rho_{\mathrm{lucite}} = 1200\ \mathrm{kg/m^3}$, t=thickness), we could obtain the transmission loss through the acoustic cavity at various frequencies, as shown in Fig. 5(a), while leaving S_1 as a constant. The results confirm that the transmission loss is almost constant over the frequency range above 1000 Hz. Figure 5(b) shows the variation in transmission loss with changing cavity inlet area S_1, while f was fixed at 1500 Hz. The size of the cavity was set by selecting the size that gave a transmission loss of around 13 dB. This value is somewhat arbitrary, and was selected on the basis of the judgment that signal reduction would be too much when the transmission loss exceeded about 13 dB. After we set a rectangular-shaped cavity of dimensions 5×2.2 mm, i.e., $S_1 = 11\times10^{-6}\ \mathrm{m^2}$ and recalculated the transmission loss, we confirmed that transmission loss was about 13.10 dB.

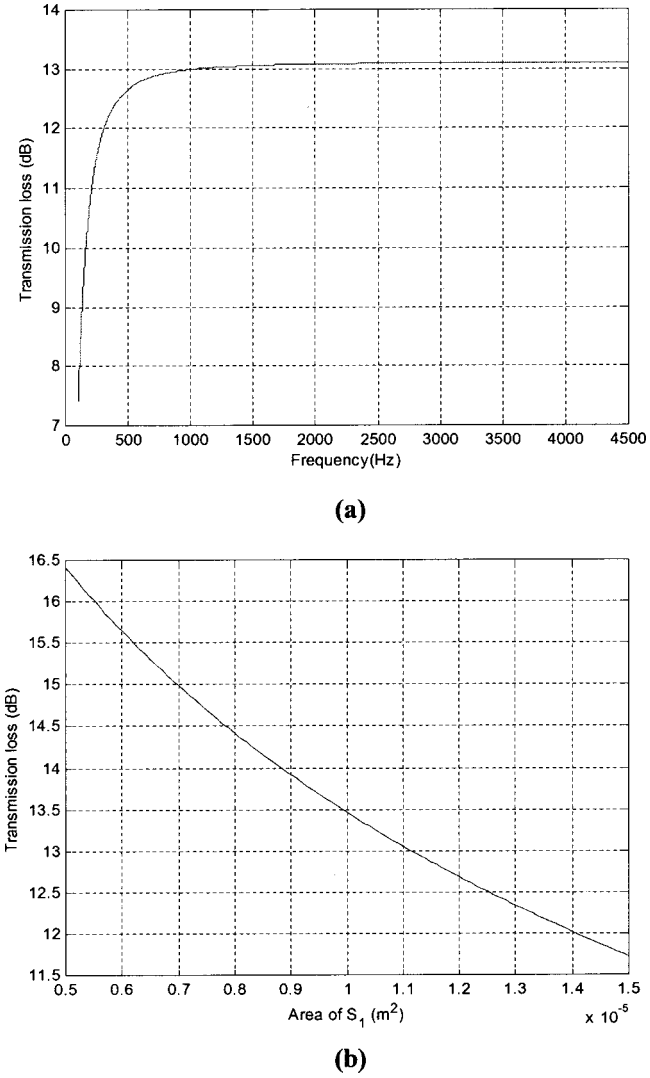


FIG. 5. Simulations for the design of the acoustic cavity; (a) transmission loss through the acoustic cavity at various frequencies; (b) variation in transmission loss with change in cavity inlet area S_1.

C. Directional characteristics

1. Directional characteristics by simple addition of signals

A power amplifier circuit with the configuration shown in Fig. 6 was used to collect the signals from the four microphones. In the proposed system, the distance between the microphones is relatively small in comparison to the wavelength of the sound waves. Hence, the phase differences be-

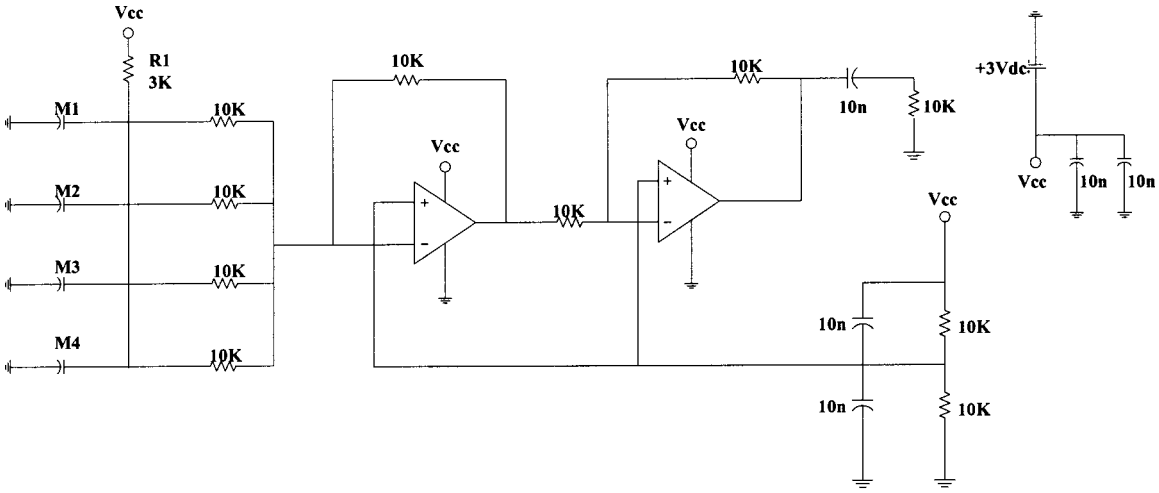


FIG. 6. Power amplifier circuits in the system.

tween sound waves arriving at the microphones installed at different locations are negligible, which means that the focusing effect is negligible when the signals obtained from the microphones are added in order to use the phase differences between the sound waves to focus on the acoustic source. To ascertain the directional characteristics of the system in this case, we used ray theory to simulate the summation of all the signals coming from the microphones. The acoustic source was assumed to be a point source, and the positions of the microphones relative to the sound source were calculated for the configuration depicted in Fig. 7. The simulation result for the directional reception characteristics to a point source at near field is shown in Fig. 8(a). As the angle becomes larger, the output signal of the microphones increases. Hence, the system does not show the focusing effect if the output signals from microphones are simply summed up. The devised system therefore needs a function that emphasizes the output signals from the direction of angle zero because the acoustic source is located on the perpendicular line at the center of system.

2. Directional characteristics by a developed signal processing

To amplify the directivity of the system, we devised a signal-processing principle using the special function, F_s. Let $P_{[i]}$ be the signal obtained at the microphone $[i]$ whose distance from the sound source is R_i, as shown in Fig. 7. The special function F_s is defined as follows:

$$F_s = \sum_{i=1}^{4} f_i \cdot P_{\mathrm{diff}[i]}. \tag{6}$$

In Eq. (6)

$$f_i = \begin{cases} 1, & \text{if } \mathrm{Amp}(P_{[i]}) \geqslant \mathrm{Amp}(P_{[i+1]}) \\ -1, & \text{if } \mathrm{Amp}(P_{[i]}) \leqslant \mathrm{Amp}(P_{[i+1]}) \end{cases}$$

$$P_{\mathrm{diff}[i]} = P_{[i]} - P_{[i+1]},$$

where $\mathrm{Amp}(P_{[i]})$ implies the amplitude of $P_{[i]}$ and $[1]$ should be assigned to $[i+1]$ if $i+1>4$. If P_{add} is defined as

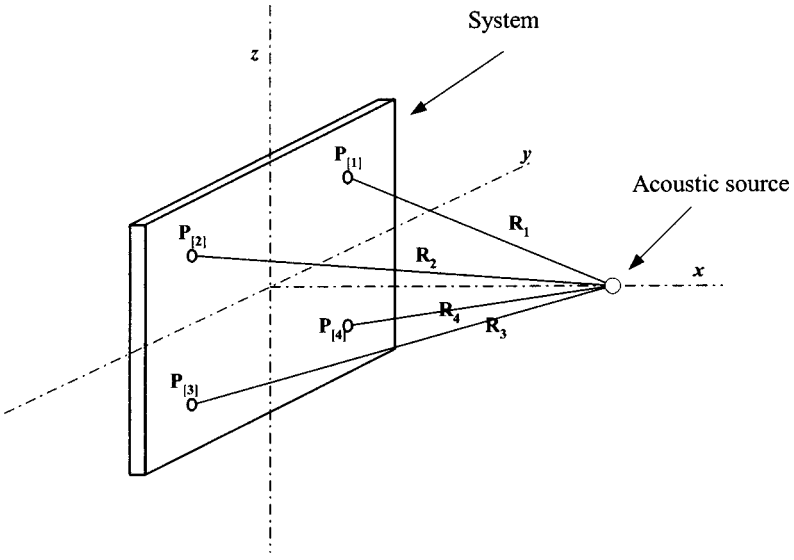


FIG. 7. Relative positions of the microphones and a sound source.

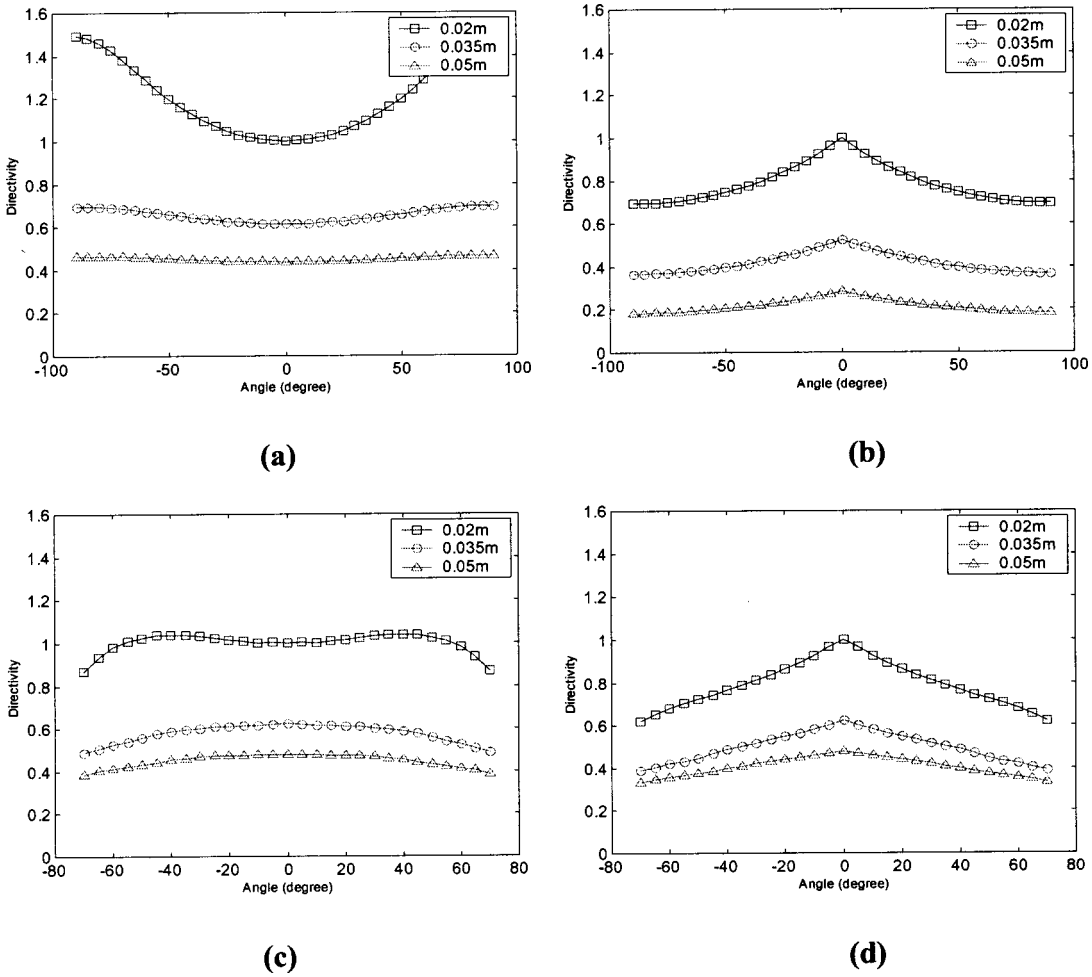


FIG. 8. Simulation and experimental results for the directional reception characteristics: (a) Directional reception characteristics to a point source at near field with simple addition of signals (simulation); (b) directional reception characteristics to a point source at near field with signal processing (simulation); (c) directional reception characteristics to a point source at near field with simple addition of measured signals (experimental measurement); (d) directional reception characteristics to a point source at near field with signal processing (experimental measurement).

$$P_{\text{add}} = P_{[1]} + P_{[2]} + P_{[3]} + P_{[4]}, \tag{7}$$

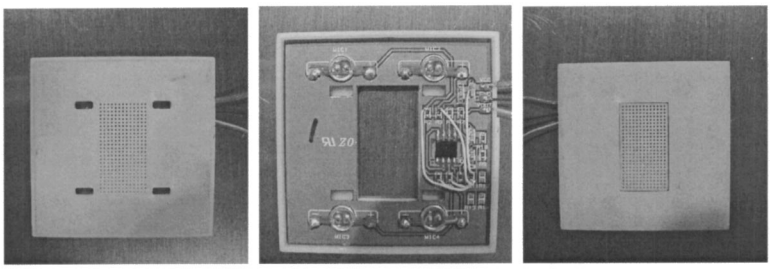
then the output signal, P_{out}, is defined as follows:

$$P_{\text{out}} = P_{\text{add}} - F_s. \tag{8}$$

The reason why the signal-processing principle described above can improve the directivity of the system can be explained as follows. If a sound point source is located near the microphone system as shown in Fig. 7, the amplitude of signal $P_{[i]}$, measured at the microphone $[i]$, decreases as the distance from the point source R_i increases. The amplitude of $P_{[i]}$ is larger than that of $P_{[j]}$ when $R_i < R_j$, while $P_{[i]}$ should be almost identical to $P_{[j]}$ when $R_i = R_j$. From the definition of F_s, the amplitude of it should have its minimum when $P_{[1]} = P_{[2]} = P_{[3]} = P_{[4]}$, which occurs if the point source lies on the x axis. Hence, in that case, the amplitude of P_{out} becomes its maximum. If the phase difference between $P_{[i]}$ and $P_{[j]}$ is sufficiently small (i.e., the wavelength is much larger than the maximum distance between the microphones), the amplitude of F_s increases while its shape is preserved, as the point source moves off from the x axis. By the definition of P_{out}, its amplitude decreases in turn. This implies that the directivity of the microphone system increases by the signal processing defined by Eqs. (6) and (8).

If the point source is located far away from the microphone system, R_i and R_j will not become considerably different even when the point source moves off from the x axis. This implies that the signal processing will not increase the directivity of the microphone system considerably. Summarizing the properties of the proposed signal processing, the directivity can be improved for a close point source. The effects on the directivity decrease as the source moves away. And, the signal processing works well for the low-frequency sounds.

The signal-processing principle described above is easy to apply for receiving a single tone in real time, but difficult to apply directly for the real-time reception of other sounds. An approximate scheme was developed to implement the developed signal-processing principle into real-time reception of various sounds, which can be implemented by use of the circuits. Since in the approximate scheme the value of f_i in Eq. (6) can be determined only with the instant values of $P_{[i]}$'s, f_i should be well defined such that $P_{\text{diff}[i]}$ should have the same sign of P_{add}



(a) Front (b) Interior (c) Backside

FIG. 9. The fabricated prototype.

$$f_i = \begin{cases} 1, & \text{sign}(P_{\text{add}}) = \text{sign}(P_{\text{diff}[i]}) \\ -1, & \text{sign}(P_{\text{add}}) \neq \text{sign}(P_{\text{diff}[i]}) \end{cases}. \quad (9)$$

In this definition, sign(P) implies the sign of P and [1] should be assigned to $[i+1]$ if $i+1>4$. Then, Eq. (8) is used again for P_{out}.

The signal-processing scheme with Eqs. (6), (8), and (9) can be applied to reception of various sounds in real time to improve the directivity of the system. But, it can bring about signal distortion to high-frequency sounds because the principle for the proposed signal-processing technique assumes that the phase difference due to the traveling distance is sufficiently small. This is the reason why the operating frequency band is set from 500 Hz to 4 kHz. Most directional microphone systems show higher directivity as the frequency of the signal becomes higher.

Figure 8(b) shows the simulation results obtained using the proposed signal-processing technique. As expected, the greatest sensitivity was found at 0° as a result of the signal processing. In addition, we found that the improvement in sensitivity becomes greater as the acoustic source comes closer to the microphones.

D. Fabrication of a prototype

In the preceding section, we verified through a numerical simulation that the microphone system with the proposed signal processing can receive sounds focused on a near-acoustic source located on the perpendicular line at the center of system. However, this result was obtained using an ideal model that assumes a point source. Therefore, we needed to experimentally verify the model results. Furthermore, we needed to test the reduction in air-blow effects achieved by the proposed system, which cannot be proved by numerical simulation. Hence, we constructed a prototype of the proposed system to test its efficiency. Figure 9 shows the fabricated prototype. The size of the entire system was 50×50 mm with a margin of 10 mm on all sides, which was used to attach the system to the supporting structures that fixed the system during performance testing. The size of the microchannels was 5 mm in length and 500 μm in diameter. The distance between microchannel centers was 1 mm and the dimension of the acoustic cavity was 5.5×2 mm. Figure 9 shows the power amplifying circuit and the microphones connected to it.

IV. PERFORMANCE TESTING AND DISCUSSION

Two experiments were conducted to evaluate the performance of the prototype system. First, the sensitivity of the system to its orientation with respect to the sound source was probed to assess the directivity of the system. Second, an experiment was carried out to evaluate the effect of air blows in front of the mouth. The details of these experiments are outlined below.

A. Reception directivity tests

Figure 10 shows a schematic diagram of the experimental setup used to measure directivity. The purpose of this experiment was to determine the angles at which the system operated efficiently and to evaluate the degree of improvement in sensitivity as the point source is moved closer to the microphone system through the vertical direction (0°). The method used was to measure the change in the signal as the system was rotated from 0° to 90° while maintaining a specific distance between the system and the acoustic source (see Fig. 10). Figure 8(c) shows the directional reception characteristics according to the angle of the system acquired from the results calculated by summing up the signals coming from the 1.5-kHz frequency. We can easily see that the system does not have the largest value of the directivity at 0°, which was also the case in the simulation result shown in Fig. 8(a). Comparing the experimental results in Fig. 8(c) with the simulation results in Fig. 8(a), we find some differences. These differences derive from the disparity between the ideal point source assumed in the simulations and the experimental acoustic source, which was made by blocking the front portion of a loudspeaker with a disk with a hole in the center of it (see Fig. 10). Unlike the ideal point source, which sends sound waves in all directions as it is located farther from the acoustic source, the acoustic source used in the experiments might have directivity. The decrease observed in the experimental output signals when the angle exceeds 50° and the distance is 2 cm occurs because the system blocks the acoustic source to some degree. This blocking arises because the radius of the system is 1.5 cm, which is close to the distance (2 cm) between the system and the acoustic source.

The directivity experiments were repeated adopting the signal-processing regime given in Eqs. (6)–(9) with the largest directivity at 0°. The result is shown in Fig. 8(d). Again, since the experimental source was not a perfect point source, we can see a few differences between the experimental result

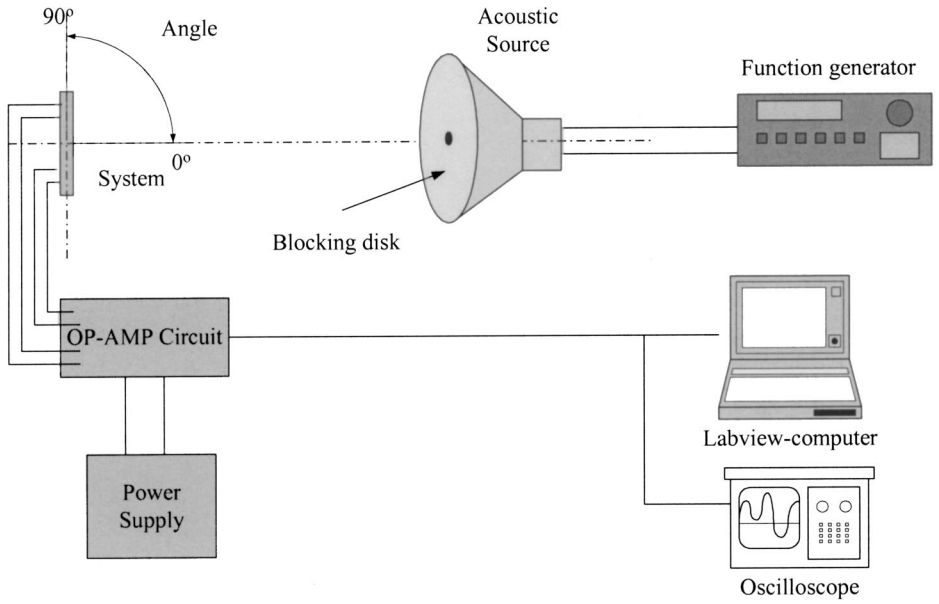


FIG. 10. Schematic diagram of experimental setup for measuring directivity.

and the simulation result shown in Fig. 8(b). However, the processed experimental signal shows that the largest directivity can be acquired at 0°. Moreover, high directivity could be obtained at 0° in the frequency range of 500–4000 Hz, the frequency in which we are interested. The experimental data also confirm that the sensitivity changes rapidly with the distance from the source.

B. Test on air-blow effects

Experiments using a wind source were carried out to confirm the proper functioning of the wind filter. The function of the wind filter is to reduce the air-blow effect in front of the mouth and thereby convey a pure sound wave to the microphone. Figure 11 shows a schematic diagram of the experimental setup used to measure air-blow effects. An air compressor was used to generate the air flow, and a speaker was used as an acoustic source. The acoustic source continuously produced sound waves of constant frequency, while the air blow was generated by the compressor. Since the larger velocity of air flow affects the acoustic signals from the microphones to a greater extent, the average air-flow velocity used in the experiments was greater than the value assumed in the simulations. The average velocity of the air flows in the experiments was about 4.6 m/s as measured with a flow meter. A frequency of 1 kHz was chosen for the acoustic source. The ability of the microchannel array to reduce the wind effect was tested for two cases: when the microchannel array was closed and when it was open. These experiments measured the signals coming from each microphone while air was continuously blown by the compressor.

Figure 12 shows the output voltage signals from the microphone system in the time domain when the microchannels in the wind filter are open or closed. In Fig. 12(a), the solid line represents the signals from the microphones measured when the microchannels in the wind filter are open, and the dashed line represents the signals when the microchannels are closed. The air-blow source is located 0.3 cm away from

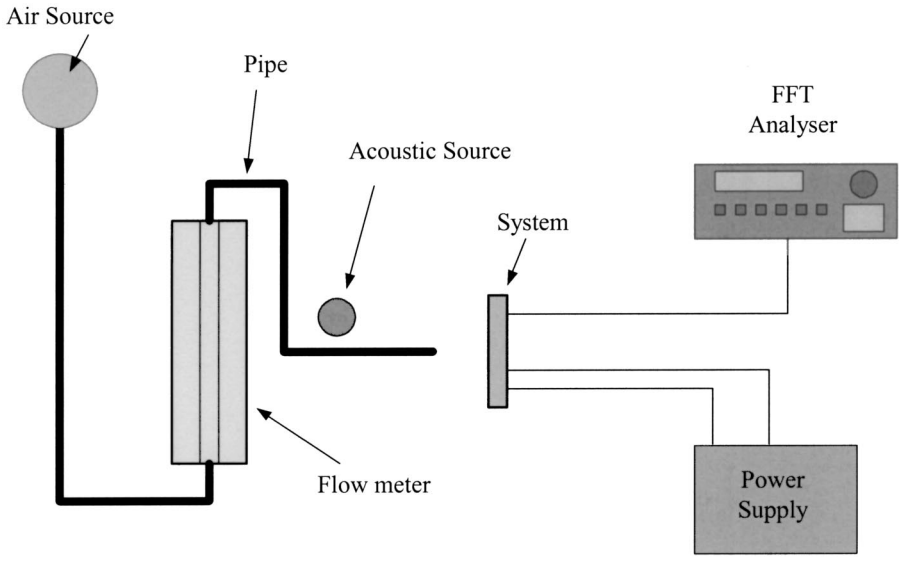
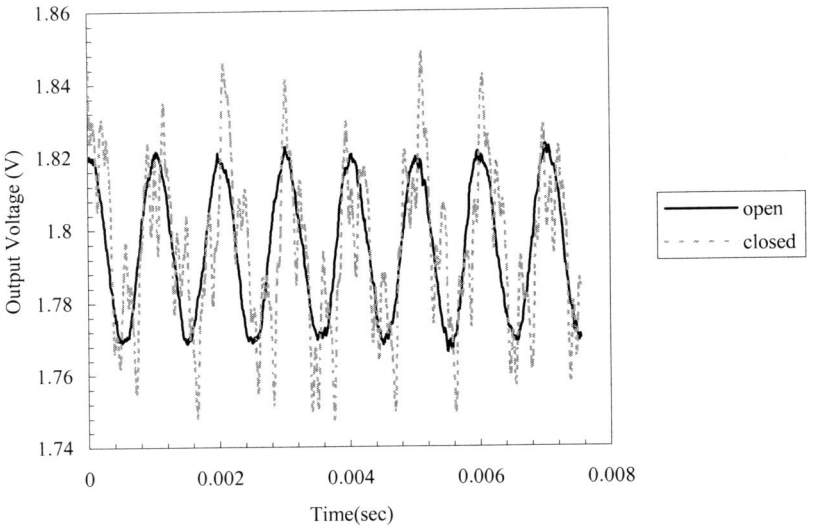
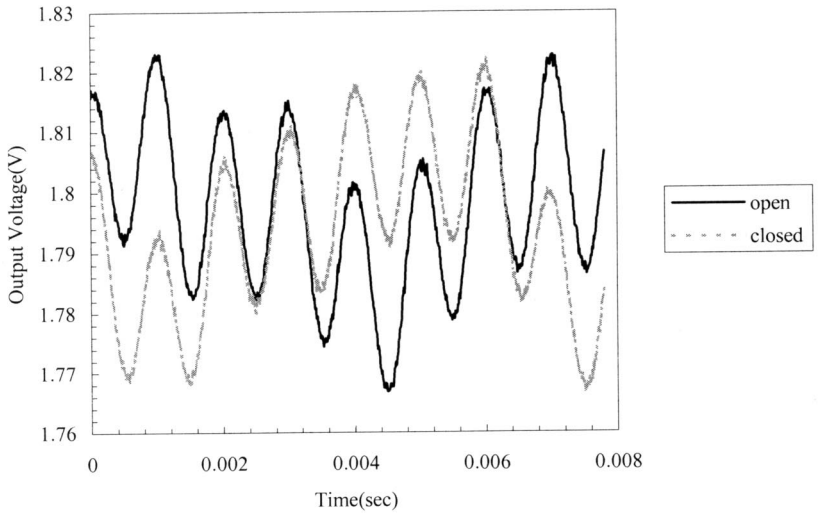


FIG. 11. Schematic diagram of experimental setup for measuring air-blow effects.



(a)



(b)

FIG. 12. Comparisons of the signals measured with microchannels open or closed at two locations of the air blow source; (a) source distance: 0.3 cm; (b) source distance: 6 cm.

the array in the perpendicular direction to the surface of wind filter, as illustrated in Fig. 11. Figure 12(b) shows the same things as Fig. 12(a) except the distance between the air-blow source and the wind filter surface, which is 6 cm in this case. From Figs. 12(a) and (b), it can be easily noticed that SNR (signal-to-noise ratio) is improved considerably when the microchannels in the wind filter are open. And, improvement of SNR is much greater when the source is located closer to the wind filter than when the source is a distance away. This implies that the wind filter shows better performance reducing the disturbances caused by the air blows from the mouth when the air-blow source (the mouth) is located close to the wind filter. Quantitatively, the SNR of the signals measured with open and closed microchannels with the air source at 0.3 cm [Fig. 12(a)] are 24.5 and 3.45 dB, respectively, confirming that the wind filter greatly reduces wind effects. The newly devised microphone system is intended for use close to the acoustic source; therefore, the microchannel array wind filter successfully fulfills the function of removing air blow.

V. CONCLUSIONS

A new microphone system was developed to receive human whispering. The system includes two special functions in addition to the usual functions of a microphone: (1) reduction of the effects of air blows from the mouth during speaking, and (2) improved the directivity in the direction to the source location. The addition of these special functions produced a microphone system with good performance, which should allow whispering communication on mobile phones in noisy places with privacy. The technical achievements in the development process are as follows:

(1) A new device referred to as the wind filter was devised to reduce the effects of air blows from the mouth. It was designed based on the existing theories on fluid flows and acoustic wave propagations in a small circular channel. The wind filter was demonstrated to significantly reduce air-blow effects in a prototype microphone system.
(2) A new signal-processing method was developed to im-

prove the directional reception characteristics from a close sound source by use of four microphones. The performance of the proposed signal-processing method was tested theoretically by numerical simulation and experimentally using a prototype microphone system. The proposed method improves the directivity considerably. The reception sensitivity of the microphone system was found to decrease rapidly with increasing distance from the sound source. This property improved the source focusing performance of the microphone system.

(3) The design process also involved the development of the acoustic cavity in the system. Experimental tests showed that the acoustic cavity fulfilled a suitable function for the system.

ACKNOWLEDGMENTS

The authors wish to thank Mr. S. Shin and Mr. Y. Ko in iCurie Lab. for their help in the fabrication of the prototype. This work was supported by iCurie Lab. Corp.

[1] V. C. Anderson, ''The first twenty years of acoustic signal processing,'' J. Acoust. Soc. Am. **51**, 1062–1065 (1971).

[2] O. M. Mracek Mitchell, C. A. Ross, and G. H. Yates, ''Signal processing for a cocktail party effect,'' J. Acoust. Soc. Am. **50**, 656–660 (1971).

[3] J. L. Flanagan and R. C. Lummis, ''Signal processing to reduce multipath distortion in small rooms,'' J. Acoust. Soc. Am. **47**, 1475–1481 (1970).

[4] J. L. Flanagan, J. D. Johnston, R. Zahn, and G. W. Elko, ''Computer-steered microphone arrays for sound transduction in large rooms,'' J. Acoust. Soc. Am. **78**, 1508–1518 (1985).

[5] J. L. Flanagan and R. A. Kubli, ''Conference microphone with adjustable directivity,'' J. Acoust. Soc. Am. **77**, 1946–1949 (1975).

[6] R. Zelinski, ''Noise reduction based on microphone array with LMS adaptive post-filtering,'' Electron. Lett. **26**, No. 24, 2036–2037 (1990).

[7] H. Teutsch and G. W. Elko, ''An adaptive close-talking microphone array,'' IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York, October 2001, pp. 21–24.

[8] P. Das and C. Lanzl, ''A surface acoustic wave sonobuoy,'' J. Acoust. Soc. Am. **66**(3), 636–640 (1979).

[9] R. D. Stoll, ''Attenuation measurement with sonobuoys,'' J. Acoust. Soc. Am. **73**(1), 163–172 (1983).

[10] Website: http://www.mod.uk/dpa/projects/nimrod/what.htm

[11] B. R. Munson and D. F. Young, *Fundamentals of Fluid Mechanics*, 3rd ed. (Wiley, New York, 1994), pp. 367–372.

[12] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 4th ed. (Wiley, New York, 2000), pp. 210–245.

[13] C. Zwikker and C. W. Kosten, *Sound Absorbing Materials* (Elsevier, Amsterdam, 1949), Chap. 2.

[14] H. Tijdeman, ''On the propagation of sound waves in cylindrical tubes,'' J. Sound Vib. **39**(1), 1–33 (1975).

Surface acoustic wave diffraction in spectral theory of interdigital transducers

Eugene J. Danicki[a)]
Polish Academy of Sciences, IPPT, 00-049 Warsaw, Poland

(Received 10 November 2002; revised 2 April 2003; accepted 4 April 2003)

A theory of interdigital transducer is presented that accounts for the surface acoustic wave diffraction. It is formulated with help of the Blotekjær, Ingebrigtsen, and Skeie expansion method used earlier in the plane-wave theory. Now, the electric field is applied to the breaks in the strips making parts of them polarized with different potentials. This way the finite aperture width transducers are modeled residing within an infinite system of periodic strips. Closed expressions are derived for system working below the Bragg condition. The theory is open for further development accounting for elastic strips, pseudo surface, leaky or even surface skimming bulk waves, for any frequency range, including Bragg reflection exploited in surface wave resonators. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1579010]

PACS numbers: 43.38.Rh, 43.20.Gp, 43.35.Pt [DEC]

I. INTRODUCTION

In surface wave (SAW) devices, the wave is generated by a finite aperture width transducer (usually, an interdigital transducer, IDT). The generated wave beam diffracts and nonuniformly insonifies the output IDT. This is why the resulting frequency response of a SAW filter departs from its predicted shape by the plane-wave theory of IDTs.

To overcome this difficulty, authors developed numerous *ad-hoc* models based on physical interpretation how the diffraction may affect the transducer generation and detection properties and the transfer function of the filter.[1] Usually, the angular spectrum theory is applied to model the SAW propagation between the generating and receiving transducer fingers. The theory, developed for bulk waves,[2] accounts for the media anisotropy and the velocity variation over the propagation angle; this is also the SAW case. The applied bulk to surface wave analogy is however, left without deeper discussion.

Another model uses the Huygen's principle,[3] again *ad-hoc* extended to the surface anisotropic case by applying the appropriate square root dependence on the propagation distance and variable wave velocity and piezoelectric coupling dependent on the propagation angle. In spite of the *ad-hoc* nature of these diffraction models, authors obtained remarkable results in designing the diffraction compensated filters. In all these models, the SAW diffraction and its generation and detection are treated separately giving no way to formulate a general unified theory. The unified model of IDTs presented here and accounting for diffraction can serve primarily for deeper theoretical understanding of the former models. Next, its final results may be incorporated in the earlier developed software for evaluation of the interaction of two transducer fingers, replacing the *ad-hoc* expressions based on either the angular spectrum or Huygen's theories. Finally, it makes certain progress in the theory of SAW devices.

a)Electronic mail: edanicki@ippt.gov.pl

The proposed is an extension of the earlier developed plane-wave theory of IDTs.[4] To model the finite aperture width transducer, the considered infinite periodic strips (the transducer fingers) may now have variable applied electric potentials along strips, taking at least two values: there is or is not the applied potential that is the source of the surface wave at the transducer fingers. This is depicted in Fig. 1, where both parts of the shown broken strip may have different electric potentials. Analogously for the receiving finger: only part of it contributes to the transducer output current.

Naturally, this generalization requires substantial reformulation of the theory. The simplest possible system is considered here: the periodic strips have rather small period to avoid Bragg scattering of SAWs, and the transducer fingers are ideal, without elastic properties and massless. Both simplifications can be easily removed on the way presented in the plane-wave theory.[4]

The paper starts with the piezoelectric substrate description by the corresponding planar harmonic Green functions, then formulation of the boundary conditions for the properly chosen wave field components is presented; they are rather nonstandard for conducting strips. Next, the solution is constructed using the method introduced by Blotekjær, Ingebrigtsen, and Skeie,[5] referred to as the BIS expansion method (see Appendix A). Finally, the transfer function between strips is evaluated and interpreted. For this reason, Appendix B presents shortly the plane-wave theory in the same notations for the readers' convenience.

II. A SAW PLANAR HARMONIC GREEN FUNCTION

The planar harmonic Green function for piezoelectric substrates supporting surface acoustic waves is rather a complicated function of a spectral variable—the wave number k for given angular frequency ω. There is however, its useful approximation accounting for only the surface waves and neglecting any bulk waves in the media. Moreover, neglecting the mechanical interaction between the piezoelectric half-space and the ideal conducting strips residing on it, the planar harmonic Green function of interest becomes[6]

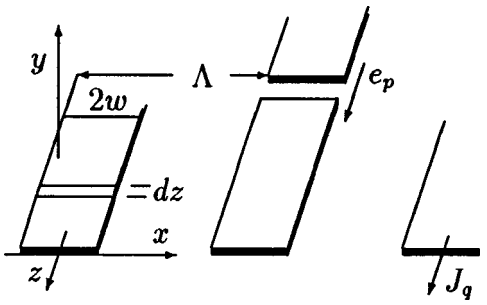


FIG. 1. Modeling of a finite aperture width transducer by periodic strips; note a break in one strip.

$$G=\Delta D_y/\phi=k\epsilon_e\frac{k^2-k_v^2}{k^2-k_o^2},\quad k=\sqrt{r^2+s^2}. \tag{1}$$

This scalar function describes the relation between the surface electric charge distribution equal to the jump of the normal electric induction ΔD_y in the substrate and vacuum above it, and electric potential ϕ, both being harmonic functions of the form $\exp(j\omega t-jrx-jsz)$ in the plane of strips $y=0$; r, s are wave numbers.

In this approximated planar harmonic Green function, ϵ_e is the equivalent surface electric permittivity, k_o is the SAW wave number for short-circuited substrate (that is at condition $\phi=0$ on its surface), and k_v is the SAW wave number on the isolated substrate surface (at $\Delta D_y=0$, that is without surface electric charge). All these parameters may depend on $\vartheta=\mathrm{atan}(s/r)$ due to the anisotropy of piezoelectrics, and can be evaluated by numerical methods,[6] taking into account what follows.

Applying rotation to the coordinate system by an angle ϑ around y axis, it is easy to transform the above spatial harmonic wave form into $\exp(j\omega t-kx')$, representing the plane wave on the surface $y=0$ (Fig. 1) that is analyzed in the discussed paper.[6] The approximation has the form

$$D_y/\phi=-k\epsilon_e(\sqrt{k^2-k_s^2}-\beta|k|)/(\sqrt{k^2-k_s^2}-\alpha|k|),$$

accounting for that $E_1=jk\phi$ and $D_3=-D_y$ in current notation. The parameters α, β, k_s are evaluated numerically. In the case of weak piezoelectric substrates, the values of α and β are very close, thus multiplying the right-hand side by

$$(\sqrt{k^2-k_s^2}+\beta|k|)/(\sqrt{k^2-k_s^2}+\alpha|k|)\approx 1,$$

one easily obtains Eq. (1) with $k_{o,v}$ resulting from k_s, α, β (note that D_y of vacuum is $k\epsilon_o\phi$).

III. SAW EXCITATION BY STRIPS

In the theory of interdigital transducers,[4] alternate voltages are applied to the system of strips which, due to the substrate piezoelectricity, excite SAWs of corresponding frequency. In simplest cases, the strip period equals half wavelength. The excitation is most efficient if the spatial period of the applied potential by means of strips equals the SAW wavelength at a given frequency. The assumed infinitely long strips excite the plane SAW; the theory is, in fact, one dimensional for the wave field in the plane of strips placed on the substrate surface (see Appendix B for details).

In the analysis presented below, we exploit the concept of composite strips depicted in Fig. 1, in which each strip is actually a stack of imaginary dz-wide ''substrips'' with possible electric field distribution between them (that is along the original strip). This approach, making the analysis two dimensional in the plane of strips, will enable us to apply the variable potential on strips. The corresponding boundary conditions for the electric field in the plane of strips are (note that E_z may vary with z)

$$\begin{aligned} &E_x=0,\ \partial_x E_z=0 \ \text{ on strips},\\ &\Delta D_y=0 \ \text{ between strips}, \end{aligned} \tag{2}$$

with the standard notation of electric field $E_i=-\partial\phi/\partial x_i$ $(x_i=x,z)$. Strips have period $\Lambda=2\pi/K$ and width $2w$. The field components involved in Eq. (2) and representing the wave field in the boundary-value problem under consideration are E_x, $\partial_x E_z$, and the surface charge distribution ΔD_y. They are all square-root singular function at the strip edges, appropriate for the BIS expansion.

Now, the condition for E_z need to be specified. Here, $E_z=0$ on the whole perfectly conducting strip except the break in it, Fig. 1. The break is infinitesimal, thus the condition of interest sounds

$$E_z(x=n\Lambda,z)=e_n\delta(z-z_n), \tag{3}$$

where e_n is the given amplitude of the Dirac δ-like electric field in the infinitesimally narrow break in the nth strip at $z=z_n$. Its z-Fourier transform is

$$e_n(s)=\int_{-\infty}^{\infty}e_n(z)e^{jsz}\,dz=e_n e^{jsz_n}. \tag{4}$$

Naturally, specifying $E_z(x=n\Lambda,z)$ over nth strip sets the potential difference between nth strip and its neighbors; for the δ-like field, the strip part $(-\infty,z_n)$ has different potential with respect to the part (z_n,∞) and with respect to other strips. Thus specifying e_n sets the corresponding E_x field between strips. Detailed discussion of this subject is presented later.

IV. BLOCH AND BIS EXPANSIONS

The system of strips under consideration is periodic, thus the appropriate field representation is the Bloch expansion

$$\begin{aligned} &\phi=\sum_n\frac{E^{(n)}}{k_n}e^{-jr_nx-jsz},\quad k_n=\sqrt{r_n^2+s^2},\\ &[E_x,E_z]=\sum_n\left[\frac{jr_n}{k_n},\frac{js}{k_n}\right]E^{(n)}e^{-jr_nx-jsz},\\ &\partial_x E_z=-jsE_x, \end{aligned} \tag{5}$$

where $r_n=r+nK$ is the wave number of nth Bloch component. The domain $0<r<K$ is chosen to avoid ambiguity of the field representation.

The wave field will satisfy the boundary conditions (2) if its Bloch components have the following BIS expansions ($\Delta=Kw$, see Appendix A for details)

$$\Delta D_y^{(n)} = \alpha_m P_{n-m}(\cos\Delta) e^{-jr_n x},$$

$$E_x^{(n)} = \frac{jr_n}{k_n} E^{(n)} = \alpha'_m S_{n-m} P_{n-m}(\cos\Delta) e^{-jr_n x} e^{-jr_n x}, \quad (6)$$

$$\partial_x E_z^{(n)} = -js \frac{jr_n}{k_n} E^{(n)} e^{-jr_n x}$$

$$= \beta_m S_{n-m} P_{n-m}(\cos\Delta) e^{-jr_n x},$$

with summation symbol over m dropped, as well as $\exp(-jsz)$. It is assumed below that the wave field amplitudes depend on both wave numbers r and s—the spectral variables of Fourier transforms over x and z, respectively. The corresponding x-harmonic dependence, being in fact the term involved in the inverse Fourier transform, is included explicitly in the above expressions for convenience of the analysis that follows: the notation follows that of earlier papers.[4,5] Note that $\partial_x E_z = \partial_z E_x$ thus

$$\alpha'_m = \frac{j}{s}\beta_m,$$

$$E^{(n)} = \frac{k_n}{sr_n}\beta_m S_{n-m} P_{n-m}(c), \quad (7)$$

$$\alpha_m P_{n-m}(\cos\Delta) = \epsilon_e \frac{k_n^2 - k_v^2}{k_n^2 - k_o^2} \frac{k_n}{sr_n} \beta_m S_{n-m} P_{n-m}(c)$$

($c = \cos\Delta$). The last equation (actually the system of equations, $-\infty < n < \infty$) results from the planar harmonic Green function (1) applied to the nth Bloch components of the tangential electric field $E^{(n)}$ and the surface electric charge $\Delta D_y^{(n)}$.

There is a doubly infinite set of equations (for negative and positive n) in which we can apply the approximation

$$\epsilon_e \frac{k_n^2 - k_v^2}{k_n^2 - k_o^2} \frac{k_n}{sr_n} = \frac{\epsilon_e}{s} S_n \quad \text{if } |n| > N, \quad (8)$$

where N is a certain large number. To satisfy all these equations with the assumed finite number of the BIS expansion coefficients α_m, β_m, one must apply that[5]

$$\beta_m = \frac{s}{\epsilon^o}\alpha_m, \quad (9)$$

where $\epsilon^o = \epsilon_e(\theta_n \to 0)$, $\theta_n = \text{atan}(s/r_n)$, $|n| \to \infty$.

This finally transforms the last of Eqs. (7) into

$$\alpha_m \left[1 - \overline{\epsilon_e}(\theta_n) \frac{k_n^2 - k_v^2(\theta_n)}{k_n^2 - k_o^2(\theta_n)} \frac{k_n}{r_n} S_{n-m} \right] P_{n-m}(\cos\Delta) = 0 \quad (10)$$

(summation symbol over m dropped, $-N \leq m \leq N+1$), where $\overline{\epsilon_e} = \epsilon_e/\epsilon^o$. There are less equations, by 1, than unknowns α_m, thus the above equations, taken for $-N \leq n \leq N$, allow us to evaluate only the relative coefficients $\overline{\alpha}_m = \alpha_m/\alpha_0$. For other n, Eq. (10) is satisfied automatically due to the approximation (8).

V. TRANSFER FUNCTION OF STRIPS

Integrating $\partial_x E_z$ over x, one obtains from (6),

$$E_z(r,s) = \sum_{m,n} j\frac{s}{\epsilon^o}\frac{\alpha_m}{r+nK} S_{n-m} P_{n-m}(c) e^{-j(r+nK)x}\big|_{x=l\Lambda}$$

$$= \frac{js\alpha_o(r,s)}{\epsilon^o K \sin\pi r/K} (-1)^m \overline{\alpha}_m P_{-m-r/K}(-c) e^{-jrl\Lambda} \quad (11)$$

($c = \cos\Delta$), evaluated with the help of Eq. (A2). The inverse x-Fourier transform for E_z is the integral

$$E_z(l\Lambda;s) = \frac{1}{K}\int_0^K \frac{js\pi\alpha_o(r,s)}{\epsilon^o K \sin\pi r/K} (-1)^m \overline{\alpha}_m \times P_{-m-r/K}(-\cos\Delta) e^{-jrl\Lambda}\, dr, \quad (12)$$

that should result in the given $e_l(s)$. This requires that

$$\alpha_0(r,s) = \frac{-j\epsilon^o K e_n(s) e^{jrn\Lambda} \sin\pi r/K}{s\pi\Sigma_m (-1)^m \overline{\alpha}_m P_{-m-r/K}(-\cos\Delta)} \quad (13)$$

(summation symbol over m included to avoid misinterpretation). Indeed, applying it in Eq. (12) results in $E_z(l\Lambda;s) = e_n(s)\delta_{nl}$ because $\int_0^K \exp(-jlr\Lambda) dr/K = \delta_{0l}$, where δ_{nl} is the Kroenecker delta.

The electric charge of kth strip (k is an integer number here) results from the integration of ΔD_y over x in the kth strip domain ($k\Lambda - w/2$, $k\Lambda + w/2$) that can be extended into ($k\Lambda - \Lambda/2$, $k\Lambda + \Lambda/2$) on the strength of Eqs. (2). Using again Eq. (A2), one obtains

$$\int_{k\Lambda-\Lambda/2}^{k\Lambda+\Lambda/2} \Delta D_y\, dx = \sum_{m,n} \alpha_m P_{n-m}(\cos\Delta) \times [e^{-jr_n(k\Lambda+\Lambda/2)} - e^{-jr_n(k\Lambda-\Lambda/2)}]/(-jr_n)$$

$$= Q_k(r,s)$$

$$= \alpha_0(r,s)\Lambda\overline{\alpha}_m P_{-m-r/K}(\cos\Delta) e^{-jrk\Lambda}. \quad (14)$$

The total current $\overline{J_k}$ flowing along the kth strip can be evaluated by applying the charge conservation law

$$-\frac{\partial}{\partial z}\overline{J_k} = j\omega Q_k, \quad \text{yielding } \overline{J_k}(r,s) = \frac{\omega}{s} Q_k(r,s). \quad (15)$$

Accounting for the result (13), one obtains

$$\overline{J_k}(r,s) = -j\omega 2\epsilon^o \frac{e_n(s)}{s^2} R(r,s) e^{-jr(k-n)\Lambda},$$

$$R(r,s) = \frac{\Sigma_m \alpha_m P_{-m-r/K}(\cos\Delta)}{\Sigma_m \alpha_m (-1)^m P_{-m-r/K}(-\cos\Delta)}. \quad (16)$$

Here, the relative $\overline{\alpha}_m$ has been replaced by α_m.

The inverse Fourier transform of $\overline{J_k}$ yields the spatial transfer function of strips (the strip transadmittance)

$$\overline{J_k}(z) = -j\omega 2\epsilon^o \int_{-\infty}^{\infty} s^{-2} e_n(s) e^{-jsz} \frac{ds}{2\pi} \times \int_0^K R(r,s) e^{-jr(k-n)\Lambda} \sin\pi r/K \frac{dr}{K}. \quad (17)$$

This is the current of kth strip resulting from the applied electric field to the nth strip. It is caused by the strip mutual capacitance and by the excited SAW that propagates along the substrate surface between strips. This latter component, J_k, is discussed further in the next section assuming that $e_n(z)$ is a δ-like function (3).

For the convenience of physical interpretation of the resulting wave component of the strip transadmittance, the applied excitation of strips is generalized here with respect to Eq. (3) by introducing the following potential distribution along the strip:

$$\phi(n\Lambda,z)=V_n\begin{cases}\exp[-a(z-z_n)], & z>z_n,\\ -\exp[a(z-z_n)], & z<z_n.\end{cases} \tag{18}$$

This corresponds to the δ-like electric field $E_z=-\partial_z\phi$ at $z=z_n$, but the potential slowly vanishes when $|z-z_n|\to\infty$ (a is assumed small). The z-Fourier transform of interest is

$$e_n(z)=-V_n\frac{2s^2}{s^2+a^2}e^{jsz_n}, \tag{19}$$

which substituted into (17) yields

$$\overline{J_k}=j\omega 4\epsilon^o V_n\int_{-\infty}^{\infty}\frac{e^{-js(z-z_n)}}{s^2+a^2}\frac{ds}{2\pi}\times\int_0^K R(r,s)e^{-jr(k-n)\Lambda}\sin\pi r/K\frac{dr}{K}. \tag{20}$$

Later, we will apply $a\to 0$ to obtain the final result, in which case the strip sections down and up z_n have constant potentials $\mp V_n$, respectively.

VI. CASE OF LARGE K

Continuing the above analysis, we would develop a general theory of IDTs that is rather complex. More instructive results of certain practical value can be obtained applying further approximations. Moreover, the resulting simplified analysis may better present the proposed model. The first assumption is that $K=2\pi/\Lambda$ is large with respect to the SAW wave numbers $k_{o,v}$. This means that there are more than two strips per SAW wavelength, like in the frequently used split-finger transducers where $K\approx 4k_v$. The other assumption is also frequently applied, $\Delta=Kw=\pi/2$ meaning equal strip width and spacing. These assumptions, and approximations that follow, have clear physical interpretation and have been already used in an earlier theory of IDT.[4,7]

For instance, if $0<r<K/2$, Eq. (10) needs only to be checked for $n=0$, with m allowed to be only 0 and 1. Noticing that $P_0(0)=1$, $P_1(0)=0$, this yields

$$\alpha_1=-\alpha_0(1-a)/(1+a),\qquad a=\overline{\epsilon_e}(\theta_0)\frac{r^2+s^2-k_v^2(\theta_0)}{r^2+s^2-k_o^2(\theta_0)}\frac{\sqrt{r^2+s^2}}{r}. \tag{21}$$

Similarly, if $K/2<r<K$, it is sufficient to check Eq. (10) at $n=-1$ only with applied $m=-1,0$:

$$\alpha_{-1}=-\alpha_0(1+b)/(1-b),\qquad b=\overline{\epsilon_e}(\theta_{-1})\frac{r_{-1}^2+s^2-k_v^2(\theta_{-1})}{r_{-1}^2+s^2-k_o^2(\theta_{-1})}\frac{\sqrt{r_{-1}^2+s^2}}{r_{-1}}. \tag{22}$$

Here, α_0 is arbitrary and the other $\alpha_m=0$. This yields

$$R(r,s)\approx\overline{\epsilon_e}(\theta_0)\frac{r^2+s^2-k_v^2(\theta_0)}{r^2+s^2-k_o^2(\theta_0)}\frac{\sqrt{r^2+s^2}}{r},\qquad R(r,s)\approx\overline{\epsilon_e}(\theta_{-1})\frac{r_{-1}^2+s^2-k_v^2(\theta_{-1})}{r_{-1}^2+s^2-k_o^2(\theta_{-1})}\frac{\sqrt{r_{-1}^2+s^2}}{-r_{-1}}, \tag{23}$$

respectively, for $0<r<K/2$ and $K/2<r<K$.

Taking both of the above into account, one obtains for $-K/2<\xi<K/2$,

$$R(\xi,s)\approx\overline{\epsilon_e}(\theta)\frac{k^2-k_v^2(\theta)}{k^2-k_o^2(\theta)}\frac{k}{\xi}S_\xi, \tag{24}$$

where $\theta=\operatorname{atan}(s/\xi)$, $k=\sqrt{\xi^2+s^2}$. Now, the inner integral (20) can be rewritten in the form

$$\int_0^K R(r,s)e^{-jr(m-n)\Lambda}\sin\pi\frac{r}{K}dr=\int_{-K/2}^{K/2}\overline{\epsilon_e}(\theta)\frac{k^2-k_v^2}{k^2-k_o^2}\frac{k}{\xi}e^{-jr(m-n)\Lambda}\sin\pi\frac{\xi}{K}d\xi.$$

Most interesting is the part of the transfer function resulting from the wave propagation between strips. This will be the part satisfying the Jordan lemma when the integration over ξ, from $-K/2$ to $K/2$ with K assumed large, is extended to $\pm\infty$. To obtain this part, we consider only the second component of

$$\frac{k^2-k_v^2}{k^2-k_o^2}=1+\frac{k_o^2-k_v^2}{k^2-k_o^2}$$

proportional to $d=(k_o-k_v)/k_o$, which quantity is called a piezoelectric coupling coefficient for SAW, being usually of an order of 1%; $k_o^2-k_v^2\approx 2k_o^2d$. (The unitary part of the above expression describes the strip mutual capacitance that is not discussed further below.) The SAW induced current is

$$J_m(z)=\frac{j\omega 4V_n}{\pi K}\int_{-\infty}^{\infty}\frac{e^{-js(z-z_n)}}{s^2+a^2}\int_{-\infty}^{\infty}\epsilon_e(\theta)\times\frac{k_o^2d(\theta)}{k^2-k_o^2}\frac{k}{\xi}e^{-j\xi(m-n)\Lambda}\sin\frac{\pi\xi}{K}d\xi\,ds. \tag{25}$$

Below, $m\neq n$ is assumed, meaning that the interaction between different strips is sought.

This result needs somewhat deeper discussion, however. It is evident that the limit value discussed in Eq. (8) depends on s: for s large, we need more equations to be accounted for in Eq. (10), perhaps rendering solutions (20)–(24) too much simplified. Happily, only small values of s are most important in the expression for current, Eq. (25), because the term of $(s^2+a^2)^{-1}$ vanish if $s\to\infty$, $a\to 0$. This proves the correctness of the applied approximation *a posteriori*, at least for large K and for the SAW-induced current J_m.

VII. SAW TRANSFER FUNCTION OF STRIPS

It is convenient to apply angular variables in the above two-dimensional integral,

$$z-z_n=z', \quad (k-n)\Lambda=x',$$
$$z'=p\sin\alpha, \quad x'=p\cos\alpha, \tag{26}$$
$$\xi=k\cos\theta, \quad s=k\sin\theta,$$

with the transformation Jakobian equal to k.

Noticing that $\epsilon_e(\theta+\pi)=\epsilon_e(\theta)$ due to the material symmetry, and similarly for the other parameters of the planar harmonic Green function, one obtains

$$J^R_{m\neq n}(z')=\frac{j\omega 8V_n}{2\pi K}\int_{-\pi/2}^{\pi/2}d\theta\int_{-\infty}^{\infty}\frac{\epsilon_e(\theta)k_o^2(\theta)d(\theta)}{k^2\sin^2\theta+a^2}$$
$$\times k^2\frac{e^{-jpk\cos(\theta-\alpha)}}{k^2-k_o^2(\theta)}X\,dk, \tag{27}$$

where $X=[\sin(\pi k\cos\theta/K)]/(k\cos\theta)]$. This is π/K for $k\approx k_o\ll K$.

Let $x'>0$ meaning that the strip the current of which is evaluated is placed on the right-hand side of the strip supplied with external voltage source (the strip exciting SAWs in the system). To evaluate the inner integral, one must take into account that k_o has complex value in lossy materials. Assuming that k_o is the wave number of the forward propagating wave that carries acoustic power to $x\to+\infty$, $\mathrm{Im}\{k_o\}<0$ while its real part can be either positive or negative in anisotropic materials. Having this in mind and applying the Cauchy theorem, we obtain

$$J_m(z')=4\omega V_n\int_\Theta\frac{\epsilon_e k_o^3 d(\theta)/K}{k_o^2\sin^2\theta+a^2}$$
$$\times X(k_o\cos\theta)e^{-jpk_o\cos(\theta-\alpha)}\,d\theta, \quad m\neq n, \tag{28}$$

where the integration domain Θ is chosen such that the outward normal to the slowness curve k_o/ω over the integration domain is directed towards the observation point which is (x',z') appearing in the exponential function in the above integral. This selects automatically $k_o(\theta)$ that has the proper value of its imaginary part as discussed above; the wave will decay on the path from the source $(n\Lambda,z_n)$ to the observation point $(m\Lambda,z)$, when in lossy materials.

The other poles, at $k^2\sin^2\theta+a^2=0$, have been neglected in the above result because their contribution to J_m does not result from the SAW propagation between strips. Indeed, the phase shift of the exponential term resulting from these poles would be $pa\cos(\theta-\alpha)/\sin\theta$ that is far from the correct one, $\vec{k}_o\vec{p}$, resulting from the surface waves.

It is worth to compare this important result with the one resulting from the plane-wave theory of IDTs presented in Appendix B. The integrand values at $\theta\sim 0$ contribute most to the integral (28), thus applying $\sin\theta\approx\theta$, $\cos\theta\approx 1$, and noticing that $\int_0^\infty(1+x^2)^{-1}\,dx=\pi/2$, we easily obtain that

$$J_l(z'=0)\approx\frac{1}{a}2\omega V_n\epsilon^o(k_o-k_v)\Lambda e^{-jk_o|l-n|\Lambda}\sin\pi k_o/K, \tag{29}$$

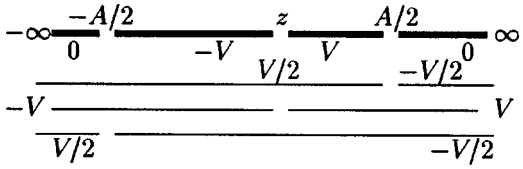


FIG. 2. Finite aperture-width transducer: modeling by superposition of three cases.

where $p\cos(\theta-\alpha)=p\cos\alpha$ that is $|l-n|\Lambda$; similarly for $\theta=0$ is $\epsilon_e\approx\epsilon^o$ and $z'=z-z_n=p\sin\theta=0$ which latter equation means that we evaluate the lth strip current at $z=z_n$.

The interpretation of this result is the following. Both parts of the nth strip having potential $\mp V_n$ excite SAWs of equal amplitudes but different signs. These almost plane SAWs propagate in imaginary neighboring acoustic channels $(-\infty,z_n)$ and (z_n,∞), and excite currents in the lth strip. At the boundary of these acoustic channels, $z=z_n$, these currents are of equal amplitudes, but flowing to the upper part of the strip and from the lower part of it; these equal currents are J_l evaluated above at $z'=0$, that is, at $z=z_n$. That is why there is nice agreement between Eqs. (29) and (B9).

VIII. REGULARIZATION ($a\to 0$)

The integral (28) is strongly singular when $a\to 0$. This corresponds to the case of $A\to\infty$ in Eq. (B10). Indeed, the strips are infinitely long in the above presented theory, corresponding to the infinite aperture width in the plane-wave theory that brings the analogous difficulty in Eq. (B10) when $A\to\infty$.

The problem arises from applying voltages to strips of infinite length. This excites the SAW of infinite beam width and, naturally, excites infinite current in the neighboring infinite grounded strip. In practice, the strips are of finite length (Fig. 2) which can be modeled by superposition of the three excitation cases described by Eq. (18) (with $a\to 0$), as shown in Fig. 2. Moreover, note that the potential difference between the strip sections is set, not the absolute section voltages, $\pm V$. Thus, the strip sections have potentials $C\pm V$ rather than $\pm V$, with the bias voltage C to be determined from the Kirchhoff laws.[4]

Similar is the case of the receiving transducer. Its output bars collect currents from its fingers: $J^R=J_l(z_l)-J_l(A/2)$ and $J^L=J_l(-A/2)-J_l(z_l)$ are lth strip contributions to the right and left transducer bars. Assuming that the transducer is short-circuited, both bars' total currents should be the same. If they are not, then the receiving transducer has certain bias potential C which can be evaluated from the condition of equal bar's currents. This potential contributes to the propagating SAWs, but in typical cases the effect is negligible and C is neglected. This corresponds to $A=0$ in the plane-wave theory. Applying this simplification, the transducer output current is simply $\Sigma_l J(x_l,z_l)$, the sum over all transducer fingers.

Introducing notation $\Delta(\theta)=4\omega\epsilon_e(k_o-k_v)\Lambda X k_o/K$, one obtains

$$J_m(z_m)=V_n\int_\Theta\Delta(\theta)\frac{k_o^2(\theta)e^{-j\vec{k}_o\vec{p}}}{k_o^2\sin^2\theta+a^2}\,d\theta, \tag{30}$$

where $\Delta(\theta)$ represents the piezoelectric property of the substrate and $k_o(\theta)$ is the SAW wave number which may depend on the propagation direction; $z'=z_m-z_n=p\sin\alpha$ and $x'=(m-n)\Lambda=p\cos\alpha$. The integration domain Θ is chosen according to the rule discussed in the section above.

To overcome the above mentioned difficulty with infinite strips but avoiding too sophisticated analysis resulting from Fig. 2, the trick is applied here by setting $A\to 0$. This is by analogy to the plane-wave theory where the troublesome component (A) is neglected with respect to this part of (B10) which depends on the transducer apodization z_n. Using the above-mentioned superposition of three cases depicted in Fig. 2, and $A=0$, one obtains[7] at $a\to 0$

$$J_m(z_m)=V_n\int_\Theta \Delta(\theta)\frac{e^{-j\vec{k}_o\vec{p}}-e^{-jk_o x'\cos\theta}}{\sin^2\theta}d\theta. \tag{31}$$

The integral principal value exists and can be evaluated numerically. Indeed, one can easily transform the integral into $\int_0^{\pi/2} f(\theta)\sin^{-2}\theta\, d\theta$, where $f(\theta)\sim z'\theta^2$ at $\theta\to 0$, for any $k_o(\theta)$. Note the linear dependence of this approximation on z. It will result in the similar dependence of the transducer current on z (i.e., on the transducer apodization), like in the plane-wave theory of IDTs (B10). Numerical examples obtained with help of this equation for IDTs counting hundreds of strips can be found elsewhere.[7] It was found convenient to sum up the contributions of all the strips to the total IDT current first, and then perform numerical integration over θ.

IX. CONCLUSIONS

The theory of IDTs accounting for the SAW diffraction has been formulated using the BIS expansion method. The same method has been used earlier in the plane-wave model[4] of IDTs that was thoroughly verified experimentally.[7] As shown in Sec. VII, both theories yield the same results in the limit case. This is certain verification of the proposed model.

The surface wave diffraction is a spurious phenomenon that distorts the frequency response of SAW filters with respect to that designed with help of the plane-wave theory (Appendix B). The diffraction may cause spurious cross-talk, for instance, between tightly arranged filters on the substrate in filter banks. Otherwise it may be exploited positively in laterally coupled SAW resonators. The presented model delivers a method for analyzing and designing of such devices, particularly in cases of narrow band selective filters and resonators exploiting systems of several hundred strips of relatively small length A, Fig. 2.

The theory is open for further development: (1) Accounting for more α_m, the system of two strips per wavelength applied in SAW resonators can be analyzed, yielding an estimation of the diffraction losses. (2) Applying an appropriate planar harmonic Green function, pseudo-SAW, leaky and surface skimming bulk waves, can be included into the theory as well. (3) And by applying the model of elastic strips,[4,7] the mechanical interaction of the transducer fingers with propagating SAWs can be accounted for. This analysis would involve the full matrix Green function.[6]

Although the presented model relies on several approximations, the results can be made as accurate as required by applying larger N in Eqs. (10), at expense of computation time. This makes it an ideal tool for verification of the other *ad-hoc* models that may appear in literature and which may be less computationally demanding. In the present theory, the computation time is to be spent primarily on evaluation of integrals like (31), due to its rapidly varying exponential function for very distant transducer fingers.

ACKNOWLEDGMENTS

This paper is dedicated to His Holiness Pope John Paul II. The author is indebted to Dr. Thomas Danicki from Rochester Hills, MI, for his kind contribution to the text of the paper that improved it considerably.

APPENDIX A: USEFUL IDENTITIES

The BIS expansion method exploits the known Fourier expansion of a complementary pair of periodic square-root singular functions that will serve as a prototype of the real wave field for further modifications. Vanishing in alternate domains, these prototype functions satisfy the mixed boundary conditions. Slightly transformed forms of these functions with respect to the identity presented in the monograph[8] are

$$\begin{aligned}
&\sum_{n=-\infty}^{\infty}\alpha_m P_{n-m}(\cos\Delta)e^{-jn\theta}\\
&\quad=\sqrt{2}\,\frac{\alpha_m\exp(-jm\theta)}{\sqrt{\cos\theta-\cos\Delta}}e^{j\theta/2},\quad |\theta|<\Delta,\\
&\sum_{n=-\infty}^{\infty}\beta_m S_{n-m}P_{n-m}(\cos\Delta)e^{-jn\theta}\\
&\quad=-jS_\theta\sqrt{2}\,\frac{\beta_m\exp(-jm\theta)}{\sqrt{\cos\Delta-\cos\theta}}e^{j\theta/2},\quad \Delta<|\theta|<\pi
\end{aligned} \tag{A1}$$

(and zero in the other domains[4]), where $S_\nu=1$ for $\nu\geqslant 0$, and -1 otherwise, and $P_\nu(\cos\Delta)$ is a Legendre function. The second form results from the first by applying substitution of θ, Δ by $\pi-\theta$, $\pi-\Delta$ and using the identity $P_n(-\cos\Delta)=S_n(-1)^n P_n(\cos\Delta)$.

The real wave field in the plane of strips is constructed as a superposition of the above prototype functions over m in certain large but finite limits. Such a constructed wave field satisfies the above mentioned boundary conditions at the plane of strips automatically like the prototype functions in Eqs. (A1), although having different values, dependent on the expansion coefficients α_m and β_m, in alternate domains where they are different from zero. This is due to the resulting summation over m (a finite Fourier series) in the numerator of the right-hand sides of Eqs. (A1) involving arbitrary coefficients α_m, β_m. This secondary finite expansion over m is called the BIS expansion, in contrast to the infinite Bloch expansion over n in Eqs. (A1).

The unknown expansion coefficients (α_m,β_m) will be determined by checking the differential equations describing the medium in which the strips are embedded, for each of harmonic (Bloch) components of the wave field. Conveniently, these equations can be sufficiently represented by the

corresponding planar harmonic Green function describing both homogeneous half-spaces above and below the plane of strips. The asymptotic property of this function of spectral variable makes the above-mentioned checking significant only for certain finite number of the lowest Bloch components. Moreover, careful analysis of the resulting infinite number of equations for higher Bloch components shows that β_m depends on α_m, thus eliminating half unknowns.

The resulting system of equations has very specific form depicted below

$$\begin{bmatrix} \cdot & & & & & & & \\ \cdot & \cdot & & & & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \\ & & & & & & \cdot & \end{bmatrix}.$$

It is evident that the upper two unknowns of this system of equations and the lowest one have trivial solutions equal to zero. Apparently, including more equations taken for higher positive and negative Bloch components does not change the nontrivial solution for α_m. The strength of the BIS expansion method relies on this feature, that it results in a small system of equations that must be solved numerically.

Summarizing, as opposed to many approximate methods, the BIS expansion method tends to satisfy the boundary conditions exactly. This is achieved at the cost of certain inaccuracy in the differential equations describing the medium, the inaccuracy dependent on how many coefficients α_m, β_m are included in the wave field representation.

Two useful identities, resulting from the Dougall expansion[8] are, after some easy transformations

$$\sum_{n=-\infty}^{\infty} \alpha_m \frac{(-1)^n P_{n-m}(x)}{r+nK} = \frac{\pi \alpha_m}{K \sin \pi r/K} P_{-m-r/K}(x),$$
$$\sum_{n=-\infty}^{\infty} \beta_m \frac{S_{n-m} P_{n-m}(x)}{r+nK} = \frac{\pi (-1)^m \beta_m}{K \sin \pi r/K} P_{-m-r/K}(-x), \tag{A2}$$

$x=\cos\Delta$, which help us to present the final results of the analysis in quite simple forms.

APPENDIX B: PLANE-WAVE THEORY OF IDT

The theory of IDTs based on the BIS expansion method is presented in the earlier papers.[4,7] Here, we present it shortly for the readers' convenience neglecting the strip elasticity and the resulting SAW stopbands. In spite of it, the results remain valid for most applications.

For plane-wave approximation, strips are infinitely long and the field is assumed independent of z, thus

$$E^{(n)} = j r_n \phi_n, \quad r_n = r + nK, \quad k_n = \sqrt{r_n^2},$$
$$\Delta D_y^{(n)} = -j \epsilon^o \frac{r_n^2 - k_v^2}{r_n^2 - k_o^2} \frac{k_n}{r_n} E^{(n)}, \tag{B1}$$

for nth Bloch wave field component; k_n/r_n can be replaced by S_n (assuming $0<r<K$). The BIS expansion is (summation symbol over m dropped)

$$\Delta D_y^{(n)} = \alpha_m P_{n-m}(\cos\Delta) e^{-j r_n x},$$
$$E^{(n)} = \alpha'_m S_{n-m} P_{n-m}(\cos\Delta) e^{-j r_n x}. \tag{B2}$$

Applying (B1), one finds that $\alpha'_m = j\alpha_m/\epsilon^o$ which results in

$$\alpha_m \left[1 - S_n S_{n-m} \frac{r_n^2 - k_v^2}{r_n^2 - k_o^2} \right] P_{n-m}(\cos\Delta) = 0. \tag{B3}$$

It is satisfied automatically when $|n| \to \infty$ for any α_m and m in finite limits (this results from the form of the system matrix discussed in the preceding section).

This set of equations, applied for $|n| \leqslant N$ and $-N \leqslant m \leqslant N+1$, can be solved for $\overline{\alpha}_m = \alpha_m/\alpha_0$, then the Fourier transforms can be evaluated for the current $\overline{J}$ flowing to the strips, and of the strip potential V, both of which are dependent on r

$$\overline{J}(r) = j\omega\alpha_0 \int_{-\Lambda/2}^{\Lambda/2} \Delta D_y \, dy = j\omega\Lambda \alpha_0(r) \overline{\alpha}_m P_{-m-r/K}(x),$$
$$V(r) = -\int E \, dx = \alpha_0(r) \frac{\pi(-1)^m \overline{\alpha}_m}{\epsilon^o K \sin \pi r/K} P_{-m-r/K}(-x) \tag{B4}$$

($x=\cos\Delta$). The derivation is analogous to that presented in the preceding sections.

The inverse Fourier transform to $V(r)$ must yield the given strip potential V_n, hence it must be

$$\alpha_0(r) = V_n \frac{\epsilon^o K e^{-jrn\Lambda} \sin \pi r/K}{\pi(-1)^m \overline{\alpha}_m P_{-m-r/K}(-\cos\Delta)}. \tag{B5}$$

This, when applied in the inverse Fourier transform for the current of lth strip, yields ($x=\cos\Delta$)

$$\overline{J}_l = \frac{1}{K} \int_0^K \overline{J}(r) e^{-jr\Lambda} \, dr$$
$$= j\omega V_n 2\epsilon^o \frac{1}{K} \int_0^K R(r) e^{-jr(l-n)\Lambda} \sin \pi r/K \, dr,$$
$$R(r) = \frac{\alpha_m P_{-m-r/K}(x)}{(-1)^m \alpha_m P_{-m-r/K}(-x)} \approx \frac{r-r_v}{r-r_o} \frac{K-r-r_v}{K-r-t_o}, \tag{B6}$$

where r_v and r_o are zero and the pole of $R(r)=R(K-r)$, which functional symmetry was accounted for in the above approximation to R; moreover $r_{o,v} \approx k_{o,v}$ if $K \gg k_o$.

The part of the current caused by the SAWs is

$$J_l = j\omega 2\epsilon^o V_n \frac{1}{K} \int_0^K \left(\frac{r_o - r_v}{r - r_o} + \frac{r_o - r_v}{K - r - r_o} \right)$$
$$\times e^{-jr(l-n)\Lambda} \sin \pi r/K \, dr, \tag{B7}$$

which integration is performed by (1) extending limits to infinity (assuming K large), and (2) applying the Cauchy theorem. The result is for $l \neq n$

$$J_l = 2\omega V_n \epsilon^o (r_o - r_v) \Lambda e^{-jr_o|l-n|\Lambda} \sin \pi r_o/K. \tag{B8}$$

This is the fundamental relation of the theory of IDTs: the voltage V_n applied to nth strips excites a plane SAW that

propagates along the substrate surface with wave number r_o and excites the current of density J_l (per strip length) in a grounded lth strip.

The contribution of a strip of length A to the IDT current is AJ, neglecting its part caused by the strip mutual capacitance.[4] Applying still the plane-wave theory, suppose that the voltage V_n varies along the strip like $\exp(-az)$, $z \in (0,\infty)$, a small. The total current excited in the lth strip evaluated from the plane-wave theory, is

$$I_l = \int_0^\infty J_l e^{-az}\, dz = J_l/a. \tag{B9}$$

This means that the equivalent aperture width of the applied excitation is $A = 1/a$.

For the so-called apodized IDTs which strips spans between the transducer bars (separated by the distance A), and having a break at z_n between bars, the transducer total current has been evaluated[4] to be

$$I = yU \sum_{n,m} [A - |z_n - z_m|] e^{-jr_o|n-m|\Lambda}, \tag{B10}$$

where U is the potential difference applied to the bars. The current depends on ω mainly due to $k_o \sim \omega$ in the exponential terms involved (y is a certain coefficient weakly dependent on ω). The transducer frequency response $I(\omega)$ is known to be dependent on the transducer apodization z_n rather than on the transducer bar separation A, which term is usually neglected in the analysis under the assumption that the IDT has such a typical construction that the terms with A cancel each other in the frequency band of interest.[4]

[1] F. Z. Bi and K. Hansen, "Investigation of diffraction compensation at the trap position for TV SAW filters," IEEE Ultrasonics Symposium Proceedings, 1996, pp. 139–144.

[2] M. S. Kharusi and G. W. Farnell, "Plane ultrasonic transducer diffraction field in highly anisotropic crystals," J. Acoust. Soc. Am. **48**, 665–670 (1970).

[3] M. Jungwirth, T. Gneifeneder, K. Scheiblhofer, A. Stögmüller, R. Weigel, D. C. Malocha, W. Ruile, and C. C. W. Ruppel, "Improved time domain diffraction analysis for the SAW transducers of arbitrary shape," IEEE Ultrasonics Symposium Proceedings, 1999, pp. 97–100.

[4] E. Danicki, "Generation and Bragg reflection of surface acoustic waves in nearly periodic system of elastic metal strips on piezoelectric half-space," J. Acoust. Soc. Am. **93**, 116–131 (1993).

[5] K. Bløtekjær, K. A. Ingebrigtsen, and H. Skeie, "A method for analyzing waves in structures consisting of metal strips on dispersive media," IEEE Trans. **ED-20**, 1133–1138 (1973).

[6] E. J. Danicki, "An approximation to the planar harmonic Green's function at branch points in wave-number domain," J. Acoust. Soc. Am. **104**, 651–663 (1998).

[7] E. Danicki, "Spectral theory for IDTs," IEEE Ultrasonics Symposium Proceedings, 1994, pp. 213–222.

[8] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 1, Chap. 3.

The predicted barrier effects in the proximity of tall buildings

Kai Ming Li[a] and Siu Hong Tang
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

(Received 27 November 2002; revised 18 May 2003; accepted 30 May 2003)

A ray model is developed and validated for the prediction of the insertion loss of barriers that are placed in front of a tall building in high-rise cities. The model is based on the theory of geometrical acoustics for sound diffraction at the edge of a barrier and multiple reflections by the barrier and façade surfaces. It is crucial to include the diffraction and multiple reflection effects in the ray model, as they play important roles in determining the overall sound pressure levels for receivers located between the façade and barrier. Comparisons of the ray model with indoor experimental data and wave-based boundary element formulation show reasonably good agreement over a broad frequency range. Case studies are also presented that highlight the significance of positioning the barrier relative to the noise-sensitive receivers in order to achieve improved shielding efficiency of the barrier. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1593060]

PACS numbers: 43.50.Gf, 43.28.En, 43.50.Rq [DKW]

I. INTRODUCTION

The use of acoustic barriers to shield land transportation noise sources from neighborhood residents has been implemented extensively in many countries worldwide. Since the late 1960s,[1–6] much effort has been put into predicting the sound levels behind barriers by using theoretical studies, scaled-model experiments, and other more costly full-scale field experiments. Despite this widespread interest, there are relatively few studies that consider the barrier effect on the overall sound pressure levels in front of a tall building. This typical urban scenario is particularly important in high-rise cities where noise barriers are sometimes built close to tall residential buildings.

In the absence of other reflecting surfaces, diffraction over the top edge is often the only path for sound waves to reach receivers located at the opposite side of the barrier. In this simplest case, rules and charts are available to aid the practical design of noise barriers such that their acoustic performance can be assessed readily. Theoretical approximations and field experience were combined heuristically to devise these design rules and charts.[4,7,8] Much of this early theoretical work on the study of noise barriers was based on the classical diffraction theory.[9] More recent theoretical work has allowed the presence of the impedance ground, the use of sound-absorbent materials on the barrier surface, and multiple diffraction/reflection between two barriers.[10–12] However, with a barrier situated in front of a tall building, a series of image sources are formed behind the façade and barrier surfaces due to multiple reflections. Most standard prediction methodologies, for example, the CRTN[13] and FHWA programs,[14] do not take into account the prediction of the sound field due to these multiple reflections occurring between a façade–barrier system.

Sakurai *et al.*[15] used a time–domain method to investigate the sound field of the façade–barrier system. The prediction of the sound field around the barrier was based on the Kirchhoff–Rubinowicz theory but their theoretical model was not in accord with their experimental data for sound fields measured behind the barrier. Walerian *et al.*[16] developed a computer simulation program to evaluate the time-averaging sound pressure levels in an outdoor environment. However, their program involved too many different parameters, which made it somewhat complex to apply in the present façade–barrier system.

Cheng and Ng[17] studied the acoustic performance of a parallel barrier in front of a building façade. They used an empirical diffraction model but they did not consider the multiple reflections between the façade and barrier and only one image source was included in their numerical formulation. Godinho *et al.*[18] used the boundary element method (BEM) to determine the sound field produced by an infinitely long barrier in the vicinity of building façades. The BEM numerical scheme is an accurate numerical method. One of its main disadvantages is that it is very computationally intensive, especially at high frequencies. In Godinho's investigations, no experiments were conducted, but they compared their numerical results with those obtained by other rather simplified prediction models.[14,19] These simplified models were mainly used for outdoor situations and did not show much agreement with their BEM results.

In this paper, we aim to develop a ray-based model for the prediction of the effect of a noise barrier in the vicinity of tall buildings. The theoretical formulation for the ray model is outlined in Sec. II. Details of numerical and indoor experiments for the validation of the theoretical model are presented in Sec. III. In Sec. IV, typical outdoor situations are simulated that investigate the façade effect on the acoustic performance of noise barriers. Concluding remarks are offered in Sec. V.

II. THEORY—FORMULATION OF A RAY MODEL

We consider a typical scenario of a high-rise city in which a noise barrier of height H is aligned parallel to a row

[a]Author to whom correspondence should be addressed. Electronic mail: mmkmli@polyu.edu.hk

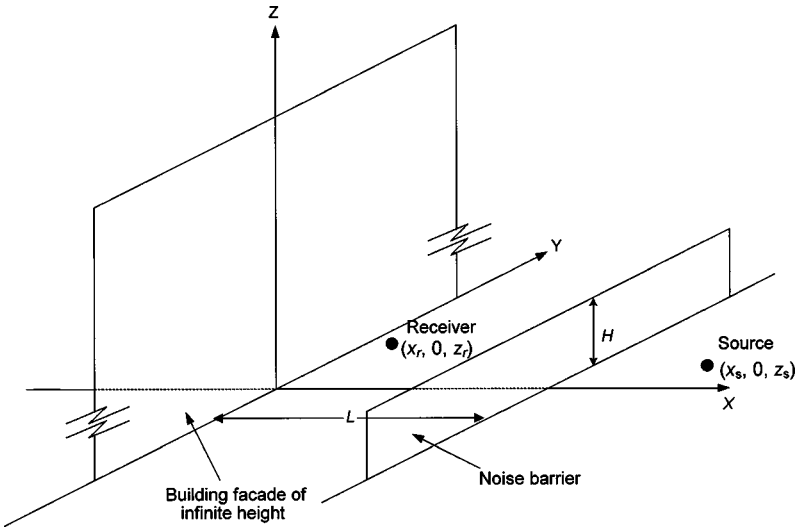


FIG. 1. Schematic diagram of the source/receiver configuration in a façade–barrier system.

of tall buildings. The barrier is designed to shield residents in the buildings from noise sources that are located on the opposite side of the barrier. To model the problem, we assume that the tall buildings are replaced by a plane façade where the barrier is built at a distance L in front of it, above an absorbing ground. Figure 1 shows a schematic diagram of the specified problem. The façade is situated along the y axis at the $x=0$ plane. Furthermore, the façade is assumed to be much higher than the barrier and receiver so that diffraction of sound at the top of the façade can be ignored. We are primarily interested in a three-dimensional problem where the barrier and building façade are placed on the ground surface at the plane of $z=0$. They are extended to infinity in both directions along the y axis, i.e., $-\infty \leqslant y \leqslant +\infty$, so that sound diffracted at their side edges is omitted. To simplify the problem, let the source and receiver be located at the same vertical plane at $y=0$. A time-dependent factor $e^{-i\omega t}$ is understood in this paper.

A rectangular coordinate system is used where the receiver is positioned at $\mathfrak{R} \equiv (x_r,0,z_r)$ between the barrier and façade. A noise source is placed on the opposite side of the barrier at $\Psi_0 \equiv (x_s,0,z_s)$, where $x_s \geqslant L$ and z_s, $z_r \geqslant 0$. Using the image source concept, an image source above the impedance boundary at $\Psi_0' \equiv (x_s,0,-z_s)$ can be identified immediately. In the present study, we wish to investigate the case where the receivers are located in front of the façade but are separated from the source by the barrier, i.e., $L \geqslant x_r \geqslant 0$. Let us consider a more general situation where the ground surfaces have the specific normalized admittance of β_1 at the receiver side and β_2 at the source side. Since acoustically hard materials are commonly used for the façade and barrier surfaces in most metropolitan areas, the specific normalized admittance of the façade and barrier surfaces are assumed to be zero.

When the barrier blocks the direct line-of-sight contact between the receiver and noise source, the receiver itself is located in the shadow zone. In this case, the diffraction of sound at the top edge of the barrier is the only transmission path for the propagation of noise toward the receiver, as the transmission of sound through the barrier is ignored. On reaching the top edge of the barrier, part of the diffracted waves propagate toward the receiver directly. Other parts of the diffracted waves are reflected at the ground and façade surfaces before they reach the receiver. The point of diffraction at the barrier edge, which is located at H above the ground surface, may be treated as a secondary noise source. Since this secondary noise source is located at the top edge of the barrier's surface, multiple reflections take place between the façade and barrier before the diffracted waves arrive at the reception point. Images of the primary and secondary sources are formed in the quarters of $x>0$, $z>0$ and $x<0$, $z>0$. By virtue of the impedance boundary, ground-reflected image sources are also formed in the quarters of $x<0$, $z<0$ and $x>0$, $z<0$. The positions of these primary and secondary sources and their corresponding images can be identified as located at $\Psi_n=(x_n,0,z_s)$, $\mathbf{S}_n=(\overline{x_n},0,H)$, $\Psi_n'=(x_n,0,-z_s)$, and $\mathbf{S}_n'=(\overline{x_n},0,-H)$, respectively. Here, n is the order of the image sources that can be regarded as the number of reflections from the two parallel surfaces. When the order n is an odd number (equal to $2m+1$, say), the image source is located at the left side behind the façade surface ($x<0$). In this particular image source, the ray can be identified as hitting the façade surface $m+1$ times and interacting with the barrier surface m times. When n is even, the image source is situated at the right side behind the barrier surface ($x>0$). The corresponding ray will hit the façade and barrier surfaces an equal number of times. In addition, an extra reflection occurs for those image sources located below the ground surface.

A close examination of the positions of the secondary source and its images reveals that the distance in the x direction of the consecutive images in each quarter differs by a length of $2L$. It is straightforward to generalize the x coordinate, x_n of the secondary source and its images. It is interesting to note that the even- and odd-order rays (order $2m$ and $2m+1$ for $m=0,1,2,...$) are of equal horizontal distance from the origin:

$$\begin{aligned} x_{2m} &= -x_{2m+1} = (2m+1)x_s, \\ \overline{x_{2m}} &= -\overline{x_{2m+1}} = (2m+1)L, \end{aligned} \quad \text{for } m=0,1,2,3,\ldots. \tag{1}$$

A negative sign is required for the x coordinates of all odd-

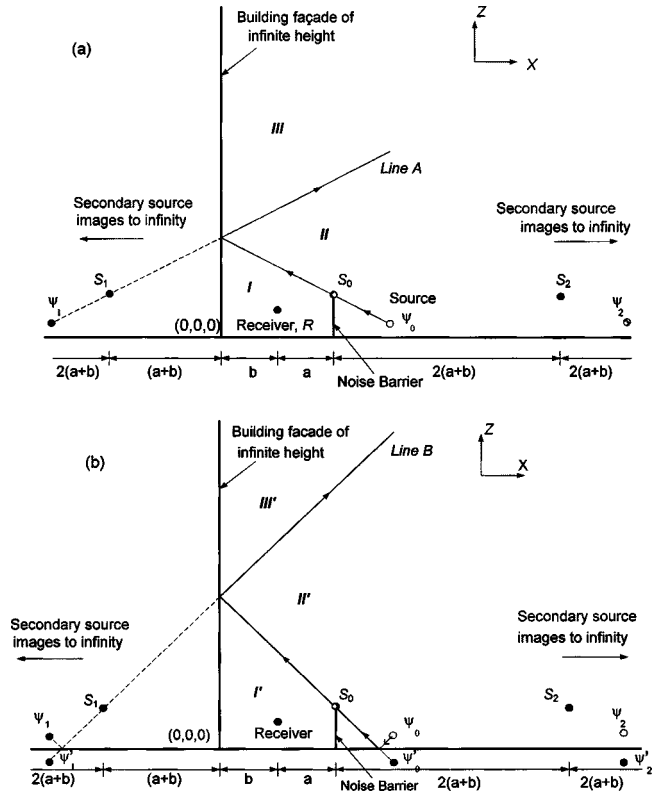


FIG. 2. The source/receiver geometry in a façade–barrier system in the x-z plane. The distance between the receiver and barrier is denoted as "a." The distance between the receiver and building façade is denoted as "b." Subsequent secondary source images are separated by a distance of $2(a+b)$. The three regions of the sound field are formed by (a) the primary noise source, ψ_0; and (b) the image source, ψ_0', due to the reflection from the ground.

order rays because these sources are located at the left side behind the façade surface ($x<0$).

The geometrical configuration of the problem is shown in Fig. 2(a), illustrating the positions of receiver, primary noise source, secondary noise source, and its images in the $x-z$ plane. For reasons of clarity, the image sources below ground are not shown in Fig. 2(a). In principle, the number of images tends to infinity as they line up in the $\pm x$ directions. However, in practice, only a finite number of image sources contribute to yield the total sound field. Obviously, if the separation between the barrier and façade is smaller, more image sources will be generated within a specified distance because of the increased number of possible reflections between these two vertical surfaces. This may lead to a more intense noise level as more image sources contribute to the total sound field. We reiterate that both source and receiver are assumed to locate at the same vertical plane at $y=0$ and the primary source is located below the barrier top edge, but the receiver is allowed to vary along the height of the façade above the barrier top edge. This arrangement permits us to simulate situations where the noise-sensitive receivers reside on different floors in the buildings.

Let us investigate different zones of the sound field in the proximity of the façade–barrier system. The current setup is quite complicated due to the presence of a building façade behind the noise barrier. When a receiver is located in the shadow zone [i.e., Region I of Fig. 2(a)], the line of sight between the source and receiver is blocked. The sound field is mainly dominated by the diffracted waves and their subsequent multiple reflections between the façade and barrier surfaces. An image source method, see, for example, Ref. 20, is used to account for the effects of multiple reflections. According to this concept, each reflected wave is replaced by a wave emitted from an image source. In Region I, the total sound field is determined by summing all diffraction terms formed by these secondary image sources. To compute the diffracted wave terms, it is useful to consider the geometrical configuration, as shown in Fig. 3(a). For a point source located at $\Psi_0 \equiv (x_s,0,z_s)$, receiver at $\Re \equiv (x_r,0,z_r)$, and the point of diffraction at $\mathbf{D} \equiv (x_d,0,z_d)$, Pierce's formulation[21] may be used to compute the diffracted sound field by using the following formula:

$$P(\mathbf{S},\Re,\mathbf{D}) = \left(\frac{e^{i\pi/4}}{\sqrt{2}}\right)\left[\frac{e^{ik(d_S+d_R)}}{4\pi(d_S+d_R)}\right] \times [A_D(X_+)+A_D(X_-)], \tag{2}$$

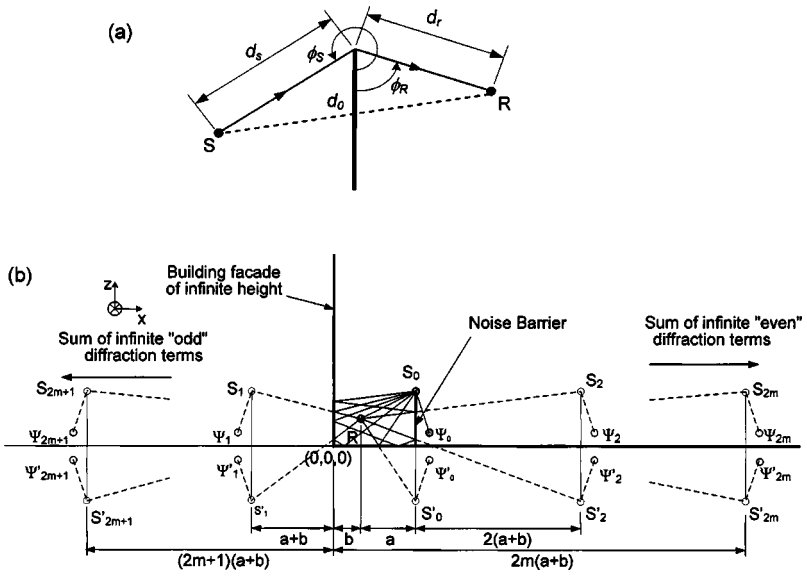


FIG. 3. (a) The geometrical configuration for the diffraction of sound by a thin barrier. (b) Schematic diagram showing multiple reflections between the façade and barrier surfaces for the receiver height is less than that of the barrier.

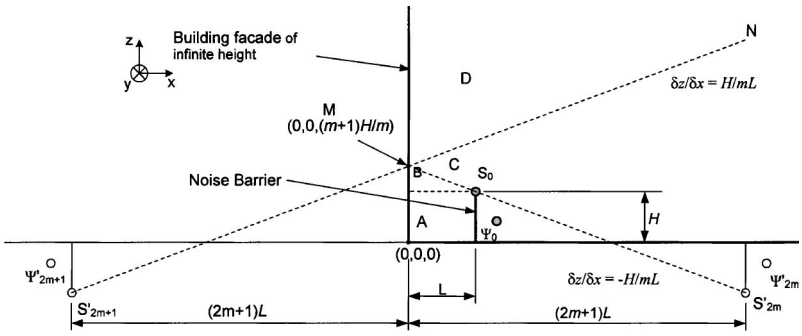


FIG. 4. The source/receiver geometry in a façade–barrier system in the x-z plane. Four different regions A, B, C, and D, are identified due to positions of the diffraction edge and the images of the even and odd virtual sources, S_{2m} and S'_{2m}.

where d_S and d_R are the respective distances from the source and receiver to the diffraction point. The function, $A_D(X)$, is the diffraction integral[21] given by

$$A_D(X)=\mathrm{sgn}(X)[f(|X|)-ig(|X|)], \tag{3}$$

where sgn(X) is the sign function, and $f(X)$ and $g(X)$ are the auxiliary Fresnel functions[22] of real argument X. The arguments of the diffraction integral, X_+ and X_-, are determined by

$$X_\pm = X(\phi_R \pm \phi_S), \tag{4}$$

where

$$X(\Phi)=\left[-2\cos\left(\frac{\Phi}{2}\right)\right]\sqrt{\frac{2\cdot d_S\cdot d_R}{\lambda(d_S+d_R)}}, \tag{5}$$

λ is the wavelength of the diffracted sound, and the angles ϕ_R and ϕ_S are defined in the surface of the screen, as shown in Fig. 3(a). The argument Φ in Eq. (5) is either $(\phi_R+\phi_S)$ or $(\phi_R-\phi_S)$ for X_+ and X_-, respectively.

The total diffracted sound field due to the primary sources can be calculated as follows:

$$P_D = P_A + P_B \tag{6a}$$

where P_A and P_B are corresponding contributions from the secondary sources (even- and odd-order image sources, respectively) above and below the ground surfaces. They are determined by

$$P_A=\sum_m P(\Psi_{2m},\Re,\mathbf{S}_{2m})+\sum_m P(\Psi_{2m+1},\Re,\mathbf{S}_{2m+1}) \tag{6b}$$

and

$$P_B=\sum_m Q(\mathbf{S}'_{2m},\Re,\beta_1)P(\Psi_{2m},\Re,\mathbf{S}'_{2m}) + \sum_m Q(\mathbf{S}'_{2m+1},\Re,\beta_1)P(\Psi_{2m+1},\Re,\mathbf{S}'_{2m+1}), \tag{6c}$$

where the function Q is the spherical wave reflection coefficient for a given source and receiver position and the acoustical characteristics of the plane boundary. Generally speaking, for a given source position $\mathbf{S}$, receiver position $\Re$, and specific normalized admittance of the impedance boundary β, the spherical wave reflection coefficient $Q(\mathbf{S},\Re,\beta)$ can be computed by[23]

$$Q(\mathbf{S},\Re,\beta)=V(\theta)+[1-V(\theta)]F(w), \tag{7a}$$

where the plane wave reflection coefficient $V(\theta)$, the boundary loss factor $F(w)$, and the numerical distance w are determined according to

$$V(\theta)=\frac{\cos\theta-\beta}{\cos\theta+\beta}, \tag{7b}$$

$$F(w)=1+i\sqrt{\pi}we^{-w^2}\,\mathrm{erfc}(-iw), \tag{7c}$$

and

$$w=+\sqrt{\tfrac{1}{2}ikd}(\cos\theta+\beta). \tag{7d}$$

The function erfc(X) is the complementary error function of complex argument X, d is the distance between the image source and receiver, and θ is the angle of incidence of the reflected wave. It is relatively simple to determine d and θ for a given geometrical configuration of the source $\mathbf{S}$ and receiver $\Re$.

Similarly, the total diffracted sound field due to the images of the primary sources Ψ'_n are given below:

$$P'_D = P'_A + P'_B, \tag{8a}$$

where

$$P'_A=\sum_m Q(\Psi'_{2m},\mathbf{S}_{2m},\beta_2)P(\Psi'_{2m},\Re,\mathbf{S}_{2m}) + \sum_m Q(\Psi'_{2m+1},\mathbf{S}_{2m+1},\beta_2)P(\Psi'_{2m+1},\Re,\mathbf{S}_{2m+1}) \tag{8b}$$

and

$$P'_B=\sum_m Q(\mathbf{S}'_{2m},\Re,\beta_1)Q(\Psi'_{2m},\mathbf{S}'_{2m},\beta_2)P(\Psi'_{2m},\Re,\mathbf{S}'_{2m}) + \sum_m Q(\mathbf{S}'_{2m+1},\Re,\beta_1)Q(\Psi'_{2m+1},\mathbf{S}'_{2m+1},\beta_2) \times P(\Psi'_{2m+1},\Re,\mathbf{S}'_{2m+1}). \tag{8c}$$

All series in Eqs. (6b), (6c) and (8b), (8c) start from $m=0$, but the number of terms required in each series is different. It is independent of the source position but depends only on the receiver position between the façade and barrier surfaces.

We need to determine the number of terms required in the series for the calculation of the diffraction fields [cf. Eqs. (6a)–(6c) and (8a)–(8c)]. If the receiver height is less than that of the barrier, as shown in the Region A of Fig. 4, then

multiple reflections between the façade and barrier surfaces are possible, as they are aligned parallel to each other [see also Fig. 3(b)]. Hence, in principle, a sum of infinite diffraction terms is required for each series in Eqs. (6b), (6c) and (8b), (8c). However, in practice, only a finite number of terms are required because the higher-order rays have longer distances to travel between the barrier edge and the receiver. Consequently, higher-order diffracted rays that are smaller in magnitude can be neglected in the total sound field. This could be verified by the convergence of the calculated sound field in the simulations. On the other hand, if the receiver is located at a higher position than the top edge of the barrier, then the number of diffraction terms is dependent on the positions of virtual sources (either $\mathbf{S}_n$ or $\mathbf{S}'_n$) and the receiver $\mathfrak{R}$. This is because multiple reflections between the two surfaces can only be sustained as long as the height of the ray path is less than the top edge of the barrier. As a result, only some diffraction terms from the secondary sources and their images contribute to the total sound field. A close examination of Eqs. (8b) and (8c) reveals that only zeroth- and first-order diffracted rays, $\mathbf{S}_0$ and $\mathbf{S}_1$, can contribute to the total field if the receiver is located in regions B, C, and D, as shown in Fig. 4. Hence, only the first term is required in the series.

To determine the required terms for the series in Eqs. (6c) and (8c), let us consider a pair of consecutive even- and odd-order diffracted rays, $\mathbf{S}'_{2m}$ and $\mathbf{S}'_{2m+1}$, where $m=0,1,2$, 3, etc. We can easily identify different regions where these two diffracted rays can contribute to the total field. In the $y=0$ plane, a line can be drawn from the virtual source $\mathbf{S}'_{2m}$ to the barrier edge $\mathbf{S}_0$ and its slope can be determined as $-H/mL$. The line can be extended farther until it reaches point M at $(0,0,(m+1)H/m)$ on the façade. By considering the ray path's geometrical configuration, we can see that the diffracted ray can only penetrate Regions A and B in Fig. 4. No diffracted rays from the virtual source $\mathbf{S}'_{2m}$ can reach noise-sensitive receivers that are located at Regions C and D.

The equation of this limiting ray can be determined that leads to the following condition for the receiver position $(x_r,0,z_r)$, where the virtual source $\mathbf{S}'_{2m}$ contributes to the total field,

$$z_r \leqslant -\frac{Hx_r}{mL}+\frac{(m+1)H}{m}. \tag{9}$$

Alternatively, we can determine the maximum number of terms $\hat{m}$ from Eq. (9) as

$$\hat{m}=\mathrm{int}\left[\frac{H(L-x_r)}{L(z_r-H)}\right]. \tag{10}$$

Similarly, there is an analogous limiting ray to determine the region where there is a contribution of the odd-order ray from the virtual source, $\mathbf{S}'_{2m+1}$. We can construct a reflected ray MN, which has a slope of H/mL and passes through the virtual source $\mathbf{S}'_{2m+1}$. It is easy to identify that there is no diffracted ray from the virtual source $\mathbf{S}'_{2m+1}$ in Region D. The corresponding condition for the existence of the diffracted ray is

$$z_r \leqslant \frac{Hx_r}{mL}+\frac{(m+1)H}{m}. \tag{11}$$

In this case, the maximum number of terms $\hat{m}'$ is simply

$$\hat{m}'=\mathrm{int}\left[\frac{H(L+x_r)}{L(z_r-H)}\right]. \tag{12}$$

We remark that further diffraction by the secondary sources and their images, i.e., multiple diffractions at the barrier edge, are ignored in the current formulation. The expression for the diffracted fields can be simplified considerably if the ground surfaces are acoustically hard such that $\beta_1=\beta_2=0$. In this case, all spherical wave reflection coefficients are equal to 1 in Eqs. (6) and (8).

When the receiver height increases above the shadow zone and enters Region II as indicated in Fig. 2(a), the direct line-of-sight contact will be established with the noise source. As a result, the overall sound fields are composed of not only the diffracted wave but also a direct wave and a wave reflected from the façade. Finally, when the receiver height is raised further, only the direct and diffracted waves can reach the receiver [see Region III of Fig. 2(a)]. From the geometrical consideration, the direct line-of-sight sound field P_1 can be written as

$$P_1=\begin{cases} 0, & \text{in Region I,} \\ e^{ikR_a}/4\pi R_a+Q(\Psi_0,\mathfrak{R},\beta_f)e^{ikR_b}/4\pi R_b, & \\ \quad \text{in Region II,} & \\ e^{ikR_a}/4\pi R_a, & \text{in Region III,} \end{cases} \tag{13}$$

where R_a and R_b are, respectively, the pathlengths of the direct wave and the specularly reflected wave on the façade surface.

Similar zones of the sound field can be identified in the vicinity of the barrier due to the image of the primary noise source in the quarter of $x>0$, $z<0$. Three regions of the sound field can be identified, i.e., Regions I′, II′, and III′, as shown in Fig. 2(b). The direct line-of-sight sound field P_2, due to the image of the primary source, can be written analogously as

$$P_2=\begin{cases} 0, & \text{in Region I}', \\ Q(\Psi'_0,\mathfrak{R},\beta_2)[e^{ikR_c}/4\pi R_c+Q(\Psi'_0,\mathfrak{R},\beta_f) & \\ \quad \times e^{ikR_d}/4\pi R_d, & \text{in Region II}', \\ Q(\Psi'_0,\mathfrak{R},\beta_2)e^{ikR_d}/4\pi R_d, & \text{in Region III}', \end{cases} \tag{14}$$

where R_c and R_d are, respectively, the pathlengths of the specularly reflected waves on the ground surface (in the source side) and that reflected on both the ground and façade surfaces.

As can be seen in this section, different regions in the vicinity of a façade–barrier system have different formulas for calculating the sound pressure levels. They have to be identified carefully by considering the geometrical configuration of the barrier, source, and receiver. By incorporating the identified regions in the façade–barrier system from Figs. 1, 2(a), 2(b), and 4, we can compute the sound field P_T in front of a building façade as follows:

$$P_T = P_D + P_D' + P_1 + P_2, \tag{15}$$

where P_D, P_D', P_1, and P_2 can be calculated by Eqs. (6), (8), (13), and (14), respectively. In fact, the region between the façade and barrier is the region of interest because we wish to investigate the effect of the barrier on the change in the overall sound field in front of a tall building.

III. VALIDATION OF THE RAY MODEL

A. Numerical comparisons

The ray model derived in the last section has been implemented for the computation of sound fields in front of a tall building behind a hard screen. A two-dimensional (2-D) Boundary Element Method (BEM) is also used to predict the acoustic sound field of this façade–barrier system. The BEM model is developed and its details of implementation are described elsewhere.[24] As the BEM model is an accurate numerical scheme to handle the scattering and diffraction of sound by obstacles, the predicted results from the BEM model provide useful benchmarking data to validate the ray method. The computational time of a BEM formulation increases with the number of "boundary" elements. As a rule of thumb, six elements per wavelength are required to ensure an acceptable numerical accuracy, although Marburg[25] has pointed out that use of biquadratic or even higher-order boundary elements can lead to a more accurate numerical solution. Obviously, a traffic noise prediction model with a long barrier along the road in front of tall buildings (often over 100 m in height) requires an exceedingly large number of elements, even for a simple 2-D BEM model let alone for the full 3-D model. Nevertheless, to reduce the required computational time, a 2-D BEM model is used in the present paper to obtain the benchmarking results for numerical comparison with the ray model. The use of a 2-D BEM is justifiable because Ouis[26] has showed theoretically that the effect of wave divergence is insignificant such that the 2-D prediction results for a coherent line source are equivalent to a more general 3-D case of a point source. Moreover, the view of Ouis has been supported by scale model[27] and full scale[28] experimental results using point sources.

In this section we present two typical sets of numerical examples for a reflecting screen on a hard ground surface. In this investigation, a realistic outdoor configuration is used in the BEM and ray models. The barrier height is taken to be 3 m and it is situated at 4 m in front of a building façade. We also assume that the contributions from sound diffracted at the top edge of the façade should be negligibly small if the height ratio of façade to barrier is greater than 7:1, i.e., for the height of the façade of about 20 m or more. Obviously, this height ratio is dependent on the frequency range of interest. Using a larger height ratio in the numerical analysis can lead to more accurate BEM results at the expense of a longer computational time. We have no attempt to optimize the height ratio, but the close agreement of numerical comparisons, see Figs. 5 and 6, between the BEM and ray model predictions justifies our choice of the façade height.

The rectangular coordinates, (x,y,z), are used to represent the situation, assuming that the origin (0.0,0.0,0.0) is at the point where the face of the façade meets the ground surface; the top edge of the barrier will be at (4.0,0.0,3.0). The noise source is fixed at a position of 0.3 m high above the hard ground; see Fig. 1 for the general geometrical configuration of the problem. This is used to simulate the engine exhaust noise of a normal vehicle. It is placed 2 m away from the barrier's outer surface, i.e., at point (6.0,0.0,0.3). Four different receiver positions at coordinates of (1.0,0.0,2.0), (1.0,0.0,4.0), (3.0,0.0,2.0), and (3.0,0.0,4.0) are simulated. These receivers are all located within the shadow zone behind the barrier. A frequency range from 100 to 2000 Hz is used in the numerical comparison. Figure 5 displays the results of using the ray model described in Sec. II and the BEM simulation of relative sound pressure levels with the

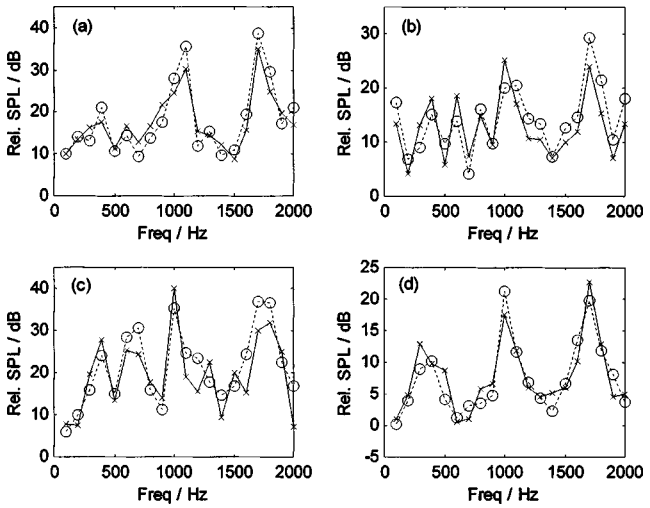


FIG. 5. Theoretical predictions and BEM data for sound propagation of a façade-barrier system. The distance between the building façade and barrier =4.0 m. The barrier height=3.0 m, source height=0.3 m at 2.0 m from the barrier, and the receiver position (the origin is at the base of the barrier where it meets the ground): (a) (1.0,2.0); (b) (1.0,4.0); (c) (3.0,2.0); (d) (3.0,4.0) (dotted line: analytic formulation; solid line: BEM prediction).

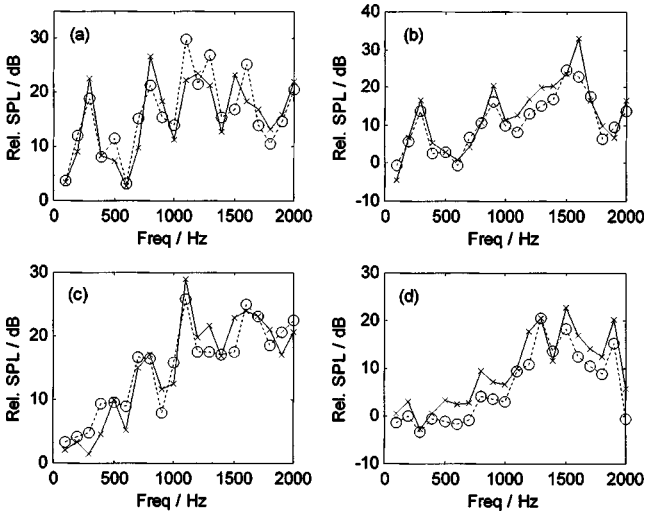


FIG. 6. Theoretical predictions and BEM data for sound propagation of a façade-barrier system. The distance between the building façade and barrier =4.0 m. The barrier height=3.0 m, source height=0.1 m at 4.0 m from the barrier, and the receiver position (the origin is at the base of the barrier where it meets the ground): (a) (1.0,2.0); (b) (1.0,4.0); (c) (3.0,2.0); (d) (3.0,4.0) (dotted line: analytic formulation; solid line: BEM prediction).

reference free-field level. These geometrical configurations are chosen according to the corresponding configuration of the measurement setup shown in the next section. They show good agreement with each other with discrepancies less than 5 dB in general. These discrepancies are due to the numerical deficiency of calculating the BEM data with insufficient ''elements.'' Another simulation is carried out when the source is situated farther away from the barrier surface. It is located 4 m away from the outer barrier's surface and at 0.1 m above the hard ground, i.e., at coordinates (8.0,0.0,0.1). The source can be used to simulate the noise produced near the ground with the interaction with the vehicle. The receivers are kept at the same positions as before between the façade and barrier, i.e., at (1.0,0.0,2.0), (1.0,0.0,4.0), (3.0,0.0,2.0), and (3.0,0.0,4.0). Here, the receivers are located in the shadow and illuminated zones. Figure 6 shows the comparison of the numerical results obtained from the BEM model and the mathematical prediction model developed in the previous section. Again, good agreement between the two curves is achieved with discrepancies less than 5 dB in general. We note that approximately 30 terms are used in the ray model in order to achieve converged results in all predictions shown in Figs. 5 and 6.

B. Experimental validation

The ray model for the prediction of sound pressure levels of the façade–barrier system involves the use of a somewhat intricate set of formulations; cf. Eqs. (1), (13)–(15), which is dependent on the source/receiver position and the relative location of the barrier. To further validate the numerical formulation, indoor experiments were carried out in an anechoic chamber of size 6 m×6 m×4 m. In the measurements, two thin hardwood boards 2 cm thick were used to represent the building façade and barrier surfaces. Both hardwood boards have a nominal width of 2.4 m, but their respective heights are 1.8 and 0.36 m for the simulation of the building façade and the barrier, respectively. The height ratio of the façade to barrier is 5:1 such that we assume the contribution from the sound diffracted at the top edge of the façade does not affect the measured sound fields significantly. Excellent agreements between the experimental results (see Figs. 7 and 8 below) and theoretical predictions justify this assumption.

Two different types of ground surface were simulated in the experiment. Hardwood boards were also used to provide the simulation of a hard ground. All hardwood boards were varnished to create smooth surfaces and were sealed to prevent sound leakage for all measurements. In the experimental measurements, a hard ground was used initially. It was then followed by putting a layer of carpet (about 1 cm thick) on top of the hard ground to create a surface of finite impedance. The barrier base was aligned with the lower face of the carpet. In all experiments, the façade and barrier surfaces were aligned parallel to each other and perpendicular to the ground.

A BSWA TECH MK224 $\frac{1}{2}''$ condenser microphone and a BSWA TECH MA201 preamplifier were used together as the receiver. The microphone was placed between the façade and barrier. A Tannoy driver with a tube with an internal diameter

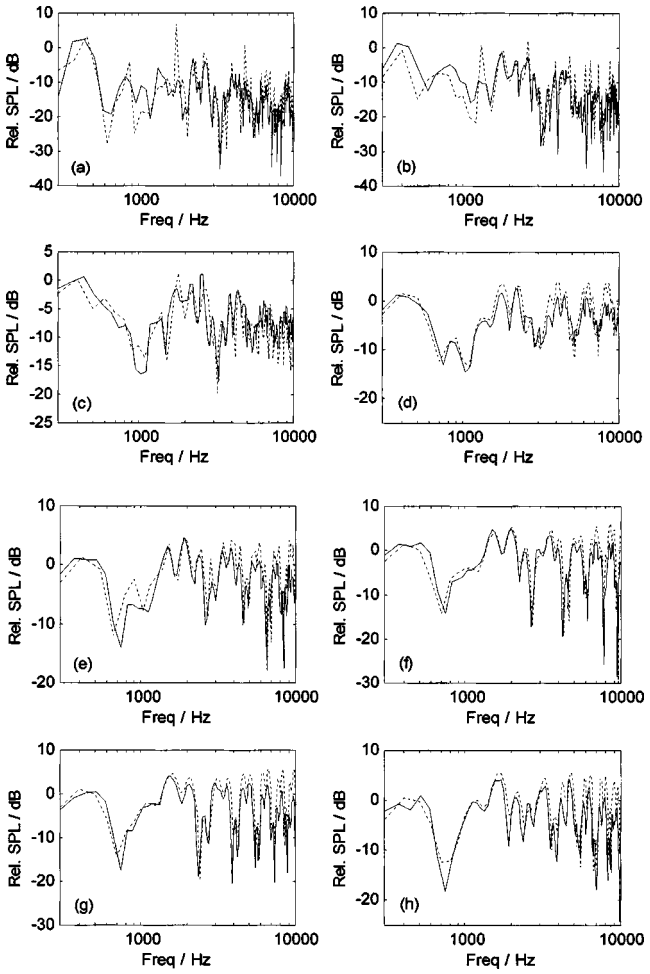


FIG. 7. Experimental data and theoretical predictions for sound propagation of a façade–barrier system. The ground, façade, and barrier are all hard surfaces. The distance between the façade and the barrier=0.79 m. Barrier height=0.36 m. Source height=0.15 m, and the receiver is positioned at 0.395 m from the façade. The height of the receiver from the surface of the ground: (a) 0.2 m; (b) 0.3 m; (c) 0.5 m; (d) 0.6 m; (e) 0.8 m; (f) 0.9 m; (g) 1.0 m; (h) 1.1 m (dotted line: theoretical prediction; solid line: experimental result).

of 3 cm and a length of 1 m was used as a point source in all indoor experiments. The noise source was placed at the opposite side of the barrier. This setup was to simulate the situation in which the barrier was used to protect the noise-sensitive receivers in the building from traffic noise (i.e., to block the line of sight between source and receiver). The Tannoy driver was connected to a maximum length sequence system analyzer (MLSSA) with a MLS card installed in a PC computer.[29] The analyzer was connected to a B&K 2713 amplifier. The MLSSA system was used both as the signal generator for the source and as the signal processing analyzer. The Tannoy driver and the condenser microphone were placed in a fixed position by means of a stand and clamps, and the position of the receiver was adjusted for different sets of measurements. In the experiment, the source and receiver were both located in the same vertical plane mutually perpendicular to the ground and barrier surfaces.

Two sets of experiments were conducted separately above different ground surfaces (hard and carpet-covered). For the hard ground surface, the distance between the façade

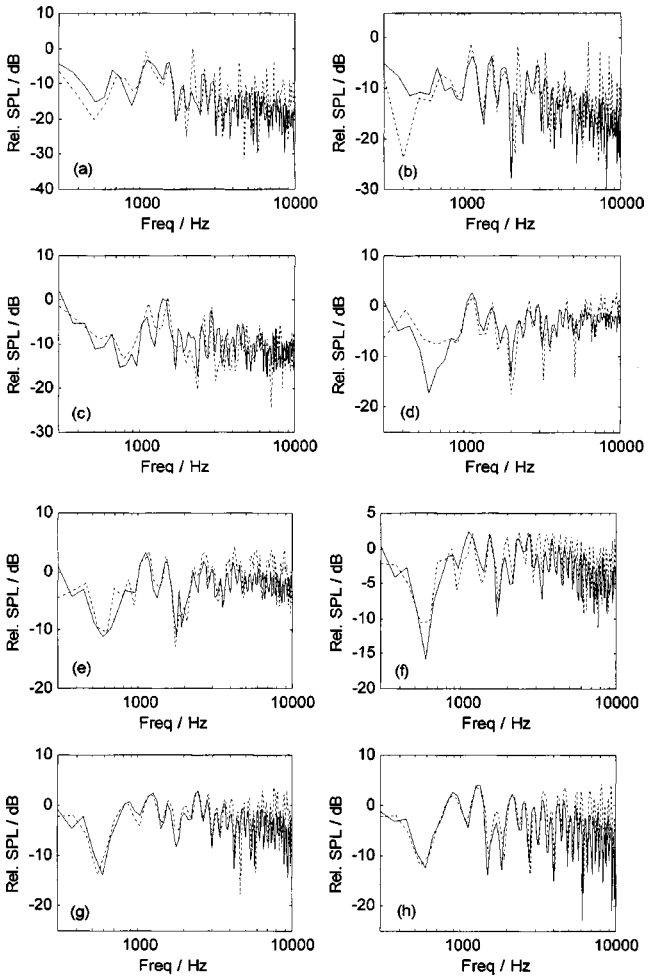


FIG. 8. Experimental data and theoretical predictions for sound propagation of a façade–barrier system. The façade and barrier are of hard surfaces. The hard ground is covered with a layer of carpet. The distance between the façade and the barrier=0.94 m. The barrier height=0.36 m. Source height =0.20 m, and the receiver is positioned at 0.47 m from the façade. The height of the receiver from the surface of the carpet: (a) 0.2 m; (b) 0.3 m; (c) 0.4 m; (d) 0.6 m; (e) 0.7 m; (f) 0.8 m; (g) 0.9 m; (h) 1.1 m (dotted line: theoretical prediction; solid line: experimental result).

and barrier was kept constant at 0.79 m. The receiver was positioned in the middle between these two vertical planes (i.e., 0.395 m from the inner surface of both the façade and barrier). The source height was kept constant at 0.15 m above the ground surface and its distance away from the outer barrier surface was fixed at 0.6 m. The position of the receiver was allowed to vary vertically above the ground surface from 0.2 to 1.1 m. Measurements of the sound field were taken at intervals of 0.1 m. With this geometrical configuration, the receiver position was allowed to vary from locations in the shadow zone to locations in the illumination zone, where direct line-of-sight contact was possible between the source and receiver. Figure 7 displays the experimental results and theoretical predictions of the relative sound pressure levels with the reference free-field level measured at a horizontal distance of 1 m from the source. They both show very good agreement with each other.

For the impedance ground surface, a layer of carpet was put on top of the hardwood board in our indoor experiments. Prior measurements were conducted to characterize the acoustical impedance of the carpet.[30] A two-parameter model[31] was used to predict its effective normalized admittance, as follows:

$$\beta = \frac{1}{0.436(1+i)(\sigma_e/f)^{0.5} + 19.48 i \alpha_e / f}, \tag{16}$$

where σ_e is the effective flow resistivity and α_e is the effective rate of change of porosity with the depth of the material used in the experiment, i.e., carpet. Best-fit parameters with $\sigma_e = 10\,000\ \mathrm{Pa\,s\,m^{-2}}$ and $\alpha_e = 80\ \mathrm{m^{-1}}$ were obtained (see Ref. 30 for typical experimental results of the ground characterization). These paremeter values were used in our subsequent predictions based on the ray method. We remark that Busch *et al.*[32] have reported an improved technique for simultaneously selecting both an optimal scale factor and optimal model materials for indoor scale model experiments. However, there is no attempt to select the most appropriate materials for modeling an outdoor porous ground surface in our present study. However, the use of carpet-covered ground will allow a validation of the ray model by comparing theoretical predictions with precise indoor measurements for ground surfaces with finite impedance.

After the characterization of the ground surface, the second set of experiments was conducted. In these measurements, the separation between the façade and barrier was fixed at 0.94 m. The receiver was placed in the middle between these two vertical planes (i.e., 0.47 m from each of the inner surfaces). The noise source was located on the other side of the barrier. Its position was 0.62 m away from the outer barrier surface and 0.2 m above the carpet covered ground. The position of receiver was raised vertically above the carpet surface from 0.2 to 1.1 m. Experimental results were recorded at 0.1 m intervals. The comparison of the results obtained from the theoretical prediction and experiment is shown in Fig. 8. They show reasonably good agreement for all the measured frequencies. Generally speaking, the developed ray model was able to predict the sound pressure levels in the façade–barrier system.

The total sound pressure levels in the region between the façade and barrier is the summation of the different wave terms produced by the primary and secondary noise sources and their images, as described in Sec. II. For a receiver located below the height of the barrier, the sound pressure level is made up of the multiple reflections of the secondary noise source between the façade and barrier. In this case, it would be ideal to sum up all diffraction terms produced by the secondary source images. In the theoretical prediction using the dimensions of the experimental setup, up to 30 secondary source images are found to be sufficient for predicting the sound pressure levels. However, it should be noted that if the distance between the façade and barrier is very short, a stronger reverberant sound field is formed in the area. Therefore, a large number of secondary source images will be required for the computation. Nevertheless, in a more realistic environment where the distance between the façade and barrier is much greater—of the order of 10 m or more, only a few of these terms are needed for the computation of the total sound field, as higher-order terms are negligibly small. This is due

to the fact that these higher-order terms are weaker in strength when they are located farther from the receiver. Here, we remark that no attempts have been made to explore the effect of absorptive barriers in the current investigations. This area will be an area of future studies.

IV. APPLICATION OF THE RAY MODEL IN URBAN ENVIRONMENTS

With the numerical and experimental validations, we are confident about further applying the ray model to predict the sound field of a façade–barrier system so that we can assess the acoustic performance of a barrier in high-rise cities. Indeed, a barrier aligned parallel to the front of a building is often used to block the noise produced by the traffic on the other side of the barrier. Over the years, it has been proven to be an effective measure. However, little attention has been focused on the distribution of the sound field between the façade and barrier. Therefore, we shall investigate these effects using the ray model described in Sec. II for a variety of source heights and distances away from the barrier.

In the first simulation, the setup of the façade–barrier system is similar to that of Fig. 1. The receiver is placed between the façade and barrier and the source is situated on the other side of the barrier. The source and receiver are located in the same plane in which they are mutually perpendicular to the ground and building façade. The ground, façade, and barrier all have hard and reflecting surfaces. The façade is infinitely high compared with the height of the barrier, and hence the diffraction at the top of the building is ignored. The barrier height is 4 m and is situated 6 m away from the building façade. The source is kept at a distance of 2 m away from the outer barrier's surface (which is the surface facing away from the façade) and its height chosen at 0.5 and 1.5 m above the ground for each simulation.

Figure 9 displays the results of the prediction of the barrier insertion loss. It is calculated as the difference in the A-weighted sound pressure level with and without the presence of the barrier. A white noise spectrum is used for the A-weighting process (see Ref. 31 for details). The region of sound pressure levels in front of the façade of up to 6 m away from its surface and from 0 to 16 m above ground is investigated. The two different heights of noise source considered all obtained a similar range of values of sound pressure level attenuation in the shadow zone from about 0 to 30 dB(A), when the barrier was erected. The most obvious feature of these figures is the variation of the size of the shadow zone behind the barrier. As the position of the source is lifted, the area of the shadow zone on the other side of the barrier decreases. Therefore, people who live on high floors will be protected by the barrier from the noise produced near the ground, such as noise from tires. If the source height is higher, for example, the engine exhaust noise from a large wagon that has the exhaust pipe situated at the top of the truck, the direct line-of-sight contact can be established, even for receivers located at the low to middle floors in a building. The maximum noise attenuation level of about 35 dB(A) is obtained close to the ground behind the barrier when the source is at the lowest position (i.e., 0.5 m). This is because it is easier for the sound waves from the higher source position to travel around and diffract at the top edge of the barrier. Hence, the sound pressure levels attained on the other side of the barrier will be slightly higher compared with the case where the source position is close to the ground. Moreover, the noise attenuation level is higher in the vicinity of the barrier because it is deep inside the shadow zone. As the receiver proceeds toward the façade, the attenuation level becomes less.

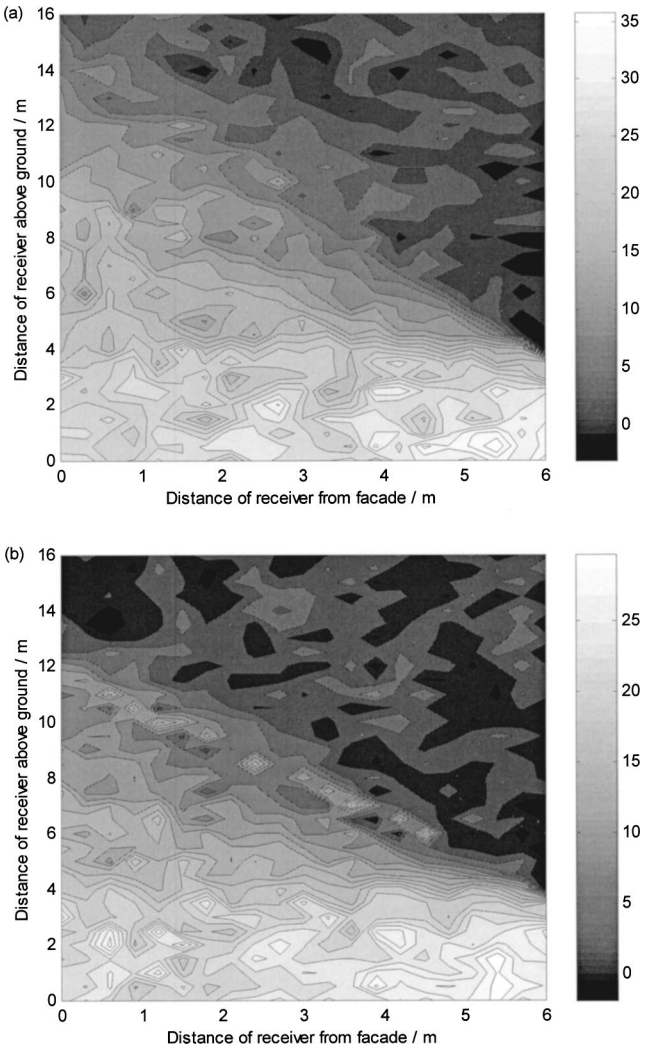


FIG. 9. Prediction of the barrier insertion loss of a façade–barrier system. The façade is erected on the left-hand side. The distance between the façade and the barrier is 6 m. The barrier height is 4 m. The source is located at 2 m from the outer barrier's surface. Source height: (a) 0.5 m; (b) 1.5 m.

Another case is considered for the determination of the sound field of a façade–barrier system when the source is moving away from the barrier. The source is fixed at a height of 0.5 m and its distance away from the outer barrier surface is 2 and 4 m, respectively, for each simulation. Figure 10 shows the prediction of the barrier insertion loss in the same region discussed above (a distance of up to 6 m from the façade surface and from 0 to 16 m above-ground). The two distances considered show similar noise attenuation, from about 0 to 30 dB(A) in the shadow zone, with the presence of a barrier. A slightly higher sound pressure attenuation of 35 dB(A) is found when the source is closest to the barrier at 2 m [Fig. 10(a)]. As expected, the area of the shadow zone

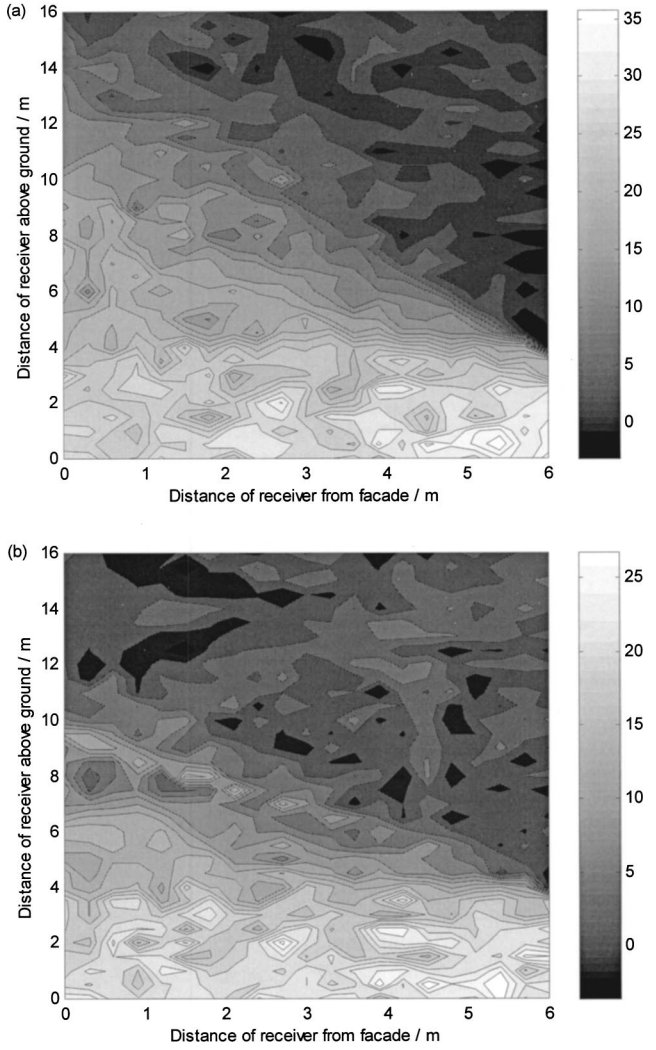


FIG. 10. The prediction of the barrier insertion loss of a façade–barrier system. The façade is erected on the left-hand side. The distance between the façade and the barrier is 6 m. The barrier height is 4 m. The source height is situated at 0.5 m above the ground, and at (a) 2.0 m; (b) 4.0 m from the outer barrier's surface.

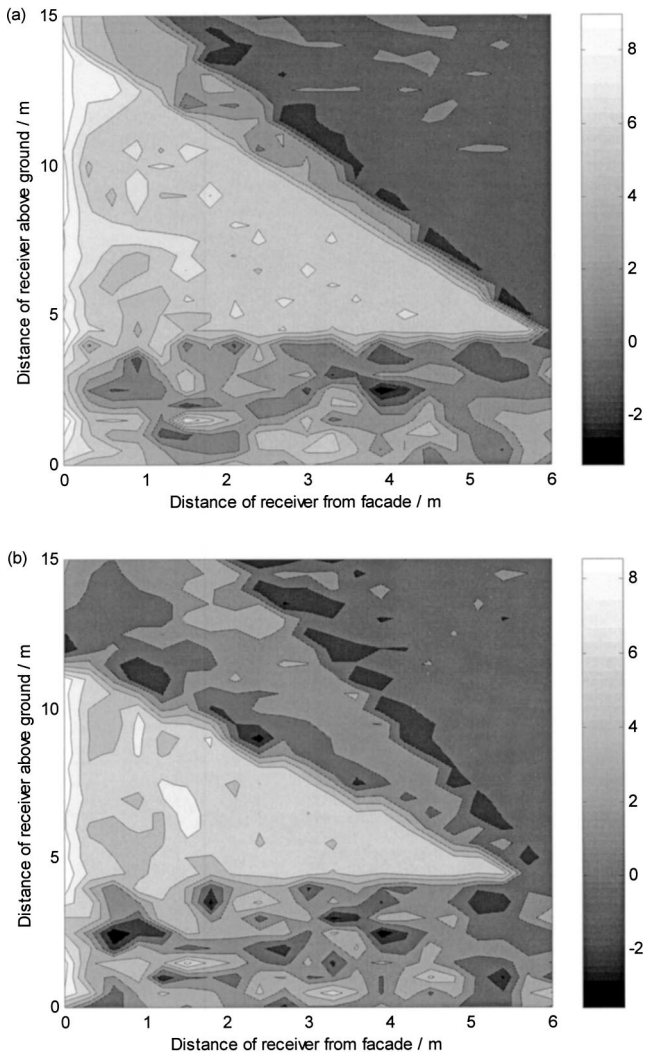


FIG. 11. The prediction of the insertion loss degradation of the building façade in a façade–barrier system. The barrier is erected on the right-hand side. The configuration of the noise source, receiver, barrier, and façade is the same as described in the setup in Fig. 9.

behind the barrier diminishes with an increase in the distance of source away from the barrier. Again, better noise attenuation is found close to the barrier and ground surface, i.e., deep inside the shadow zone. Therefore, the residents on higher floors are better protected by the barrier when traffic is operating close to it.

The presence of a building next to a road leads to the reflection of noise on the façade's surface. It was shown earlier in this section that the insertion of a barrier creates a shadow zone that can protect the residents of low to middle floors from traffic noise. In the region close to the ground [region I in Fig. 2(a)], multiple reflections of the diffracted sound occur between the façade and barrier. Above the height of the barrier within region I, diffracted sound waves impinge on the façade and barrier surfaces before reaching the receiver. Above the shadow zone, direct, reflected, and diffracted sound waves contribute to the total sound field. The presence of a building façade seems to be rather significant in determining the sound fields in front of it because of its nature to reflect noise. Here, we investigate the effects of the sound pressure levels behind a barrier due to the presence of a tall building, i.e., the insertion loss degradation[20] of the building façade. The insertion loss degradation (DIL) represents the difference in sound pressure levels before and after the erection of the building. It can be calculated using Eq. (13):

$$\mathrm{DIL} = 20 \log \left| \frac{P_{\text{total,ground+barrier}}}{P_{\text{total,ground+facade+barriers}}} \right|. \tag{17}$$

Two simulations are considered to predict the insertion loss degradation of a building façade (assumed to be infinitely high) behind a barrier. A barrier of 4 m high is aligned parallel at 10 m in front of a building. A noise source is fixed at 2 m from the outer barrier's surface. The receiver position varies inside the region between the façade and barrier, that is, 0–6 m from the inner barrier surface and 0–15 m above the hard ground. In the first simulation, the configuration of the noise source, receiver, barrier, and building façade is the same as the setup described in Fig. 9. The two contour plots in Fig. 11 show the insertion loss degradation of the building façade, when the noise source is 0.5 and 1.5 m above the

hard ground, respectively. This is calculated as the difference in the A-weighted sound pressure levels with and without the consideration of the building façade. The range of the overall insertion loss degradation is from about −6 to 8 dB(A). Positive readings indicate an increase in the sound pressure levels due to insertion of the building façade. From the two plots shown in Fig. 11, the maximum degradation occurs near the façade surface within the shadow zone. A receiver situated in the region above the height of the barrier and inside the shadow zone also experiences quite a large insertion loss degradation of about 5 to 8 dB(A), in general. The area of this insertion loss degradation is inversely proportional to the height of the noise source. Furthermore, the magnitude of the degradation is slightly more apparent when the position of the noise source is high. Hence, a noisier environment is created above the level of barrier. In the region close to the ground, the change in insertion loss degradation is smaller. It ranges from about −2 to 3 dB(A), which indicates that the increase in sound pressure levels due to the multiple reflections of sound between the façade and barrier is not too significant. At some positions, especially near the barrier, a negative insertion loss degradation is found. This implies that the presence of the building leads to a quieter environment at those locations.

In the region above the shadow zone, there is almost no change in the insertion loss degradation, whether or not the building is erected. This indicates that the direct line of sight between the source and receiver is dominant in the overall sound field. A more negative insertion loss degradation is found just above the shadow zone. The variation of the sound field is due to the interference effects between the diffracted waves at the barrier's top edge and the reflected waves from the façade's surface. It is noteworthy that significant variability in the barrier insertion loss over short spatial scales is evident in Figs. 9–11. This is an artifact of the calculation method. The contours in these figures will have smoother appearance if more points are used in the designated regions or more frequencies had been used in the calculations.

V. CONCLUSIONS

Our aim in the present study is to investigate the acoustical performance of noise barriers in a high-rise city. A ray model was developed for a barrier placed closed to high-rise buildings in many urban areas. The ray model was validated by extensive numerical studies and precise indoor experimental measurments. Simulations of sound fields were then carried out for different configurations in typical urban environments. In particular, the sound field in the vicinity of a façade–barrier system was investigated in which a barrier was aligned parallel to the front of a building. The receiver was positioned between the building façade and the barrier and a noise source was situated on the other side of the barrier. The secondary noise source was formed at the top edge of the barrier due to the diffraction at this position. A series of secondary source images were formed due to the multiple reflections between the façade and barrier. The sound pressure levels obtained by the receiver when it is below the top edge of the barrier consists of the summation of the diffracted sound waves of the secondary source images. When the receiver is lifted up to a position above the shadow zone of the barrier, the direct sound wave from the noise source to the receiver and/or noise reflection from the ground surface may be established. Different regions of the sound field can be identified in the vicinity of the façade–barrier system. The size and position of these regions are dependent on the configuration of the barrier, source, and receiver.

In predicting the insertion loss of a noise barrier, the maximum noise attenuation is obtained close to the ground immediately behind the barrier, deep within the shadow zone. In general, slightly less attenuation of the noise is found when the receiver is close to the building façade. The erection of a barrier close to the traffic has a significant effect for the people who live on the higher floors of a building. The effect is greatest for noise produced close to the ground, such as tire noise, as this creates a bigger shadow zone behind the barrier.

The presence of a building façade behind a noise barrier causes an increase in sound pressure levels in the region between these two vertical planes, especially close to the façade surface and above the barrier's top edge within the shadow zone. When a receiver is close to the ground, multiple reflections, which take place between the façade and barrier, lead to an increase in the sound field. The effect due to reflection on the façade surface also increases the sound pressure levels in the region within the shadow. However, above the shadow zone region, the change in sound pressure levels is not significant because of the domination of the direct sound waves.

ACKNOWLEDGMENTS

This research was supported in part by the Research Grants Council of the Hong Kong Special Administrative Region (Project No. PolyU 5151/99E) and the Hong Kong Polytechnic University. The financial support of the Research Committee of the Hong Kong Polytechnic University is gratefully acknowledged for the award of a Research Studentship to one of the authors (S.H.T.). The manuscript was prepared while K.M.L. was on a study leave at the Department of Mechanical, Materials, and Mechatronics Engineering, The University of Western Australia (UWA). One of the authors (K.M.L.) wishes to thank Dr. Jie Pan (UWA), Professor W. S. Siu (PolyU), and Professor R. M. C. So (PolyU) for the opportunity to visit UWA.

[1] R. H. Bolt and E. A. G. Shaw, "Initial program of the Co-ordinating Committee on environmental acoustics," J. Acoust. Soc. Am. **50**, 443–445 (1971).
[2] W. E. Scholes, "Noise reduction by barriers," Proceedings of the British Acoustics Society Meeting, Paper 70/67, April 1970.
[3] U. J. Kurze and G. S. Anderson, "Sound attenuation by barriers," Appl. Acoust. **4**, 35–53 (1971).
[4] Z. Maekawa, K. Fujiwara, and M. Morimoto, "Some problems of noise reduction by barriers," *Symposium On Noise Prevention*, Miskolc, Paper No. 4.8, 1971.
[5] U. J. Kurze, "Noise reduction by barriers," J. Acoust. Soc. Am. **55**, 504–518 (1974).
[6] A. D. Pierce, "Diffraction of sound around corners and over wide barriers," J. Acoust. Soc. Am. **55**, 941–944 (1974).

[7] Z. Mawkawa, ''Noise reduction by screens,'' Appl. Acoust. **1**, 157–173 (1968).
[8] S. Hayek, ''Mathematical modeling of absorbent highway noise barriers,'' Appl. Acoust. **31**, 77–100 (1990).
[9] J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi, *Electromagnetic and Acoustic Scattering by Simple Shapes* (North-Holland, Amsterdam, 1969).
[10] J. Nicolas, T. F. W. Embleton, and J. E. Piercy, ''Precise model measurements versus theoretical prediction of barrier insertion loss in the present of the ground,'' J. Acoust. Soc. Am. **73**, 44–54 (1983).
[11] A. L'Espérance, J. Nicolas, and G. A. Daigle, ''Insertion loss of absorbent barriers on ground,'' J. Acoust. Soc. Am. **86**, 1060–1064 (1989).
[12] E. M. Salomons, A. C. Geerlings, and D. Duhamel, ''Comparison of a ray model and a Fourier-Boundary Element Method for traffic noise situations with multiple diffractions and reflections,'' Acustica **83**, 35–47 (1997).
[13] Department of Transport and Welsh Office, *Calculation of Road Traffic Noise*, HMSO, 1988.
[14] T. M. Barry and J. A. Reagan, FHWA Highway Traffic Noise Prediction Model, U.S. Federal Highway Administration, Report FHWA-RD-7-108, Washington, DC, 1978.
[15] Y. Sakurai, E. Walerian, and H. Morimoto, ''Noise barrier for a building façade,'' J. Acoust. Soc. Jpn. (E) **11**, 257–265 (1990).
[16] E. Walerian, R. Janczur, and M. Czechowicz, ''Sound levels forecasting for city-centers. Part I: Sound level due to a road within an urban canyon,'' Appl. Acoust. **62**, 359–380 (2001).
[17] W. F. Cheng and C. F. Ng, ''The acoustic performance of an inclined barrier for high-rise residents,'' J. Sound Vib. **242**, 295–308 (2001).
[18] L. Godinho, J. Antonio, and A. Tadeu, ''3D sound scattering by rigid barriers in the vicinity of tall buildings,'' Appl. Acoust. **62**, 1229–1248 (2001).
[19] L. L. Beranek and I. L. Ver, *Noise and Vibration Control Engineering* (Wiley Interscience, New York, 1992).
[20] R. Panneton, A. L'Espérance, and G. A. Daigle, ''Development and validation of a model predicting the performance of hard or absorbent parallel noise barriers,'' J. Acoust. Soc. Jpn. (E) **14**, 251–258 (1993).
[21] W. J. Hadden and A. D. Pierce, ''Diffraction of sound around corners and over wide barriers,'' J. Acoust. Soc. Am. **69**, 1266–1276 (1981).
[22] M. Abramowitz and A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover, New York, 1970), Chap. 7, p. 300.
[23] K. Attenborough, ''Review of ground effects on outdoor sound propagation from continuous broadband sources,'' Appl. Acoust. **24**, 289–319 (1988).
[24] G. R. Watt, S. N. Chandler-Wilde, and P. A. Morgan, ''The combined effects of porous asphalt surfacing and barriers on traffic noise,'' Appl. Acoust. **58**, 351–377 (1999).
[25] S. Marburg, ''Six boundary elements per wavelength: Is that enough?'' J. Comput. Acoust. **10**, 25–51 (2002).
[26] D. Ouis, ''Noise shielding by simple barriers: Comparison between the performance of spherical and line sound sources,'' J. Comput. Acoust. **8**, 25–51 (2000).
[27] D. C. Hothersall, S. N. Chandler-Wilde, and N. M. Hajmirzae, ''Efficiency of single noise barriers,'' J. Sound Vib. **146**, 303–321 (1991).
[28] G. R. Watt, D. H. Crombie, and D. C. Hothersall, ''Acoustic performance of new designs of traffic noise barriers: full scale tests,'' J. Sound Vib. **177**, 289–305 (1994).
[29] D. D. Rife and J. Vanderkooy, ''Transfer-function measurement with maximum length sequences,'' J. Audio Eng. Soc. **37**, 419–444 (1989).
[30] S. H. Tang and K. M. Li, ''The prediction of façade effects from a point source above an impedance ground,'' J. Acoust. Soc. Am. **110**, 278–288 (2001).
[31] K. Attenborough, ''Ground parameter information for propagation modelling,'' J. Acoust. Soc. Am. **92**, 418–427 (1992).
[32] T. A. Busch and M. R. Hodgson, ''Improved method for selecting scale factors and model materials for scale modeling of outdoor sound propagation,'' J. Sound Vib. **243**, 173–181 (2001).

Robustness of multichannel equalization in an acoustic reverberant environment

Fotios Talantzis[a] and Darren B. Ward[b]
Department of Electrical and Electronic Engineering, Imperial College of Science, Technology and Medicine, Exhibition Road, London SW7 2BT, United Kingdom

(Received 25 October 2002; revised 20 March 2003; accepted 2 June 2003)

A theoretical framework is established, for the robustness of multichannel sound equalization in reverberant environments. Using results from statistical room acoustics, a closed-form expression is derived that predicts the degradation in performance of an equalization system as the sound source moves from its nominal position inside the enclosure. The presented analysis also provides means of identifying the performance bounds that can be expected when using such a system in an actual room. Using extensive computer simulations, the effect of physical parameters such as the relative positions of the source and the receivers, as well as effects of different design parameters are investigated. Based on the conditions imposed by these parameters, it is shown that, depending on the array geometry and the exact form of the equalizers, slight performance gains can be expected as the number of receivers is increased. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1594189]

PACS numbers: 43.55.Br, 43.60.Cg, 43.38.Hz [MK]

I. INTRODUCTION

The problem of acoustic equalization in a reverberant environment arises in two classes of application. First, when a microphone cannot be placed close to the mouth of someone speaking, the signal received by the microphone can be equalized to remove distortion caused by the room transfer characteristics. Alternatively, when a loudspeaker is used to deliver sound to a listener, an equalizer can be used to prefilter the source signal to compensate for its subsequent room transmission.

The sound wave produced by an acoustic source in a room propagates through a direct path and a set of multiple reflections that impinge on the walls at random angles while traveling with the speed of sound c (generally defined as being equal to 343 m/s at room temperature). The receiver first detects the direct sound, followed by reflections of the signal off nearby surfaces, which are called early echoes. After a few tens milliseconds, the number of the reflected waves becomes very large, and the remainder of the reverberant decay is characterized by a dense collection of echoes traveling in all directions, whose intensity is relatively independent of the location within the room. This is called diffuse reverberation. In the frequency domain, the nature of this multipath distortion is the appearance of peaks and troughs in the room transfer function. Applications requiring a high-quality, undistorted version of the original source signal must attempt to equalize the undesired multipath effects of reverberation.

Assuming that one has designed an equalizer that performs this task for a specific source and receiver location, the question arises as to how well this equalizer performs if the source (or receiver) moves. This question was addressed by Radlovic *et al.*,[1] who showed that movement by as little as a few tenths of a wavelength can render the equalizer ineffective. However, the analysis in Ref. 1 was restricted to the case of a single source and single receiver.

The performance of a multichannel equalizer (with a single source and multiple receivers) was recently addressed to some extent by Bharitkar *et al.* in Refs. 2 and 3. They considered a specific class of equalizers, designed such that the impulse response is measured at a number of locations, and the equalizer is chosen to compensate for the spatially averaged impulse response. The analysis presented was restricted to a geometry where the source and receivers were confined to lie in a plane.

In this paper we aim to generalize these previous studies by considering the robustness of a very general class of multichannel equalizers. Note that we are not specifically concerned with how to design such equalizers; rather, our focus is on the fundamental question of whether the performance of multichannel equalizers is robust enough to make them feasible in a time-varying environment where the source location cannot be fixed. Specifically, we derive a closed-form expression for the expected degradation in performance when the source moves from its nominal position (where the multichannel equalizer performs well). Initially, we consider the performance of exact equalizers (which compensate precisely for the reverberant room transfer functions), and then extend our results to the case of beamformer-based equalizers. Next, we attempt to quantify the performance of an ''optimum'' equalizer, which is optimum in the sense that it maximizes robustness. We also consider the effect of design parameters such as the number of receivers and the geometry of the receiver array, and thereby provide design guidelines on how to maximize the robustness of multichannel equalizers.

Because the acoustic behavior in real rooms is too complex to model explicitly, as in Refs. 1 and 2, we make use of

[a]Electronic mail: fotios.talantzis@ic.ac.uk
[b]Electronic mail: d.ward@ic.ac.uk

the classical tools of statistical room acoustics (SRA). SRA provides a statistical description of the transfer function between a source and receiver in terms of a few key quantities. More specifically, SRA considers the distribution of amplitudes and phases of individual plane waves being so close to random that the sound field can be considered uniform throughout the volume of the room. The validity of this model is subject to a set of conditions that must be satisfied to ensure the accuracy of calculations.[4] Our analysis therefore implicitly assumes that the following conditions hold:[1] (i) the dimensions of the room are relatively large compared to the wavelength; (ii) the frequencies considered exceed the Schroeder frequency $f_s = 2000\sqrt{T_{60}/V}$, where V is the room volume (in m^3) and T_{60} is the reverberation time in seconds (defined as the time for the reverberation level to decay to 60 dB below the initial level); and (iii) the source and receivers are at least one half-wavelength away from the walls, where $\lambda = c/f$ is the wavelength and f is the frequency of the source signal.

A. Notation

Throughout this paper we use the following notation. Matrices and vectors are represented by upper and lower case bold face, respectively, e.g., $\mathbf{A}$ and $\mathbf{a}$. The (n,m)th element of a matrix is denoted by $[\mathbf{A}]_{n,m}$, and similarly the nth element of a vector is $[\mathbf{a}]_n$. The symbol $j = \sqrt{-1}$ is used to denote the imaginary part of a complex number. The vector 2-norm of a vector $\mathbf{a}$ is denoted by $\|\mathbf{a}\|$. Matrix transpose is denoted by ^{T}, and Hermitian (complex conjugate) transpose by ^{H}.

II. PROBLEM FORMULATION

Consider a source located in a reverberant room at a position $\mathbf{y}_1$, which we hereafter refer to as the *nominal source position*. Without loss of generality we assume a Cartesian coordinate system with $\mathbf{y}_1 = [0,0,0]^T$ at the origin. Let $G(\mathbf{y}_1, \mathbf{x}_n)$ denote the complex steady-state transfer function (TF) from $\mathbf{y}_1$ to a receiver located at the arbitrary position

$$\mathbf{x}_n = [\mathcal{X}_n, \mathcal{Y}_n, \mathcal{Z}_n]^T. \tag{1}$$

Although $G(\mathbf{y}_1, \mathbf{x}_n)$ is a function of frequency, to simplify notation we will not express this frequency dependence explicitly. Assume that the transmission path between $\mathbf{y}_1$ and $\mathbf{x}_n$ is equalized by a causal equalizer

$$H(\mathbf{x}_n) = \frac{e^{j2\pi f \tau_n}}{G(\mathbf{y}_1, \mathbf{x}_n)}, \tag{2}$$

where τ_n is a delay chosen to ensure causality.

The robustness of this single-channel system to movement of the source was recently considered in Ref. 1 (although the equalizer used was noncausal).

Now, consider the general N-channel equalization system shown in Fig. 1, in which the signal at the nth receiver is filtered by an equalizer $H(\mathbf{x}_n)$ as in Eq. (2), and is weighted by a scalar constant w_n (one obvious choice for the scalar weights is $w_n = 1/N$, $\forall n$). These filtered and weighted signals are then summed to form the output signal.

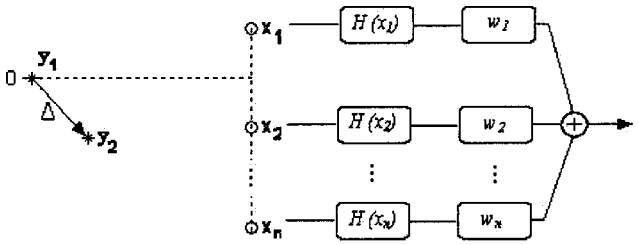


FIG. 1. Block diagram of an N-channel equalization system with the source placed at the origin $\mathbf{y}_1 = [0,0,0]^T$ and N receivers placed at positions $\mathbf{x}_n = [\mathcal{X}_n, \mathcal{Y}_n, \mathcal{Z}_n]^T$, $n = 1,...,N$.

Without loss of generality, assume that the causal delay on each equalizer is the same, i.e., $\tau_n = \tau_0$, $\forall n$. Ideally, with the source in the nominal position, the overall transfer function from the source to the equalizer output is a delay, which we will set to τ_0 (again, without loss of generality).

Assume the source moves to a position $\mathbf{y}_2$ that is a distance Δ from the nominal position $\mathbf{y}_1$, i.e., $\Delta = \|\mathbf{y}_1 - \mathbf{y}_2\|$ (see Fig. 2). One can quantify the sensitivity of this multichannel equalization system in terms of the *mean-squared error* (MSE) function, defined as

$$\mathrm{MSE}(\Delta) \triangleq E\left\{ \left| \sum_{n=1}^{N} w_n G(\mathbf{y}_2, \mathbf{x}_n) H(\mathbf{x}_n) - e^{j2\pi f \tau_0} \right|^2 \right\}, \tag{3}$$

where

$$\mathbf{y}_2 = \Delta[\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta]^T \tag{4}$$

is the displaced source position, and

$$E\{g(\mathbf{y}_2)\} = \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} g(\mathbf{y}_2)\sin\theta\, d\theta\, d\phi \tag{5}$$

is the spatial expectation operator [with $g(\cdot)$ an arbitrary function of the displaced position]. Hence, Eq. (3) measures the expected degradation in performance when the source moves a distance Δ from the nominal position *in an arbitrary direction*.

In the sequel we derive a closed-form expression for the MSE (3), and use this expression to investigate various aspects of the multichannel equalization problem.

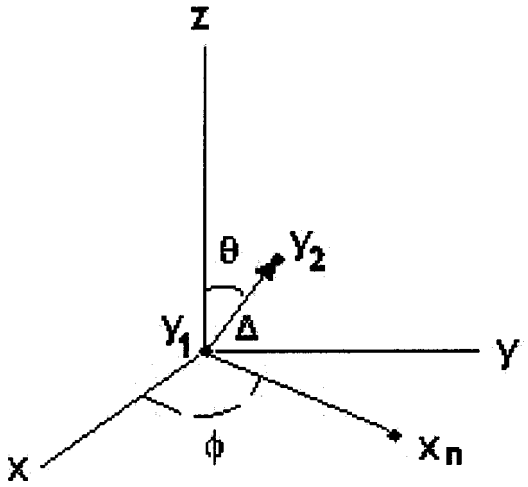


FIG. 2. Geometry used for the analysis of the multichannel equalizer.

III. ROBUSTNESS ANALYSIS

A. Exact channel equalization

The main result of this paper is summarized in the following theorem.

Theorem 1: Let $R_n = \|\mathbf{y}_1 - \mathbf{x}_n\|$ be the distance between the nominal source position $\mathbf{y}_1$ and the nth receiver position $\mathbf{x}_n$. Let $k = 2\pi/\lambda$ denote the wave number (with $\lambda = c/f$ being the wavelength). If the source is displaced from the nominal position by a distance Δ in an arbitrary direction, then the MSE (3) at a given frequency is

$$\mathrm{MSE}(\Delta) \approx \mathbf{w}^H \mathbf{Q}_\Delta \mathbf{w} - 2\frac{\sin k\Delta}{k\Delta}\mathrm{Re}\{\mathbf{w}^H \mathbf{1}\} + 1, \tag{6}$$

where $\mathrm{Re}\{\cdot\}$ denotes the real part of a complex quantity, $\mathbf{1}$ denotes the N vector of ones

$$\mathbf{w} = [w_1^* \quad w_2^* \quad \cdots \quad w_n^*]^T, \tag{7}$$

is the vector of weights, and $\mathbf{Q}_\Delta$ is a Hermitian symmetric matrix whose diagonal and off-diagonal elements are, respectively,

$$[\mathbf{Q}_\Delta]_{n,n} = \frac{\gamma_n \frac{R_n}{2\Delta}\ln\left|\frac{R_n+\Delta}{R_n-\Delta}\right| + 1}{\gamma_n + 1} \tag{8}$$

and

$$[\mathbf{Q}_\Delta]_{\substack{n,m \\ n\neq m}} = \frac{\sqrt{\gamma_n\gamma_m}e^{jk(R_n - R_m + \Delta^2\mu_{nm})}\frac{\sin(k\Delta\zeta_{nm})}{(k\Delta\zeta_{nm})} + \frac{\sin(k\|\mathbf{x}_n - \mathbf{x}_m\|)}{k\|\mathbf{x}_n - \mathbf{x}_m\|}}{\sqrt{\gamma_n\gamma_m}e^{jk(R_n - R_m)} + \frac{\sin(k\|\mathbf{x}_n - \mathbf{x}_m\|)}{k\|\mathbf{x}_n - \mathbf{x}_m\|}}, \tag{9}$$

where

$$\zeta_{nm} = \frac{\sqrt{(R_m\mathcal{X}_n - R_n\mathcal{X}_m)^2 + (R_m\mathcal{Y}_n - R_n\mathcal{Y}_m)^2 + (R_m\mathcal{Z}_n - R_n\mathcal{Z}_m)^2}}{R_n R_m}, \tag{10}$$

and

$$\mu_{nm} = \frac{1}{2}\left(\frac{1}{R_n} - \frac{1}{R_m}\right). \tag{11}$$

The direct-to-reverberant energy ratio at the nth receiver, γ_n, is defined as

$$\gamma_n \triangleq \frac{1}{16\pi R_n^2}\left(\frac{S\alpha}{1-\alpha}\right), \tag{12}$$

where S is the total surface area of the walls and α is the average absorption coefficient of the walls.

Proof: See the Appendix.

As expected, if the source is at the nominal source position, i.e., $\Delta = 0$, the MSE (6) reduces to zero, provided that the equalizer weights are normalized to sum to unity. Also, in the special case where $N = 1$, Eq. (6) reduces to the expression give in Ref. 1.

Equation (6) demonstrates that the MSE is a function of both the source displacement Δ, and several environmental variables (which include the relative positions of the source and the receivers, the weights of the equalizers, and the room parameters). The diagonal elements of $\mathbf{Q}_\Delta$ express the autocorrelation of the signals detected at any one of the receivers, and the off-diagonal elements express the cross correlations between each pair of receivers. So, the MSE becomes larger as the values of these correlations become larger. The weight vector $\mathbf{w}$ also affects the overall system performance by determining the contribution of each receiver equalizer to the overall system equalization.

B. Direct-path equalization

1. MSE of direct-path equalizer

In deriving the MSE in Eq. (6), we assumed that the room TF was equalized exactly according to Eq. (2). As a special case, we now consider a system in which only the direct-path TF is equalized, i.e.,

$$H_{\mathrm{DP}}(\mathbf{x}_n) = \frac{e^{j2\pi f\tau_0}}{G_d(\mathbf{y}_1, \mathbf{x}_n)}, \tag{13}$$

where

$$G_d(\mathbf{y}, \mathbf{x}) \triangleq \frac{e^{jk\|\mathbf{y}-\mathbf{x}\|}}{4\pi\|\mathbf{y}-\mathbf{x}\|}, \tag{14}$$

is the free-space Green's function between $\mathbf{y}$ and $\mathbf{x}$.

Corollary 1: If the equalizers are designed to equalize the direct-path transfer function according to Eq. (13), then the MSE (3) is

$$\mathrm{MSE}_{\mathrm{DP}}(\Delta) \approx \mathbf{w}^H \mathbf{Q}_\Delta^{\mathrm{DP}} \mathbf{w} - 2\frac{\sin k\Delta}{k\Delta}\mathrm{Re}\{\mathbf{w}^H \mathbf{1}\} + 1, \tag{15}$$

where $\mathbf{Q}_\Delta^{\mathrm{DP}}$ is a Hermitian symmetric matrix whose diagonal and off-diagonal elements are, respectively,

$$[\mathbf{Q}_\Delta^{\mathrm{DP}}]_{n,n} = \frac{R_n}{2\Delta}\ln\left|\frac{R_n+\Delta}{R_n-\Delta}\right| + \frac{1}{\gamma_n}, \tag{16}$$

and

$$[\mathbf{Q}_{\Delta}^{\mathrm{DP}}]_{n\neq m,m} = e^{jk\Delta^2\mu_{nm}}\frac{\sin(k\Delta\zeta_{nm})}{(k\Delta\zeta_{nm})} + \frac{\frac{\sin(k\|\mathbf{x}_n-\mathbf{x}_m\|)}{k\|\mathbf{x}_n-\mathbf{x}_m\|}}{\sqrt{\gamma_n\gamma_m}e^{jk(R_n-R_m)}}. \tag{17}$$

In contrast with Eq. (6), for direct-path equalization the MSE does not reduce to zero when the source is at its nominal location. Instead, the minimum MSE is

$$\mathrm{MSE}_{\mathrm{DP}}(\Delta=0)\approx\mathbf{w}^H\mathbf{Q}_0^{\mathrm{DP}}\mathbf{w}-2\,\mathrm{Re}\{\mathbf{w}^H\mathbf{1}\}+1, \tag{18}$$

where

$$[\mathbf{Q}_0^{\mathrm{DP}}]_{n,m}=1+\frac{\frac{\sin(k\|\mathbf{x}_n-\mathbf{x}_m\|)}{k\|\mathbf{x}_n-\mathbf{x}_m\|}}{\sqrt{\gamma_n\gamma_m}e^{jk(R_n-R_m)}}, \quad \forall n,m. \tag{19}$$

2. Relation to beamforming

One can interpret the direct-path equalizer (13) as a beamformer. The well-known delay-and-sum beamformer (DSB) delays the signal received at each sensor such that a signal from the chosen look direction adds constructively.[5] Since steered beamforming is often used as the basis of microphone arrays[6] in reverberant rooms, we now apply our results to this particular case.

To steer a DSB towards the nominal source position $\mathbf{y}_1$, the nth equalizer would therefore be chosen as

$$H_{\mathrm{DSB}}(\mathbf{x}_n)=e^{j2\pi f(\tau_0-\|\mathbf{y}_1-\mathbf{x}_n\|/c)}, \tag{20}$$

where τ_0 is a delay chosen to ensure causality. Observe that the signals received at all sensors are weighted equally in the DSB.

Comparing with Eq. (13) shows that

$$H_{\mathrm{DP}}(\mathbf{x}_n)=4\pi\|\mathbf{y}_1-\mathbf{x}_n\|H_{\mathrm{DSB}}(\mathbf{x}_n). \tag{21}$$

Hence, the direct-path equalizer weights the sensor signals proportionally to the distance from the source to the sensor. This may be undesirable from a beamforming perspective, since sensors that are further from the source will have lower direct-to-reverberant energies, and thus will provide poorer estimates of the source signal.

In general, the MSE for any steered beamformer can be predicted using Eq. (15) as follows:

$$\mathrm{MSE}_{\mathrm{BF}}(\Delta)\approx\mathbf{w}_{\mathrm{BF}}^H\mathbf{Q}_{\Delta}^{\mathrm{DP}}\mathbf{w}_{\mathrm{BF}}-2\,\frac{\sin k\Delta}{k\Delta}\,\mathrm{Re}\{\mathbf{w}_{\mathrm{BF}}^H\mathbf{1}\}+1, \tag{22}$$

where $\mathbf{Q}_{\Delta}^{\mathrm{DP}}$ is defined in Eqs. (16) and (17), and $\mathbf{w}_{\mathrm{BF}}$ is a specific set of equalizer weights that implements the desired beamformer response [after undoing the implicit weighting of the direct-path equalizer, evident in Eq. (21)]. In particular, if b_n, $n=1,\ldots,N$ is a set of beamforming weights that implements a desired beamformer response, the MSE performance of this beamformer is given by Eq. (22) with

$$[\mathbf{w}_{\mathrm{BF}}]_n\propto\frac{b_n}{4\pi\|\mathbf{y}_1-\mathbf{x}_n\|}, \tag{23}$$

where the constant of proportionality is typically chosen to ensure that the beamformer has a unity response to the chosen direction.

C. Choice of the equalization weights

We have already considered the choice of the equalization weights for the specific case of steered beamforming above. For the case of exact equalization, an obvious question to ask is whether the robustness of the equalization system can be improved if the amount of equalization is sacrificed. In other words, is there a different equalizer that reduces the MSE? Since the equalizer on each sensor is simply a complex-valued scalar (at any given frequency), and since each equalizer is weighted by a scalar weight (see Fig. 1), the question of using a different equalizer can easily be incorporated into the existing MSE expressions by attempting to optimize the equalizer weights under certain conditions.

With this motivation in mind, we consider the optimization problem of choosing $\mathbf{w}$ in order to minimize the MSE over some range of displacements Λ. For this process we will not specifically impose the constraint that the equalization weights should sum to unity, since this requirement is only imposed if we are aiming for exact equalization. To avoid numerical instability we impose an additional constraint on the norm of $\mathbf{w}$, resulting in the following problem:

$$\min_{\mathbf{w}}\int_{\Delta\in\Lambda}\mathrm{MSE}(\Delta)d\Delta, \quad \text{such that } \mathbf{w}^H\mathbf{w}\leqslant 1. \tag{24}$$

This has a solution

$$\mathbf{w}_o=(\tilde{\mathbf{Q}}+\epsilon\mathbf{I})^{-1}\tilde{\mathbf{q}}, \tag{25}$$

where

$$\tilde{\mathbf{Q}}=\int_{\Delta\in\Lambda}\mathbf{Q}_{\Delta}d\Delta, \tag{26}$$

$$\tilde{\mathbf{q}}=\mathbf{1}\int_{\Delta\in\Lambda}\frac{\sin(k\Delta)}{k\Delta}d\Delta, \tag{27}$$

and ϵ is a Lagrange multiplier chosen to satisfy the constraint $\mathbf{w}^H\mathbf{w}\leqslant 1$.

For a given range of displacements, the integration in Eq. (26) and Eq. (27) can be carried out numerically. Our approach for finding the Lagrange multiplier is to initially solve the unconstrained optimization problem with $\epsilon=0$, and then test if $\mathbf{w}_o$ satisfies the constraint. If it does not, we use a numerical root-finding technique to solve for ϵ in

$$\tilde{\mathbf{q}}^H(\tilde{\mathbf{Q}}+\epsilon\mathbf{I})^{-1}(\tilde{\mathbf{Q}}+\epsilon\mathbf{I})^{-1}\tilde{\mathbf{q}}=1, \tag{28}$$

and then find $\mathbf{w}_o$ from Eq. (25).

IV. RESULTS

A. Verification of MSE expression

Having established the theoretical framework, we now verify the results against a representative experiment using the well-known image model.[7] The model was implemented in the frequency domain as described in Ref. 1. For the purposes of the analysis we consider an example room of volume 128 m^3 and of dimensions [length,width,height] $=[6.4,5,4]$ in meters (these room dimensions are the same as used in Ref. 1). The average absorption coefficient α was

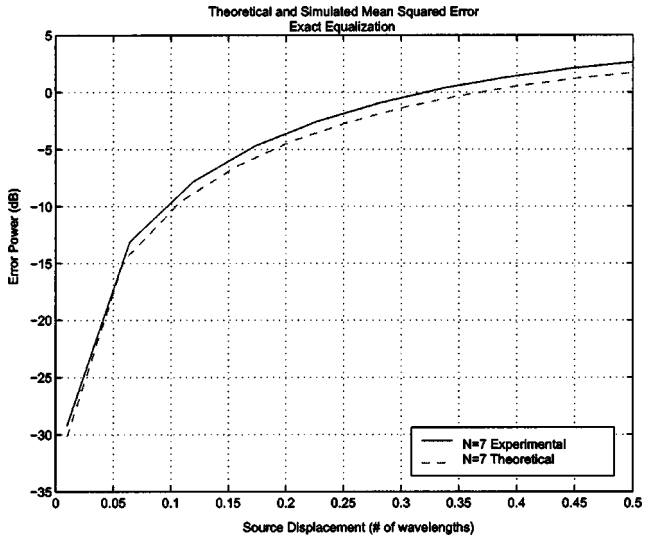


FIG. 3. Seven-channel reverberant error power versus the displacement of the source from the equalization point, at frequency 2 Khz. Exact equalization scheme is used. The source is nominally located at $\mathbf{y}_1=[0,0,0]^T m$ and the receivers at $\mathbf{x}_n=[\mathcal{X}_n,1,0]m$, where $\mathcal{X}_n=\{-1,-0.67,-0.33,0,0.33,0.67,1\}m$.

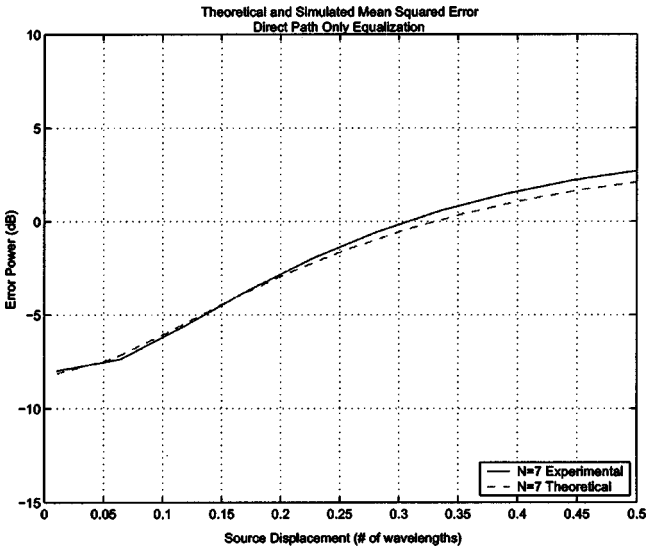


FIG. 4. As in Fig. 3, but using direct-path-only equalization.

chosen as $\alpha=0.29$, resulting in a reverberation time of $T_{60}=0.45$ s. We perform the measurements in two steps.

First, we determine the frequency response between random positions of the source and each one of the receivers for fixed distances between them. The second step involves the movement of the source from its original position $\mathbf{y}_1$ to a new arbitrary one $\mathbf{y}_2$, for which the frequency response is again calculated. This process is repeatedly performed for various displacements up to half a wavelength. For each of these steps the error function is calculated according to Eq. (3). We repeat this series of actions M times for a random set of source and receiver positions inside the room. The average normalized error at frequency f was then given as

$$\mathrm{MSE}=\frac{1}{M}\sum_{m=1}^{M}\left|\sum_{n=1}^{N} w_n G(\mathbf{y}_2,\mathbf{x}_n)H(\mathbf{x}_n)-e^{j2\pi f\tau_0}\right|^2, \tag{29}$$

where m denotes the simulation number for which the corresponding frequency response is calculated. The averaging took place over $M=100$ different simulations for a series of experiments that varied the positions of the source and the two receivers. These same experiments allowed for the verification of the theoretical results as shown in Fig. 3 and Fig. 4 for exact and direct-only equalization respectively (with $N=7$). The graphs compare the tendency of the experimental error calculations (solid line) over M runs with the theoretical error (dashed line) as this is calculated from Eq. (6). Observe that the theoretical curves are very close to the experimental ones, thereby validating our theoretical expressions in Sec. III. In the remainder, all plots are of the theoretical MSE only.

B. Effect of number of elements

As discussed earlier, using Eq. (6) to predict the MSE of a random source displacement is subject to the choice of an appropriate set of weights for each of the equalizers. If these weights are chosen to be equal to $1/N$, we then effectively weight the equalizers uniformly. The gain in performance then showed sensitivity only to the geometry. For the array configuration the geometry is a function of two control parameters, i.e., the size of the array and the distance of the source from the receivers.

Figure 5 shows the MSE for a series of simulations in which the number of receivers was increased while keeping the total array size constant (extra receivers were added uniformly between the original two receivers). Observe that the two-channel case gives the most robust performance (i.e., the MSE is lowest as the source displacement increases). Somewhat surprisingly, adding more receivers does not increase performance when uniform weights are used and the control parameters are kept identical. This can be explained if we note that when uniform weighting is used the elements of $\mathbf{w}$ are all equal to $1/N$. This means that as we add more receiv-

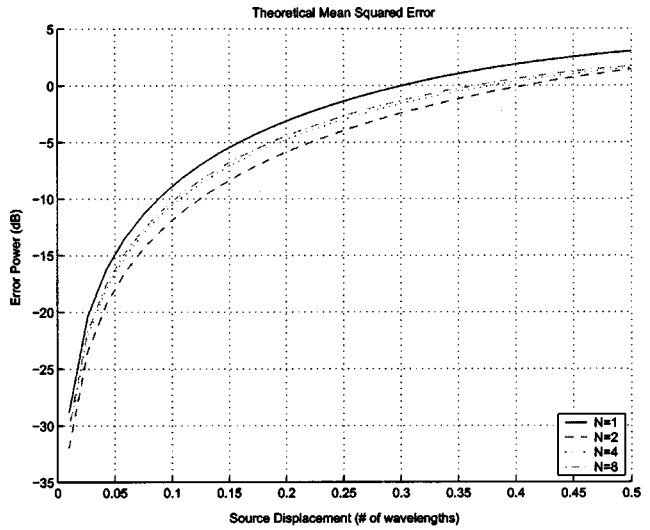


FIG. 5. MSE of equalizers with uniform weights $w_n=1/N$ for source displacements Δ up to 0.5λ. Results are shown for increasing number of receivers. Total length of the array is 2 m and the distance of the source from the midpoint of the array is 1 m.

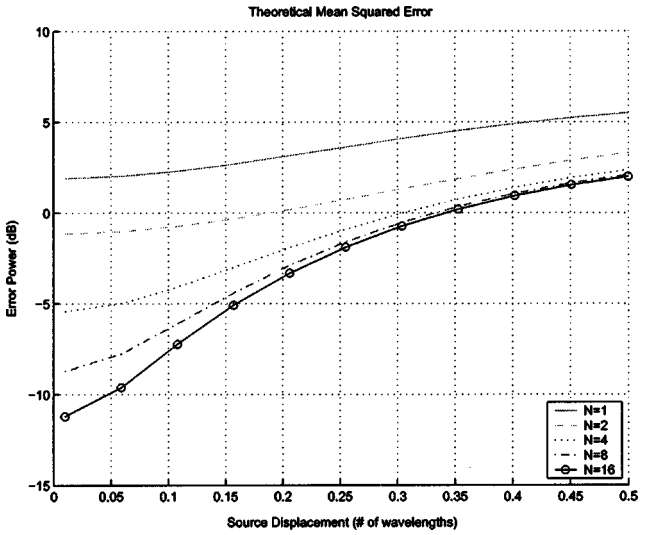


FIG. 6. As in Fig. 5, but using direct-path-only equalization.

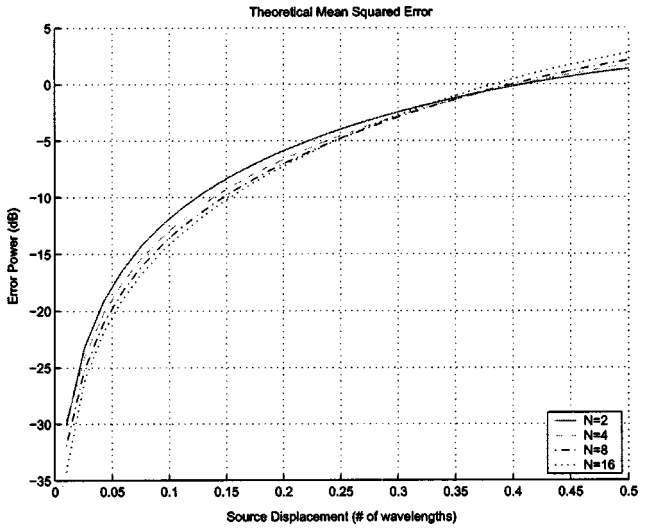


FIG. 7. As in Fig. 5, but using optimum weights [given by Eq. (25)]. Optimized over the range $\Delta \in [0.01\lambda, 0.3\lambda]$.

ers the weights become smaller. This same addition of receivers, though, causes the elements of $[\mathbf{Q}_\Delta]_{\substack{n,m \\ n\neq m}}$ to become significantly large and consequently the MSE becomes larger, as shown in Fig. 5. The elements of $[\mathbf{Q}_\Delta]_{n,n}$ are not affected by the addition of receivers, as they remain approximately equal to 1 for all cases. To reduce the MSE the system requires a set of weights that will not necessarily attempt to become small themselves, but rather to make the MSE as small as possible.

Using Eq. (15), Eq. (16), and Eq. (17) we may derive the expected error if only the direct paths between the source and the receivers were equalized. As can be seen in Fig. 6 the performance of such a system degrades very quickly and becomes inappropriate for most applications. However, in this case the addition of more receivers *does* improve robustness. This difference from exact equalization can be explained as follows. For direct-path equalization, the off-diagonal terms of $\mathbf{Q}_\Delta^{\mathrm{DP}}$ are close to those for exact equalization. The diagonal terms, however, are no longer approximately unity (as in the exact equalization system). It is thus the difference in the diagonal terms that explains the difference in MSE performance. Recall that for exact equalization, the contribution of the weighted diagonal terms is $\Sigma_n \mathbf{w}_n^2$, whereas for the direct-path equalization these terms become $\Sigma_n \mathbf{w}_n^2(1+1/\gamma_n)$. But, $1/\gamma_n \propto R_n^2$ [see Eq. (12)], where R_n^2 is the squared distance from the nominal source position to the nth receiver. In the example we consider, the nominal source position is located perpendicular to the midpoint of the receiver array. Thus, if extra receivers are added between the original outside receivers, then for each additional receiver the distance to the source R_n is reduced. It is for this reason that adding more receivers reduces the MSE when direct-path equalization is used.

C. Effect of equalizer weights

Despite the fact that performance can be improved by using a two-channel setup, the equalization quality still remains highly restricted in the range of a fraction of a wavelength. There are, though, alternative means of optimizing the robustness offered by Eq. (6). If the equalizers are weighted according to the optimal scheme of Eq. (25), then gains in performance can be achieved over both the one-channel and two-channel systems. This demonstrates that exact equalizers are not the optimum choice to the presented equalization problem. Figure 7 shows the resulting MSE variation for increasing number of receivers. For this case the two-channel system (which was again weighted with an 1/2 factor per channel) serves as a lower performance bound. Note from Eq. (25) that the optimum weights are a function of the source displacement Δ. For all of the optimized plots, the optimized weights were calculated for an example displacement interval of the source, i.e., for $\Delta \in \Lambda$ where $\Lambda = [0.01\lambda, 0.3\lambda]$. Comparing this to Fig. 5 (which considers uniform weighting), observe in Fig. 7 that increasing the number of receivers *does* give a more robust equalizer. The difference in performance can be explained by analyzing the optimization process. Optimal weighting alters Eq. (6) so that the weights are not necessarily equal. $\mathbf{w}$ is now chosen to compensate for the rising elements of $[\mathbf{Q}_\Delta]_{\substack{n,m \\ n\neq m}}$ as more receivers are added. The elements of this matrix are dominated by the $[\sin(k\zeta_{nm}\Delta)]/[(k\zeta_{nm}\Delta)]$ term. This term results from the cross correlation between the direct-path transfer functions from the perturbed source position $\mathbf{y}_2$ to the nth and mth receivers. The magnitude of this term is predominantly determined by the perturbation distance Δ and ζ_{nm}. ζ_{nm} is solely dependent on the positions of the source and the receivers. The addition of more receivers creates more cross-correlation effects that need to be made as small as possible. The optimum weighting function chooses the weights affecting these terms according to this condition.

D. Effect of geometry

Results presented up to now are for linear array with equally spaced elements. We now consider alternative geometries, specifically nonuniform spacings and circular arrays. Figure 8 shows an example of a 4-channel linear array with

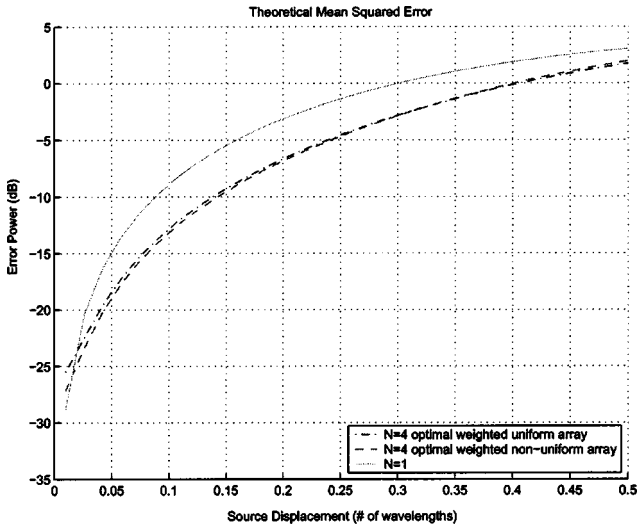


FIG. 8. Demonstration of robustness improvement when receivers are placed nonuniformly on the array. Geometrical parameters as the ones of Fig. 5. For the nonuniform case, the receivers are placed at distances {1,0.1, −0.1,1} (in m) from the center point of the array.

nonuniform spacings $\mathbf{x}_n=\{1,-0.1,0.1,1\}$. The MSE performance is marginally improved over the equally spaced array.

If we place all of the receivers equidistantly from the source and in a circular geometry, we witness a significant improvement in exact equalization robustness over the linear array geometry. As can be seen in Fig. 9 performance degrades much slower even when four receivers are used. In this case we make all of the R_n-R_m terms in $[\mathbf{Q}_\Delta]_{n\neq m}^{n,m}$ zero, i.e., we perform the minimization of the dominant sinc term directly. The radius of the used circle affects performance as before, i.e., the smaller the radius the slower the degradation. Optimal weighting does not generally improve performance further.

V. CONCLUSIONS

We have established the robustness performance of multichannel equalization systems when used in reverberant environments. Specifically, we developed a closed-form expression for the expected mean-square error when the source moves from the position for which the equalizer was designed.

Three classes of equalizer were considered: (i) exact equalizers that attempt to precisely compensate for the transfer function from the source to each receiver (this is the direct extension of Radlovic *et al.*[1]); (ii) direct-path equalizers (of which beamformers are a special case); and (iii) optimum equalizers that aim to minimize the MSE.

Results are parametrized by physical quantities such as the room size, reverberation time, distance of the source from the array, and array size. Our results show that for an equally spaced, uniformly weighted linear array, adding more receivers *does not* improve the MSE when exact equalization is used; $N=2$ gives the best performance. If optimum weighting is used, however, then the MSE *does* improve as more receivers are added, although the improvement is not significant.

For an equally spaced linear array, when direct-path equalization (such as beamforming) is used, adding more receivers always improves the MSE. The best performance we found is when a circular array is used with the nominal source position at its center.

Although we have shown that the robustness of an acoustic equalizer can be improved by adding more receivers (depending on the geometry), the improvement is not significant, and the region in which the MSE is below 10 dB is restricted to a fraction of a wavelength.

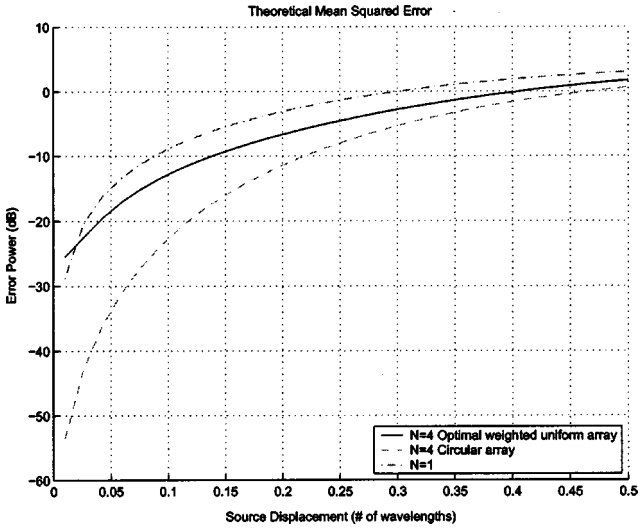


FIG. 9. Demonstration of robustness improvement for $N=4$ when circular geometry is used. The radius of the array is 1 m.

ACKNOWLEDGMENTS

This work was supported by the Engineering and Physical Sciences Research Council. The authors also wish to thank Bob Williamson for valuable discussions.

APPENDIX: PROOF OF THEOREM 1

Here, we present the proof of *Theorem 1*, which follows similar lines to that given in Ref. 1. Equation (3) can be written

$$\begin{aligned}\mathrm{MSE}&=E\left\{\left|\sum_{n=1}^{N} w_n\frac{G(\mathbf{y}_2,\mathbf{x}_n)}{G(\mathbf{y}_1,\mathbf{x}_n)}e^{j2\pi f\tau_0}-e^{j2\pi f\tau_0}\right|^2\right\}\\&=E\left\{1+\sum_{n=1}^{N}\sum_{m=1}^{N} w_n w_m^*\frac{G(\mathbf{y}_2,\mathbf{x}_n)G^*(\mathbf{y}_2,\mathbf{x}_m)}{G(\mathbf{y}_1,\mathbf{x}_n)G^*(\mathbf{y}_1,\mathbf{x}_m)}\right.\\&\quad\left.-\sum_{n=1}^{N}\left[w_n\frac{G(\mathbf{y}_2,\mathbf{x}_n)}{G(\mathbf{y}_1,\mathbf{x}_n)}+w_n^*\frac{G^*(\mathbf{y}_2,\mathbf{x}_n)}{G^*(\mathbf{y}_1,\mathbf{x}_n)}\right]\right\}.\end{aligned}\tag{A1}$$

Using a zeroth-order Taylor expansion for the expectations, i.e., $E\{g(x)\}\approx g(\overline{x})$, this becomes

$$\mathrm{MSE}\approx 1+\sum_{n=1}^{N}\sum_{m=1}^{N} w_n w_m^* \frac{E\{G(\mathbf{y}_2,\mathbf{x}_n)G^*(\mathbf{y}_2,\mathbf{x}_m)\}}{E\{G(\mathbf{y}_1,\mathbf{x}_n)G^*(\mathbf{y}_1,\mathbf{x}_m)\}} - \sum_{n=1}^{N}\left[w_n \frac{E\{G(\mathbf{y}_2,\mathbf{x}_n)G^*(\mathbf{y}_1,\mathbf{x}_n)\}}{E\{G(\mathbf{y}_1,\mathbf{x}_n)G^*(\mathbf{y}_1,\mathbf{x}_n)\}} + w_n^* \frac{E\{G^*(\mathbf{y}_2,\mathbf{x}_n)G(\mathbf{y}_1,\mathbf{x}_n)\}}{E\{G^*(\mathbf{y}_1,\mathbf{x}_n)G(\mathbf{y}_1,\mathbf{x}_n)\}}\right], \tag{A2}$$

where we have multiplied and divided the last summation terms by $G^*(\mathbf{y}_1,\mathbf{x}_n)$ and $G(\mathbf{y}_1,\mathbf{x}_n)$, respectively. This can be written in matrix form as

$$\mathrm{MSE}\approx 1+\mathbf{w}^H\mathbf{Q}_\Delta\mathbf{w}-\mathbf{w}^H\mathbf{q}_\Delta-\mathbf{q}_\Delta^H\mathbf{w}, \tag{A3}$$

where $\mathbf{w}$ is defined in Eq. (7)

$$[\mathbf{Q}_\Delta]_{n,m}=\frac{E\{G(\mathbf{y}_2,\mathbf{x}_n)G^*(\mathbf{y}_2,\mathbf{x}_m)\}}{E\{G(\mathbf{y}_1,\mathbf{x}_n)G^*(\mathbf{y}_1,\mathbf{x}_m)\}}, \tag{A4}$$

and

$$[\mathbf{q}_\Delta]_n=\frac{E\{G(\mathbf{y}_2,\mathbf{x}_n)G^*(\mathbf{y}_1,\mathbf{x}_n)\}}{E\{G(\mathbf{y}_1,\mathbf{x}_n)G^*(\mathbf{y}_1,\mathbf{x}_n)\}}. \tag{A5}$$

The frequency response between two points $\mathbf{x}$, $\mathbf{y}$ in a reverberant room can be expressed as the superposition of the direct and reverberant frequency components, i.e., $G(\mathbf{y},\mathbf{x})=G_d(\mathbf{y},\mathbf{x})+G_r(\mathbf{y},\mathbf{x})$, where the direct-path component is

$$G_d(\mathbf{y},\mathbf{x})=\frac{e^{jk\|\mathbf{y}-\mathbf{x}\|}}{4\pi\|\mathbf{y}-\mathbf{x}\|}. \tag{A6}$$

Under the conditions presented in Sec. I, the direct and reverberant sound pressure at any point is uncorrelated, giving

$$[\mathbf{Q}_\Delta]_{n,m}=\frac{E\{G_d(\mathbf{y}_2,\mathbf{x}_n)G_d^*(\mathbf{y}_2,\mathbf{x}_m)\}+E\{G_r(\mathbf{y}_2,\mathbf{x}_n)G_r^*(\mathbf{y}_2,\mathbf{x}_m)\}}{E\{G_d(\mathbf{y}_1,\mathbf{x}_n)G_d^*(\mathbf{y}_1,\mathbf{x}_m)\}+E\{G_r(\mathbf{y}_1,\mathbf{x}_n)G_r^*(\mathbf{y}_1,\mathbf{x}_m)\}}, \tag{A7}$$

$$[\mathbf{q}_\Delta]_n=\frac{E\{G_d(\mathbf{y}_2,\mathbf{x}_n)G_d^*(\mathbf{y}_1,\mathbf{x}_n)\}+E\{G_r(\mathbf{y}_2,\mathbf{x}_n)G_r^*(\mathbf{y}_1,\mathbf{x}_n)\}}{E\{G_d(\mathbf{y}_1,\mathbf{x}_n)G_d^*(\mathbf{y}_1,\mathbf{x}_n)\}+E\{G_r(\mathbf{y}_1,\mathbf{x}_n)G_r^*(\mathbf{y}_1,\mathbf{x}_n)\}}. \tag{A8}$$

We now consider each of the expectations in turn.

1. Direct-path terms

Consider

$$E\{G_d(\mathbf{y}_2,\mathbf{x}_n)G_d^*(\mathbf{y}_2,\mathbf{x}_m)\}=E\left\{\frac{e^{jk\|\mathbf{y}_2-\mathbf{x}_n\|}}{4\pi\|\mathbf{y}_2-\mathbf{x}_n\|}\frac{e^{-jk\|\mathbf{y}_2-\mathbf{x}_m\|}}{4\pi\|\mathbf{y}_2-\mathbf{x}_m\|}\right\}. \tag{A9}$$

Let $R_n=\|\mathbf{y}_1-\mathbf{x}_n\|$ and $R_n'=\|\mathbf{y}_2-\mathbf{x}_n\|$. Using Eq. (1) and Eq. (4), we have

$$R_n'^2=\mathcal{X}_n^2+\mathcal{Y}_n^2+\mathcal{Z}_n^2+\Delta^2-2\Delta(\mathcal{X}_n\sin\theta\cos\phi+\mathcal{Y}_n\sin\theta\sin\phi+\mathcal{Z}_n\cos\theta)=R_n^2+\Delta^2-2\Delta(\mathcal{X}_n\sin\theta\cos\phi+\mathcal{Y}_n\sin\theta\sin\phi+\mathcal{Z}_n\cos\theta). \tag{A10}$$

If $R_n\gg\Delta$, then the Taylor series approximation $\sqrt{1+x}\approx 1+x/2$ can be used to yield

$$R_n'\approx R_n+\frac{\Delta^2}{2R_n}-\frac{\Delta}{R_n}(\mathcal{X}_n\sin\theta\cos\phi+\mathcal{Y}_n\sin\theta\sin\phi+\mathcal{Z}_n\cos\theta)=\hat{R}_n'. \tag{A11}$$

Substituting this into the numerator of Eq. (A9), and approximating the denominator as $R_n'\approx R_n$, $\forall n$, we have

$$E\{G_d(\mathbf{y}_2,\mathbf{x}_n)G_d^*(\mathbf{y}_2,\mathbf{x}_m)\}=\frac{1}{(4\pi)^2R_nR_m}E\{e^{jk(\hat{R}_n'-\hat{R}_m')}\}, \tag{A12}$$

where

$$E\{e^{jk(\hat{R}_n'-\hat{R}_m')}\}=\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}e^{jk(\hat{R}_n'-\hat{R}_m')}\sin'\theta\,d\theta\,d\phi, \tag{A13}$$

and

$$\hat{R}_n'-\hat{R}_m'=R_n-R_m+\frac{\Delta^2}{2}\left(\frac{1}{R_n}-\frac{1}{R_m}\right)-\frac{\Delta}{R_nR_m}\times[(R_m\mathcal{X}_n-R_n\mathcal{X}_m)\sin\theta\cos\phi+(R_m\mathcal{Y}_n-R_n\mathcal{Y}_m)\sin\theta\sin\phi+(R_m\mathcal{Z}_n-R_n\mathcal{Z}_m)\cos\theta]. \tag{A14}$$

The integral in Eq. (A13) can be solved by noting that[8]

$$\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}g(c_1\cos\theta+c_2\sin\theta\cos\phi+c_3\sin\theta\sin\phi)\sin\theta\,d\theta\,d\phi=\frac{1}{2}\int_{-1}^{1}g(At)dt, \tag{A15}$$

where $A=\sqrt{c_1^2+c_2^2+c_3^2}$. Hence, after lengthy but straightforward manipulations, we have

$$E\{G_d(\mathbf{y}_2,\mathbf{x}_n)G_d^*(\mathbf{y}_2,\mathbf{x}_m)\}=\frac{e^{jk(R_n-R_m+\Delta^2\mu_{nm})}}{(4\pi)^2R_nR_m}\frac{\sin(k\zeta_{nm}\Delta)}{k\zeta_{nm}\Delta},\quad \forall n,m,n\neq m, \tag{A16}$$

where ζ_{nm} and μ_{nm} are given by Eq. (10) and Eq. (11), respectively.

In the special case of $n=m$, we have

$$E\{G_d(\mathbf{y}_2,\mathbf{x}_n)G_d^*(\mathbf{y}_2,\mathbf{x}_n)\}=E\left\{\frac{1}{(4\pi)^2\|\mathbf{y}_2-\mathbf{x}_n\|^2}\right\}, \tag{A17}$$

which can be solved exactly using Eq. (A10) and Eq. (A15), to give

$$E\{G_d(\mathbf{y}_2,\mathbf{x}_n)G_d^*(\mathbf{y}_2,\mathbf{x}_n)\}=\frac{1}{(4\pi)^2}\frac{1}{2\Delta R_n}\ln\left|\frac{R_n+\Delta}{R_n-\Delta}\right|,\quad \forall n. \tag{A18}$$

Next, consider

$$E\{G_d(\mathbf{y}_2,\mathbf{x}_n)G_d^*(\mathbf{y}_1,\mathbf{x}_n)\}$$
$$=E\left\{\frac{e^{jk\|\mathbf{y}_2-\mathbf{x}_n\|}}{4\pi\|\mathbf{y}_2-\mathbf{x}_n\|}\frac{e^{-jk\|\mathbf{y}_1-\mathbf{x}_n\|}}{4\pi\|\mathbf{y}_1-\mathbf{x}_n\|}\right\}=\frac{e^{-jkR_n}}{(4\pi)^2R_n}E\left\{\frac{e^{jkR_n'}}{R_n'}\right\}. \tag{A19}$$

Using Eq. (A11) and Eq. (A15), we obtain

$$E\left\{\frac{e^{jkR_n'}}{R_n'}\right\}=\frac{e^{jkR_n}}{R_n}\frac{\sin k\Delta}{k\Delta}. \tag{A20}$$

Substitution gives

$$E\{G_d(\mathbf{y}_2,\mathbf{x}_n)G_d^*(\mathbf{y}_1,\mathbf{x}_n)\}=\frac{1}{(4\pi)^2R_n^2}\frac{\sin k\Delta}{k\Delta},\quad \forall n. \tag{A21}$$

Finally,

$$E\{G_d(\mathbf{y}_1,\mathbf{x}_n)G_d^*(\mathbf{y}_1,\mathbf{x}_m)\}=\frac{e^{jk(R_n-R_m)}}{(4\pi)^2R_nR_m},\quad \forall n,m. \tag{A22}$$

2. Reverberant-path terms

Using results from Ref. 9 it immediately follows that the reverberant path terms are

$$E\{G_r(\mathbf{y}_2,\mathbf{x}_n)G_r^*(\mathbf{y}_2,\mathbf{x}_m)\}=\left(\frac{1-\alpha}{\pi S\alpha}\right)\frac{\sin k\|\mathbf{x}_n-\mathbf{x}_m\|}{k\|\mathbf{x}_n-\mathbf{x}_m\|},\quad \forall n,m, \tag{A23}$$

$$E\{G_r(\mathbf{y}_1,\mathbf{x}_n)G_r^*(\mathbf{y}_1,\mathbf{x}_m)\}=\left(\frac{1-\alpha}{\pi S\alpha}\right)\frac{\sin k\|\mathbf{x}_n-\mathbf{x}_m\|}{k\|\mathbf{x}_n-\mathbf{x}_m\|},\quad \forall n,m, \tag{A24}$$

and

$$E\{G_r(\mathbf{y}_2,\mathbf{x}_n)G_r^*(\mathbf{y}_1,\mathbf{x}_n)\}=\left(\frac{1-\alpha}{\pi S\alpha}\right)\frac{\sin k\Delta}{k\Delta},\quad \forall n. \tag{A25}$$

Substituting these expressions into Eq. (A7) yields

$$[\mathbf{Q}_\Delta]_{\substack{n,m\\n\neq m}}\approx\frac{\dfrac{e^{jk(R_n-R_m+\Delta^2\mu_{nm})}}{(4\pi)^2R_nR_m}\dfrac{\sin(k\zeta_{nm}\Delta)}{k\zeta_{nm}\Delta}+\left(\dfrac{1-\alpha}{\pi S\alpha}\right)\dfrac{\sin k\|\mathbf{x}_n-\mathbf{x}_m\|}{k\|\mathbf{x}_n-\mathbf{x}_m\|}}{\dfrac{e^{jk(R_n-R_m)}}{(4\pi)^2R_nR_m}+\left(\dfrac{1-\alpha}{\pi S\alpha}\right)\dfrac{\sin k\|\mathbf{x}_n-\mathbf{x}_m\|}{k\|\mathbf{x}_n-\mathbf{x}_m\|}}$$
$$=\frac{\sqrt{\gamma_n\gamma_m}e^{jk(R_n-R_m+\Delta^2\mu_{nm})}\dfrac{\sin(k\zeta_{nm}\Delta)}{k\zeta_{nm}\Delta}+\dfrac{\sin k\|\mathbf{x}_n-\mathbf{x}_m\|}{k\|\mathbf{x}_n-\mathbf{x}_m\|}}{\sqrt{\gamma_n\gamma_m}e^{jk(R_n-R_m)}+\dfrac{\sin k\|\mathbf{x}_n-\mathbf{x}_m\|}{k\|\mathbf{x}_n-\mathbf{x}_m\|}}, \tag{A26}$$

where the last equality follows by multiplying the numerator and denominator by $(\pi S\alpha)/(1-\alpha)$, and using Eq. (12). Similarly, the diagonal elements of $\mathbf{Q}_\Delta$ are

$$[\mathbf{Q}_\Delta]_{n,n}=\frac{\gamma_n\dfrac{R_n}{2\Delta}\ln\left|\dfrac{R_n+\Delta}{R_n-\Delta}\right|+1}{\gamma_n+1}. \tag{A27}$$

Finally, the elements of $\mathbf{q}_\Delta$ become

$$[\mathbf{q}_\Delta]_n=\frac{\dfrac{1}{(4\pi)^2R_n^2}\dfrac{\sin k\Delta}{k\Delta}+\left(\dfrac{1-\alpha}{\pi S\alpha}\right)\dfrac{\sin k\Delta}{k\Delta}}{\dfrac{1}{(4\pi)^2R_n^2}+\left(\dfrac{1-\alpha}{\pi S\alpha}\right)}$$
$$=\frac{(\gamma_n+1)\dfrac{\sin k\Delta}{k\Delta}}{(\gamma+1)}=\frac{\sin k\Delta}{k\Delta}, \tag{A28}$$

thus completing the proof.

[1] B. D. Radlovic, R. C. Williamson, and R. A. Kennedy, "Equalization in an acoustic reverberant environment: Robustness results," IEEE Trans. Speech Audio Process. **8**(3), 311–319 (2000).
[2] S. Bharitkar, P. Hilmes, and C. Kyriakakis, "Robustness of spatial averaging equalization methods: A statistical approach," in 36th IEEE Asilomar Conf. on Signals, Systems, & Computers, Nov. 2002, CA.
[3] S. Bharitkar, P. Hilmes, and C. Kyriakakis, "Robustness of multiple listener equalization with magnitude response averaging," in *Audio Engineering Society, 113th Convention*, Oct. 2002.
[4] M. R. Schroeder, "Frequency correlation functions of frequency responses in rooms," J. Acoust. Soc. Am. **34**, 1819–1823 (1962).
[5] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," IEEE ASSP Magazine, April 1988.
[6] *Microphone Arrays: Signal Processing Techniques and Applications*, edited by M. S. Brandstein and D. B. Ward (Springer, Berlin, 2001).
[7] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943–950 (1979).
[8] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, London, 1980).
[9] D. B. Ward, "On the performance of acoustic crosstalk cancellation in a reverberant environment," J. Acoust. Soc. Am. **110**, 1195–1198 (2001).

Application of cylindrical near-field acoustical holography to the visualization of aeroacoustic sources

Moohyung Lee,[a] J. Stuart Bolton, and Luc Mongeau
Ray W. Herrick Laboratories, School of Mechanical Engineering, Purdue University, 140 South Intramural Drive, West Lafayette, Indiana 47907-2031

(Received 24 January 2003; revised 25 April 2003; accepted 12 May 2003)

The purpose of this study was to develop methods for visualizing the sound radiation from aeroacoustic sources in order to identify their source strength distribution, radiation patterns, and to quantify the performance of noise control solutions. Here, cylindrical Near-field Acoustical Holography was used for that purpose. In a practical holographic measurement of sources comprising either partially correlated or uncorrelated subsources, it is necessary to use a number of reference microphones so that the sound field on the hologram surface can be decomposed into mutually incoherent partial fields before holographic projection. In this article, procedures are described for determining the number of reference microphones required when visualizing partially correlated aeroacoustic sources; performing source nonstationarity compensation; and applying regularization. The procedures have been demonstrated by application to a ducted fan. Holographic tests were performed to visualize the sound radiation from that source in its original form. The system was then altered to investigate the effect of two modifications on the fan's sound radiation pattern: first, leaks were created in the fan and duct assembly, and second, sound absorbing material was used to line the downstream duct section. Results in all three cases are shown at the blade passing frequency and for a broadband noise component. In the absence of leakage, both components were found to exhibit a dipole-like radiation pattern. Leakage was found to have a strong influence on the directivity of the blade passing tone. The increase of the flow resistance caused by adding the acoustical lining resulted in a nearly symmetric reduction of sound radiation.
 [DOI: 10.1121/1.1587735]

PACS numbers: 43.60.Sx, 43.28.Ra, 43.50.Nm [EGW]

I. INTRODUCTION

Since Near-field Acoustical Holography (NAH) was first proposed by Williams and others in the 1980s, this sound field visualization technique has been applied to many sound radiation problems.[1] However, most applications have been focused on the study of sound radiation from vibrating structures. At the same time, many alternative, array-based measurement techniques have been developed to identify and locate aeroacoustic sources (e.g., *the acoustic mirror*,[2] the acoustic telescope,[2,3] and polar correlation[2,4]). In contrast with NAH measurements that are made in a source's near-field, these other techniques are based on farfield measurements, thus imposing a limit on the spatial resolution with which a source can be visualized.

In the present study, cylindrical NAH was applied to the visualization of a stationary aeroacoustic source. Due to the spatially random nature of aeroacoustic sources, the use of multiple reference microphones is required to enable both a scanning measurement and the decomposition of the field into mutually incoherent partial fields. In the past, the NAH procedure for the visualization of multiple incoherent sources has been implemented using two different approaches to partial field decomposition: *the virtual coherence method and the partial coherence method*.[5,6] Although those methods differ in the way the sound field is decomposed, the composite holographic projections that result are the same in either case under ideal circumstances.[6]

In the NAH reconstruction process, especially when projecting back toward the source, results can be degraded by the amplification of measurement noise associated with evanescent wave components.[1] A number of regularization procedures for ill-posed, inverse problems have been applied recently to address this problem. Among the various regularization techniques considered, e.g., Tikhonov-based methods,[7,8] Landweber iteration,[8,9] and the conjugate gradient approach,[7,8] Tikhonov-based methods have been most frequently adopted in recent work.[10,11] Those regularization techniques are used to determine the shape of a *k*-space reconstruction filter and are used in conjunction with a parameter selection technique to determine the optimal break point of the *k*-space filter. The parameter selection technique is classified in two ways depending on whether the variance of the measurement noise is known. When the Mozorov Discrepancy Principle (MDP)[7,8] is used, knowledge of the variance of the measurement noise is required, whereas that information is not required when the Generalized Cross Validation (GCV)[7,12] or L-Curve Criterion[7,13] is used. Sureshkumar and Raveendra have compared the performance of three techniques for the selection of the regularization parameter in conjunction with the standard Tikhonov regularization method.[14]

[a]Author to whom correspondence should be addressed. Phone: +1-765-494-2147; fax: +1-765-494-0787; electronic mail: leemoohy@ecn.purdue.edu

In the present study, a NAH procedure based on the use of the virtual coherence method and a modified Tikhonov regularization method in conjunction with the Mozorov Discrepancy Principle was applied to visualize the sound field of a small ducted fan assembly. The frequency range of interest in this case was below 400 Hz, in which range a small fan is usually considered a compact dipole source. In a recent paper by Jordan and Fitzpatrick, it was found that conventional farfield beamforming methods fail when applied to the identification of the location and spectral characteristics of a simple dipole.[15] In contrast, here the applicability of NAH to an aeroacoustic dipole is demonstrated. The theory underlying the current NAH implementation is described in Sec. II. The experimental apparatus and procedures used here are described in Sec. III, which is followed by the presentation of the results in Sec. IV. A discussion and conclusions are presented in the final two sections.

II. THEORY

A. Cross-spectral measurement by using multiple references

An aeroacoustic source is typically only partially spatially correlated. So consider here the case of a composite noise source that comprises multiple uncorrelated subsources. If it were possible to measure the contribution of each incoherent source separately, each partial field could be projected independently by using NAH: they could then be added on an energy basis to reconstruct the total field. However, this approach is usually not practical since it is typically not possible to measure the partial fields individually. Alternatively, a cross-spectral procedure based on using a set of reference signals can be used.[5] The latter signals can then be used to decompose the composite sound field into a set of linearly independent partial fields, as will be shown next.

1. Partial field decomposition

Here the virtual coherence method was used to perform the partial field decomposition.[16] In particular, the following cross-spectral relations were used:[17]

$$\mathbf{C}_{rp}=\mathbf{C}_{rr}\mathbf{H}_{rp}, \tag{1}$$

$$\mathbf{C}_{pp}=\mathbf{H}_{rp}^{\mathrm{H}}\mathbf{C}_{rr}\mathbf{H}_{rp}. \tag{2}$$

In Eqs. (1) and (2), $\mathbf{H}_{rp}$ is the transfer matrix that relates the reference signals and the measured field signals on the hologram surface, and $\mathbf{C}_{rr}$, $\mathbf{C}_{rp}$, and $\mathbf{C}_{pp}$ are cross-spectral matrices defined as

$$\mathbf{C}_{rr}=\mathrm{E}[\mathbf{r}^{*}\mathbf{r}^{\mathrm{T}}], \tag{3}$$

$$\mathbf{C}_{rp}=\mathrm{E}[\mathbf{r}^{*}\mathbf{p}^{\mathrm{T}}], \tag{4}$$

$$\mathbf{C}_{pp}=\mathrm{E}[\mathbf{p}^{*}\mathbf{p}^{\mathrm{T}}], \tag{5}$$

where $\mathbf{r}=[r_1\cdots r_N]^{\mathrm{T}}$ and $\mathbf{p}=[p_1\cdots p_M]^{\mathrm{T}}$ are the Fourier transforms of the time histories measured by the references and by the field microphones on the hologram surface, respectively, N is the number of references, M is the number of measurement points, the superscript H denotes the Hermitian operator, the superscript T denotes the transpose operator, and the superscript * denotes the complex conjugate. In these equations, $\mathrm{E}[\cdot]$ represents the expectation operator that averages the elements of $\mathbf{r}$ and $\mathbf{p}$ calculated based on the time history record length T in the limit $T\rightarrow\infty$.

Singular Value Decomposition (SVD) plays a crucial role in this method since it can be used both to diagonalize the reference cross-spectral matrix and to determine the number of reference signals required to describe the sound field. The reference cross-spectral matrix, $\mathbf{C}_{rr}$, can be decomposed in the following way:

$$\mathbf{C}_{rr}=\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}=\mathbf{U}\mathbf{\Sigma}\mathbf{U}^{\mathrm{H}}, \tag{6}$$

where $\mathbf{\Sigma}$ is a diagonal matrix composed of the singular values, $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices that are composed of the left and right singular vectors, respectively, and $\mathbf{U}$ and $\mathbf{V}$ are identical since $\mathbf{C}_{rr}$ is a positive semidefinite Hermitian matrix.

To perform an effective decomposition, the number of references should be equal to or larger than the number of singular values greater than the background noise level.[18] Note that, in practical cases, the number of references determined by SVD does not exactly correspond to the number of the actual uncorrelated sources creating the total sound field. Since measured spectra may be in error, especially due to bias errors introduced by finite discrete Fourier transform (DFT) operations, additional artificial sources may be introduced. That is, the number of singular values greater than the background noise level may be greater than the number of actual sources.[19] The rearrangement of Eq. (6) and the substitution of Eq. (6) into Eq. (2) gives

$$\mathbf{\Sigma}=\mathbf{U}^{\mathrm{H}}\mathbf{C}_{rr}\mathbf{U}=\mathrm{E}[\mathbf{v}^{*}\mathbf{v}^{\mathrm{T}}]=\mathbf{C}_{vv}, \tag{7}$$

$$\mathbf{C}_{pp}=\mathbf{H}_{rp}^{\mathrm{H}}\mathbf{U}\mathbf{\Sigma}\mathbf{U}^{\mathrm{H}}\mathbf{H}_{rp}=\mathbf{H}_{vp}^{\mathrm{H}}\mathbf{\Sigma}\mathbf{H}_{vp}, \tag{8}$$

where $\mathbf{v}$ is the virtual reference vector, $\mathbf{C}_{vv}$ is the cross-spectral matrix between the virtual references, and $\mathbf{H}_{vp}=\mathbf{U}^{\mathrm{H}}\mathbf{H}_{rp}$ is the transfer matrix that relates the virtual references and the measured signals on the hologram surface. Equation (7) indicates that the singular values of the reference cross-spectral matrix represent the autospectral amplitudes of the virtual references. By using Eqs. (1) and (6), the virtual transfer matrix, $\mathbf{H}_{vp}$, can be calculated as

$$\mathbf{H}_{vp}=\mathbf{\Sigma}^{-1}\mathbf{U}^{\mathrm{H}}\mathbf{C}_{rp}. \tag{9}$$

The decomposed partial fields, $\hat{\mathbf{P}}$, can then be calculated as the product of the virtual transfer matrix and the square root of the diagonal matrix composed of the singular values: i.e.,

$$\hat{\mathbf{P}}=\mathbf{H}_{vp}^{\mathrm{T}}\mathbf{\Sigma}^{1/2}=\mathbf{H}_{rp}^{\mathrm{T}}\mathbf{U}^{*}\mathbf{\Sigma}^{1/2}=\mathbf{C}_{rp}^{\mathrm{T}}\mathbf{U}^{*}\mathbf{\Sigma}^{-1/2}, \tag{10}$$

where the ith column vector of $\hat{\mathbf{P}}$ represents the ith partial field. Note that the decomposed partial fields, $\hat{\mathbf{P}}$, are subject to the condition that $\mathbf{C}_{pp}=\hat{\mathbf{P}}^{*}\hat{\mathbf{P}}^{\mathrm{T}}$.[16]

The sufficiency of the reference set determined by applying SVD can be verified by calculating the virtual coherence, γ_{ij}^2, at each frequency of interest:[6] i.e.,

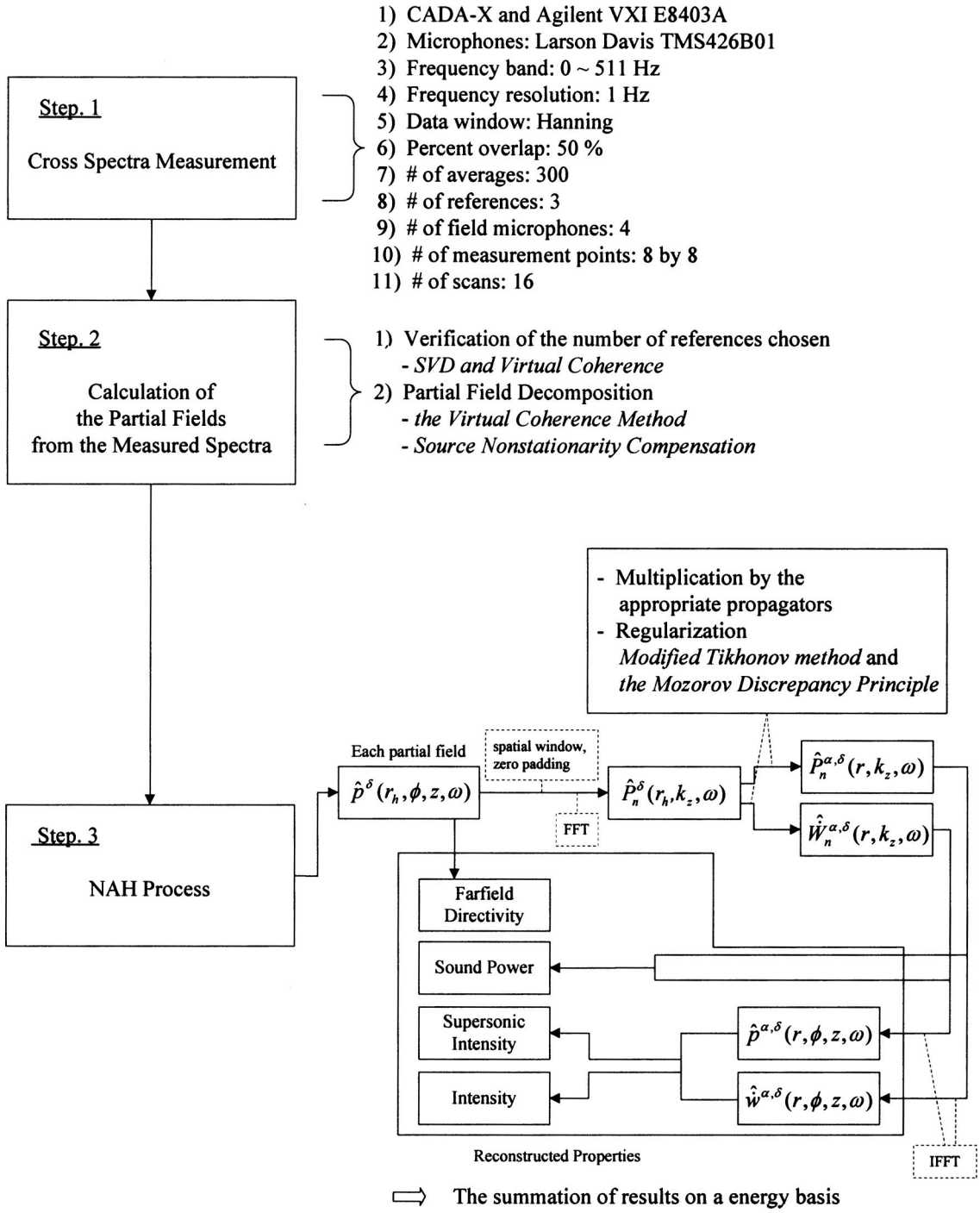


FIG. 1. Flow chart of the data analysis procedures.

$$\gamma_{ij}^2 = \frac{|C_{v_j p_i}|^2}{C_{p_i p_i} C_{v_j v_j}}, \tag{11}$$

the sound field: if its value is nearly unity, the reference set is sufficient.

where $\mathbf{C}_{vp} = \mathbf{U}^{\mathrm{H}} \mathbf{C}_{rp}$ is the cross-spectral matrix between the virtual references and the measured signals on the hologram surface, $C_{v_j p_i}$ represents the (j,i)th element of $\mathbf{C}_{vp}$, i denotes the ith measurement point, and j denotes the jth partial field. The sum of the virtual coherences, $\Sigma_{j=1}^{N} \gamma_{ij}^2$, indicates how well a subset of virtual references describes

2. Source nonstationarity compensation

When the sound field on the hologram surface is measured by scanning a subarray over a number of patches in sequence, as is typically the case, the reference cross-spectral matrix may vary from scan to scan owing to the slight nonstationarity that is characteristic of real sources. In this case,

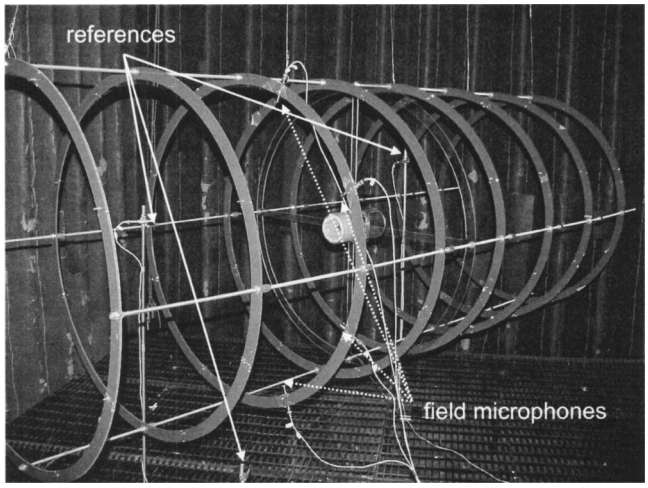


FIG. 2. Holography measurement array (ducted fan located at the center of the frame).

the partial fields calculated by the direct application of Eq. (10) may be corrupted by spatial noise in the shape of the subarray. In contrast, the transfer matrix, $\mathbf{H}_{rp}$, calculated for each patch is more likely to be independent of source levels since it represents amplitude and phase information *relative* to the sources.[20] When $\mathbf{H}_{rp}$ is independent of source level, the effects of source nonstationarity can be compensated for by modifying Eq. (10) and using the relation $\mathbf{H}_{rp}=\mathbf{C}_{rr}^{-1}\mathbf{C}_{rp}$ to form

$$\hat{\mathbf{P}}=\mathbf{H}_{rp(\mathrm{scan})}^{\mathrm{T}}\mathbf{U}_{(\mathrm{avg})}^{*}\boldsymbol{\Sigma}_{(\mathrm{avg})}^{1/2}$$
$$=(\mathbf{C}_{rr(\mathrm{scan})}^{-1}\mathbf{C}_{rp(\mathrm{scan})})^{\mathrm{T}}\mathbf{U}_{(\mathrm{avg})}^{*}\boldsymbol{\Sigma}_{(\mathrm{avg})}^{1/2}, \tag{12}$$

where the subscripts (scan) and (avg) denote estimates calculated during each scan and estimates averaged over all scans, respectively.[20]

The partial fields calculated using Eq. (12) can be used as input to the NAH equations presented in the following two sections. It is necessary to perform as many holographic projections to the reconstruction surface as there are meaningful partial fields: the total sound field is then obtained by adding the projected partial fields together on an energy basis.

B. Equations of cylindrical holography

When implementing exterior, cylindrical NAH in an anechoic space, the homogeneous, time-independent Helmholtz equation in cylindrical coordinates for the sound pressure, $\hat{p}(r,\phi,z)$, is solved by the method of separation of variables, and the pressure field is expressed in terms of cylindrical modal functions:[21] i.e.,

$$\hat{p}(r,\phi,z)=\sum_{n=-\infty}^{\infty}e^{in\phi}\frac{1}{2\pi}\int_{-\infty}^{\infty}A_n(k_z,\omega)$$
$$\times H_n^{(1)}(k_r r)e^{ik_z z}\,dk_z, \tag{13}$$

where $k_r=\sqrt{k^2-k_z^2}$, k_r and k_z are the radial and axial wave numbers, respectively, $k=\omega/c$, $A_n(k_z,\omega)H_n^{(1)}(k_r r)$ is the two-dimensional Fourier transform, $\hat{P}_n(r,k_z)$, of $\hat{p}(r,\phi,z)$

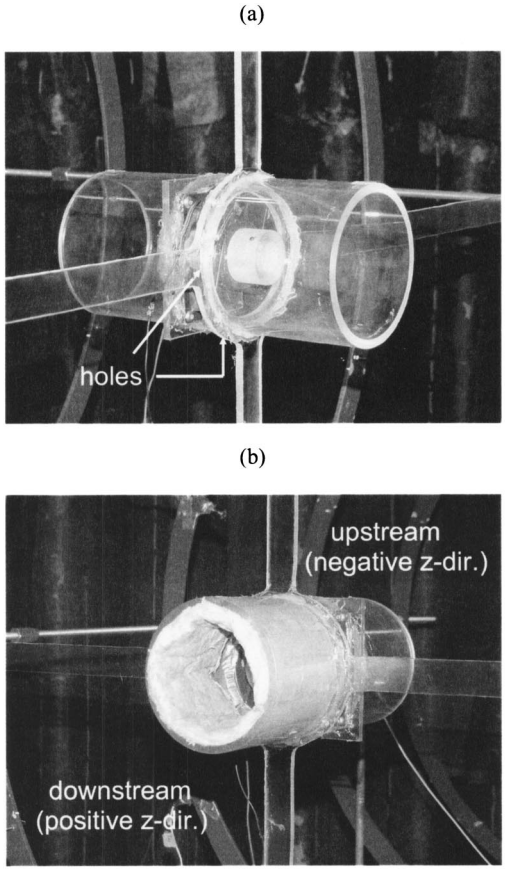


FIG. 3. Ducted fan system configurations: (a) locations of the holes in fan housing; (b) downstream fiber glass duct lining.

(i.e., the *helical wave spectrum*), $H_n^{(1)}$ is the Hankel function of the first kind, ω is the angular frequency, and c is the ambient sound speed. Here $\hat{p}(r,\phi,z)$ represents one of the partial fields. Once the helical wave spectrum of each partial field on the cylindrical hologram surface is known, the acoustical properties associated with each of them (i.e., pressure, particle velocity, and acoustic intensity) can be calculated in a three-dimensional volume by multiplying the helical wave spectrum by an appropriate propagator and then taking an inverse Fourier transform to obtain[21]

$$\hat{p}(r,\phi,z)$$
$$=\sum_{n=-\infty}^{\infty}e^{in\phi}\frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{P}_n(r_h,k_z)\frac{H_n^{(1)}(k_r r)}{H_n^{(1)}(k_r r_h)}e^{ik_z z}\,dk_z, \tag{14}$$

$$\hat{w}(r,\phi,z)=\sum_{n=-\infty}^{\infty}e^{in\phi}\frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{P}_n(r_h,k_z)$$
$$\times\frac{-ik_r}{\rho ck}\frac{H_n^{(1)'}(k_r r)}{H_n^{(1)}(k_r r_h)}e^{ik_z z}\,dk_z, \tag{15}$$

where $\hat{w}$ is the radial particle velocity, r is the radius of the surface of interest, r_h is the radius of the hologram surface, ρ

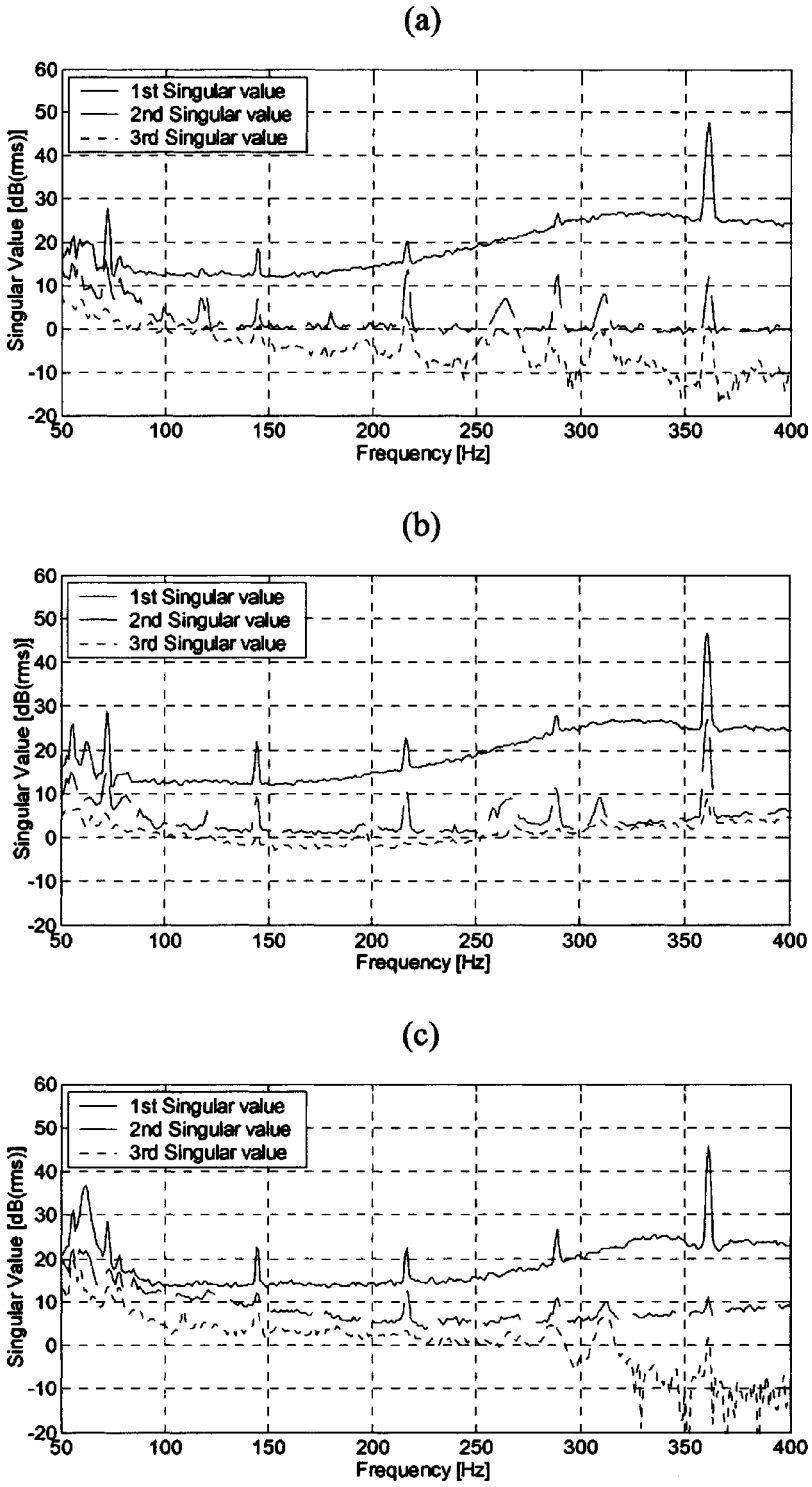


FIG. 4. Singular values of the reference cross-spectral matrices, $\mathbf{C}_{rr}$: (a) the first case (without leakage); (b) the second case (with leakage); (c) the third case (without leakage and with the acoustical lining).

is the ambient density, and $\{\cdot\}'$ denotes a derivative with respect to the argument.

In practice, the helical wave spectrum and the projected fields are calculated using two-dimensional, discrete Fourier transforms. Equations (14) and (15) can be expressed in matrix form as

$$\hat{\mathbf{p}}_i(r) = \mathbf{T}_p^{-1}\hat{\mathbf{p}}_i(r_h), \tag{16}$$

$$\hat{\mathbf{w}}_i(r) = \mathbf{T}_{\dot{w}}^{-1}\hat{\mathbf{p}}_i(r_h), \tag{17}$$

where $\hat{\mathbf{p}}_i(r_h)$ represents the ith partial field vector on the hologram (i.e., the ith column of $\hat{\mathbf{P}}$ [Eq. (12)], $\hat{\mathbf{p}}_i(r)$ and $\hat{\mathbf{w}}_i(r)$ represent the reconstructed pressure field and the radial particle velocity vector on the surface of interest, respectively, $\mathbf{T}_p$ is the transfer matrix that relates the pressures on different surfaces, and $\mathbf{T}_{\dot{w}}$ is the transfer matrix that relates the pressures and the particle velocity.

The matrices, $\mathbf{T}_p$ and $\mathbf{T}_{\dot{w}}$, can be decomposed by using an eigenvalue decomposition such as

$$\mathbf{T}_p = \overline{\mathbf{U}}\mathbf{G}_p\overline{\mathbf{U}}^{\mathrm{H}} = \overline{\mathbf{U}}\,\mathrm{diag}(\tau_1,\ldots,\tau_M)\overline{\mathbf{U}}^{\mathrm{H}}, \tag{18}$$

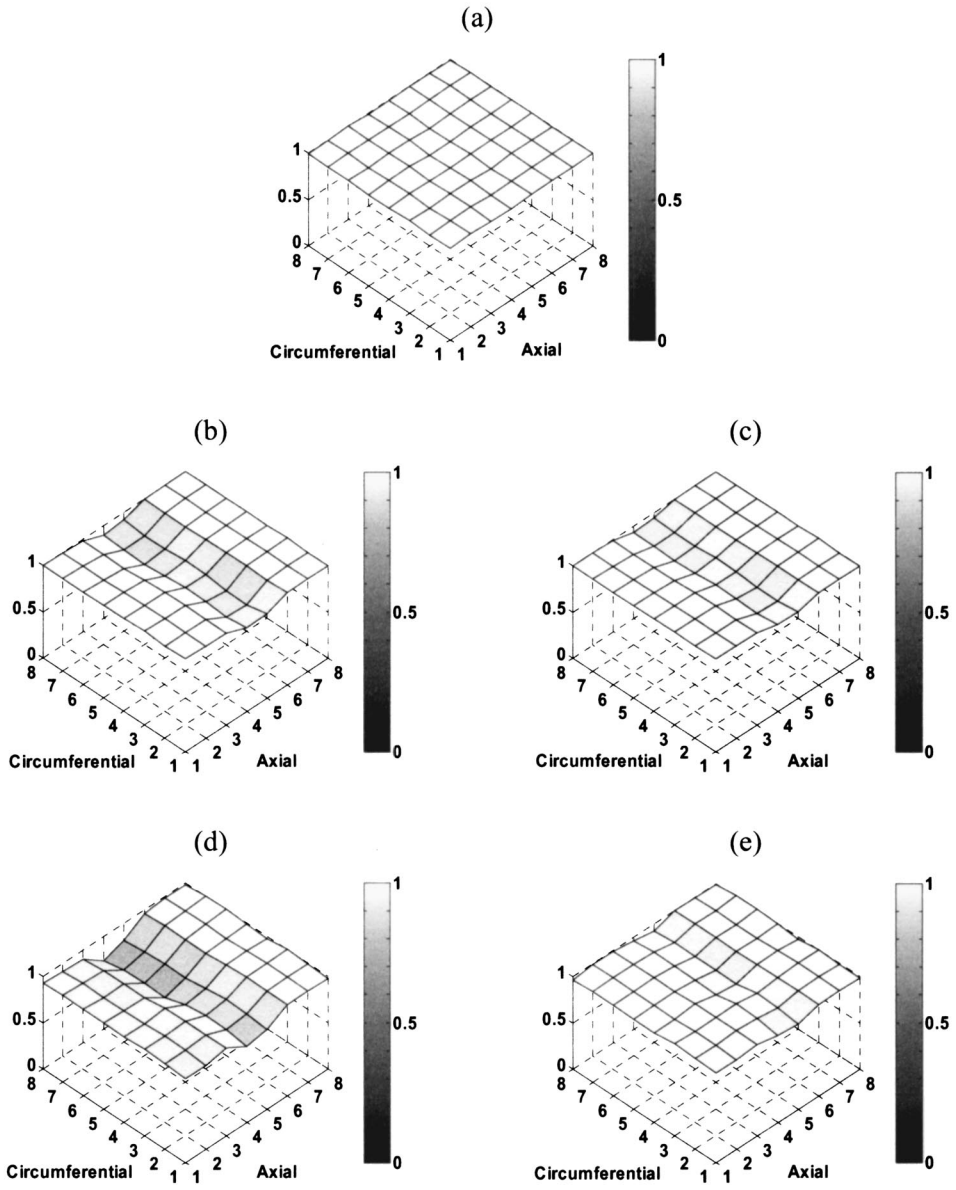


FIG. 5. Virtual coherences at 330 Hz versus measurement position: (a) γ_{i1}^2 for the first case; (b) γ_{i1}^2 for the second case; (c) $\gamma_{i1}^2+\gamma_{i2}^2$ for the second case; (d) γ_{i1}^2 for the third case; (e) $\gamma_{i1}^2+\gamma_{i2}^2$ for the third case.

$$\mathbf{T}_{\dot{w}}=\overline{\mathbf{U}}\mathbf{G}_{\dot{w}}\overline{\mathbf{U}}^{\mathrm{H}}=\overline{\mathbf{U}}\,\mathrm{diag}(\lambda_1,\ldots,\lambda_M)\overline{\mathbf{U}}^{\mathrm{H}}, \tag{19}$$

where $\overline{\mathbf{U}}$ is an M by M matrix whose columns are the corresponding eigenvectors, the τ_j's and λ_j's are the eigenvalues of the transfer matrices, $\mathbf{G}_p$ and $\mathbf{G}_{\dot{w}}$ are diagonal matrices that comprise the eigenvalues (their inverses are referred to as the pressure propagator and the velocity propagator, respectively), and diag$(\cdots)$ denotes a diagonal matrix. Note that M, previously defined as the number of the measurement points, may here be larger than the actual number of measurement points, owing to zero padding of the hologram data in the axial direction.

The substitution of Eqs. (18) and (19) into Eqs. (16) and (17) and rearrangement of the equations gives

$$\overline{\mathbf{U}}^{\mathrm{H}}\hat{\mathbf{p}}_i(r)=\mathrm{diag}\left(\frac{1}{\tau_1},\ldots,\frac{1}{\tau_M}\right)\overline{\mathbf{U}}^{\mathrm{H}}\hat{\mathbf{p}}_i(r_h), \tag{20}$$

$$\overline{\mathbf{U}}^{\mathrm{H}}\hat{\mathbf{w}}_i(r)=\mathrm{diag}\left(\frac{1}{\lambda_1},\ldots,\frac{1}{\lambda_M}\right)\overline{\mathbf{U}}^{\mathrm{H}}\hat{\mathbf{p}}_i(r_h). \tag{21}$$

In the eigenvalue decomposition, $\overline{\mathbf{U}}^{\mathrm{H}}$ corresponds to the two-dimensional Fourier transform operator. Thus $\overline{\mathbf{U}}^{\mathrm{H}}\hat{\mathbf{p}}_i$ and $\overline{\mathbf{U}}^{\mathrm{H}}\hat{\mathbf{w}}_i$ represent the helical wave spectra of the pressure and the particle velocity of the ith partial field, respectively. The forms of the eigenvalues, τ_j and λ_j, depend on the coordinate system, and, in a cylindrical geometry, are defined by[1,21]

$$\tau_{nk}=\frac{H_n^{(1)}(k_{rk}r_h)}{H_n^{(1)}(k_{rk}r)}, \tag{22}$$

$$\lambda_{nk}=\frac{i\rho ck}{k_{rk}}\frac{H_n^{(1)}(k_{rk}r_h)}{H_n^{(1)'}(k_{rk}r)}, \tag{23}$$

where $k_{rk}\equiv\sqrt{k^2-k_{zk}^2}$, the subscript k represents the discretization of the helical wave spectrum, and τ_{nk} and λ_{nk} correspond to τ_j and λ_j, respectively, depending on the way that the rows of $\mathbf{C}_{rp}$ are ordered. Note that the magnitudes of the τ_{nk}'s and λ_{nk}'s depend on the relative back-projection dis-

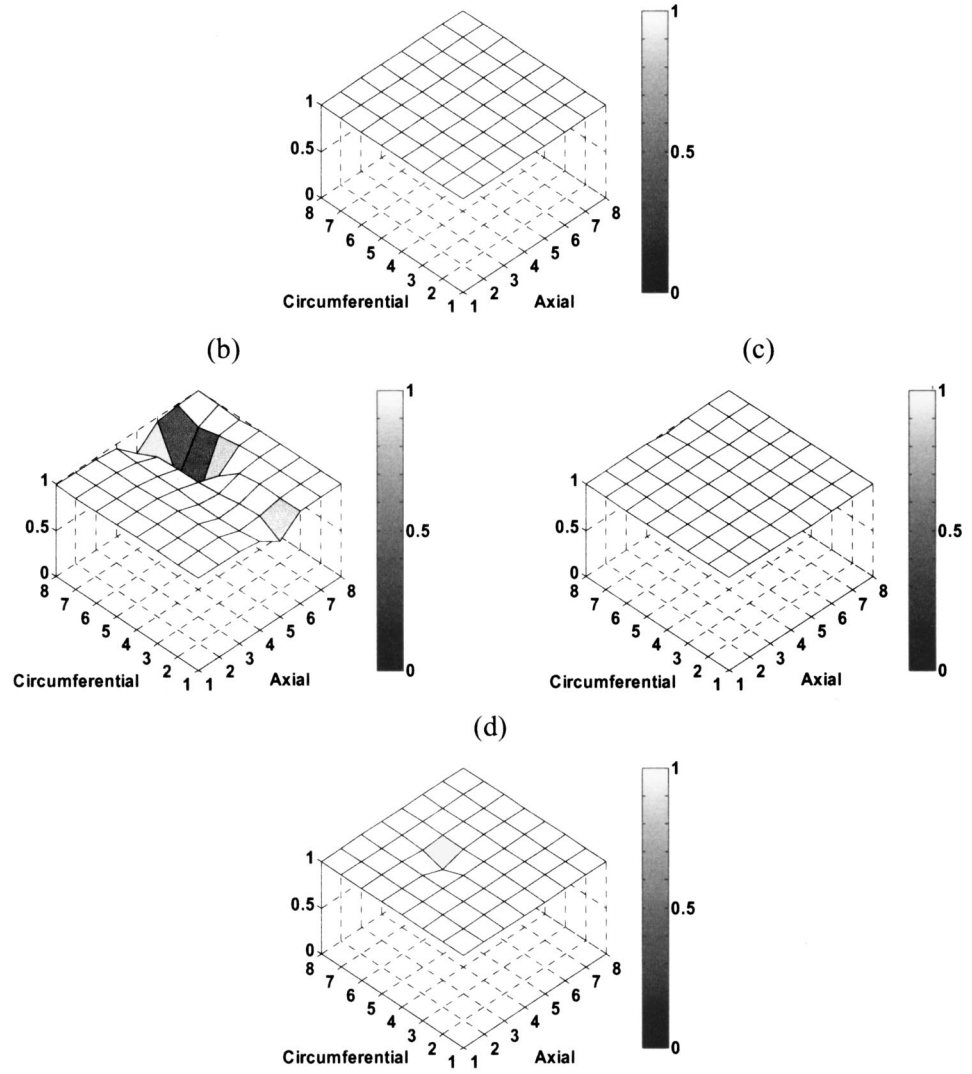


FIG. 6. Virtual coherences at 361 Hz versus measurement position: (a) γ_{i1}^2 for the first case; (b) γ_{i1}^2 for the second case; (c) $\gamma_{i1}^2+\gamma_{i2}^2$ for the second case; (d) γ_{i1}^2 for the third case.

tance. By using Eqs. (22) and (23), Eqs. (20) and (21) can be expressed as

$$\begin{Bmatrix} \vdots \\ \hat{P}_n(r,k_{zk})_i \\ \vdots \end{Bmatrix} = \begin{bmatrix} \ddots & & \\ & \dfrac{H_n^{(1)}(k_{rk}r)}{H_n^{(1)}(k_{rk}r_h)} & \\ & & \ddots \end{bmatrix} \begin{Bmatrix} \vdots \\ \hat{P}_n(r_h,k_{zk})_i \\ \vdots \end{Bmatrix}, \tag{24}$$

$$\begin{Bmatrix} \vdots \\ \hat{W}_n(r,k_{zk})_i \\ \vdots \end{Bmatrix} = \begin{bmatrix} \ddots & & \\ & \dfrac{k_{rk}}{i\rho ck}\dfrac{H_n^{(1)'}(k_{rk}r)}{H_n^{(1)}(k_{rk}r_h)} & \\ & & \ddots \end{bmatrix} \times \begin{Bmatrix} \vdots \\ \hat{P}_n(r_h,k_{zk})_i \\ \vdots \end{Bmatrix}. \tag{25}$$

Once the helical wave spectra are calculated on the reconstruction surface by using Eqs. (24) and (25), the spatial distribution of the pressure field and the particle velocity can be obtained by applying the inverse discrete Fourier transform.

C. Regularization

Before the partial fields are projected, however, they must almost always be filtered to remove poorly determined, typically evanescent, components of the sound field. The latter process is referred to as regularization. The particular method used to perform filtering in the present work is described in this section.

When noise is present in a measurement, the relation between particle velocity and the measured sound pressure, Eq. (17), can be rewritten as

$$\hat{\mathbf{w}}_i^{\delta} = \mathbf{T}_w^{-1}\hat{\mathbf{p}}_i^{\delta}, \tag{26}$$

where $\hat{\mathbf{p}}_i^{\delta}$ represents the ith partial field vector, which is assumed to contain superimposed measurement noise having zero mean and a standard deviation of $\overline{\sigma}$, and $\hat{\mathbf{w}}_i^{\delta}$ is the corresponding, noise-contaminated particle velocity field.[8] When a field is projected back toward the source by using Eqs. (16) and (17), components associated with subsonic axial components and large circumferential orders are ampli-

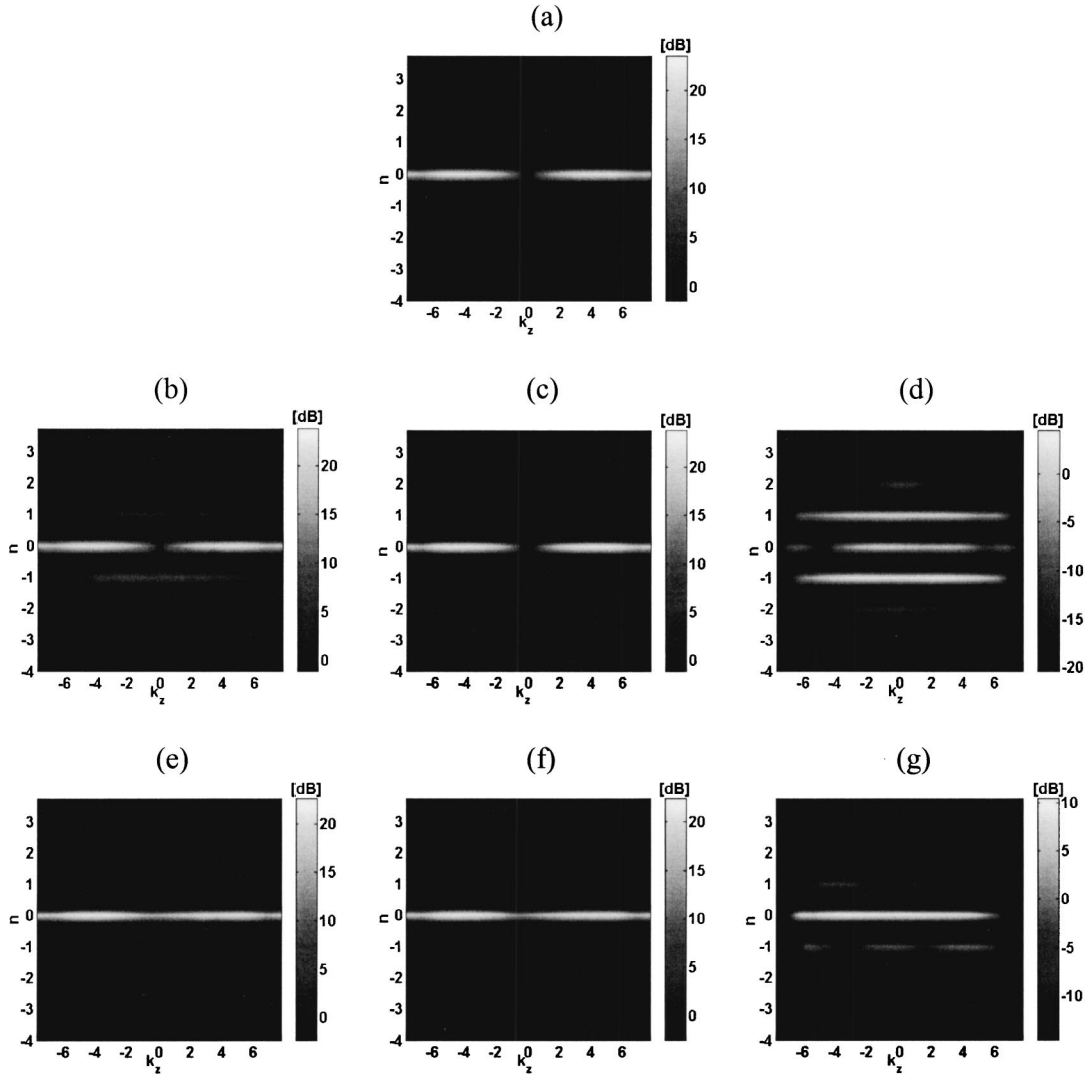


FIG. 7. The helical wave spectra at 330 Hz: (a) $\langle \hat{P}_n^{\alpha,\delta}(r_h,k_z)_1 \rangle$ for the first case; (b) $\langle \hat{P}_n^{\alpha,\delta}(r_h,k_z)_{1+2} \rangle$ for the second case; (c) $\langle \hat{P}_n^{\alpha,\delta}(r_h,k_z)_1 \rangle$ for the second case; (d) $\langle \hat{P}_n^{\alpha,\delta}(r_h,k_z)_2 \rangle$ for the second case; (e) $\langle \hat{P}_n^{\alpha,\delta}(r_h,k_z)_{1+2} \rangle$ for the third case; (f) $\langle \hat{P}_n^{\alpha,\delta}(r_h,k_z)_1 \rangle$ for the third case; (g) $\langle \hat{P}_n^{\alpha,\delta}(r_h,k_z)_2 \rangle$ for the third case.

fied exponentially and geometrically, respectively. This inverse process is ill-posed since the latter components often cannot be accurately estimated during a measurement, owing to their rapid radial decay. As a result, reconstruction quality can be poor unless additional processing is performed, as described next.

Here, the modified Tikhonov regularization method, used in conjunction with the Mozorov Discrepancy Principle, served as the basis for *k*-space filtering. The procedure followed here is based closely on that described in Ref. 8 and is based on a knowledge of the variance of the noise components. When applying the Tikhonov regularization method, the regularized inverse of $\mathbf{T}_{\dot{w}}$ is determined by minimizing the Tikhonov functional, J_α, with respect to $\hat{\mathbf{w}}_i^\delta$ for a fixed regularization parameter, α. Here J_α is defined as[7,8]

$$J_\alpha(\hat{\mathbf{w}}_i^\delta) = \|\mathbf{T}_{\dot{w}}\hat{\mathbf{w}}_i^\delta - \hat{\mathbf{p}}_i^\delta\|^2 + \alpha\|\mathbf{L}\hat{\mathbf{w}}_i^\delta\|^2, \tag{27}$$

where $\|\cdot\|$ denotes the L2 norm, the second term is a penalty function, and $\mathbf{L}$ denotes the regularization matrix. The functional J_α is found to be minimized when

$$\hat{\mathbf{w}}_i^{\alpha,\delta} = (\alpha\mathbf{I} + \mathbf{T}_{\dot{w}}^{\mathrm{H}}\mathbf{T}_{\dot{w}})^{-1}\mathbf{T}_{\dot{w}}^{\mathrm{H}}\hat{\mathbf{p}}_i^\delta = \mathbf{R}_\alpha\hat{\mathbf{p}}_i^\delta, \quad \text{for } \alpha > 0, \tag{28}$$

where the superscript α denotes the dependence on α, $\mathbf{R}_\alpha$ is the regularized inverse of $\mathbf{T}_{\dot{w}}$, and $\mathbf{L}$ has been set to the identity matrix, $\mathbf{I}$ (the standard form).[7]

The substitution of Eq. (19) into Eq. (28) gives

$$\hat{\mathbf{w}}_i^{\alpha,\delta} = \overline{\mathbf{U}}\mathbf{F}^\alpha \,\mathrm{diag}\left(\frac{1}{\lambda_1},\ldots,\frac{1}{\lambda_M}\right)\overline{\mathbf{U}}^{\mathrm{H}}\hat{\mathbf{p}}_i^\delta, \tag{29}$$

where

$$\mathbf{F}^\alpha = \mathrm{diag}\left(\frac{|\lambda_1|^2}{\alpha+|\lambda_1|^2},\ldots,\frac{|\lambda_M|^2}{\alpha+|\lambda_M|^2}\right) \tag{30}$$

is a *filter factor* that acts as a low-pass *k*-space filter, and where the regularization parameter α is determined based on a knowledge of the measurement noise level (see below). When Eq. (23) was implemented here, r was set equal to a: i.e., to the nominal radius of the source surface.

The filtered pressure on the hologram surface is then obtained by premultiplication of both sides of Eq. (29) by $\mathbf{T}_{\dot{w}}$: i.e.,

$$\hat{\mathbf{p}}_i^{\alpha,\delta} = \overline{\mathbf{U}}\mathbf{F}^{\alpha}\overline{\mathbf{U}}^{\mathrm{H}}\hat{\mathbf{p}}_i^{\delta}. \tag{31}$$

Equation (31) is the main results of this section: it is applied in turn to each column of $\hat{\mathbf{P}}$ to obtain the filtered partial fields, each of which is then projected using Eq. (16) or (17) as appropriate.

It has been suggested that the Tikhonov method could be improved by increasing the filter's slope.[8] In that case, a modified filter is obtained by substituting $\mathbf{L}=(\mathbf{I}-\mathbf{F}^{\alpha})\overline{\mathbf{U}}^{\mathrm{H}}$ into Eq. (27), instead of the standard form, to obtain

$$\mathbf{F}^{\alpha,1} = \mathrm{diag}\left\{\cdots |\lambda_j|^2 \Big/ \left[|\lambda_j|^2 + \alpha\left(\frac{\alpha}{\alpha+|\lambda_j|^2}\right)^2\right]\cdots\right\}. \tag{32}$$

In Eqs. (30) and (32), the unknown parameter, α, is determined by an application of the Mozorov Discrepancy Principle based on a knowledge of the standard deviation of the noise: i.e.,

$$\|\mathbf{T}_{\dot{w}}\hat{\mathbf{w}}_i^{\alpha,\delta} - \hat{\mathbf{p}}_i^{\delta}\| = \overline{\sigma}\sqrt{M}. \tag{33}$$

Since $\mathbf{T}_{\dot{w}}\hat{\mathbf{w}}_i^{\alpha,\delta}$ is the filtered pressure, $\hat{\mathbf{p}}_i^{\alpha,\delta}$, Eq. (33) indicates that α should be chosen to remove an amount of energy equal to that of the noise contained in the measurement. In practice, α can be determined by varying α iteratively until Eq. (33) is satisfied. The standard deviation of the measurement noise in each partial field can be estimated by using the equation[8]

$$\|\overline{\mathbf{u}}_q^{\mathrm{H}}\hat{\mathbf{p}}_i^{\delta}\|/\sqrt{Q} \approx \overline{\sigma}, \quad q=1,...,Q, \tag{34}$$

where $\overline{\mathbf{u}}_q$ is a column vector of $\overline{\mathbf{U}}$, and Q is the number of the vectors that are used when evaluating Eq. (34). The norm is taken by using the appropriately chosen $\overline{\mathbf{u}}_q$'s that are associated with noise-dominated components. When Eq. (34) is implemented in a cylindrical geometry, the helical wave spectrum components that lie outside the circle defined by the radius, $k_{\max} \equiv \max(k_{z\max}, n_{\max}/a)$, (i.e., the helical wave number components, $k_h \equiv \sqrt{k_z^2+(n/a)^2}$, that are larger than $k_{\max}$), are chosen since in a well-arranged measurement, i.e., one in which the sound field is sampled spatially at a sufficiently high rate, the levels associated with the noise-free sound field should be negligible near the edges of the helical wave spectrum, leaving only the measurement noise contribution. When used in conjunction with the Mozorov Discrepancy Principle, Eq. (34) is strictly valid only when Q is not too large, and this approach does not guarantee the optimal result; however, this approach has been found to be robust and applicable in cases where other approaches fail to successfully identify a regularization parameter.[14]

D. Data treatment

In holographic formulations, the harmonic time dependence, $e^{-i\omega t}$, is usually used. In that case positive values of k_z correspond to waves propagating in the positive z direction and vice versa. Thus, in NAH as implemented here, the definition of the temporal Fourier spectrum differs subtly from the conventional one: i.e., it is taken to have the form

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{i\omega t}\,dt, \tag{35}$$

where $f(t)$ is a time history and $F(\omega)$ is its Fourier transform. The spectrum calculated by using Eq. (35) is the conjugate of the spectrum calculated by following the conventional definition:

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}\,dt. \tag{36}$$

Therefore a conjugate operation is required when preparing data for input to the holographic process when the cross-spectra are calculated using a commercial FFT analyzer in which the definition of Eq. (36) is used. There are two alternatives in the latter case: take a conjugate of the cross-spectra [i.e., Eqs. (3) and (4)] before input to the partial field decomposition, or take the conjugate of the calculated partial fields before input to the NAH procedure [i.e., Eqs. (16) and (17)]. Some authors use a different notation when defining the cross-spectra, e.g., $\mathbf{C}_{rp} = \mathrm{E}[\mathbf{r}\mathbf{p}^{\mathrm{H}}]$, in which case the temporal Fourier spectrum is calculated based on Eq. (36).[16,20] Thus, care should be taken when implementing the holographic procedure to ensure consistency in the choice of transform sign convention since phase plays a crucial role when the sound field is projected from one surface to another.

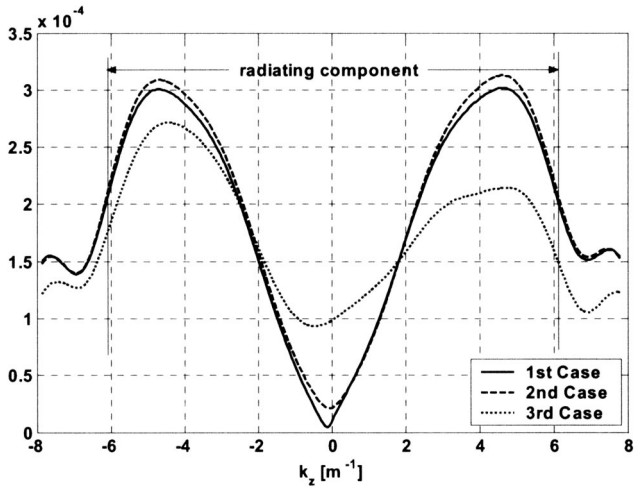


FIG. 8. A comparison of the $\langle \hat{P}_0^{\alpha,\delta}(r_h, k_z)_{\mathrm{total}} \rangle$ helical wave spectral components at 330 Hz.

E. Considerations related to the measurement of aeroacoustic sources

Unlike a structure-borne noise source, aeroacoustic sources are usually accompanied by a mean flow that may impinge on either the field or reference microphones, or both. When microphones are exposed to flow, the measured data contains the self-noise generated by the interaction of the flow and the microphone, as well as the desired signal, thus corrupting the decomposed field data. This problem can be addressed by the proper placement of the various micro-

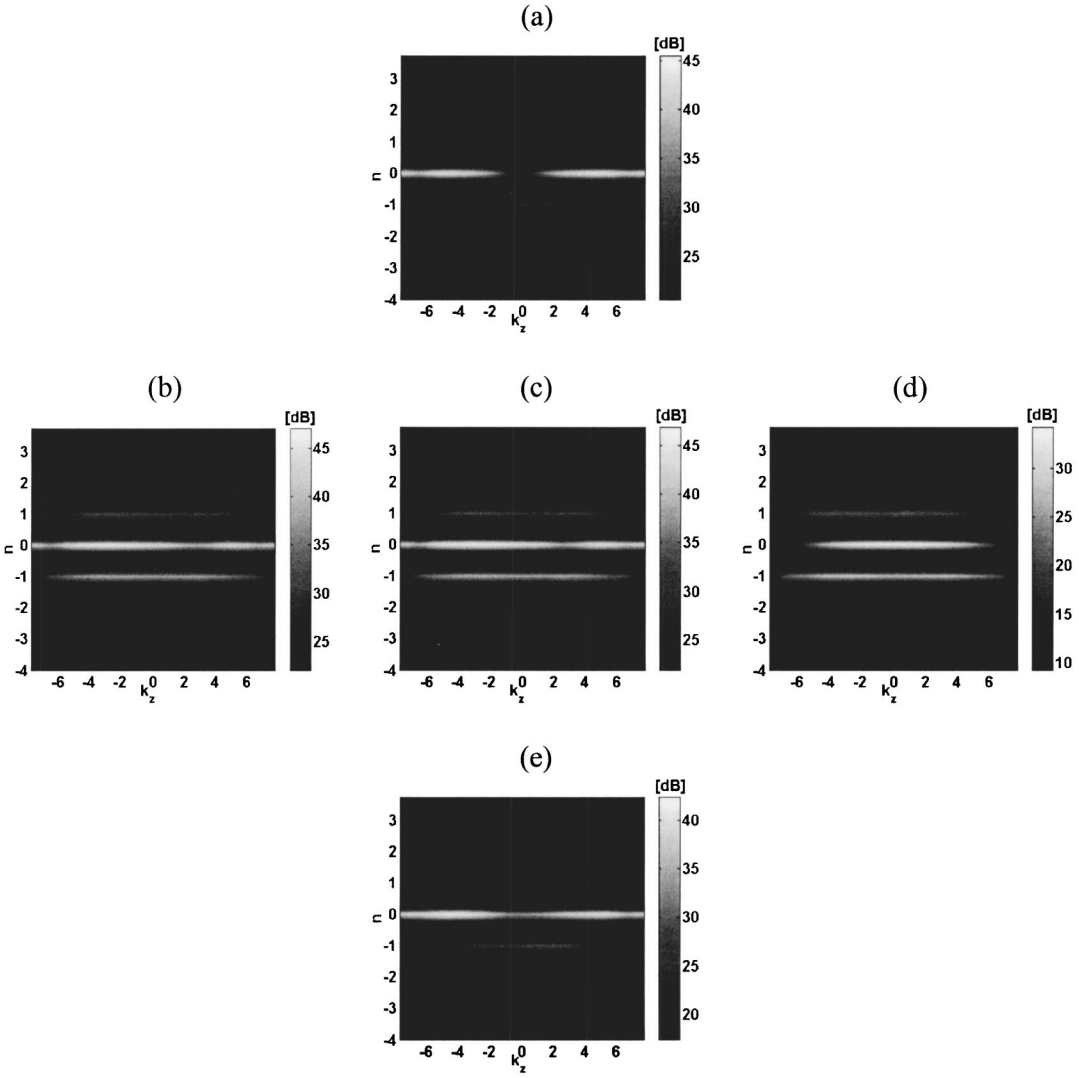


FIG. 9. The helical wave spectra at 361 Hz: (a) $\langle \hat{P}_n^{\alpha,\delta}(r_h,k_z)_1 \rangle$ for the first case; (b) $\langle \hat{P}_n^{\alpha,\delta}(r_h,k_z)_{1+2} \rangle$ for the second case; (c) $\langle \hat{P}_n^{\alpha,\delta}(r_h,k_z)_1 \rangle$ for the second case; (d) $\langle \hat{P}_n^{\alpha,\delta}(r_h,k_z)_2 \rangle$ for the second case; (e) $\langle \hat{P}_n^{\alpha,\delta}(r_h,k_z)_1 \rangle$ for the third case.

phones. When placing microphones, two conditions should be satisfied: first, the reference microphones should be placed in regions of low flow velocity to minimize the generation of self-noise, and second, the reference microphones should not sense any flow noise generated by the interaction of the flow with the field microphones.[22] These requirements can be demonstrated in the following way.

When both reference and field microphones are placed in a flow, the measured reference signal, $\mathbf{r}_{\text{total}}$, and the field microphone signal, $\mathbf{p}_{\text{total}}$, can be expressed as

$$\mathbf{r}_{\text{total}} = \mathbf{r}_s + \mathbf{r}_{sr} + \mathbf{r}_{sf}, \tag{37}$$

$$\mathbf{p}_{\text{total}} = \mathbf{p}_s + \mathbf{p}_{sr} + \mathbf{p}_{sf}, \tag{38}$$

where the subscripts *s*, *sr*, and *sf* represent the signal from the actual source, the self-noise generated by flow over the references, and the self-noise generated by flow over the field microphones, respectively. It is assumed here that all sources are mutually uncorrelated. Based on Eqs. (37) and (38), the cross-spectra between the references, and between the references and the field microphones, can be expressed as

$$\mathbf{C}_{rr,\text{total}} = \mathrm{E}[\mathbf{r}^*_{\text{total}}\mathbf{r}^{\mathrm{T}}_{\text{total}}] = \mathrm{E}[\mathbf{r}^*_s\mathbf{r}^{\mathrm{T}}_s] + \mathrm{E}[\mathbf{r}^*_{sr}\mathbf{r}^{\mathrm{T}}_{sr}] + \mathrm{E}[\mathbf{r}^*_{sf}\mathbf{r}^{\mathrm{T}}_{sf}], \tag{39}$$

$$\mathbf{C}_{rp,\text{total}} = \mathrm{E}[\mathbf{r}^*_{\text{total}}\mathbf{p}^{\mathrm{T}}_{\text{total}}] = \mathrm{E}[\mathbf{r}^*_s\mathbf{p}^{\mathrm{T}}_s] + \mathrm{E}[\mathbf{r}^*_{sr}\mathbf{p}^{\mathrm{T}}_{sr}] + \mathrm{E}[\mathbf{r}^*_{sf}\mathbf{p}^{\mathrm{T}}_{sf}]. \tag{40}$$

It is very important that the reference cross-spectra should not be corrupted by unwanted noise since the reference cross-spectra play a crucial role in the partial field decomposition. Ideally, the measured cross-spectra should include only the first terms in Eqs. (39) and (40). Since the self-noise is created locally, it is clear that $\mathbf{r}_{sf}$ and $\mathbf{p}_{sr}$ will be negligibly small as long as the reference and field microphones are not placed too close together; thus, the third term in Eq. (39) can usually be disregarded. The off-diagonal components of the second term in Eq. (39), $\mathrm{E}[\mathbf{r}^*_{sr}\mathbf{r}^{\mathrm{T}}_{sr}]$ (i.e., the cross-spectra between the references), are suppressed by spectral averaging, but the diagonal components (i.e., the reference self-noise auto-spectra) remain. To remove the latter terms, it is thus absolutely necessary that the self-noise

generated by flow over the references should be negligibly small compared to the actual source signal. The best way to satisfy the latter condition is simply to place the references outside the flow in still air, in which case $\mathbf{r}_{sr}$ is equal to zero, thus eliminating the second terms in both Eqs. (39) and (40). Unless the flow speed over the field microphones is significant, the inequality $|\mathrm{E}[\mathbf{r}_{sf}^{*}\mathbf{p}_{sf}^{\mathrm{T}}]| \ll |\mathrm{E}[\mathbf{r}_{s}^{*}\mathbf{p}_{s}^{\mathrm{T}}]|$ should hold, thus making the third term in Eq. (40) negligible. Thus, in principle, the self-noise effect can be successfully removed from the measured cross-spectra as long as the self-noise generated by flow over the reference microphones is negligibly small. Under these conditions, self-noise does not corrupt the holographic processing. Note again that there are stronger requirements placed on the positioning of the reference microphones than on the field microphones.

The cross-spectral matrix of the field pressures is

$$\mathbf{C}_{pp,\mathrm{total}} = \mathrm{E}[\mathbf{p}_{\mathrm{total}}^{*}\mathbf{p}_{\mathrm{total}}^{\mathrm{T}}] = \mathrm{E}[\mathbf{p}_{s}^{*}\mathbf{p}_{s}^{\mathrm{T}}] + \mathrm{E}[\mathbf{p}_{sr}^{*}\mathbf{p}_{sr}^{\mathrm{T}}] + \mathrm{E}[\mathbf{p}_{sf}^{*}\mathbf{p}_{sf}^{\mathrm{T}}]. \quad (41)$$

The self-noise term in Eq. (41) may be problematic since the amplitudes of the diagonal components of $\mathrm{E}[\mathbf{p}_{sf}^{*}\mathbf{p}_{sf}^{\mathrm{T}}]$ (the autospectra of field microphone self-noise) are possibly significant. In this case, the second term of Eq. (41) is equal to zero when the references are placed exterior to the flow. The off-diagonal components of $\mathrm{E}[\mathbf{p}_{sf}^{*}\mathbf{p}_{sf}^{\mathrm{T}}]$ are also small if the field microphones are spaced sufficiently far apart so that one microphone does not sense the flow noise generated at neighboring microphones. However, the diagonal components of $\mathrm{E}[\mathbf{p}_{sf}^{*}\mathbf{p}_{sf}^{\mathrm{T}}]$ are not suppressed by averaging and may be significantly large when the flow speed over the field microphones is relatively large. The latter effect does not affect the accuracy of the partial field decomposition, but causes the virtual coherence on the hologram surface to drop since $C_{p_i p_i}$ in Eq. (11) will be overestimated.[23] Therefore, when there is flow over the field microphones, care should be taken when interpreting the virtual coherence to determine whether or not one particular set of virtual references is sufficient.

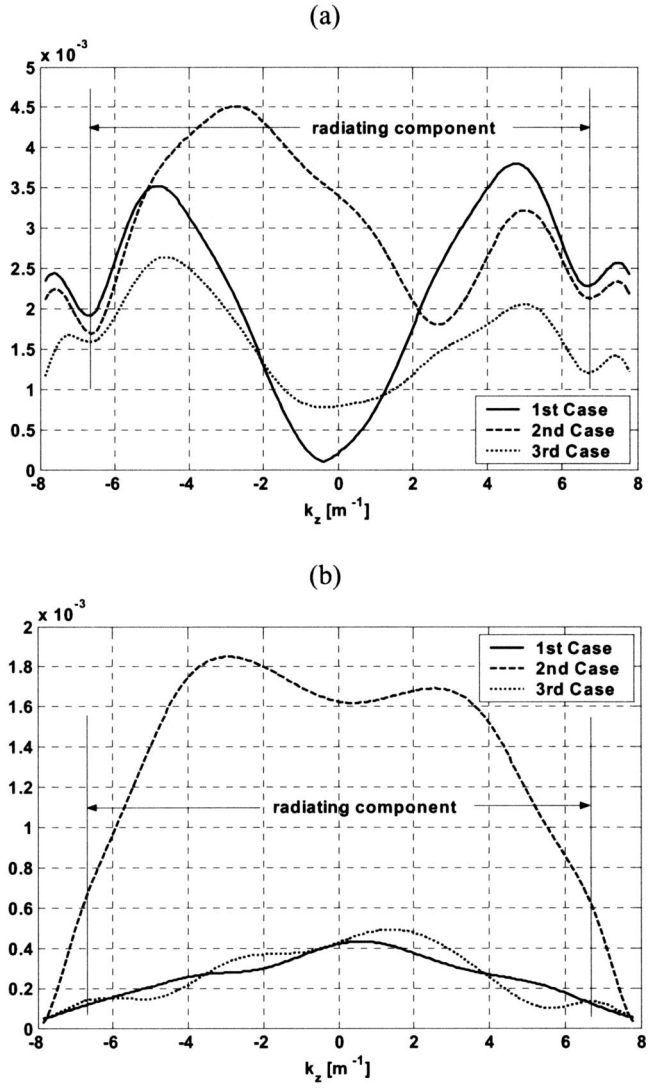


FIG. 10. A comparison of the zero and first circumferential order helical wave spectrum components at 361 Hz: (a) $\langle \hat{P}_0^{\alpha,\delta}(r_h,k_z)_{\mathrm{total}} \rangle$; (b) $(\langle \hat{P}_{-1}^{\alpha,\delta}(r_h,k_z)_{\mathrm{total}} \rangle^2 + \langle \hat{P}_1^{\alpha,\delta}(r_h,k_z)_{\mathrm{total}} \rangle^2)^{1/2}$.

III. EXPERIMENTAL APPARATUS AND PROCEDURES

The various partial field decomposition, source nonstationarity compensation, regularization, and projection procedures discussed above were implemented as shown in Fig. 1. A number of the intermediate results in this process will be described in the next section.

To perform the measurements, a cylindrical frame supporting an eight-by-eight array of measurement positions was constructed on a surface enclosing the source of interest: see Fig. 2. The measurement positions were separated by 39 cm in the z direction and 45° in the ϕ direction, the measurement radius was 50 cm, and the source was placed at the center of the frame. In the tests, three reference microphones were used. Two references were located downstream and one reference was located upstream: see Fig. 2. Both references and field microphones were located out of the flow to minimize self-noise effects.

The cross-spectra between three references, and between the references and the field microphones, were calculated. The measurements were performed by using four field microphones (Larson Davis TMS426B01) to scan the hologram surface in 16 steps. The three reference microphones were also Larson Davis TMS426B01. The measurement bandwidth was 512 Hz, the frequency resolution was 1 Hz, a Hanning window was applied to the data, and the various spectral estimates were calculated using 300 averages with 50% overlap.

Tests were performed using a 10 cm diameter ducted fan in three configurations: see Fig. 3. In the first case, the fan and duct assembly was carefully sealed so that there was no leakage. In the second case, two small holes were made on the side of the fan housing [see Fig. 3(a)], and in the third case, a sound absorbing lining made of glass fiber approximately 1.5 cm thick (identified as ''green'' in Ref. 24) was added to the inner circumference of the downstream duct section [see Fig. 3(b)]. In all cases, the upstream and downstream duct sections had an inner diameter of 10 cm and were 10.3 cm in length.

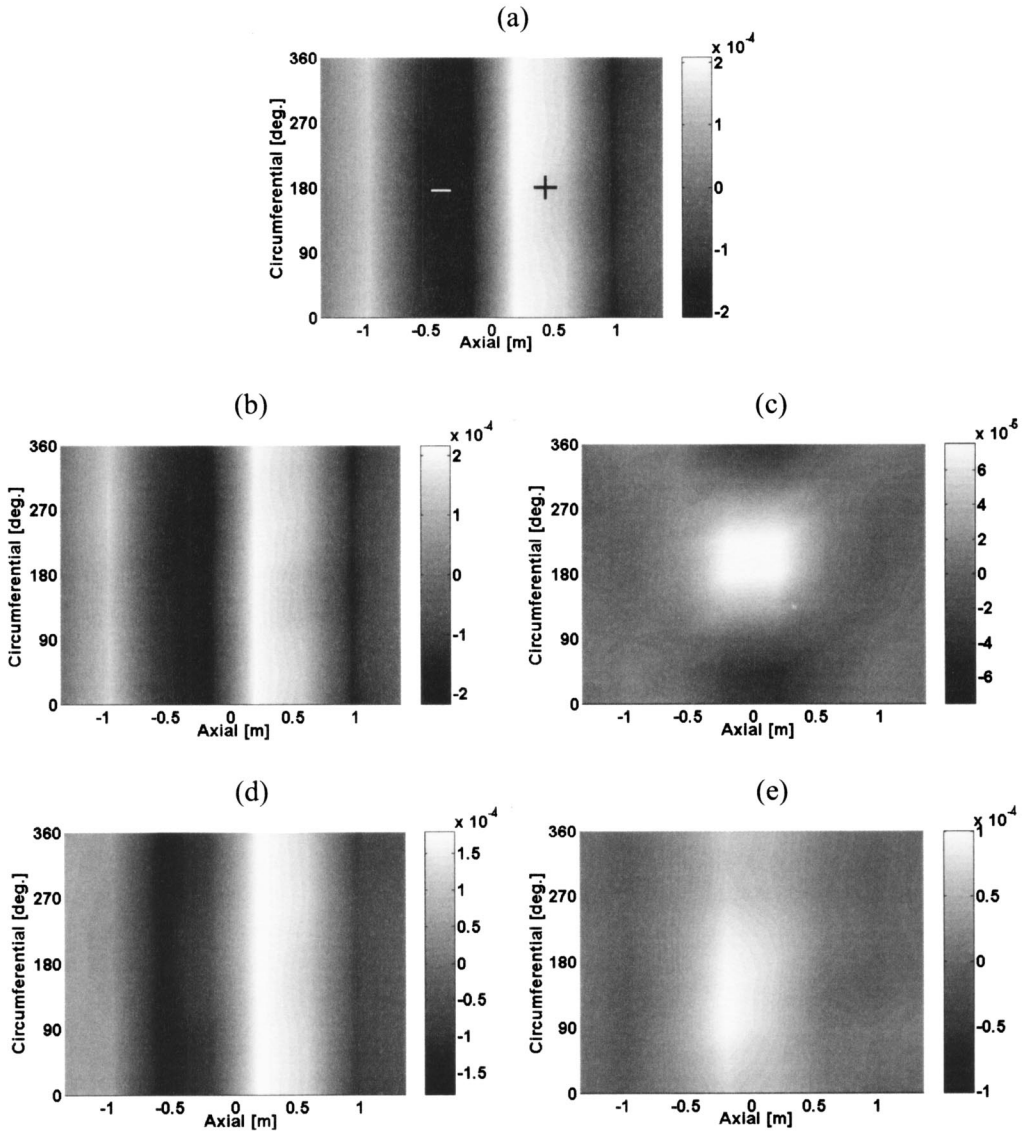


FIG. 11. Real parts of the partial fields on the hologram surface at 330 Hz: (a) $\hat{\mathbf{p}}_1^{\alpha,\delta}$ for the first case; (b) $\hat{\mathbf{p}}_1^{\alpha,\delta}$ for the second case; (c) $\hat{\mathbf{p}}_2^{\alpha,\delta}$ for the second case; (d) $\hat{\mathbf{p}}_1^{\alpha,\delta}$ for the third case; (e) $\hat{\mathbf{p}}_2^{\alpha,\delta}$ for the third case.

IV. RESULTS

Since the fan used for these tests had five blades, the blade passing tone at five times the shaft rotation frequency was the largest component in the radiated sound field. The fan was operated at the same speed in all tests. Results are presented at the blade passing frequency (361 Hz) and for one broadband noise component (330 Hz).

A. Singular value decomposition

The singular values of the reference cross-spectral matrix [Eq. (3)] were examined to establish the number of virtual references needed for the partial field decomposition [Eq. (11)] and hence the number of significant partial fields in each case. In all three cases, the levels of the third singular values at both frequencies dropped close to the background noise level. Therefore, the three references used for the measurements, in principle, were sufficient to describe the sound field.

Empirically, singular values smaller than the largest one by 30 dB may be considered negligible. Therefore, one reference was determined to be sufficient in the first (i.e., leak-free) case [see Fig. 4(a)] at both frequencies of interest. This result suggests that the sound field was generated by a single coherent source mechanism at both 330 and 361 Hz, an observation that is consistent with the compact source assumption. In the second case, in which leaks were introduced, the second singular value was less than 20 dB smaller than the first at the blade passing frequency, so at least two references were required [see Fig. 4(b)]: i.e., the introduction of the holes in the fan housing apparently causes a second source mechanism to be significant at the blade passing frequency. The leak also caused a slight increase of the second singular value at 330 Hz: thus, two references were also required in this case, even though the contribution of the second source was expected to be small. In the third case, the second singular value was generally larger than that in the first case at all frequencies [see Fig. 4(c)]. The introduction of the lining thus seems to have created a second broadband source

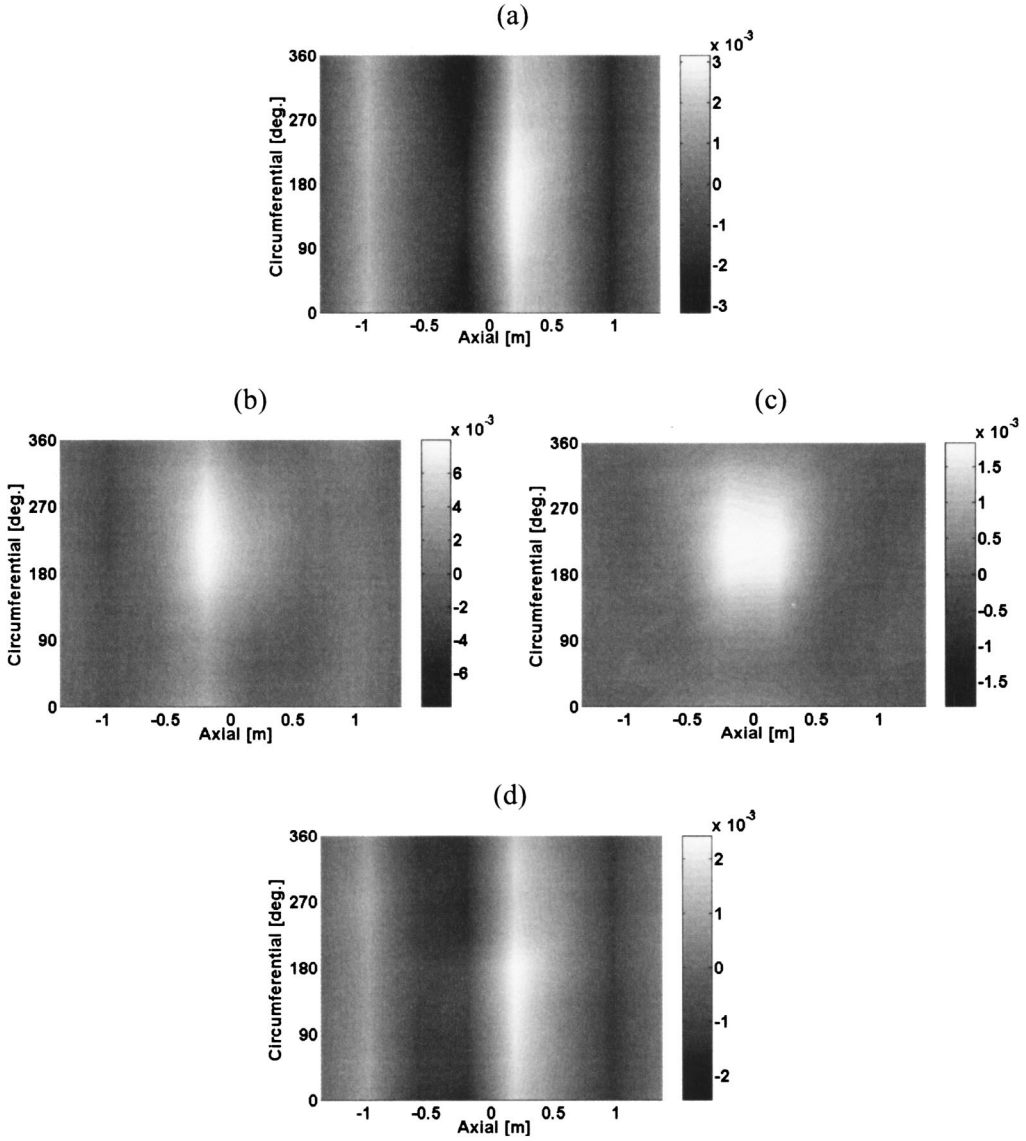


FIG. 12. Real parts of the partial fields on the hologram surface at 361 Hz: (a) $\hat{\mathbf{p}}_1^{\alpha,\delta}$ for the first case; (b) $\hat{\mathbf{p}}_1^{\alpha,\delta}$ for the second case; (c) $\hat{\mathbf{p}}_2^{\alpha,\delta}$ for the second case; (d) $\hat{\mathbf{p}}_1^{\alpha,\delta}$ for the third case.

mechanism. Note that the magnitude of the second singular value at the blade passing frequency was not affected by the presence of the lining. Two references were required to reconstruct the sound field at 330 Hz in the third case; a single reference was sufficient at the blade passing frequency.

B. Virtual coherence

The virtual coherence on the hologram surface can be used to verify that the number of virtual references determined by use of SVD is sufficient (subject to the restrictions described in Sec. II E). In the first case, the virtual coherence calculated using the first partial field only is unity at both frequencies [see Figs. 5(a) and 6(a)]. In the second case, the coherence based on a single reference was low at several locations at the blade passing frequency, in particular: see Fig. 6(b). However, the sum of the coherences calculated using the first and the second partial fields was close to unity at both frequencies [see Figs. 5(c) and 6(c)]. In the third case, the sum of the coherences calculated using the first and the second partial fields at 330 Hz and the first partial field at 361 Hz were close to unity: see Figs. 5(d) and (e), and Fig. 6(d). The virtual coherence results thus confirm that the number of references was chosen appropriately on the basis of the SVD results.

C. NAH results

In the first case, i.e., the sealed, ducted fan, it was found that both the broadband noise and the blade passing tone exhibited a clear dipole-like directivity. In contrast, the radiation patterns in the second and third cases were more complex, owing to the leakage through the two holes in the side of fan housing in one case and because of the effect of the acoustical lining within the downstream duct in the other case. Those modifications apparently broke the symmetry of the flow and caused the dipole-like sound radiation pattern to be disturbed.

When processing the partial field data, a spatial window and zero padding were applied in the axial direction to re-

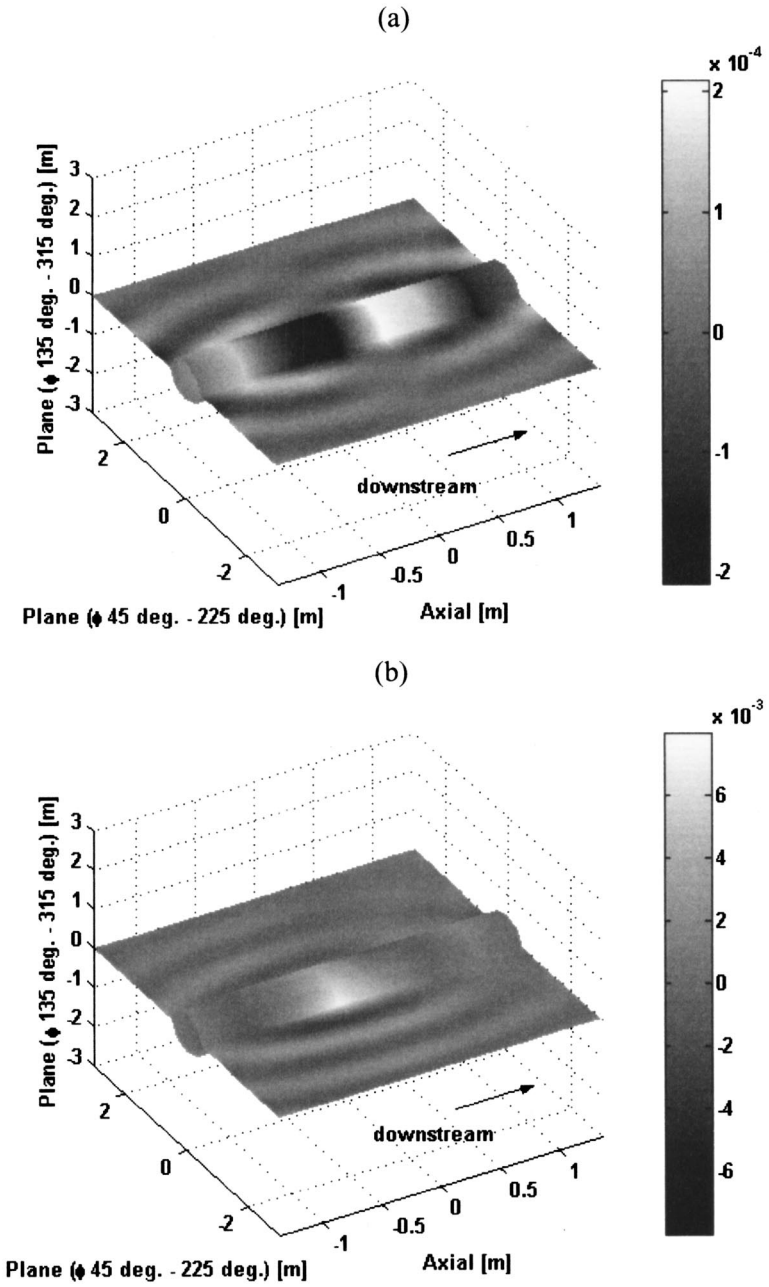


FIG. 13. Three-dimensional reconstruction of the pressure field (the real part of the first partial field plotted): (a) for the first case at 330 Hz (dipole-like radiation); (b) for the second case at 361 Hz. All results are calculated for the ϕ=(45°,225°) plane.

duce leakage and wrap-around errors, respectively. Data at the first and the last (i.e., eighth) measurement location in the axial direction were weighted by a factor of 0.5, and then 120 zeros were added to data. All the results presented in Figs. 7–13 were filtered by using the modified Tikhonov regularization method in conjunction with the filter factor of Eq. (32). In the case of aeroacoustic sources, the radius of the source surface is not well defined. In the regularization process used here, a was set to be 15 cm based on observed flow patterns near the duct inlet and exit.

1. Helical wave spectra

In the figures discussed in this section, $\langle \hat{P}_n^{\alpha,\delta}(r_h,k_z)\rangle$ denotes the root mean square helical wave spectrum on the hologram surface: each spectrum is a function of n, the circumferential order and k_z, the axial wave number. The subscript $1+2$ denotes the square root of the sum of the squared spectra of partial fields 1 and 2.

In the first, leak-free case, the helical wave spectra at both frequencies displayed dipole-like characteristics: the $n=0$ circumferential component was the main contributor to the sound radiation and the $k_z=0$ component was small, resulting in a radiation null perpendicular to the dipole axis: see Figs. 7(a) and 9(a). In addition, the $n=0$ components were symmetric with respect to $k_z=0$ [see Figs. 8 and 10(a), where the range of the axial wave number, k_z, associated with radially propagating components is also indicated].

In the second case, the higher-order circumferential components were more significant due to the existence of leakage. The effect of leakage was clearly apparent at the

blade passing frequency, there being significant contributions from the $n=\pm 1$ mode in both partial fields [see Figs. 9(c) and (d)]. At 330 Hz, higher circumferential orders were visible in the second partial field [see Fig. 7(d)], but the total field remained dipole-like in nature since the second partial field was approximately one order of magnitude smaller than the first [see Fig. 7(b)]. The zeroth- and first-order component of the helical wave spectrum at 361 Hz were significantly asymmetric with respect to $k_z=0$ [see Figs. 10(a) and (b)]. As a result, it was expected that at the blade passing frequency sound would radiate primarily in the negative-z (i.e., upstream) direction and that the radiation would be nonaxisymmetric: i.e., the blade passing tone was not expected to radiate like a dipole. At 330 Hz, the $n=0$ circumferential component remained nearly symmetric with respect to $k_z=0$ (see Fig. 8): therefore, the broadband component at this frequency was expected to display a dipole-like directivity.

In the third case, the effect of the acoustical lining was visible in two ways: an attenuation of the radiation in the downstream direction due to the sound absorbing material in the downstream duct and an overall decrease in level, owing to the reduced flow rate caused by the increase of the flow resistance resulting from the reduction of the duct area by the lining. The former effect caused the positive-going $n=0$ components to be smaller than the negative-going ones at both frequencies [see Figs. 8 and 10(a)]. However, the attenuation was relatively small, owing to the fact that the lining was thin compared to a wavelength at the frequencies of interest. The flow resistance effect was judged to have a larger impact on the sound radiation than the attenuation effect. This conclusion followed from the observation that the sound level was reduced in both the downstream and the upstream directions [compare the amplitudes of the spectra in Figs. 7(a) and (e), and 9(a) and (e)]. In contrast to the first case, it was expected that there would be some sound radiation perpendicular to the fan axis since a nonzero $k_z=0$ component appeared in the helical wave spectrum in this case [see Figs. 7(e) and 9(e)].

2. Spatial distribution of partial fields

In all cases, the first partial fields at 330 Hz showed dipole-like characteristics: i.e., the fields on the up- and downstream sides of the fan were 180° out-of-phase [see Figs. 11(a), (b), and (d)]. The second partial field at 330 Hz in the second case exhibited a clear $n=1$ circumferential mode [see Fig. 11(c)]. The second partial field at 330 Hz in the third case was approximately axially symmetric with respect to the fan location (i.e., the second partial field was monopole-like), and its amplitude was comparable to that of the first partial field [see Fig. 11(e)].

The first and third cases at 361 Hz also exhibited dipole-like patterns [see Figs. 12(a) and (d)], but owing to the contribution of the higher circumferential modes, which were probably caused by the geometrical asymmetry of a duct itself and a duct lining, it was slightly disturbed when compared with the broadband case. In the second case, the radiation pattern at the blade passing frequency was significantly changed by the effects of leakage [see Figs. 12(b) and (c)].

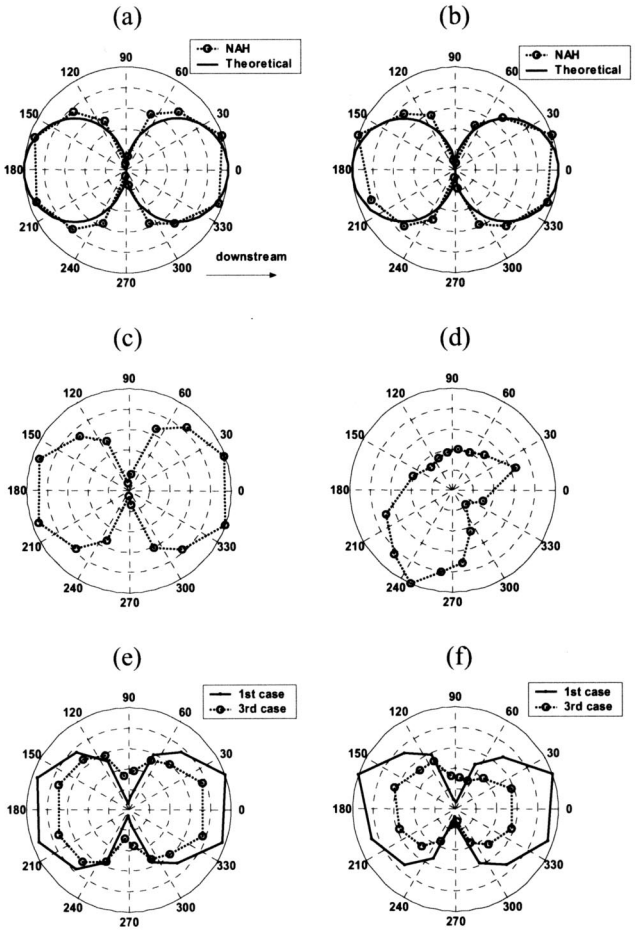


FIG. 14. Directivity plots: (a) for the first case at 330 Hz (along with the comparison to theoretical dipole directivity); (b) for the first case at 361 Hz (along with the comparison to theoretical dipole directivity); (c) for the second case at 330 Hz; (d) for the second case at 361 Hz; (e) for the third case at 330 Hz (along with the comparison to the first case); (f) for the third case at 361 Hz (along with the comparison to the first case). All results are calculated for the $\phi=(45°,225°)$ plane.

Three-dimensional reconstructions of the sound radiation patterns for the first two cases are shown in Fig. 13. The dipole-like pattern at 330 Hz in the first case is clear [see Fig. 13(a)] but the radiation pattern at 361 Hz in the second case was clearly not dipole-like, owing to the effect of the leaks [see Fig. 13(b)].

3. Source directivity

Finally, the directivity for the various source cases was calculated by reconstructing pressures at 16 points that shared the same axial locations as the measurement points on the hologram surface and were on the plane defined by $\phi=45°$ and 225°. All points were located 1.365 m from the origin (i.e., the location of the fan) and only radiating components were included in the calculation in order to represent the farfield radiation directivity. This approach is easy to implement but is valid only when the source location is known.

In the first case, the directivity patterns compare well with a theoretical dipole at both frequencies of interest [see Figs. 14(a) and (b)]. In the second case, the broadband com-

ponent at 330 Hz still radiates in a dipole-like manner [see Fig. 14(c)]. However, at the blade passing frequency, it was observed that the main lobe extended in the negative-z direction, as expected from the helical wave spectrum [see Fig. 14(d)] and the overall radiation pattern is neither axisymmetric nor axially symmetric. In the third case, the overall sound level decreased at both frequencies, but increased slightly in the direction perpendicular to the dipole axis [see Figs. 14(e) and (f)]. In the 330 Hz result, the radiation null is shifted to approximately 100° and 260° (from 90° and 270°, respectively). Finally, at the blade passing frequency, the downstream radiation is slightly reduced by comparison with the upstream radiation, presumably, because of the absorbent lining in the downstream duct section.

V. DISCUSSION

When measuring sound radiation from aeroacoustic sources, the choice of hologram aperture size and the measurement radius may be restricted owing to the constraint that the field microphones should not be exposed to significant flow if the virtual coherence is to be meaningful. In the latter case, the quantity of spatial data available decreases as the measurement radius is made smaller (in order to measure small wavelength components, for example). That is, as the measurement radius is made smaller, the degree of axial truncation increases. In particular, when a dipole axis is aligned parallel with the array axis, the aperture size should be very large since the maximum radiation occurs in the axial direction. This problem can be overcome by making the dipole axis oblique with respect to the array axis. However, another problem arises in that case since field microphones would then be exposed to the flow. These characteristics impose a limit on the frequency range since the measurement radius should be large enough so that the measurement aperture can be extended sufficiently far in the axial direction so that the sound field decays sufficiently to avoid truncation effects without the field microphones being exposed to *significant* flow. As noted earlier, however, holographic projections may be performed successfully even when the field microphones are exposed to the flow. Thus, the aperture size restriction discussed here applies only when it is wished to use the virtual coherence to confirm the sufficiency of the reference set.

VI. CONCLUSIONS

In this article, the application of cylindrical NAH procedures to aeroacoustic sources has been described. First, the procedures for determining the number of references required to describe the sound field with the help of SVD and virtual coherence analysis were described. The procedure for decomposing the sound field into partial fields by using the virtual coherence method was described next. After decomposition, each partial field can be projected to the reconstruction surface after appropriate regularization. By the comparison of the visualized sound fields corresponding to the various configurations, the ability of NAH to quantify noise source modification effects was shown. It was also shown that, unlike beamforming techniques, NAH can be successfully used to determine the farfield radiation from aeroacoustic dipole sources.

The major conclusion of the present work is that NAH may be successfully used to visualize aeroacoustic sources as long as the reference microphones are not exposed to significant flow. Further, it has been shown that successful holographic projections may be performed even if the field microphones are exposed to flow (given only that the reference microphones are not), but that in this the virtual coherence cannot then be used to assess the sufficiency of a particular set of reference microphones.

ACKNOWLEDGMENTS

This research was supported by Rolls-Royce Corporation, Indianapolis, and the Indiana 21st Century Research and Technology Fund.

[1] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic Press, London, UK, 1999).
[2] H. V. Fuchs, ''On the application of acoustic ''mirror,'' ''telescope'' and ''polar correlation'' techniques to jet noise source location,'' J. Sound Vib. **58**, 117–126 (1978).
[3] J. Billingsley and R. Kinns, ''The acoustic telescope,'' J. Sound Vib. **48**, 485–510 (1976).
[4] M. J. Fisher, M. Harper-Bourne, and S. A. L. Glegg, ''Jet engine noise source location: The polar correlation technique,'' J. Sound Vib. **51**, 23–54 (1977).
[5] J. Hald, ''STSF—A unique technique for scan-based near-field acoustic holography without restrictions on coherence,'' B&K Technical Review No. 1, 1988.
[6] D. L. Hallman and J. S. Bolton, ''A comparison of multi-reference nearfield acoustical holography procedures,'' *Proceedings of NOISE-CON 94*, pp. 929–934 (1994).
[7] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems* (SIAM, Philadelphia, PA, 1998).
[8] E. G. Williams, ''Regularization methods for near-field acoustical holography,'' J. Acoust. Soc. Am. **110**, 1976–1988 (2001).
[9] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems* (Springer-Verlag, New York, 1996).
[10] B.-K. Kim and J.-G. Ih, ''On the reconstruction of the vibro-acoustic field over the surface enclosing an interior space using the boundary element method,'' J. Acoust. Soc. Am. **100**, 3003–3016 (1996).
[11] Z. Quabili, S. Sureshkumar, and S. T. Raveendra, ''Noise source identification in a cab interior,'' SAE paper 01-1630, in *Proceedings of the SAE Noise and Vibration Conference*, 2001.
[12] G. H. Golub, M. Heath, and G. Wahba, ''Generalized cross validation as a method for choosing a good ridge parameter,'' Technometrics **21**, 215–223 (1979).
[13] P. C. Hansen, ''Analysis of discrete ill posed problems by means of the L-curve,'' SIAM Rev. **34**, 561–580 (1992).
[14] S. Sureshkumar and S. T. Raveendra, ''An analysis of regularization errors in generalized nearfield acoustical holography,'' SAE Paper 01-1616, in Ref. 11.
[15] P. Jordan and J. A. Fitzpatrick, ''Measurement of an aeroacoustic dipole using a linear microphone array,'' J. Acoust. Soc. Am. **111**, 1267–1273 (2002).
[16] H.-S. Kwon and J. S. Bolton, ''Partial field decomposition in near-field acoustical holography by the use of singular value decomposition and partial coherence procedures,'' *Proceedings of NOISE-CON 98*, pp. 649–654 (1998).
[17] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*, 3rd ed. (Wiley, New York, 2000).
[18] S. H. Yoon and P. A. Nelson, ''A method for the efficient construction of acoustic pressure cross-spectral matrices,'' J. Sound Vib. **233**, 897–920 (2000).
[19] M. S. Kompella, P. Davies, R. J. Bernhard, and D. A. Ufford, ''A technique to determine the number of incoherent sources contributing to the response of a system,'' Mech. Syst. Signal Process. **8**, 363–380 (1994).

[20] H.-S. Kwon, Y.-J. Kim, and J. S. Bolton, ''Compensation for source nonstationarity in multireference, scan-based near-field acoustical holography,'' J. Acoust. Soc. Am. **113**, 360–368 (2003).
[21] E. G. Williams, H. D. Dardy, and K. B. Washburn, ''Generalized nearfield acoustical holography for cylindrical geometry: Theory and experiment,'' J. Acoust. Soc. Am. **81**, 389–407 (1987).
[22] M. Nakamura, T. Komine, M. Tsuchiya, and J. Hald, ''Measurement of aerodynamic noise using STSF,'' B&K application note, 1990.
[23] G.-S. Kim, H.-K. Park, S.-G. Jung, and K.-D. Ih, ''Development of acoustic holography and its application in Hyundai aeroacoustic wind tunnel,'' SAE Paper 01-1497 in Ref. 11.
[24] B. H. Song and J. S. Bolton, ''A Transfer-matrix approach for estimating the characteristic impedance and wave numbers of limp and rigid porous materials,'' J. Acoust. Soc. Am. **107**, 1131–1152 (2000).

Modeling the effect of suppression on the periodicity of stimulus frequency otoacoustic emissions

Ben Lineton[a] and Mark E. Lutman
Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom

(Received 10 November 2002; revised 18 April 2003; accepted 23 April 2003)

The distributed roughness theory of the origins of spectral periodicity in stimulus frequency otoacoustic emissions (SFOAEs) predicts that the spectral period will be altered by suppression of the traveling wave (TW) [Zweig and Shera, J. Acoust. Soc. Am. **98**, 2018–2047 (1995)]. In order to investigate this effect in more detail, simulations of the variation of the spectral period under conditions of self-suppression and two-tone suppression are obtained from nonlinear cochlear models based on this theory. The results show that during self-suppression the spectral period is increased, while during high-side two-tone suppression, the period is reduced, indicating that the detailed pattern of disruption of the cochlear amplifier must be examined if the nonlinear behavior of SFOAEs is to be understood. The model results suggest that the SFOAE spectral period may be sensitive to changes in the state of the cochlear amplifier. A companion paper [Lineton and Lutman, J. Acoust. Soc. Am. **114**, 871–882 (2003)] presents experimental data which are compared with the results of the above models with a view to testing the underlying theory of Zweig and Shera. [DOI: 10.1121/1.1582175]

PACS numbers: 43.64.Bt, 43.64.Jb, 43.64.Kc [BLM]

I. INTRODUCTION

A. Spectral periodicity in otoacoustic emissions (OAEs)

Stimulus frequency otoacoustic emissions (SFOAEs) exhibit regularity, termed *spectral periodicity*, in the spacing of the peaks in their frequency spectra[1] (Kemp and Chum, 1980; Wilson, 1980; Zwicker and Schloth, 1984; Dallmayr, 1987; Zwicker, 1990; Zwicker and Peisl, 1990; Lonsbury-Martin *et al.*, 1990; Zweig and Shera, 1995). The spacing between peaks appears approximately constant when the SFOAE spectrum is plotted against the logarithm of stimulus frequency (Zweig and Shera, 1995). This interpeak spacing, measured as an interval of the logarithm of frequency, is termed here the *spectral period*. It is also often characterized by the ratio $\Delta f/f$, where Δf is the frequency spacing between two adjacent peaks and f is the geometric mean frequency of the two peaks. To a first approximation, between 0.5 and 3 kHz this ratio is independent of frequency, and has a value of 1/15 in humans (Zweig and Shera, 1995; Talmadge *et al.*, 1998).

It has been shown that the value of the spectral period of SFOAEs can be related to the group delay, $\tau_G(f)$, of transient-evoked otoacoustic emissions (TEOAEs), evoked using clicks or tone bursts (Kemp, 1986; Zweig and Shera, 1995). Several authors have reported the variation of TEOAE group delay, or related quantities, with stimulus level and/or frequency (e.g., Norton and Neely, 1987; Neely *et al.*, 1988; Bray, 1989; Wit *et al.*, 1994; Tognola *et al.*, 1997). Recently, it has been shown that changes in the group delay of the cochlear response, as evidenced by TEOAEs or the auditory brainstem response, may also occur as a result of cochlear damage (Eggermont, 1979; Don *et al.*, 1998; Avan *et al.*, 2000; Lucertini *et al.*, 2002; Sisto and Moleti, 2002). It is therefore important from both theoretical and clinical perspectives to understand the mechanisms which give rise to spectral periodicity.

B. The origins of spectral periodicity

In order to explain the existence of TEOAEs, Kemp (1978) hypothesized that local impedance irregularities exist in the mechanics of the cochlea, which lead to the reflection of the traveling wave (TW). The need for some form of reflection mechanism is borne out by examination of early cochlear models, such as Neely and Kim's (1986), which included a representation of active cochlear processes. Such models can reasonably match *in vivo* measurements of basilar-membrane (BM) vibration, but do not predict the generation of OAEs (Furst and Lapid, 1988; de Boer, 1996). Hence, the mere presence of active processes is insufficient to generate TEOAEs.

Following the work of Manley (1983), Zwicker and Lumer (1985), Peisl (1988), and Zwicker (1986, 1988), Strube (1989) proposed a theory in which the mechanical properties of the basilar membrane vary periodically with place (over and above any variation associated with place-frequency mapping). These cochlear ''corrugations'' act as reflection sites which scatter and reflect the forward TW, thereby generating OAEs that propagate in the ear canal. Moreover, the SFOAEs predicted by this theory exhibit the approximately constant $\Delta f/f$ observed experimentally (e.g., Strube, 1989; Fukazawa, 1996; Wada *et al.*, 1999). However, no anatomical structures corresponding to these corrugations

[a]Currently at: MRC Institute of Hearing Research, Royal South Hants Hospital, Southampton SO14 OYG, United Kingdom. Electronic mail: b.lineton@soton.ac.uk

have been found in primate cochleae (Zweig and Shera, 1995).

Shera and Zweig (1993a) proposed an alternative theory in which the reflection sites are fine-grained inhomogeneities distributed randomly along the BM. This has been called the distributed-roughness theory (Talmadge *et al.*, 2000). Like the corrugated-cochlea theory, this theory also predicts SFOAEs with approximately constant $\Delta f/f$ (Shera and Zweig, 1993a; Zweig and Shera, 1995). However, the two theories differ in their explanations of how the spectral period is determined. In the corrugated-cochlea theory, the period is fixed by the wavelength of the corrugations, while in the distributed-roughness theory it is determined by the wavelength of the TW near the region of the peak (Zweig and Shera, 1995). As a consequence, the distributed-roughness theory predicts that changes in the shape of the TW will give rise to changes in SFOAE spectral period. The shape of the TW may, for example, alter with stimulus intensity due to nonlinear behavior of the cochlear amplifier. No such change in period is predicted by the corrugated-cochlea model.

Of the two, the distributed-roughness theory is currently favored, partly due to anatomical considerations (though the evidence remains inconclusive) and partly for scientific parsimony. One of the aims of the current study is to present more compelling evidence by testing the predictions of the distributed-roughness theory against measurements of changes in SFOAE period induced in human subjects by suppression. Two suppression phenomena are considered: self-suppression due to increasing the level of the stimulus tone and two-tone suppression. As well as providing a test of the theory, the measurements quantify the changes in spectral period that can arise due to suppression phenomena. This reveals that self-suppression and two-tone suppression can give qualitatively different results. This may have a bearing on measured changes in spectral period (or group delay) caused by pathology, such as those reported by Avan *et al.* (2000), Lucertini *et al.* (2002), and Sisto and Moleti (2002). The theoretical model is developed in the present paper, while measurements are reported in a companion paper (Lineton and Lutman, 2003a).

C. Models of SFOAE generation

The fundamental mechanism for generating quasiperiodic OAEs proposed by Zweig and Shera (1995) is essentially linear. It requires two features to be present in the cochlear mechanics: a tall-and-broad TW envelope and distributed roughness. In addition, two further features that are present in the mechanics of the cochlea must also be considered in the present study: nonlinear effects and multiple reflections of the TW that occur when the TW is partially reflected both apically by the distributed roughness and basally by the impedance mismatch at the stapes. Though these features are not essential for generating quasiperiodic OAEs, they influence the relationship between OAE spectral period and TW shape, which is the object of this study.

Zweig and Shera (1995) present a linear longwave model of the cochlea of the squirrel monkey which includes a cochlear amplifier leading to the tall-and-broad TW envelope, distributed roughness, and multiple reflections. The effect of altering the TW shape on spectral period was simulated in the model by altering one of the parameters controlling the cochlear amplifier gain resulting in changes in SFOAE spectra that were qualitatively similar to the changes in spectra seen in measured SFOAE level series (Zweig and Shera, 1995, p. 2034). However, a quantitative measure of changes in spectral period was not attempted.

Talmadge *et al.* (1998, 2000) present a similar model of the human cochlea, but with the addition of a saturating nonlinearity in the formulation of the cochlear amplifier. This model also includes a strongly nonuniform Q-factor in the basilar membrane (BM) impedance, leading to ''symmetry violation'' (i.e., a strong deviation of the cochlear macromechanics from scaling symmetry). These additions have two consequences for spectral period. First, the deviation from scaling symmetry causes the spectral period to become frequency dependent. Second, the nonlinearity interacts with the symmetry violation to give an additional mechanism for generating OAEs that operates even in the absence of any impedance irregularities and generates OAEs of very high spectral period (i.e., a very short group delay, leading to very slowly varying fine structure). Talmadge *et al.* (2000) present a quantitative study of the changes in spectral period that are predicted by their model due to changes in stimulus level. They show that the SFOAE is dominated by the distributed roughness mechanism at low levels, and thus exhibits a low period. As stimulus level increases, the nonlinear/symmetry-violation mechanism can become dominant, leading to a increase in period. However, Talmadge *et al.* (2000, p. 2923) report that level-induced changes in period due to the distributed roughness mechanism alone, which arise from the induced changes in the TW wavelength, are small.

In the present paper the work of Zweig and Shera (1995) and of Talmadge *et al.* (2000) on the effect of TW suppression on spectral period is extended in several ways. First, the changes in period with increasing stimulus level (causing self-suppression of the TW) that Talmadge *et al.* (2000, p. 2923) found to be small are re-examined with a sensitive method of measuring of spectral period. Second, the effect of two-tone suppression is also investigated, and found to be qualitatively different from that of self-suppression. Finally, the results from various different implementations of broadly similar cochlear models are investigated, in order to assess their sensitivity to the choices made in implementing the models.

A one-dimensional longwave model of the cochlea is used that includes a model of the cochlear amplifier and of distributed roughness. A model of the measurement probe, ear canal, and middle ear ensures that reflections of the TW at the stapes also feature in the results. Self-suppression and two-tone suppression are simulated using the quasilinear method reported by Kanis and de Boer (1993, 1994). Changes in period are quantified using a parametric fitting procedure described in a companion paper (Lineton and Lutman, 2003b). Two different cochlear amplifier models were employed, one presented by Kanis and de Boer (1994) and the other by Talmadge *et al.* (1998).

The definition of SFOAEs adopted in this paper specifi-

cally excludes the nonlinear/symmetry-violation mechanism discussed by Talmadge *et al.* (2000). Though this mechanism may be present in the actual cochlea, it was found in the present study that the magnitude of the effect was sensitive to some of the less well-established parameters in the cochlear model formulation. Therefore, for the sake of simplicity, it was decided to exclude it altogether.

II. THEORY AND METHODS

A. Macromechanical models of SFOAE generation

The main objective in developing the macromechanical models described in this section was to generate predictions of the changes in SFOAE period that accompany induced changes in TW shape. Three methods of inducing these shape changes were modeled: a global reduction of cochlear amplifier gain, self-suppression, and two-tone suppression.

There is currently a high degree of uncertainty about the detailed mechanics of the cochlea (Patuzzi, 1996; de Boer, 1996). Consequently, in developing any cochlear model, a number of choices must be made and a second objective is hence to assess whether the model predictions are insensitive to these choices. To this end, a number of variants of the cochlear models are investigated, which differ significantly in their formulation.

The cochlear model variants developed here incorporate all the key elements of Shera and Zweig's theory that have a bearing on the generation of realistic SFOAEs. These are the propagation of the TW along the BM; the amplification of the TW by the cochlear amplifier; the scattering of the forward TW by a random array of inhomogeneities distributed along the BM to generate a significant backward TW; the reflection of the backward TW on encountering the stapes, and the transmission of the backward TW through the middle ear and into the ear canal.

In addition, to model the changes in TW shape induced during self-suppression and two-tone suppression (but not global cochlear amplifier gain reduction), nonlinearities in the cochlear amplifier are included.

A final aspect of the cochlear models that must be considered is the degree to which they exhibit scaling symmetry (Shera and Zweig, 1991). This determines how closely the SFOAE spectrum can be predicted by the phenomenological model described by Zweig and Shera (1995) as well as the degree to which the $\Delta f/f$ is dependent on stimulus frequency.

The cochlear models developed here follow earlier one-dimensional longwave models (e.g., Neely and Kim, 1986) and include models of the cochlear amplifier based on those described by Neely and Kim (1986), Kanis and de Boer (1993, 1994, 1996), Zweig (1991), and Talmadge *et al.* (1998). Although the longwave assumptions become invalid near the peak of the TW (de Boer, 1996), it has been argued that the longwave model still captures the essential feature of TW propagation and amplification that are required in Shera and Zweig's theory of SFOAE generation: the ''tall and broad'' TW envelope (Zweig and Shera, 1995; Zweig, 1991).

Following Zweig and Shera (1995) and Talmadge *et al.* (1998), a random array of structural inhomogeneities distributed along the BM is also included. Since the SFOAE spectrum that results from these models depends on the particular realization of the random inhomogeneities included in the model, the models are executed repeatedly with different realizations and an ensemble average calculated.

A simple model of the middle ear, ear canal, and measurement probe is also included. This ensures that reflections of the backward TW from the stapes are modeled, as well as the transmission of the OAE through the ear canal to the microphone in the probe.

All the models are formulated in the frequency domain. In the self-suppression and two-tone suppression simulations, where nonlinearities must be modeled, the quasilinear method described by Kanis and de Boer (1993) is used.

To find out whether the model predictions are sensitive to the formulation of the key elements in the cochlear models, several variants are investigated. The first model, termed here the *baseline* model, is based on the formulation of the passive cochlear structures and of the cochlear amplifier reported by Neely and Kim (1986) and adapted by Kanis and de Boer (1993). It also includes inhomogeneities modeled as a scattering impedance, which is a function of place alone.

The second model, termed here the *Zweig–Talmadge cochlear amplifier variant*, replaces the baseline formulation of the passive and active cochlear structures with that reported by Zweig (1991) and Talmadge *et al.* (1998). The main difference from the baseline is in the formulation of the cochlear amplifier, which is based on that described by Zweig (1991) and Talmadge *et al.* (1998). A second difference is that the model deviates much more strongly from perfect scaling symmetry than the baseline model.

The third model, termed the *frequency-dependent scattering variant*, replaces the baseline scattering impedance with one which varies with both place and stimulus frequency. This has implications for the applicability of the phenomenological model of Zweig and Shera (1995) and for scaling symmetry, which manifest themselves in the predictions of the SFOAE spectrum.

1. The longwave model

As in the many longwave models, the cochlea is modeled as two straight fluid-filled rectangular channels, termed here the upper and lower channels, separated by a flexible partition, termed here the cochlear partition (CP) (e.g., Neely and Kim, 1986; Kanis and de Boer, 1993). The longwave model is represented by the wave equation

$$\frac{d^2\mathbf{p_d}(x,\omega)}{dx^2}+\mathbf{k}^2(x,\omega)\mathbf{p_d}(x,\omega)=0, \tag{1}$$

where the $\mathbf{k}$ is the complex wave number, given by

$$\mathbf{k}^2(x,\omega)=\frac{2\rho\omega^2}{i\omega H\mathbf{Z_{CP}}(x,\omega)}, \tag{2}$$

$\mathbf{p_d}(x,\omega)$ is the semidifference pressure, defined as half the pressure difference across the cochlear partition; x is the distance along the CP; ω is the radian stimulus frequency; ρ is the cochlear fluid density; H is the height of the upper channel (here assumed constant); and $\mathbf{Z_{CP}}$ is the CP impedance. $\mathbf{Z_{CP}}$, $\mathbf{k}$, and $\mathbf{p_d}$ are complex numbers (which will be denoted

here by bold typeface) and can be functions of both x and ω.

The semidifference pressure must also satisfy two boundary conditions. At the base of the cochlea, it is related to the stapes velocity, $\mathbf{u_{St}}$, by

$$\left.\frac{d\mathbf{p_d}(x,\omega)}{dx}\right|_{x=0} = -i2\omega\rho\mathbf{u_{St}}(\omega), \tag{3}$$

while at the helicotrema, the semidifference pressure is assumed to be zero. The CP velocity, $\mathbf{v_{CP}}$, is obtained from the CP impedance once the semidifference pressure is known

$$\mathbf{v_{CP}}(x,\omega) = -2\frac{\mathbf{p_d}(x,\omega)}{\mathbf{Z_{CP}}(x,\omega)}. \tag{4}$$

The CP impedance, $\mathbf{Z_{CP}}$, comprises three components

$$\mathbf{Z_{CP}}(x,\omega) = \mathbf{Z_{Pass}}(x,\omega) + \mathbf{Z_{CA}}(x,\omega) + \mathbf{Z_{Sc}}(x,\omega), \tag{5}$$

where $\mathbf{Z_{Pass}}$, $\mathbf{Z_{CA}}$, and $\mathbf{Z_{Sc}}$ are termed the passive impedance, the cochlear-amplifier impedance, and the scattering impedance, respectively. The cochlear-amplifier impedance can be interpreted as an active device that, at any given point on the CP, applies a force to the CP determined by the CP velocity. The device injects energy into the TW at all points and frequencies for which the real part of $\mathbf{Z_{CA}}(x,\omega)$ is negative.

2. The linear baseline model

The passive impedance is modeled by a single spring–mass–damper system at each cochlear location, such that

$$\mathbf{Z_{Pass}}(x,\omega) = \frac{m_0}{i\omega}\omega_C^2(x) + m_0\delta_0\omega_C(x) + i\omega m_0, \tag{6}$$

where m_0 is the mass per unit area of the CP (assumed independent of x), δ_0 is twice the critical damping ratio (assumed independent of x), and $\omega_C(x)$ is the local CP natural frequency given by

$$\omega_C(x) = \omega_{C_0}\exp(-x/l), \tag{7}$$

where ω_{C_0} is the basal natural frequency and l is the place-frequency length parameter.

The cochlea amplifier impedance for the baseline model is given by

$$\mathbf{Z_{CA}}(x,\omega) = e_0 d_0 \omega_C(x)\frac{1+i\gamma(x,\omega)}{\delta_{SC}+i[\gamma(x,\omega)-\sigma^2/\gamma(x,\omega)]}, \tag{8}$$

where $\gamma(x,\omega)$ is the nondimensional stimulus frequency given by

$$\gamma(x,\omega) \equiv \omega/\omega_C(x); \tag{9}$$

and where e_0, d_0, δ_{SC}, and σ are parameters, all independent of x, whose values are chosen to yield a region of negative damping at and basal to the peak of the TW.

The scattering impedance has been defined as a perturbation in the passive damping of the CP. This ensures that it is independent of the stimulus frequency, ω, thereby bringing the cochlear model closer to the simplified phenomenological model reported by Zweig and Shera (1995). To model a random distribution of reflection sites, the scattering impedance is defined as

$$\mathbf{Z_{Sc}}(x) = r(x)a_{Sc}b(x), \tag{10}$$

where $r(x)$ is the passive damping, given by

$$r(x) = m_0\delta_0\omega_C(x); \tag{11}$$

a_{Sc} is the scattering amplitude parameter; and $b(x)$ is a spatially random Gaussian signal with an rms amplitude of 1, and with spatial wavelengths down to the discretization length of the model, which was set below 10 μm (i.e., the typical width of a hair cell). To obtain a particular realization of $b(x)$, a random number generator was used to obtain a digital white-noise sequence.

To allow SFOAEs in the ear canal to be predicted, the cochlear model is connected to a model of the ear canal and middle ear given by Kringlebotn (1988), and driven by a perfect volume velocity source representing the measurement probe earphone. At the base of the cochlea, the pressure gradient is related to the stapes velocity, $\mathbf{u_{St}}$, by Eq. (3).

Assuming negligible round window impedance, the pressure at the stapes, $\mathbf{p_{St}}$, is simply related to the basal semidifference pressure by

$$\mathbf{p_{St}} = 2\mathbf{p_d}|_{x=0}. \tag{12}$$

The stapes pressure and velocity are related to the ear canal pressure, $\mathbf{p_{EC}}$, and ear canal volume velocity, $\mathbf{U_{EC}}$, by a two-port network

$$\begin{bmatrix}\mathbf{p_{St}}\\ \mathbf{u_{St}}\end{bmatrix} = \begin{bmatrix}\mathbf{T_{11}} & \mathbf{T_{12}}\\ \mathbf{T_{21}} & \mathbf{T_{22}}\end{bmatrix}\begin{bmatrix}\mathbf{p_{EC}}\\ \mathbf{U_{EC}}\end{bmatrix}, \tag{13}$$

where the matrix elements $\mathbf{T_{ij}}$ are determined by the properties of the middle ear and ear canal. Finally, the ear-canal pressure and volume velocity are related to the properties of the measurement probe by

$$\mathbf{U_{EC}} = \mathbf{U}_0 - \mathbf{p_{EC}}\mathbf{Y}_0, \tag{14}$$

where $\mathbf{U}_0$ is the OAE probe volume velocity with no acoustic load, and $\mathbf{Y}_0$ is the measurement probe acoustic admittance. Equations (12) to (14) can be manipulated to derive the boundary condition at the base of the cochlea, required to solve Eq. (1).

In the baseline model, the matrix elements $\mathbf{T_{ij}}$ are obtained by adopting the middle-ear model reported by Kringlebotn (1988), setting the probe acoustic admittance, $\mathbf{Y}_0$, to zero and holding the probe volume velocity at no acoustic load, $\mathbf{U}_0$, constant for a given nominal stimulus level in the ear canal.

3. The nonlinear baseline model

The treatment of nonlinearity follows the quasilinear method described by Kanis and de Boer (1993). This method works in the frequency domain and is highly efficient when, as in our case, only one or two primary tones are presented simultaneously. The method works iteratively by successively modifying the ''quasilinear'' cochlear amplifier impedance depending on the calculated CP velocity response. A hyperbolic tangent function is used to represent the saturation of the active force applied by the cochlear amplifier at high CP velocities. The reader is referred to Kanis and de Boer (1993) for details of this method.

B. Calculating the normalized SFOAE frequency spectrum

The SFOAE pressure has been defined here as the contribution to the ear-canal pressure that arises from the presence of the scattering impedance. Thus, the SFOAE pressure is given by

$$\mathbf{p_{SF}}(f) \equiv \mathbf{p_{EC}}(f) - \mathbf{p_{EC:R=0}}(f), \tag{15}$$

where $\mathbf{p_{EC}}(f)$ is the ear-canal pressure obtained with the complete cochlear model, and $\mathbf{p_{EC:R=0}}(f)$ is the zero-reflectance ear-canal pressure, obtained from the cochlear model with the scattering impedance set to zero (the ''smooth'' cochlear model) while holding the measurement probe volume velocity, $\mathbf{U}_0$, constant.

Following Shera and Zweig (1993b), it is useful to define the *normalized SFOAE pressure*, $\Lambda(f)$, as

$$\Lambda(f) \equiv \frac{\mathbf{p_{SF}}(f)}{\mathbf{p_{EC:R=0}}(f)}. \tag{16}$$

In all cases the normalized SFOAE pressure is obtained from a frequency sweep from 1.5 to 3 kHz at 4-Hz intervals holding a constant probe volume velocity. This frequency range falls within the range where SFOAEs are readily measurable in human ears and where psychoacoustic measures suggest that the auditory filter obeys approximate scaling symmetry (Moore and Glasberg, 1987).

The normalized SFOAE spectrum, $\Lambda(f)$, depends on the particular realization of the scattering impedance. Thus, for any given scattering impedance, the spectrum, $\Lambda(f)$, must be viewed as a single realization arising from the random process that defines the infinite set of all possible realizations. The random process itself is determined by the entire cochlear model, including the statistical properties of the random number generator, which in this case is simply a Gaussian white-noise generator. In order to estimate the properties of the random process better, the frequency sweep was repeated 32 times with different realizations of the scattering impedance to obtain an ensemble of realizations of $\Lambda(f)$.

It should be noted that the definition given in Eq. (15) ensures that the quantity, $\mathbf{p_{SF}}(f)$, does not contain the purely nonlinear contribution to SFOAEs which was investigated by Talmadge *et al.* (2000). To obtain this additional contribution, an alternative definition of $\mathbf{p_{SF}}(f)$ in which $\mathbf{p_{EC:R=0}}(f)$ is defined in terms of the passive, rather than the smooth cochlea, should be used. This would have the advantage that it is closer to the quantity commonly used in SFOAE measurements (as in the companion paper, Lineton and Lutman, 2003a). However, a preliminary investigation found that the magnitude of the nonlinear effect depended greatly on the particular cochlear model used, with the baseline model showing virtually no effect. This indicates that the nonlinear mechanism is sensitive to the choice of model. Therefore, in order to simplify the interpretation of the results, the definition in Eq. (15) is adopted, thereby limiting the investigation to effects arising from the scattering mechanism proposed by Zweig and Shera (1995). The validity of this simplification is confirmed by the experimental data presented in the companion paper (Lineton and Lutman, 2003a).

C. Estimating the spectral period of the SFOAE frequency spectrum

Because of the random nature of the scattering impedance, calculating the spectral period of Λ over a given frequency interval is not straightforward. This problem is addressed in a companion paper (Lineton and Lutman, 2003b) in which a parametric model of the SFOAE spectrum, termed the four-parameter model, is developed. This is based on the scaling-symmetric phenomenological model presented by Zweig and Shera (1995) and yields parameters which are related to the underlying TW properties.

To apply the four-parameter model to data from the cochlear models, Λ is first expressed as a function of the logarithmic frequency variable, η, defined by

$$\eta \equiv -\ln(f/1\ \text{kHz}), \tag{17}$$

where the minus sign in Eq. (17) is included for compatibility with Zweig and Shera (1995). The quantity termed the raw autocovariance function of Λ is then calculated from

$$\mathbf{C}_{\Lambda\Lambda}^{(\text{raw})}(\eta') \equiv \frac{1}{\eta_2 - \eta_1} \int_{\eta_1}^{\eta_2 - \eta'} \Lambda^*(\eta)\Lambda(\eta + \eta')\, d\eta \tag{18}$$

for $0 < \eta' < \eta_2 - \eta_1$, where the η_1 and η_2 indicate the range of logarithmic frequencies over which Λ has been obtained, and $*$ denotes complex conjugation. The four-parameter model then uses a fitting procedure to obtain the fitted autocovariance function, denoted $\mathbf{C}_{\Lambda\Lambda}^{(fit)}(\eta')$. For reasons explained in the companion paper on parametric fitting (Lineton and Lutman, 2003b), the fit was performed over the range $0 < \eta' < \eta'_{\text{Lim}}$, where $\eta'_{\text{Lim}} = 0.125(\eta_2 - \eta_1)$.

It is useful to calculate the inverse Fourier transform (denoted $\mathbb{F}^{-1}$) of the raw and fitted autocovariance functions to yield the functions termed here the raw and fitted ϕ spectra, respectively

$$S_{\Lambda\Lambda}^{(\text{raw})\text{or}(\text{fit})}(\phi) \equiv \mathbb{F}^{-1}\{\mathbf{C}_{\Lambda\Lambda}^{(\text{raw})\text{or}(\text{fit})}(\eta')\}. \tag{19}$$

The independent variable, ϕ, can be interpreted as the envelope delay of a single frequency component, expressed as the number of periods of the carrier frequency (Zweig and Shera, 1995). $S_{\Lambda\Lambda}^{(\text{fit})}(\phi)$ is a pulse-shaped function which is fully characterized by four parameters denoted here ϕ_C, ϕ_{BW}, α, and $\sigma_{\Lambda\Lambda}$. The most important of these is ϕ_C (termed the ϕ-center value), which gives the location of the peak of the pulse and which Zweig and Shera (1995) show is approximately determined by the wavelength near the peak of the TW. They also show that ϕ_C (equal to the quantity they denoted $\hat{\varphi}/2\pi$) is roughly related to the spectral period by

$$\phi_C \approx \frac{f}{\Delta f}, \tag{20}$$

where f is defined as the geometric mean frequency and Δf the frequency interval of two adjacent spectral peaks in the amplitude of the ear-canal pressure. The value of ϕ_C can also be interpreted as the average OAE delay, expressed in terms of cycles of the stimulus frequency (e.g., Kemp, 1986; Wilson, 1980; Zweig and Shera, 1995). Typically $\phi_C \approx 15$ in humans over the 0.5- to 3-kHz region (Zweig and Shera, 1995; Talmadge *et al.*, 1998).

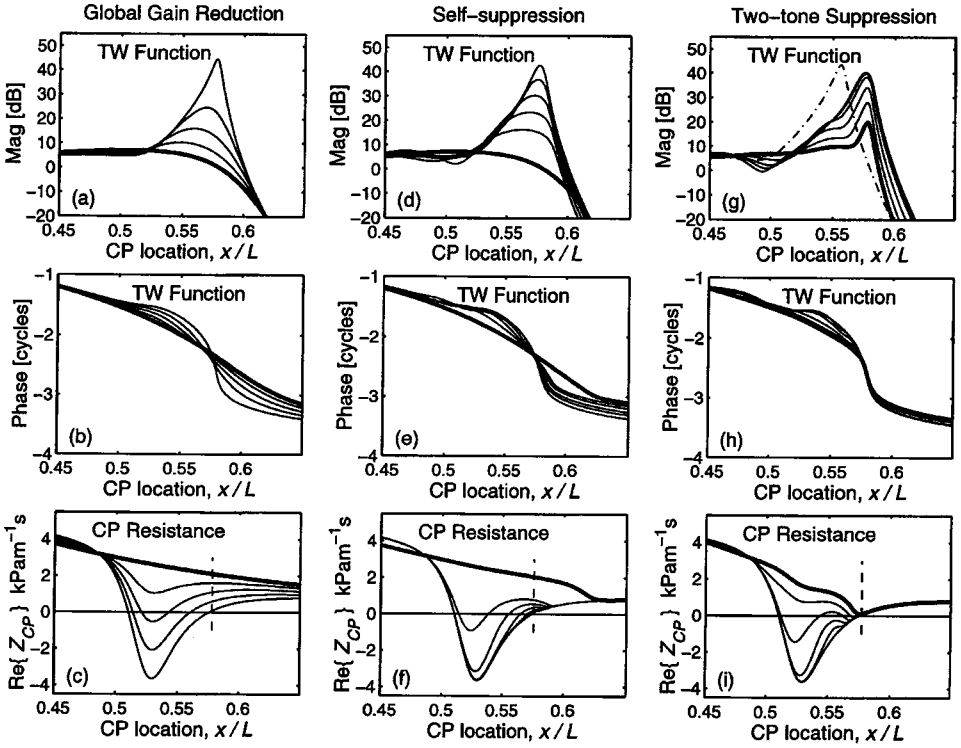


FIG. 1. Variation of the TW function, defined as the normalized CP velocity, and the CP resistance during three types of suppression: global gain reduction, self-suppression, and two-tone suppression. All functions are obtained from the baseline cochlear models (with no scattering impedance included) and are shown against CP location for a constant stimulus frequency of 1.5 kHz. The top two rows show the magnitude and phase of the TW function, respectively. The bottom row shows CP resistance. In (a), (b), and (c) the results are shown for global attenuation of the cochlear amplifier of 0, 2.5, 6, 12, and ∞ dB. In (e), (f), and (g) the results are shown for self-suppression with stimulus levels of 40, 50, 60, 70, 80, and 120 dB SPL. In (h), (i), and (j) the results are shown for two-tone suppression with a probe level of 45 dB SPL, and suppressor levels of 30, 50, 60, 70, 80, and 90 dB SPL. Solid lines show response at the stimulus frequency. Chain line [in (g) only] shows the response at the suppressor frequency for the 30-dB SPL suppressor case. Thick solid lines show the most highly suppressed response in each of the three types of suppression. The vertical dashed lines in (c), (f), and (i) indicate the location of the TW peak at the stimulus frequency in the unsuppressed condition. At points of negative CP resistance, TW amplification occurs.

As well as ϕ_C, the fitting procedure yields three other parameters: ϕ_{BW}, α, and $\sigma_{\Lambda\Lambda}$. The parameter ϕ_{BW} characterizes the width of the pulse, while $\sigma_{\Lambda\Lambda}$ is simply the variance of Λ between η_1 and η_2, given by

$$\sigma^2_{\Lambda\Lambda} \equiv \mathbf{C}_{\Lambda\Lambda}(\eta')|_{\eta'=0}, \tag{21}$$

and therefore characterizes the height of peak. The parameter α characterizes the extent to which Λ is influenced by reflections of the backward TW at the stapes. Such basal reflections lead to the appearance in $S^{(\mathrm{fit})}_{\Lambda\Lambda}(\phi)$ of additional pulses centered at $2\phi_C$, $3\phi_C$, etc. However, in general it was found that basal reflections had little influence on the estimate of spectral period (via the parameter ϕ_C). A fuller discussion of the role of α in the characterization of Λ is given in the companion paper on parametric fitting (Lineton and Lutman, 2003b).

III. RESULTS

A. The effect of global reduction of the cochlear amplifier gain on SFOAE spectral period

A simple way of altering the TW shape is to deactivate the cochlear amplifier by reducing the gain globally along the entire CP. For the baseline linear model, this is achieved by introducing a single scale factor into the cochlear amplifier impedance equation, Eq. (8). The effect of this on the TW function is shown in Fig. 1, panels (a) and (b), for a single stimulus frequency. As might be expected, this shows that decreasing the gain causes both a broadening of the TW peak and an increase in the TW wavelength in the peak region. (The wavelength is inversely proportional to the gradient of the phase curve.) Panel (c) shows how the CP resistance in the TW peak region increases, eventually becoming positive everywhere as the gain is reduced. Note that to emphasize the changes in TW shape, the results in Fig. 1 are obtained with higher gain attenuations than required to obtain realistic degrees of SFOAE suppression.

SFOAE spectra from 1.5 to 3 kHz were generated for a number of different settings of the cochlear amplifier gain. Results for a single realization of the scattering impedance for the baseline model are shown in Figs. 2 and 3(a). In order to characterize the random process more accurately, SFOAE spectra were calculated for 32 different realizations of the scattering impedance and the raw autocovariance functions then averaged across the ensemble to give the ensemble-averaged autocovariance function, denoted $\langle \mathbf{C}^{(\mathrm{raw})}_{\Lambda\Lambda}(\eta') \rangle_{\mathrm{ensemble}}$ shown normalized in Fig. 3(b). Applying the four-parameter signal characterization model to the ensemble-averaged autocovariance functions gives the fitted autocovariance functions shown normalized in Fig. 3(c). Inverse Fourier transforming the ensemble-averaged and fitted autocovariance functions gives the ϕ spectra shown in Figs. 4(a) and (b), respectively. These clearly show that a reduc-

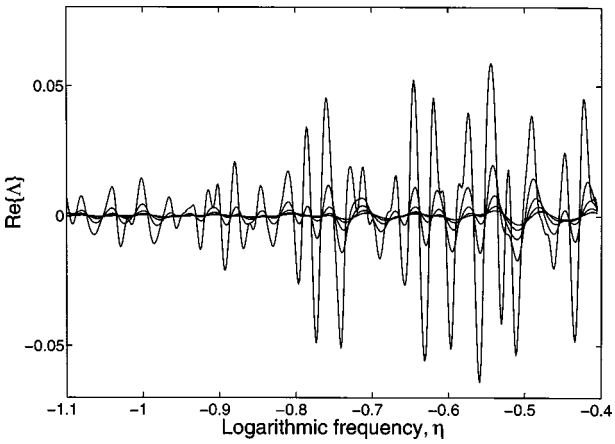


FIG. 2. Variation of the normalized SFOAE spectrum during a global reduction of cochlear amplifier gain. The function $\mathrm{Re}\{\Lambda(\eta)\}$ obtained from the baseline cochlear model with a single realization of the scattering impedance and with cochlear amplifier attenuations of 0, 0.4, 0.8, 1.2, and 1.6 dB.

tion in cochlear amplifier gain leads to a reduction in ϕ-center value, ϕ_C. Figure 5(a) shows the variation of ϕ_C with the level of the normalized SFOAE, $L_{\Lambda\Lambda}$, defined as the rms amplitude, $\sigma_{\Lambda\Lambda}$, expressed in dB

$$L_{\Lambda\Lambda} \equiv 20 \log_{10}(\sigma_{\Lambda\Lambda}). \tag{22}$$

B. The effect of self-suppression on SFOAE period

To model the effects of increasing the stimulus level, the quasilinear cochlear model was used. In this model, the cochlear amplifier saturation depends on the local CP response leading to a local reduction in gain which is greatest near the peak of the TW [Fig. 1(f)], rather than the uniform reduction seen in Fig. 1(a). Despite these differences, the changes in the TW function, shown for a single frequency in Figs. 1(d) and (e), are qualitatively similar to those seen for the global gain reduction in panels (a) and (b).

To simulate self-suppression, SFOAE spectra from 1.5 to 3 kHz were generated at a number of different nominal stimulus levels, the source volume velocity being held constant at each level. Results for the nonlinear models require much more computation time than those for the linear model. Consequently, at each stimulus level, the SFOAE spectrum was obtained for only two realizations of the scattering impedance, from which two "raw" (i.e., unaveraged) autocovariance functions were calculated. The four-parameter model was then applied to both raw autocovariance functions, giving two fitted autocovariance functions, and hence two estimates of ϕ_C for each frequency sweep. Results for the raw and fitted ϕ spectra for one of these realizations are shown in Figs. 4(c) and (d). As with the global gain reduction, Fig. 4(d) indicates a reduction in the ϕ-center value, ϕ_C, with increasing SFOAE suppression. This can be seen more clearly in Fig. 5(a), which shows the variation of ϕ_C for both realizations. The difference between the results for the two realizations gives an indication of the random error in the estimates obtained by the four-parameter model.

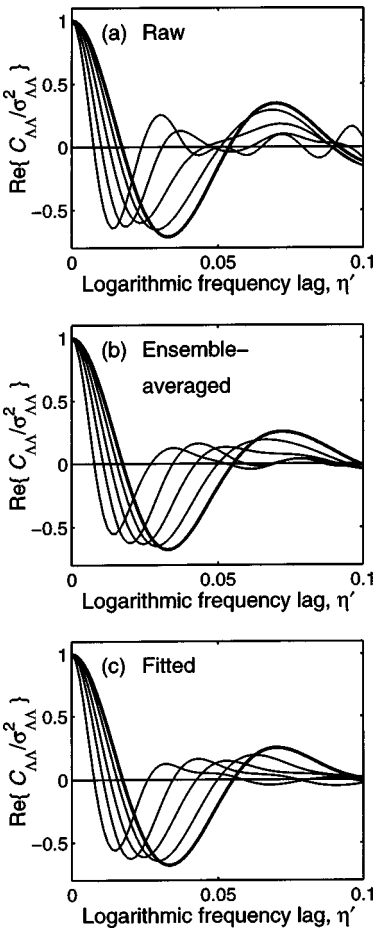


FIG. 3. Variation of the normalized autocovariance function of $\Lambda(\eta)$ during a global reduction of the cochlear amplifier gain. The cochlear amplifier attenuation is 0, 0.4, 0.8, 1.2 (thin lines), and 1.6 dB (thick lines). Three different estimates of the normalized autocovariance function are shown. Panel (a) shows the raw autocovariance function, estimated from a single realization of the normalized SFOAE spectrum, is shown. Panel (b) shows the ensemble-averaged autocovariance function obtained by averaging the raw autocovariance function across 32 realizations. Panel (c) shows the fitted autocovariance function from the four-parameter fit to the ensemble-averaged autocovariance function. In all cases the real part of the autocovariance function, normalized by $\sigma^2_{\Lambda\Lambda}$, is shown.

C. The effect of two-tone suppression on SFOAE spectral period

The quasilinear cochlear model was used to simulate the interaction of two pure tones, termed here the probe tone, with a frequency, f_1, and the suppressor tone, with a frequency, f_2. The frequency ratio was held constant at $f_2/f_1 = 1.1$ throughout the simulations. Because $f_2 > f_1$, this is termed high-side suppression.

The effect of the high-side suppressor tone on the TW is quite different from that of either the global gain reduction or self-suppression, because the cochlear amplifier does not saturate uniformly along the CP, but instead tends to saturate first at those points where the CP velocity amplitude is highest (Fig. 1). Thus, in contrast to the self-suppression case where saturation begins near the peak of the probe tone and then spreads basally as the probe level increases [Fig. 1(f)], in the two-tone suppression case saturation begins near the peak of the suppressor tone, which lies basal to the peak of the probe tone [Fig. 1(i)].

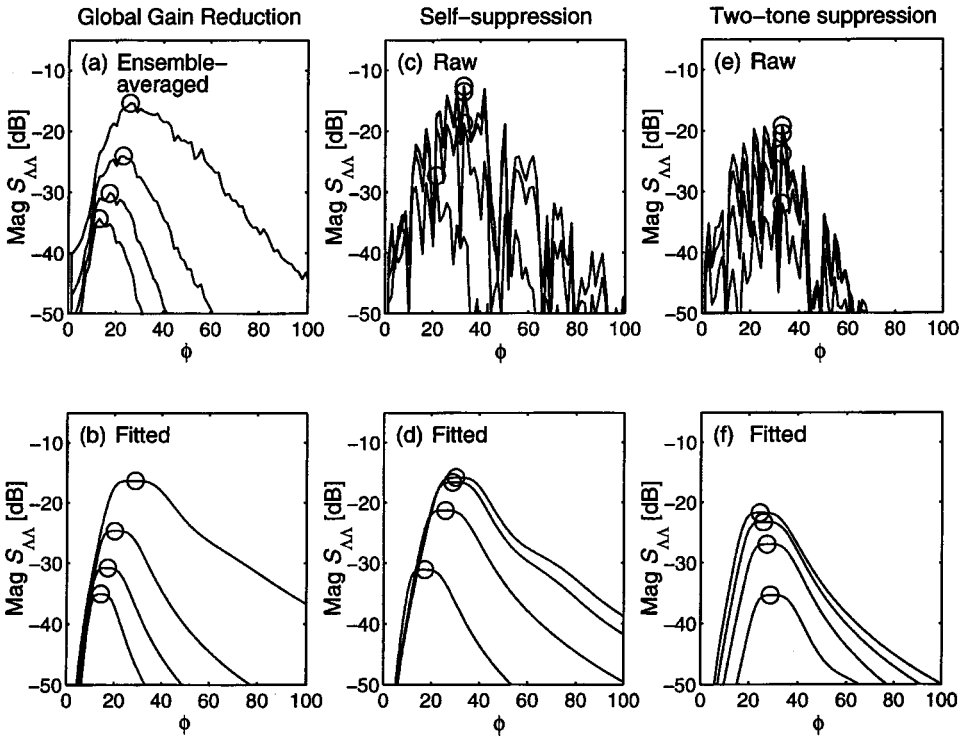


FIG. 4. Variation of the ϕ-spectrum of $\Lambda(\eta)$ during three types of suppression: global gain reduction, self-suppression, and two-tone suppression. In all cases the ϕ spectrum is obtained from the Fourier transform of an estimate of the autocovariance function. In the left-hand column, the ϕ spectra at global cochlear amplifier attenuations of 0, 0.4, 0.8, and 1.2 dB are shown. The estimated autocovariance functions are obtained in (a) from an ensemble average over 32 realizations, and in (b) from the four-parameter fit to the ensemble-averaged data. In the middle column, ϕ spectra under self-suppression at stimulus levels of 20, 40, 50, and 60 dB SPL are shown. The estimated autocovariance functions are obtained in (c) from a single realization, and in (d) from the four-parameter fit to the single realization. In the right-hand column, ϕ spectra under two-tone suppression at suppressor levels of 30, 55, 65, and 75 dB SPL are shown. The estimated autocovariance functions are obtained in (e) from a single realization, and in (f) from the four-parameter fit to the single realization. The peak of the ϕ spectrum is identified with a circle. The value of ϕ at the peak of the fitted ϕ spectrum is termed the ϕ-center value, is denoted ϕ_C, and is reciprocally related to the spectral period, $\Delta f/f$.

The effect on the TW function itself is seen in Figs. 1(g) and (h). Panel (g) shows that the width of the peak remains relatively sharp, despite the reduction in level. Similarly, panel (h) shows that the TW wavelength in the peak region changes little with increasing suppressor level. This has important consequences for the SFOAE spectral period, since in Zweig and Shera's theory period is related to the wavelength of the TW in the peak region.

To simulate high-side suppression, the probe frequency was swept from 1.5 to 3 kHz, while the suppressor frequency was altered to maintain the frequency ratio of $f_2/f_1=1.1$. For each frequency sweep, the nominal probe and suppressor levels (as determined by the source volume velocity) were both held constant. Frequency sweeps were obtained for a series of different nominal suppressor levels, while the nominal probe level was held constant throughout. As for the self-suppression simulation, for each stimulus setting the SFOAE spectrum was obtained for two realizations of the scattering impedance, thus giving two raw and two fitted autocovariance functions for each frequency sweep.

The results for one of the realizations are shown in Figs. 4(e) and (f). In contrast to self-suppression, the ϕ-center value, ϕ_C, shows a slight increase during high-side, two-tone suppression [Fig. 4(f)]. This can be seen more clearly in Fig. 5(a), which shows the variation of the ϕ_C for both the self- and two-tone suppression cases. Taking the two modes of suppression together, the pattern of variation of ϕ_C follows a distinctive *y* shape, with the longer limb of the *y* arising from self-suppression, and the shorter limb from two-tone suppression. The intersection point of the *y* shape occurs at the common probe-stimulus level of 45 dB SPL, where the suppressor level is too weak to cause significant suppression additional to the self-suppression.

D. Variations of ϕ bandwidth

As well as yielding a measure of the SFOAE spectral period via the estimate of the ϕ_C, the four-parameter model also gives a measure of the spread of periods, via the estimate of the ϕ bandwidth, ϕ_{BW}. Figure 5(b) shows that the variation of ϕ_{BW} during suppression resembles the *y*-shaped pattern seen for ϕ_C. However, the variability of ϕ_{BW} across the two realizations is much greater than that seen for ϕ_C. Further investigation reported in the companion paper on parametric fitting (Lineton and Lutman, 2003b) revealed that the reliability of the estimate of ϕ_{BW} was poorer than that of ϕ_C, making ϕ_{BW} less useful as an indicator of changes in TW shape. To a first approximation, the relative bandwidth of $S_{\Lambda\Lambda}(\phi)$, as indicated by ϕ_{BW}/ϕ_C appears independent of $L_{\Lambda\Lambda}$.

E. Sensitivity of results to alternative formulations of the cochlear models

To assess the sensitivity of the main results presented above to the formulation of the cochlear model, the predic-

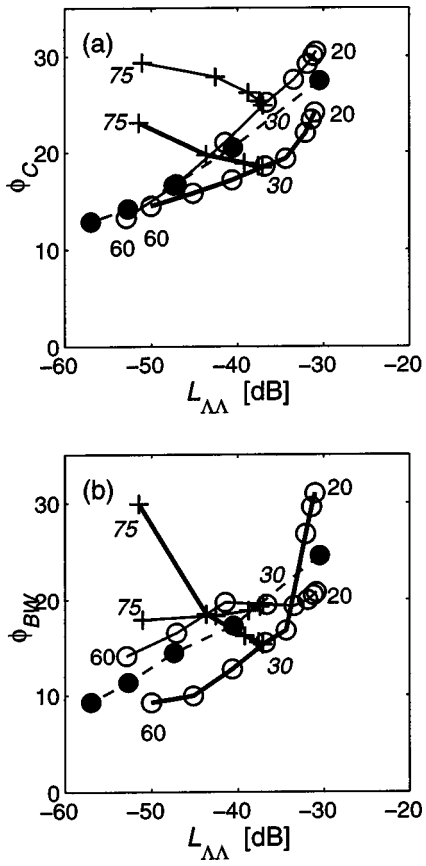


FIG. 5. Variation of fitted parameters during three types of suppression: global gain reduction, self-suppression, and two-tone suppression. For the global gain reduction (filled circles, dashed lines), each point is obtained from the four-parameter fit to the ensemble-averaged autocovariance function over 32 realizations. Cochlear amplifier attenuations of 0, 0.4, 0.8, 1.2 and 1.6 dB are used. For the self-suppression (open circles) and two-tone suppression conditions (+symbols) each point is obtained from a four-parameter fit to a single realization of the raw autocovariance function. The results for two realizations of the CP scattering impedance are overlaid (results from the first indicated by a thick solid line, those from the second by a thin solid line) to illustrate the random variability of the estimates. Self-suppression is shown for stimulus levels of 20, 30, 35, 40, 45, 50, 55, and 60 dB SPL (indicated beside selected points). Two-tone suppression is shown for a probe level of 45 dB SPL and suppressor levels of 30, 45, 55, 65, and 75 dB SPL (indicated beside selected points in italic typeface). In (a), ϕ_C is plotted against $L_{\Lambda\Lambda}$ revealing a *y*-shaped pattern for the self and two-tone suppression conditions taken together. In (b), ϕ_{BW} is plotted against $L_{\Lambda\Lambda}$.

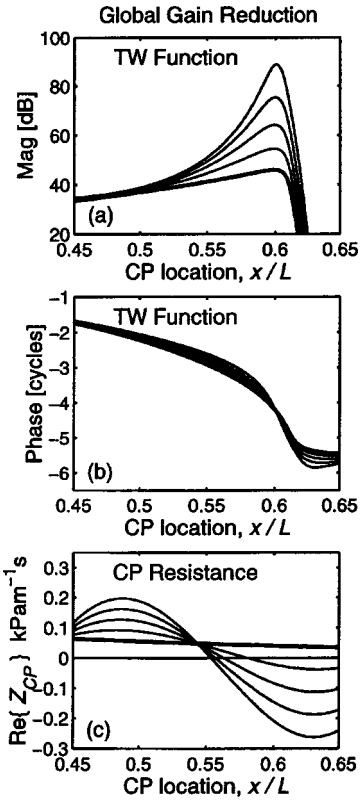


FIG. 6. Variation of the TW function and the CP resistance for the Zweig–Talmadge cochlear amplifier variant during global gain reduction. The functions are obtained from the baseline cochlear models (with no scattering impedance included) and are shown against CP location for a constant stimulus frequency of 1.5 kHz, and for global attenuation of the cochlear amplifier of 0, 2.5, 6, 12, and ∞ dB. Panels (a) and (b) show the magnitude and phase of the TW function, respectively. Panel (c) shows the CP resistance. At points of negative CP resistance, TW amplification occurs. Thick solid lines show the response at full attenuation.

tions have been repeated for several model variants. Results for two of the major model variants are described in this section.

1. The Zweig–Talmadge cochlear amplifier variant

This model variant is based on the passive and active elements reported by Zweig (1991) and Talmadge *et al.* (1998). The passive impedance is given by

$$\mathbf{Z_{Pass}}(x,\omega)=\frac{m_0}{i\omega}\omega_C^2(x)+m_0\delta(x)\omega_C(x)+i\omega m_0, \qquad (23)$$

with the natural frequency given by

$$\omega_C(x)=\omega_{C_1}\exp(-x/l)+\omega_{C_2}, \qquad (24)$$

the damping variable given by

$$\delta(x)=\frac{\delta_0+\delta_1\exp(x/l)}{1+\mu\exp(x/l)}, \qquad (25)$$

and with the parameter values given by $l=7.24$ mm; $\omega_{C_1}=2\pi\times 20.8$ kHz; $\omega_{C_2}=-2\pi\times 0.1455$ kHz; $m_0=0.05$ kgm^{-2}; $\delta_0=0.0385$; $\delta_1=0.000\,765$; $\mu=-0.007$. The variation in the damping ratio along the cochlea seen in Eq. (25) leads to a much greater deviation from scaling symmetry than seen in the baseline model.

The variant cochlear amplifier impedance is given by

$$\mathbf{Z_{CA}}(x,\omega)=\frac{m_0\rho_f\omega_C^2(x)}{i\omega}\exp[-i\psi_f\gamma(x,\omega)] + \frac{m_0\rho_s\omega_C^2(x)}{i\omega}\exp[-i\psi_s\gamma(x,\omega)], \qquad (26)$$

where m_0 and ω_C are defined as for the passive model in Eqs. (23) and (24) and γ is the nondimensional frequency defined in Eq. (9). The parameter values are taken from Talmadge *et al.* (1998): $\rho_f=0.16$; $\rho_s=0.1416$; $\psi_f=2\pi\times 0.24$; $\psi_s=2\pi\times 1.742$ (the subscripts *s* and *f* referring to slow and fast feedback terms).

For comparison with the results from the baseline model, the effect of a global gain reduction on both the TW response, shown in Fig. 6, and the SFOAE spectrum (Fig. 7)

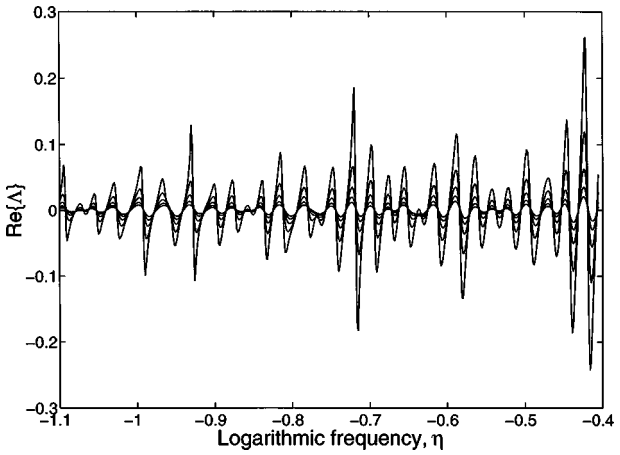


FIG. 7. Variation of the normalized SFOAE spectrum for the Zweig–Talmadge cochlear amplifier variant during global reduction of the cochlear amplifier gain. The function Re{$\Lambda(\eta)$} obtained from the Zweig–Talmadge cochlear amplifier variant with a single realization of the scattering impedance and with cochlear amplifier attenuations of 0, 0.4, 0.8, 1.2, and 1.6 dB.

have been obtained for this model variant. The TW response [Fig. 6(a)] and the CP resistance [Fig. 6(c)] appear quite different from those of the baseline model (Fig. 1). However, the change in TW wavelength with gain is qualitatively similar [Fig. 6(b)], and consequently, as was seen with the baseline model, ϕ_C clearly reduces as the cochlear amplifier gain is reduced (Fig. 8). Note also that in Fig. 8 that an additional pulse centered at $2\phi_C$ can be seen when the attenuation is low, indicating the presence of multiple reflections.

2. The frequency-dependent scattering variant

A variant of the cochlear model was used in which the baseline scattering impedance given in Eq. (10) was replaced by a frequency-dependent impedance given by

$$\mathbf{Z_{Sc}}(x,\omega)=[\mathbf{Z_{Pass}}(x,\omega)+\mathbf{Z_{CA}}(x,\omega)]a_{Sc}b(x), \qquad (27)$$

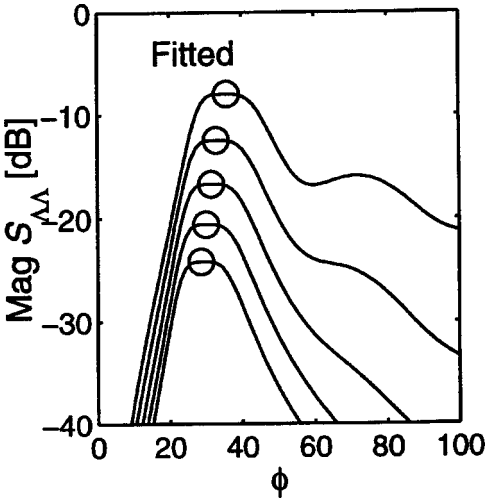


FIG. 8. Variation of the fitted ϕ spectrum of $\Lambda(\eta)$ for the Zweig–Talmadge cochlear amplifier variant during global gain reduction. The fitted ϕ spectrum is obtained from the four-parameter fit to data averaged over an ensemble of 32 realizations. The global cochlear amplifier attenuation was 0, 0.4, 0.8, 1.2, and 1.6 dB.

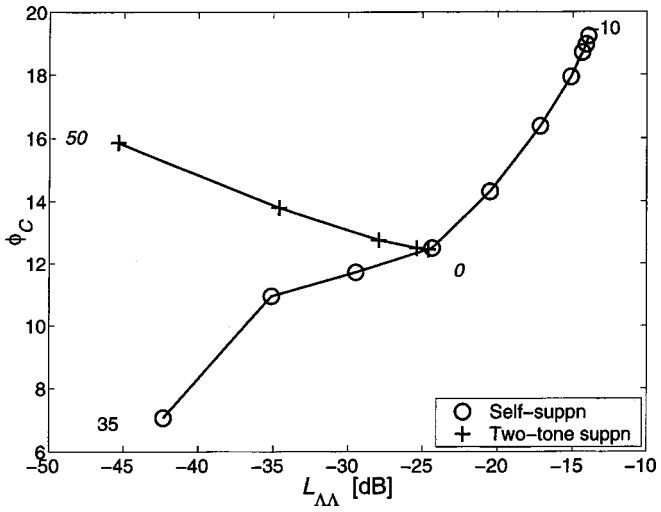


FIG. 9. Variation of the ϕ-center value for the frequency-dependent scattering variant during self- and two-tone suppression. The CP scattering impedance dependents on both CP place and stimulus frequency. Results from a single realization of the scattering impedance are shown. Each point is obtained from the four-parameter fit to a single realization of the raw autocovariance function. Self-suppression (open circles) is shown for stimulus levels of −10 to 35 dB SPL in 5-dB steps (indicated beside selected points). Two-tone suppression (+symbols) is shown for a probe level of 20 dB SPL and suppressor levels of 0, 20, 30, 40, and 50 dB SPL (indicated beside selected points in italic typeface). The graph of ϕ_C against $L_{\Lambda\Lambda}$ reveals a *y*-shaped pattern for the self- and two-tone suppression conditions taken together.

where a_{Sc} and $b(x)$ have the same meaning as in Eq. (10). This introduces frequency dependence into the scattering potential, causing the cochlear model to deviate still further from the assumptions underlying the simple spatial filtering model (Zweig and Shera, 1995, p. 2026).

The absolute value of ϕ_C predicted by this variant was found to be slightly lower than that from the baseline model. However, the predicted trends in ϕ_C during self- and two-tone suppression were found to trace out a similar *y*-shape to that seen in the baseline model (Fig. 9).

IV. DISCUSSION

The results of simulations of global gain reduction and of self-suppression, showing a clear reduction in ϕ_C with increased suppression, may have been anticipated. They are consistent with the findings of modeling studies reported by Zweig and Shera (1995) and Talmadge *et al.* (2000) and are explained by the reduction in TW delay that accompanies a broadening of the TW peak as stimulus levels are increased. The parametric fit to the data revealed a reduction in ϕ_C of around 50% for a suppression of the normalized SFOAE of around 20 dB.

Less expected are the results of the high-side two-tone suppression simulation, which show a slight increase in ϕ_C with increased suppression. The explanation for this is that the high-side suppressor disrupts the active processes basal to, rather than at the peak of the TW of the probe tone. Though the probe TW is suppressed, there is no corresponding increase in the wavelength of the TW, and hence no reduction in ϕ_C.

Although the changes in ϕ_C were revealed most clearly when the parametric fit was applied to ensemble averaged data, as in the simulation of the global gain reduction, changes in ϕ_C could still be discerned even when only a single realization of the SFOAE spectrum was obtained, as in the self- and two-tone suppression simulations. This is encouraging, since it suggests that changes in ϕ_C may be estimated for a single ear (see also the experimental data in the companion paper by Lineton and Lutman, 2003a).

The results of the global gain reduction simulations were found to be insensitive to major changes in the formulation and the parameters of the cochlear models, suggesting that they are a general feature of a broad class of cochlear model. Note, however, that the predicted changes in ϕ_C occur only for models which include some form of random scattering impedance. Cochlear models which contain a corrugated scattering impedance, as proposed by Strube (1989), cannot show such changes in ϕ_C, since in this case the spectral period is determined by the wavelength of the corrugations, rather than by the wavelength of the TW.

The predicted *y*-shaped patterns of variation of ϕ_C and ϕ_{BW} seen in Fig. 5 for the self- and two-tone suppression cases are readily testable results. An investigation of changes in the period of SFOAEs measured in human subjects is reported in Lineton (2001) and Linton and Lutman (2003a).

Recently, several studies have examined the relationship between spectral period and mild cochlear hearing loss (Don *et al.*, 1998; Avan *et al.*, 2000; Lucertini *et al.*, 2002; Sisto and Moleti, 2002). The implications of the results presented here for such an approach are threefold. First, the variability of the estimate of the spectral period arising from the inherently random nature of the scattering impedance (as evidenced by the differences in estimates for two realizations) makes such effects difficult to find in cross-sectional studies (necessitating large sample sizes). Second, the complex origins of the high-side two-tone suppression result cautions against applying a simple rule such as ''reduced cochlear amplifier gain causes increased spectral period''. Though this rule applies in self-suppression of SFOAEs, it does not apply in two-tone suppression, because the reduction in cochlear amplifier gain does not result in an increase in TW wavelength. While the rule also applies to a global reduction in cochlear amplifier gain, the effect of a localized reduction in cochlear amplifier gain due, for example, to a frequency-specific hearing loss may not be in every case so simply described. Third, caution must be applied when comparing the latency of TEOAEs with latency measures derived from the period of SFOAEs. Since TEOAEs employ broadband stimuli, the TW at any given frequency will, in general, experience high-side, low-side, and self-suppression, making the overall suppressive effect difficult to predict.

V. CONCLUSIONS

The results from the cochlear models presented here, when analyzed with a sensitive signal-processing technique described by Lineton and Lutman (2003b), show characteristic changes in SFOAE period due to ipsilateral acoustic suppression. The mechanism for these changes is that described by Zweig and Shera (1995): disruption of the cochlear amplifier due to suppression causes changes in the shape of the TW function, which in turn leads to changes in SFOAE spectral period.

In the case of self-suppression, an increase in spectral period was found with increasing stimulus level. This can be explained using the approximate relationship between TW wavelength and period presented by Zweig and Shera (1995): increasing stimulus level causes a broadening of the TW envelope, an increase in TW wavelength near the peak, and consequently an increase in SFOAE period. Interestingly, a different effect was found during high-side two-tone suppression: SFOAE spectral period is reduced rather than increased as suppression increases. The explanation for this is that the high-side suppressor disrupts the cochlear amplifier in a region basal to the peak of TW due to the probe tones. Consequently, it does not simply cause an overall broadening of the TW envelope, but instead leads to a complex change in its shape including localized narrowing.

It is important to recognize that different patterns of cochlear amplifier disruption along the BM can lead to quite different changes in SFOAE spectral period. The complexity of these results highlights the potential dangers of making general predictions about the effects of cochlear amplifier disruption on OAEs without examining the pattern of disruption in detail. This may be important in interpreting studies of the effects of hearing loss on OAE spectral period or group delay. Furthermore, the predicted patterns of variation in period form the basis of testable hypotheses to examine the veracity of different models.

ACKNOWLEDGMENTS

We wish to thank Torsten Marquardt, Amanda Hall, Sarosh Kapadia, Phil Nelson, and Bob Allen for their assistance and for many stimulating discussions. Christopher Shera and an anonymous reviewer also provided useful comments. The research was funded by the University of Southampton.

[1]In this paper, the SFOAE is defined as the component of the ear-canal pressure due to a backward-traveling wave originating in the cochlea. With this definition, no periodicity is seen in the magnitude spectrum of the SFOAE. Instead, periodicity is seen in both the real and imaginary parts of the frequency spectrum of the SFOAE. Other authors (e.g., Talmadge *et al.*, 2000) characterize SFOAEs using the magnitude spectrum of the ear-canal pressure. This spectrum shows a similar periodicity owing to the pattern of interference between the emission component and the stimulus components of the ear-canal pressure.

Avan, P., Wit, H. P., Guitton, M., Mom, T., and Bonfils, P. (**2000**). ''On the spectral periodicity of transient-evoked otoacoustic emissions from normal and damaged cochleas,'' J. Acoust. Soc. Am. **108**, 1117–1127.

Bendat, J. S., and Piersol, A. G. (**1966**). *Measurement and Analysis of Random Data* (Wiley, New York).

Bray, P. (**1989**). ''Click evoked otoacoustic emissions and the development of a clinical otoacoustic hearing test instrument,'' Ph.D. thesis, University of London.

Dallmayr, C. (**1987**). ''Stationary and dynamical properties of simultaneous evoked otoacoustic emissions (SEOAE),'' Acustica **63**, 223–255.

de Boer, E. (**1996**). ''Mechanics of the cochlea: Modeling efforts,'' in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), pp. 258–317.

Don, M., Ponton, C. W., Eggermont, J. J., and Kwong, B. (**1998**). ''The effects of sensory hearing loss on cochlear filter times estimated from auditory brainstem response latencies,'' J. Acoust. Soc. Am. **104**, 2280–2289.

Eggermont, J. J. (**1979**). ''Compound action potentials: Tuning curves and delay times,'' in *Models of the Auditory System and Related Signal Processing Techniques*, edited by M. Hoke and E. de Boer, Scand. Audiol. Suppl. **9**, 129–139.
Fukazawa, T., and Tanaka, Y. (**1996**). ''Spontaneous otoacoustic emissions in an active feed-forward model of the cochlea,'' Hear. Res. **95**, 135–143.
Furst, M., and Lapid, M. (**1988**). ''A cochlear model for otoacoustic emissions,'' J. Acoust. Soc. Am. **84**, 222–229.
Kanis, L. J., and de Boer, E. (**1993**). ''Self-suppression in a locally active nonlinear model of the cochlea: A quasilinear approach,'' J. Acoust. Soc. Am. **94**, 3199–3206.
Kanis, L. J., and de Boer, E. (**1994**). ''Two-tone suppression in a locally active nonlinear model of the cochlea,'' J. Acoust. Soc. Am. **96**, 2156–2165.
Kanis, L. J., and de Boer, E. (**1996**). ''Comparing frequency domain with time-domain solutions for a locally active nonlinear model of the cochlea,'' J. Acoust. Soc. Am. **100**, 2543–2546.
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, traveling waves and cochlear mechanisms,'' Hear. Res. **22**, 95–104.
Kemp, D. T., and Chum, R. A. (**1980**). ''Observations on the generation mechanism of stimulus frequency acoustic emissions—two tone suppression,'' in *Psychophysical, Physiological and Behavioural Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen, *Proceedings of the 5th International Symposium on Hearing* (Delft University Press, Delft, The Netherlands), pp. 34–42.
Kringlebotn, M. (**1988**). ''Network model for the human middle ear,'' Scand. Audiol. **17**, 75–85.
Lineton, B. (**2001**). ''Testing a model of stimulus frequency otoacoustic emissions in humans,'' Ph.D. thesis, University of Southampton.
Lineton, B., and Lutman, M. E. (**2003a**). ''The effect of suppression on the periodicity of stimulus frequency otoacoustic emissions: Experimental data,'' J. Acoust. Soc. Am. **114**, 871–882.
Lineton, B., and Lutman, M. E. (**2003b**). ''A parametric model of the spectral periodicity of stimulus frequency otoacoustic emissions,'' J. Acoust. Soc. Am. **114**, 883–895.
Lonsbury-Martin, B. L., Harris, F. P., Stagner, B. B., Hawkins, D. H., and Martin, G. K. (**1990**). ''Distortion product emissions in humans II. Relations to acoustic immittance and stimulus frequency and spontaneous otoacoustic emissions in normally hearing subjects,'' Ann. Otol. Rhinol. Laryngol. Suppl. **147**, 15–29.
Lucertini, M., Moleti, A., and Sisto, R. (**2002**). ''On the detection of early cochlear damage by otoacoustic emission analysis,'' J. Acoust. Soc. Am. **111**, 972–978.
Manley, G. A. (**1983**). ''Frequency spacing of acoustic emissions: A possible explanation,'' in *Mechanisms of Hearing*, edited by R. A. Webster and L. M. Aitken (Monash University Press, Clayton, Australia), pp. 36–39.
Moore, B. C., and Glasberg, B. R., (**1987**). ''Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns,'' Hear. Res. **28**, 209–225.
Neely, S. T., and Kim, D. O. (**1986**). ''A model for active elements in cochlear biomechanics,'' J. Acoust. Soc. Am. **79**, 1472–1480.
Neely, S. T., Norton, S. J., Gorga, M. P., and Jesteadt, W. (**1988**). ''Latency of auditory brainstem responses and otoacoustic emissions using tone-burst stimuli,'' J. Acoust. Soc. Am. **83**, 652–656.
Norton, S. J., and Neely, S. T. (**1987**). ''Tone-burst-evoked otoacoustic emissions from normal-hearing subjects,'' J. Acoust. Soc. Am. **81**, 1860–1872.
Patuzzi, R. (**1996**). ''Cochlear micromechanics and macromechanics,'' in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), pp. 186–257.
Peisl, W. (**1988**). ''Simulation von zeitverzögerten evozierten otoakustischen Emissionen mit hilfe eines digitalen Innenohrmodells,'' in *Fortschritte der Akustik—DAGA'88* (DPG-GmbH, Bad onnef), pp. 553–556.
Shera, C. A., and Zweig, G. (**1991**). ''A symmetry suppresses the cochlea catastrophe,'' J. Acoust. Soc. Am. **89**, 1276–1289.
Shera, C. A., and Zweig, G. (**1993a**). ''Order from chaos: Resolving the paradox of periodicity in evoked otoacoustic emissions,'' in *Biophysics of Hair-Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 54–63.
Shera, C. A., and Zweig, G. (**1993b**). ''Noninvasive measurement of the cochlear traveling-wave ratio,'' J. Acoust. Soc. Am. **93**, 3333–3352.
Sisto, R., and Moleti, A. (**2002**). ''On the frequency dependence of otoacoustic emission latency in hypoacoustic and normal ears,'' J. Acoust. Soc. Am. **111**, 297–308.
Strube, H. W. (**1989**). ''Evoked otoacoustic emissions as cochlear Bragg reflections,'' Hear. Res. **38**, 35–46.
Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). ''Modeling otoacoustic emission and hearing threshold fine structures,'' J. Acoust. Soc. Am. **104**, 1517–1543.
Talmadge, C. L., Tubis, A., Long, G. R., and Tong, C. (**2000**). ''Modeling the combined effects of basilar membrane nonlinearity and roughness on stimulus frequency otoacoustic emission fine structures,'' J. Acoust. Soc. Am. **108**, 2911–2932.
Tognola, G., Grandori, F., and Ravazzani, P. (**1997**). ''Time-frequency distributions of click-evoked otoacoustic emissions,'' Hear. Res. **106**, 112–122.
Wada, H., Nakajima, T., and Ohyama, K. (**1999**). ''Effect of irregularities in basilar membrane impedance on TEOAEs: Theoretical considerations,'' J. Oto-rhino-laryngology Relat. Special. **61**, 252–258.
Wilson, J. P. (**1980**). ''Evidence for a cochlear origin for acoustic re-emissions, threshold fine-structure, and tonal tinnitus,'' Hear. Res. **2**, 233–252.
Wit, H. P., van Dijk, P., and Avan, P. (**1994**). ''Wavelet analysis of real ear and synthesised click evoked otoacoustic emissions,'' Hear. Res. **73**, 141–147.
Zweig, G. (**1991**). ''Finding the impedance of the organ of Corti,'' J. Acoust. Soc. Am. **89**, 1229–1254.
Zweig, G., and Shera, C. A. (**1995**). ''The origin of periodicity in the spectrum of evoked otoacoustic emissions,'' J. Acoust. Soc. Am. **98**, 2018–2047.
Zwicker, E. (**1986**). ''Otoacoustic emissions in a nonlinear cochlear hardware model with feedback,'' J. Acoust. Soc. Am. **80**, 154–162.
Zwicker, E. (**1988**). ''Otoacoustic emissions and cochlear traveling waves,'' in *Cochlear Mechanisms—Structure, Function, and Models*, edited by J. P. Wilson and D. T. Kemp, NATO ASI Series (Plenum, New York), pp. 359–366.
Zwicker, E. (**1990**). ''On the frequency separation of simultaneously evoked otoacoustic emissions' consecutive extrema and its relation to cochlear traveling waves,'' J. Acoust. Soc. Am. **88**, 1639–1641.
Zwicker, E., and Lumer, G. (**1985**). ''Evaluating traveling wave characteristics in man by an active nonlinear cochlea preprocessing model,'' in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis, Lecture Notes in Biomathematics, Vol. 64 (Springer, Berlin), pp. 250–257.
Zwicker, E., and Peisl, W. (**1990**). ''Cochlear preprocessing in analog models, in digital models, and in human inner ear,'' Hear. Res. **44**, 209–216.
Zwicker, E., and Schloth, E. (**1984**). ''Interrelation of different otoacoustic emissions,'' J. Acoust. Soc. Am. **75**, 1148–1154.

The effect of suppression on the periodicity of stimulus frequency otoacoustic emissions: Experimental data

Ben Lineton[a] and Mark E. Lutman
Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom

(Received 10 November 2002; revised 18 April 2003; accepted 23 April 2003)

In a companion paper [Lineton and Lutman, J. Acoust. Soc. Am. **114**, 859–870 (2003)], changes in the spectral period of stimulus frequency otoacoustic emissions (SFOAEs) during self-suppression and two-tone suppression were simulated using a nonlinear cochlear model based on the distributed roughness theory of otoacoustic emission generation [Zweig and Shera, J. Acoust. Soc. Am. **98**, 2018–2047 (1995)]. The current paper presents the results of an experimental investigation of SFOAE suppression obtained from 20 human subjects. It was found that, in most subjects, the spectral period increased during self-suppression, but reduced during high-side two-tone suppression. This pattern of results is in close agreement with the predictions of the cochlear model, and therefore strongly supports the distributed roughness theory of Zweig and Shera. In addition, the results suggest that the SFOAE spectral period is sensitive to changes in the state of the cochlear amplifier. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1582437]

PACS numbers: 43.64.Jb, 43.64.Kc [BLM]

I. INTRODUCTION

Current theories of cochlear mechanics predict that the spectral period of otoacoustic emissions (OAEs) is determined predominantly by the shape of the cochlear traveling wave (TW) (Shera and Zweig, 1993a; Zweig and Shera, 1995; Talmadge *et al.*, 1998, 2000) and that it may therefore be altered by suppression of the TW. Using a model based on the theory of Shera and Zweig (1993a), simulations of the suppression of stimulus frequency OAEs (SFOAEs) are presented in a companion paper (Lineton and Lutman, 2003a). These simulations show that changes in SFOAE spectral period can be induced by both self-suppression and two-tone suppression. However, a clear difference is seen in the pattern of variation of spectral period in the two types of suppression: spectral period is seen to increase during self-suppression, but decrease during high-side two-tone suppression. These results from the cochlear models are summarized by the y-shaped pattern shown in Fig. 1. This difference between the effect of the two types of suppression on the SFOAE may be explained by the differences in the pattern of progressive saturation of the cochlear amplifier along the basilar membrane (BM).

Several researchers have investigated the effect on SFOAEs of increasing stimulus level (i.e., self-suppression). For pure tones below 5 dB above the threshold of hearing, SFOAEs were found to grow linearly with stimulus level (Shera and Zweig, 1993b). However, as stimulus level is increased further, SFOAEs exhibit compressive nonlinearity. Several authors report this nonlinear effect on amplitude without reporting any accompanying effect of level on spectral period (Kemp and Chum, 1980a; Zwicker and Schloth, 1984; Zwicker, 1990). Dallmayr [1987, Fig. 8(b)] explicitly reported an invariance of the SFOAE phase spectrum (which is directly related to the period) with stimulus level. In contrast, a few authors have reported small effects of level on SFOAE period (or related quantities). Zweig and Shera (1995, Fig. 13) remark that they have recorded changes in SFOAE spectral period with level which qualitatively match the predictions of their model (though the measured data are not presented). Kemp and Brown (1983) reported a reduction in SFOAE group delay with increasing stimulus level, which is consistent with the increase in spectral period predicted by Zweig and Shera (1995).

Similarly, reductions of the latency of click-evoked and tone-burst evoked OAEs with increasing stimulus level have also been reported (Wilson, 1980; Norton and Neely, 1987; Tognola *et al.*, 1997). Although these changes in OAE latency with stimulus level are in the same direction as those predicted for the SFOAE group delay, the results from cochlear models presented by Lineton and Lutman (2003a) caution against the assumption that these two different measures of delay will behave similarly under all experimental conditions.

The effects of suppressor tones on OAEs have also been studied extensively (Kemp and Chum, 1980a, 1980b; Brass and Kemp, 1993; Zwicker and Wesel, 1990; Sutton, 1985; Dallmayr, 1987; Tavartkiladze *et al.*, 1994). However, none of these has looked for, or reported, any change in SFOAE period with suppressor level.

The effects of other suppressive agents on OAEs has also been investigated, without any changes in spectral period being reported. These include ototoxic drugs (Long and Tubis, 1988; Karlsson *et al.*, 1991; Brown *et al.*, 1993), contralateral acoustic stimulation (e.g., Collet *et al.*, 1994; Lind, 1994; Giraud *et al.*, 1996) and of acoustic overstimulation (Kemp, 1986). There is, however, evidence that changes in group delay may occur as a result of cochlear damage (Don

[a]Currently at: MRC Institute of Hearing Research, Royal South Hants Hospital, Southampton SO14 OYG, United Kingdom. Electronic mail: b.lineton@soton.ac.uk

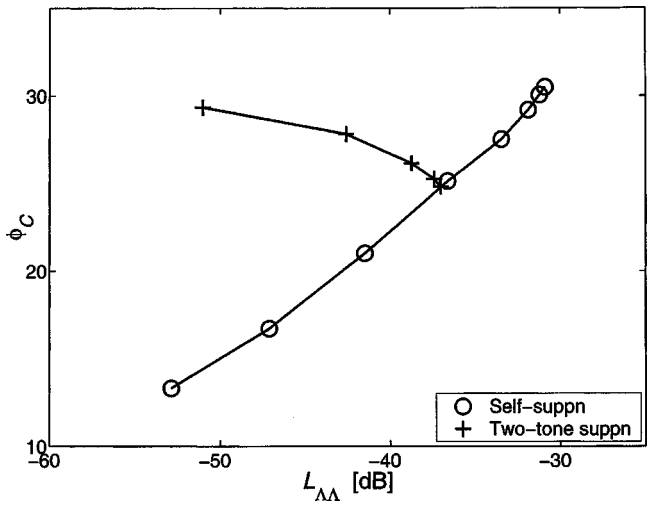


FIG. 1. Variation of ϕ_C during self- and two-tone suppression, obtained from cochlear models based on Zweig and Shera's theory (1995) [adapted from the companion paper on cochlear modeling, Lineton and Lutman, 2003a, Fig. 5(a)]. For each point, the quantity ϕ_C (approximately the reciprocal of $\Delta f/f$) is obtained from the four-parameter fit to a single realization of the raw autocovariance function. Self-suppression (open circles) is shown for stimulus levels of 20, 30, 35, 40, 45, 50, 55, and 60 dB SPL. Two-tone suppression (+ symbols) is shown for a probe level of 45 dB SPL and suppressor levels of 30, 45, 55, 65, and 75 dB SPL, and with a suppressor-to-probe frequency ratio of 1.1.

et al., 1998; Avan *et al.*, 2000; Lucertini *et al.*, 2002; Sisto and Moleti, 2002).

The aim of this paper is to test the detailed predictions of the behavior of spectral period during self- and two-tone suppression made in the companion paper on cochlear modeling (Lineton and Lutman, 2003a), thereby providing a test of the distributed roughness theory of the origin of spectral periodicity proposed by Zweig and Shera (1995). Experimental data are presented from extensive measurements of SFOAEs from 20 subjects under conditions of self- and two-tone suppression and the results then compared directly with predictions made in the companion paper (Lineton and Lutman, 2003a). A sensitive parametric fitting technique for detecting changes in SFOAE period described in a second companion paper (Lineton and Lutman, 2003b) is applied to the results.

II. METHODS

A. Subjects

Subjects were normal-hearing adults aged between 18 and 40 years. Measurements were in one ear per subject, with 20 ears being tested in total (10 male and 10 female). Subjects were screened to ensure that the test ear was normal on otoscopy and tympanometry, and that the pure-tone hearing threshold level was 10 dB or better at each of the frequencies 0.5, 1, 2, 4, and 8 kHz, and was 15 dB or better at 0.25 kHz. Measurements of spontaneous OAEs (SOAEs) or evoked OAEs did not form part of the subject selection criteria.

All subjects were paid volunteers, and the experiments were performed in accordance with the guidelines of the Declaration of Helsinki.

B. Equipment

The equipment used is specifically designed for the measurement of distortion product OAEs (DPOAEs) and SFOAEs. It comprises a PC containing a Loughborough Sound Images DSP card that controls a remote converter module, which is connected to the two earphones and microphone of an Etymotic ER-10B+ probe assembly.

The Etymotic ER-10B+ probe assembly comprises a low-noise microphone which is connected through a preamplifier to the input of the remote converter module, and two ER-2 insert earphones delivering tones to the ear via coupling tubes running through the body of the probe. The two earphones are connected to two output channels on the remote converter module. This allows two tones to be presented to the ear simultaneously via the two earphones, thus minimizing nonlinear artifacts due to the measurement system during two-tone suppression experiments.

The remote converter module comprises stereo 16-bit DACs, stereo 20-bit ADCs, and a sample rate generator. Tones are generated digitally in the DSP card with a sample rate of 32.768 kHz at frequencies which are multiples of 16 Hz. Each stereo output sample pair is sent to the remote converter module via a fast serial connection on the DSP and output under the timing of the sample rate generator. The input sample pair is acquired synchronously and sent via the serial connection to the DSP. The ADCs employ 8× oversampling which obviates the need for antialiasing filters. Only the most significant 16 bits of the ADCs are used by the DSP as implemented for the present work. The tones are segmented into epochs of 62.5 ms (2048 sample points) containing a whole number of periods of the stimulus tones. The sampled microphone signal is recorded via one ADC in corresponding epochs and loaded into buffers of 2048 points. On completion of the epoch and before the next epoch begins, an in-place FFT with 16-Hz resolution is calculated. Since the buffer holds a whole number of periods of the tones, truncation effects are eliminated. The FFT is then transferred to the PC for further processing. For simple recording of SFOAEs, only one DAC channel is used, while for recording DPOAEs and SFOAEs during two-tone suppression, both channels are used.

For DPOAEs and SFOAEs, the first buffer is always discarded since it contains the transient response of the transducers and the ear. Averaging of the complex FFT is performed across subsequent buffers to improve signal-to-noise ratio. Noisy buffers, as determined by the noise power in a number of spectral lines adjacent to the stimulus frequency, may be excluded from the average.

For recording SOAEs, no stimulus tones are presented and averaging of the FFT power is performed across a number of buffers, giving an estimate of the power spectrum of the ear-canal sound pressure with a resolution of 16 Hz.

For the present experiments, the measurement system was used to measure SOAEs as described above and SFOAEs, either with or without a suppressor tone. Noisy epochs were rejected when the average level in the ten spectral lines on either side of the SFOAE frequency (but excluding the spectral line at the frequency of the suppressor tone, if present) exceeded a preset level, termed the rejection level.

Averaging of the epochs continued until a set number of epochs had been accepted. The choice of the rejection level and the number of averaged epochs involves a trade-off between two noise components termed here the additive noise and drift noise (Sec. II C 1). To minimize the overall noise, both the rejection level and the number of averaged epochs were altered with stimulus level. Depending on the stimulus level, the rejection level was set between 10 and 15 dB SPL and the number averaged epochs was set between 8 and 128. During frequency sweeps, the controlling software on the PC automatically changed the stimulus frequency as soon as the stopping criteria for the previous frequency had been satisfied.

The measurement system was calibrated using a Bruel & Kjaer type 4157 ear simulator conforming to IEC 711-1981 at 16-Hz intervals across the measured frequency range. Throughout this paper, the earphone level is defined as the sound-pressure level that would be recorded in the ear simulator at the same applied voltage level. Since the earphones and their tubing provided a high acoustic source impedance, a given earphone voltage gives a volume velocity that is approximately independent of the ear-canal impedance loading the probe. Thus, to a first approximation, earphone level corresponds to volume velocity. Phase calibration was not required, because a normalized measure of the SFOAE is calculated (Sec. III A 1) which is insensitive to the phase response of the system at the stimulus frequency.

C. Stimuli and procedure

Each subject sat for two sessions on different days. In each session, SFOAEs were measured following both the self-suppression and two-tone suppression protocols described below. SOAEs were also recorded at the beginning and end of each session.

Subjects were seated comfortably within a sound-treated test booth, with the PC and remote converter module located outside the booth. They were asked to remain as still as possible, and to swallow as infrequently as was comfortable. Each session lasted approximately 1.5 h.

1. Self-suppression protocol

In the self-suppression protocol, frequency sweeps were performed at seven stimulus levels, L, ranging from 14 to 50 dB SPL in 6-dB steps, where the stimulus level refers to the level measured in the ear simulator for the same earphone voltage. The stimulus frequency, f, was decreased from 2816 down to 1376 Hz in 16-Hz intervals, giving 91 points in the frequency spectrum.

In order to obtain an estimate of the complex SFOAE pressure, $\mathbf{p}_{SF}(f,L)$, at each point, a version of the high-level suppressor method (e.g., Kemp and Brown, 1983; Kemp *et al.*, 1990) was used whereby recordings are made in two different conditions (termed here the *unsuppressed* and *maximally suppressed* conditions). In the unsuppressed condition, the stimulus tone is presented alone to give a measurement of the complex ear-canal pressure, denoted $\mathbf{p}_{\mathbf{EC}}(f,L)$, while in the maximally suppressed condition, the stimulus tone is presented together with a high-level suppressor tone to give an estimate of the passive ear-canal pressure, denoted $\mathbf{p}_{\mathbf{EC:R=0}}(f,L)$. A suppressor tone at 58 dB SPL at a frequency 16 Hz above the stimulus tone was used throughout. The SFOAE pressure is then given by

$$\mathbf{p}_{\mathbf{SF}} \equiv \mathbf{p}_{\mathbf{EC}} - \mathbf{p}_{\mathbf{EC:R=0}}, \tag{1}$$

where the bold typeface denotes complex quantities, and where the dependence on f and L has been omitted. The term "$R=0$" in the subscript indicates that the cochlear reflectance is assumed to be zero in this measurement condition.

This technique gives rise to two separate components of noise in $\mathbf{p}_{SF}(f,L)$, termed here *additive noise* and *drift noise*. Additive noise is simply the random error in $\mathbf{p}_{\mathbf{SF}}(f,L)$ arising from acoustic or electrical noise contamination. Drift noise arises from the fact that the two measurements of ear-canal pressure used to calculate $\mathbf{p}_{SF}(f,L)$ cannot be obtained simultaneously. It is defined as the spurious component in $\mathbf{p}_{SF}(f,L)$ that arises from any slow change in ear-canal impedance between the two measurement, due, for example, to changes in probe fit or middle-ear pressure. Additive noise is reduced by averaging over more epochs and is most problematic at low stimulus levels where the SFOAE component is weakest. Conversely, drift noise is minimized by reducing the time interval between the unsuppressed and maximally suppressed measurements (e.g., by reducing the number of epochs in the average) and is most problematic at high stimulus levels where the magnitude of the SFOAE component is a smaller proportion of the ear-canal pressure. Therefore, in an attempt to optimize the overall signal-to-noise ratio (SNR), both the rejection level and number of averaged epochs were altered with stimulus level. Typically the two measurements were separated by about 8 s at the lowest stimulus level and 0.5 s at the highest level.

To eliminate any possible order effects, frequency sweeps were performed in order of increasing stimulus level in the first session, and in order of reducing stimulus level in the second session (although pilot studies had found that the measured ear-canal pressure was unaffected by either the level or the frequency of the preceding measurement).

2. Two-tone suppression protocol

In the two-tone suppression protocol, the stimulus comprised a probe tone that evokes the SFOAE, and a suppressor tone. To obtain a frequency sweep, the probe frequency, f_1, was decreased from 2816 down to 1376 Hz in 16-Hz intervals (as in the self-suppression protocol), while the suppressor frequency, f_2, was $1.3\times f_1$ (rounded to the nearest 16 Hz). A high value of the suppressor-to-probe frequency ratio is desirable in order to maximize the spatial separation of the two TW peaks, thereby maximizing the difference in the pattern of cochlear amplifier saturation between the self- and two-tone suppression conditions. The ratio of 1.3 was chosen as a compromise between this desire and the need to achieve a high degree of SFOAE suppression. The discrepancy in suppressor-to-probe frequency ratio between the experiment and the cochlear models is discussed in Sec. IV.

The probe level, L_1, and suppressor level, L_2, were held constant over each sweep. Frequency sweeps were performed at four suppressor levels, L_2, ranging from 26 to 62 dB SPL in 12-dB steps, while the probe level, L_1, was held

constant at 26 dB SPL for all sweeps. In addition to the four partially suppressed frequency sweeps, a fifth unsuppressed sweep was also recorded at a probe level of 26 dB SPL. This was an exact repeat of one of the self-suppression sweeps, but was included to give a baseline measurement made close in time to the other four partially suppressed sweeps.

In order to obtain an estimate of the partially suppressed SFOAE pressure, $\mathbf{p_{SF}}(f_1,L_1,f_2,L_2)$, at each probe frequency, a variant of the high-level suppressor method described above was again used, in which the ear-canal pressure was measured in two conditions. In the first (*partially suppressed*) condition, the suppressor tone is presented with the frequency and level specified above to give a measurement of the complex ear-canal pressure, $\mathbf{p_{EC}}(f_1,L_1,f_2,L_2)$. In the second (*maximally suppressed*) condition, the suppressor tone is set to 58 dB SPL at a frequency 16 Hz above the probe tone to give an estimate of the passive ear-canal pressure, $\mathbf{p_{EC:R=0}}(f_1,L_1)$. The probe tone is the same in both conditions. The partially suppressed SFOAE pressure is calculated from Eq. (1), but with $\mathbf{p_{SF}}$ and $\mathbf{p_{EC}}$ now dependent on f_1, L_1, f_2, and L_2. All complex pressures are defined as the components at the probe frequency. Note that in the maximally suppressed condition, rather than presenting both the partially suppressing and the maximally suppressing tones together, only the single high-level suppressing tone is presented. It is assumed that once the SFOAE has been maximally suppressed, the partially suppressing tone would have no significant additional effect. This is true if the passive ear-canal pressure, $\mathbf{p_{EC:R=0}}(f_1,L_1)$, is dependent only on the probe tone. As in the self-suppression case, the measurements in the two conditions were made within a few seconds of each other to minimize drift error. In the first session, frequency sweeps were performed in order of increasing suppressor level, while in the second session, this order was reversed.

III. RESULTS

A. Analysis of results

1. Estimating the period using a four-parameter fit

The following analysis procedure applies to the data obtained from both the self- and two-tone suppression protocols. Since the two-tone suppression protocol was only performed at a single probe level and a fixed suppressor: probe frequency ratio, the SFOAE pressure $\mathbf{p_{SF}}$ can be specified by only two independent variables: the probe frequency, f_1, and the suppressor level, L_2. To allow a single equation to refer to both the self- and two-tone suppression protocols, the SFOAE pressure $\mathbf{p_{SF}}$ is expressed as functions of a single frequency, f_1 and the level, L_V, of the variable tone. Thus, f_1 and L_V refer to the stimulus frequency and stimulus level in the self-suppression protocol, and to the probe frequency and suppressor level in the two-tone suppression protocol.

To quantify the variations in SFOAE period, the data were analyzed in four stages. In the first stage, the normalized SFOAE spectrum is calculated from

$$\Lambda(f_1,L_V)\equiv\frac{\mathbf{p_{SF}}(f_1,L_V)}{\mathbf{p_{EC:R=0}}(f_1,L_V)}, \tag{2}$$

for each frequency sweep, where $\mathbf{p_{SF}}$ is obtained from the measured ear-canal pressure in the two conditions via Eq. (1).

In the second stage, the normalized SFOAE is obtained as a function of logarithmic frequency, η, defined by

$$\eta\equiv-\ln(f_1/1\ \text{kHz}). \tag{3}$$

This is achieved by first resampling the 91-point spectrum using a low-pass interpolation algorithm, and then further interpolating at 512 equal logarithmic frequency intervals, to give $\Lambda(\eta,L_V)$. In the following analysis, the discrete form of the normalized SFOAE sampled at these equal logarithmic frequency intervals is denoted by $\Lambda(n,L_V)$ where n indexes the 512 points.

In the third stage, an estimate of the autocovariance function of Λ is obtained for each frequency sweep. A direct estimate of this function can be obtained from

$$\mathbf{C}_{\Lambda\Lambda}^{(\text{raw})}(m)\equiv\frac{1}{N}\sum_{n=0}^{N-m-1}\Lambda^*(n)\Lambda(n+m), \tag{4}$$

where m indexes the logarithmic frequency lag, * denotes complex conjugation, the superscript "*raw*" denotes an estimate based on a single realization of the random process, and where for brevity the dependence on the level has been omitted. However, an alternative method, described in the Appendix, which gives an improved estimate in the presence of noise, was used for the experimental data.

In the fourth stage, a set of four parameters was extracted from each of the SFOAE autocovariance functions, using the same four-parameter signal characterization method that was used for the cochlear model results in the companion paper (Lineton and Lutman, 2003a) and which is explained in detail in the second companion paper (Lineton and Lutman, 2003b). The first of these parameters, termed the *ϕ-center value* and denoted ϕ_C, is used in this paper as the measure of the spectral period, taken over the entire frequency interval of interest. It is roughly related to the interpeak spacing by

$$\phi_C\approx\frac{f}{\Delta f}, \tag{5}$$

where f is defined as the geometric mean frequency and Δf the frequency interval of two adjacent spectral peaks in the amplitude of the ear-canal pressure.

The second parameter, termed the *ϕ bandwidth* and denoted, ϕ_{BW}, quantifies the spread of periodicities within the spectrum. The third parameter, α, quantifies the effects of multiple apical and basal reflections of the TW within the cochlea. The final parameter, $\sigma_{\Lambda\Lambda}$, is the rms amplitude of Λ calculated across the measured spectral interval. When expressed in decibels, it is termed the *normalized SFOAE level* and denoted $L_{\Lambda\Lambda}$.

The parametric fit was performed using the same procedure as in the analysis of the cochlear model results. In summary, this procedure minimizes the mean-square error between the estimated and fitted autocovariance functions. For reasons explained in the companion paper on parametric fit-

TABLE I. Number of frequency sweeps recorded from 20 subjects.

	Target	Recorded	Rejected	Accepted
Self-suppn.	280	273	53	220
Two-tone suppn.	200	170	47	123

ting (Lineton and Lutman, 2003b), the fit was performed over the range $0<m<m_{\mathrm{Lim}}$, where $m_{\mathrm{Lim}}=64$ (giving $\eta'_{\mathrm{Lim}}=0.087$). This corresponds to about 1/8 of an octave, or roughly 1–2 spectral periods. This equals the value used for the cochlear model results (Lineton and Lutman, 2003a). The results of the parametric fit were found to be insensitive to this choice over a wide range of values.

In order to illustrate any changes in spectral period better, it is useful to inverse Fourier transform the autocovariance functions into the ϕ domain, where ϕ is the Fourier conjugate of the logarithmic frequency, η (or logarithmic frequency lag, η'). The inverse Fourier transform of the estimated autocovariance function is termed here the *raw ϕ spectrum*, and is defined

$$S_{\Lambda\Lambda}^{(\mathrm{raw})}(\phi)\equiv\mathbb{F}^{-1}\{\mathbf{C}_{\Lambda\Lambda}^{(\mathrm{raw})}(\eta')\}$$
$$=\int_{-\infty}^{\infty}\mathbf{C}_{\Lambda\Lambda}^{(\mathrm{raw})}\exp(i2\pi\phi\eta')d\eta', \tag{6}$$

where the estimated autocovariance function is assumed to be zero for values of η' outside the range obtained from the measurements. [Note that $S_{\Lambda\Lambda}^{(\mathrm{raw})}(\phi)$ is approximately a scaled version of the squared magnitude of the inverse FFT of $\Lambda(\eta)$.] The Fourier transform of the fitted autocovariance function, termed here the *fitted ϕ spectrum*, is defined in the same way. The fitted ϕ spectrum, which is a function of the four parameters only, is a pulse-shaped function centered at ϕ_C, and with a pulse width characterized by ϕ_{BW}.

2. Rejection of frequency sweeps with a poor SNR

Each frequency sweep was analyzed to assess the SNR. The SNR was estimated from the ratio of the frequency-averaged power of the SFOAE pressure, $\mathbf{p_{SF}}(f)$, to the frequency-averaged power of the estimated noise, where the noise estimate is based on the power in the ten spectral lines adjacent to the probe frequency. Frequency sweeps with an estimated SNR of less than 6 dB were rejected from the analysis. Table I shows the breakdown of frequency sweeps according to whether they were accepted or rejected on the basis of their estimated SNR. The total recorded sweeps fell short of the target, due to some subjects leaving before all measurements were complete.

3. Treatment of SOAEs

SOAEs were identified by visual inspection of the background recordings of ear-canal pressure. Any sharp peak of bandwidth 32 Hz or less which clearly stood out above the surrounding background noise by at least 1 dB was designated an SOAE. If at any frequency the level of the SOAE was of similar amplitude to that of the SFOAE, then it may be argued that the measured OAE originates predominantly as a spontaneous rather than an evoked emission. To test whether the presence of SOAEs influenced the results, each frequency sweep was given a *spontaneous-to-evoked OAE* (S/EOAE) rating, defined as follows. The level difference obtained by subtracting the SOAE level from the SFOAE level at the SOAE frequency was calculated for each SOAE present. The minimum level difference over all the SOAEs was then obtained and the frequency sweep given a S/EOAE rating of ''low'' if this minimum level difference exceeded 6 dB, and ''high'' if it did not. If no measurable SOAEs were present, the S/EOAE rating was ''zero.'' The S/EOAE rating was used only to classify the data, not to reject data. Table II shows the breakdown of accepted frequency sweeps according to their S/EOAE rating.

TABLE II. Number of accepted frequency sweeps broken down by S/EOAE rating.

		S/EOAE rating		
	Total	Zero	Low	High
Self-suppn.	220	96	74	50
Two-tone suppn.	123	41	45	37

B. Variation of SFOAE frequency spectrum

An example of the variation in SFOAEs under self-suppression for a subject with relatively strong SFOAEs is shown in Fig. 2. In panel (a), the magnitude of the ripples in the spectrum of the ear-canal pressure is seen to reduce as stimulus level is increased, indicating the compressive nonlinear growth in the SFOAE as stimulus level is increased. This is more clearly seen in the magnitude of the normalized SFOAE spectrum, $\Lambda(f)$, shown in panel (b). As expected, the SFOAE shows a peak in the magnitude spectrum at all the SOAE frequencies.

The SFOAE spectral period is seen in the ripple pattern of the real and imaginary parts of $\Lambda(f)$ shown in panels (d) and (e). At low stimulus levels, the magnitude of the $\Lambda(f)$ shown in panel (b) shows a similar pattern in the spacing of its peaks (coinciding with the SOAEs). However, as discussed by Shera and Zweig (1993b), this pattern arises from multiple reflections of the TW, rather than from a single reflection, and thus becomes less clear at high stimulus levels. Spectral period is also related to the slope of the phase of the normalized SFOAE shown in panel (c). The question of how changes in period across stimulus level manifest themselves in these representations is addressed in a later section.

The variation in SFOAEs under two-tone suppression is presented in Fig. 3 for the same subject. Apart from the degree of suppression achieved, which is less than that under self-suppression experiment, there are no obvious differences between the two series.

C. Variation of SFOAE ϕ spectrum

Changes in spectral period are revealed most clearly by examining the ϕ spectra. [Recall from Eq. (5) that ϕ_C and spectral period are reciprocally related.] The raw and fitted ϕ spectra calculated from the SFOAEs in Figs. 2 and 3 are shown in Fig. 4. While the raw ϕ spectra shown in panels (a)

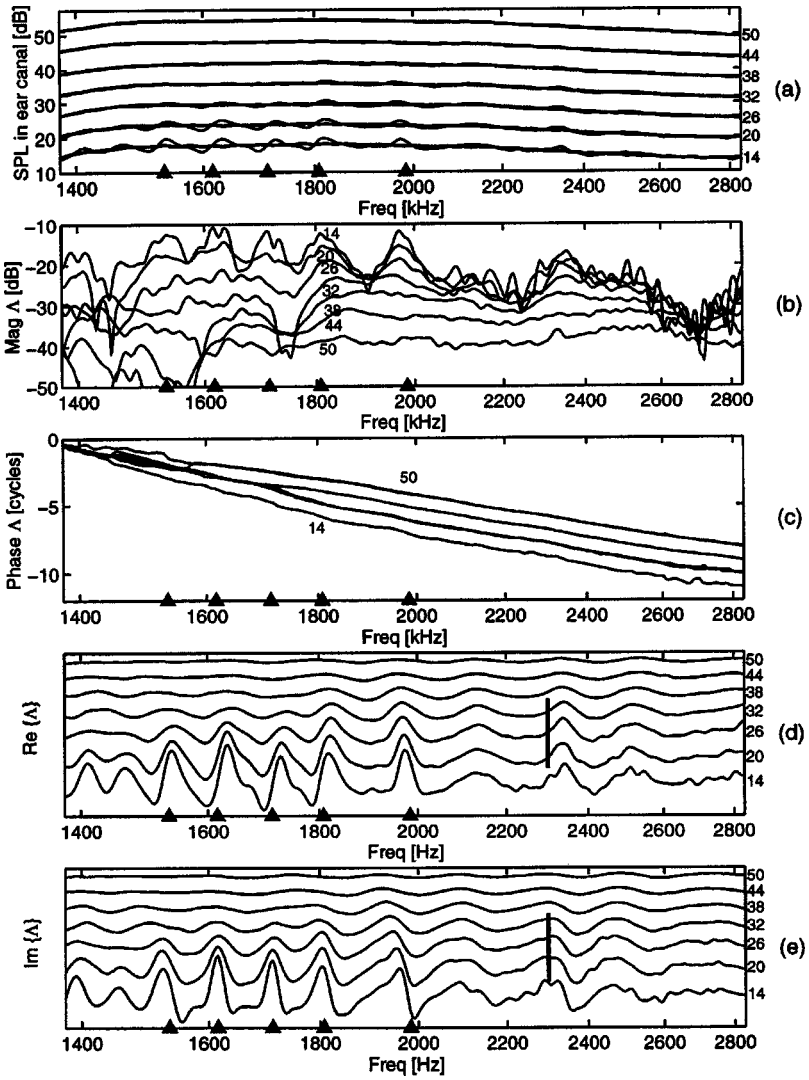


FIG. 2. Measured SFOAE frequency sweeps in self-suppression for ear S20. Each frequency sweep is obtained by stepping the stimulus frequency down from 2816 to 1376 Hz in 16-Hz intervals, holding the nominal stimulus level constant. The nominal stimulus level (indicated beside selected sweeps) ranges from 14 to 50 dB SPL in 6-dB steps. Frequencies where SOAEs are present are indicated by filled triangles on the frequency axis. In (a), the sound-pressure level measured in the ear canal is shown. The thick line indicates the maximally suppressed ear canal pressure, $|\mathbf{p}_{EC:R=0}|$, and the thin line the unsuppressed ear canal pressure, $|\mathbf{p}_{EC}|$. In (b)–(e) the normalized SFOAE, defined by $\Lambda=(\mathbf{p}_{EC}-\mathbf{p}_{EC:R=0})/\mathbf{p}_{EC:R=0}$, is shown. Panels (b) and (c) show the magnitude and phase. In panels (d) and (e), the real and imaginary parts are shown, with the frequency sweeps offset vertically for clarity. The vertical scale bars in (d) and (e) both indicate a range of 0.5.

and (b) clearly resemble the fitted ϕ spectra in panels (b) and (d), changes in period (as indicated by changes in the central location) can only be clearly seen in the fitted ϕ spectra. Similar differences between raw and fitted ϕ spectra are seen in the results from cochlear models (Lineton and Lutman, 2003a), where the locations of the peaks of the raw ϕ spectra were found to be much less sensitive to changes in the TW wavelength than those of the fitted ϕ spectra.

As the stimulus level is increased in the self-suppression experiment, the fitted ϕ spectrum in panel (b) shows a reduction in amplitude, a narrowing of the pulse width, and a clear reduction in the ϕ-center value, ϕ_C. In contrast, as the suppressor level is increased in the two-tone suppression experiment, the fitted ϕ spectrum in panel (d) shows a reduction in amplitude together with a slight increase in ϕ_C.

In order to present both self- and two-tone suppression data together, the variation of ϕ_C is plotted against the normalized SFOAE level, $L_{\Lambda\Lambda}$, under both conditions, for nine of the 20 subjects in Fig. 5. Most subjects show a similar pattern of results: ϕ_C is reduced in self-suppression and increased in two-tone suppression. This tends to give a characteristic y-shaped pattern in these plots. The results from both sessions are overlaid, and in most cases show good agreement. The remaining 11 subjects had considerably weaker SFOAEs than those in Fig. 5, with typical maximum values of $L_{\Lambda\Lambda}$ of -35 to -50 dB, and often with too many frequency sweeps being rejected on the basis of the poor SNR criterion for any trends in ϕ_C to be determined. Where sufficient frequency sweeps were retained in these subjects, the results were consistent with the trends to those in Fig. 5.

The variation of ϕ_{BW} with $L_{\Lambda\Lambda}$ is presented in Fig. 6 and shows a similar, though less consistent pattern across subjects. The variation of α, which characterizes multiple reflections within the cochlea, is plotted against $\sigma_{\Lambda\Lambda}$ in Fig. 7. Though not clear in all subjects, this shows a roughly linear relationship between α and $\sigma_{\Lambda\Lambda}$ during suppression.

For the results of the nine subjects presented, the intersession repeatability of the parameter estimates was generally good (Figs. 5–7), indicating both the stability of the measured SFOAE spectra across sessions and the achievement of a good SNR.

IV. DISCUSSION

A. Comparisons with cochlear models

Under self- and two-tone suppression, the measured variation in ϕ_C seen in Fig. 5 generally shows the same *y*-shaped pattern that is seen in the cochlear models reported

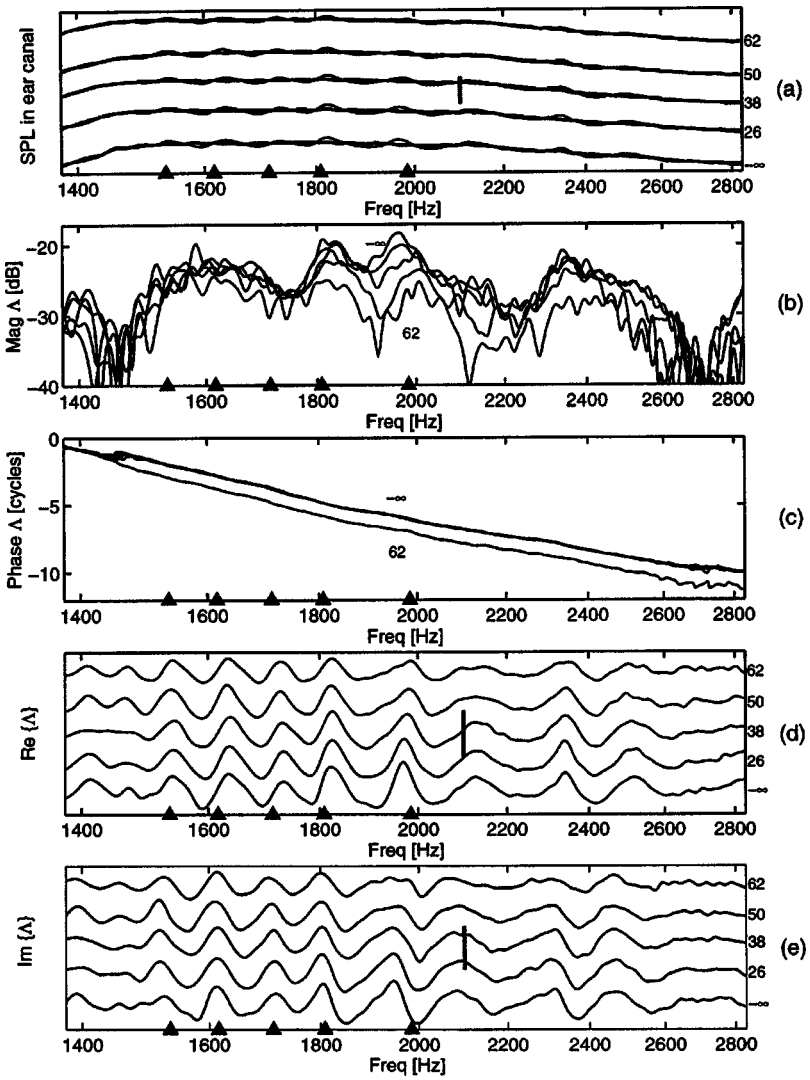


FIG. 3. Measured SFOAE frequency sweeps in two-tone suppression for ear S20. Each frequency sweep is obtained by stepping the probe frequency down from 2816 to 1376 Hz in 16-Hz intervals, holding the suppressor-to-probe frequency ratio at 1.3, and holding the nominal probe and suppressor levels constant. The nominal probe level was 26 dB SPL. The nominal suppressor level (indicated beside each sweep) ranges from 26 to 62 dB SPL in 12-dB steps. The unsuppressed sweep is also shown (indicated by a suppressor level of $-\infty$). Frequencies where SOAEs are present are indicated by filled triangles on the frequency axis. In (a), the sound-pressure level measured in the ear canal is shown. The thick line indicates the maximally suppressed ear-canal pressure, $|\mathbf{p}_{EC:R=0}|$, and the thin line the partially suppressed ear-canal pressure, $|\mathbf{p}_{EC}|$. For clarity, frequency sweeps offset vertically, with the vertically scale bar indicating a range of 5 dB SPL. In (b)–(e) the normalized SFOAE, defined by $\Lambda=(\mathbf{p}_{EC}-\mathbf{p}_{EC:R=0})/\mathbf{p}_{EC:R=0}$, is shown. Panels (b) and (c) show the magnitude and phase. In panels (d) and (e), the real and imaginary parts are shown, with the frequency sweeps offset vertically. The vertical scale bars in (d) and (e) both indicate a range of 0.2.

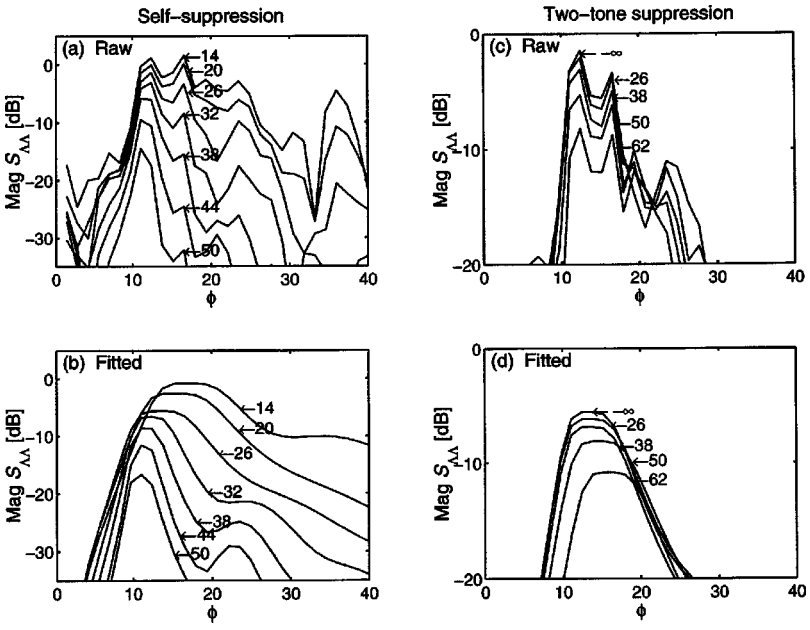


FIG. 4. Variation of the ϕ spectrum of $\Lambda(\eta)$ under self- and two-tone suppression for ear S20. Each trace is obtained from a single frequency sweep. Raw ϕ spectra were obtained from the Fourier transform of the raw autocovariance function. Fitted ϕ spectra were obtained from a four-parameter fit to the raw autocovariance function. In (a) and (b) the raw and fitted ϕ spectra under self-suppression are shown for nominal stimulus levels (indicated beside each trace) ranging from 14 to 50 dB SPL in 6-dB steps. In (c) and (d) the raw and fitted ϕ spectra under two-tone suppression are shown for a nominal probe level of 26 dB SPL and nominal suppressor levels (indicated beside each trace) ranging from 26 to 62 dB SPL in 12-dB steps. The unsuppressed sweep is also shown (indicated by a suppressor level of $-\infty$).

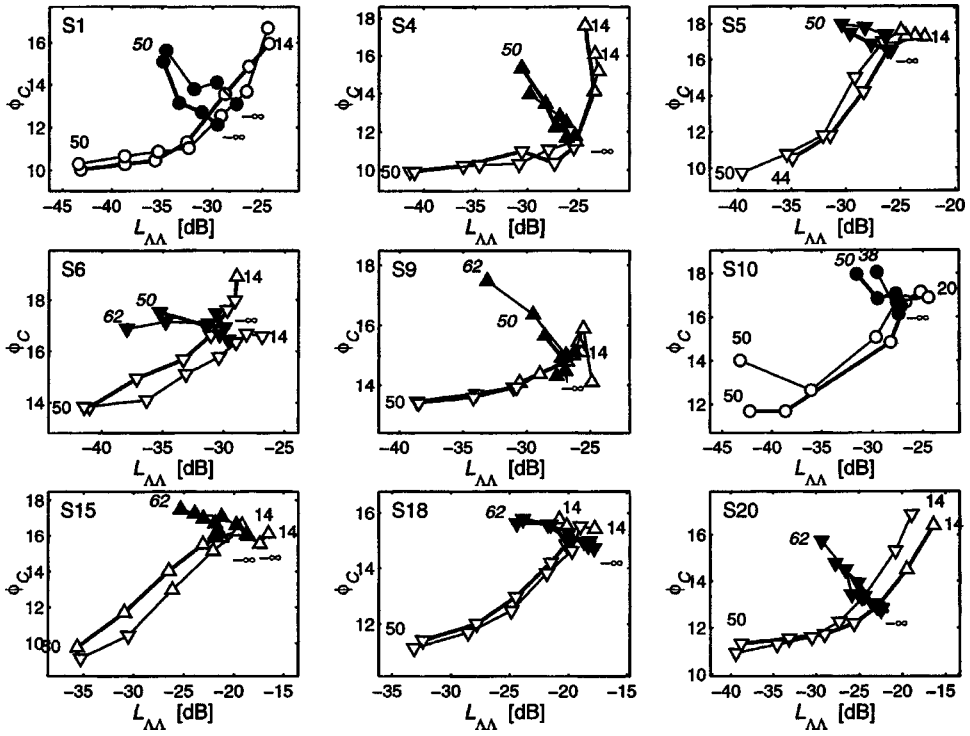


FIG. 5. Variation of the ϕ-center value (approximately the reciprocal of $\Delta f/f$) under self- and two-tone suppression for nine ears. Data for ears labeled S1, S4, S5, S6, S9, S10, S15, S18, and S20 are shown. Each point is obtained from a four-parameter fit to a single frequency sweep. Self-suppression data (indicated by open symbols) are shown for nominal stimulus levels (indicated by numbers in normal typeface beside selected points) ranging from 14 to 50 dB SPL in 6-dB steps. Two-tone suppression data (indicated by filled symbols) are shown for a nominal probe level of 26 dB SPL and nominal suppressor levels (indicated by numbers in italic typeface beside selected points) ranging from 26 to 62 dB SPL in 12-dB steps. The unsuppressed sweep is also shown (indicated by a suppressor level of $-\infty$). The shape of each point indicates the S/EOAE rating of the frequency sweep as follows: circle—*zero*; downward-pointing triangle—*low*; upward pointing triangle—*high*. Data from two repeat sessions are shown, the first identified by a solid thick line, the second by a thin solid line.

in the companion paper (Lineton and Lutman, 2003a) and reproduced here in Fig. 1. These cochlear models include both the distributed roughness model of BM reflection sites proposed by Shera and Zweig (1993a) and a saturating cochlear amplifier model described by Kanis and de Boer (1993, 1994). The self-suppression results corroborate the findings of Zweig and Shera (1995) of a qualitative match between measured and modeled SFOAE level series.

As discussed in detail by Lineton and Lutman (2003a), the y-shaped pattern arises from changes in both the wavelength and the shape of the envelope of the TW during suppression. The effect of self-suppression can be explained using the approximate relationship between TW wavelength and spectral period presented by Zweig and Shera (1995): increasing the stimulus level causes a broadening of the TW envelope, an increase in TW wavelength near the peak, and consequently an increase in SFOAE period. The effect of high-side two-tone suppression is more complex because the main site of cochlear amplifier saturation is basal to the peak of the TW due to the probe tone. Consequently, rather than causing an overall broadening of the TW envelope, the suppressor tone leads to a complex change in its shape including localized narrowing.

The close agreement between the simulated and measured y-shaped patterns thus lends strong support not only to Shera and Zweig's distributed roughness theory but also to the nonlinear model of the cochlear amplifier reported by Kanis and de Boer (1993, 1994).

Note that no such changes in ϕ_C are predicted by models based on the cochlear corrugation theory proposed by Strube (1989), in which TW reflections arise from a quasiperiodic component of the BM impedance. In Strube's model, the SFOAE period is determined by the spatial period of the corrugations and is thus unchanged during suppression. The results thus contradict Strube's cochlear-corrugation theory.

Although the nine subjects whose results are presented in Fig. 5 show a y-shaped pattern, there is considerable variation between subjects. For example, the intersection point of the ''y'' occurs at a significantly lower value of ϕ_C in subjects 1, 4, and 20 than it does in subjects 5, 10, 15, and 18. From comparisons with the model results, two explanations for these differences can be offered. One possibility is that for a given change in TW amplitude, the shape of the TW envelope may alter more rapidly in the former group of subjects than in the latter, leading to a more rapid decrease in ϕ_C for a given change in $\sigma_{\Lambda\Lambda}$. Such differences could arise from differences in the nonlinear characteristics of the cochlear amplifiers between the two groups. A second possibility is that the differences are simply due to errors in estimating ϕ_C. For example, a comparison by Lineton and Lutman (2003a, Fig. 5) of two cochlear models, differing only in the particular realization of the random inhomogeneities along the BM, shows considerable differences in the estimated y-shaped patterns.

As discussed in Lineton and Lutman (2003a), the absolute value of ϕ_C obtained from the cochlear models is unre-

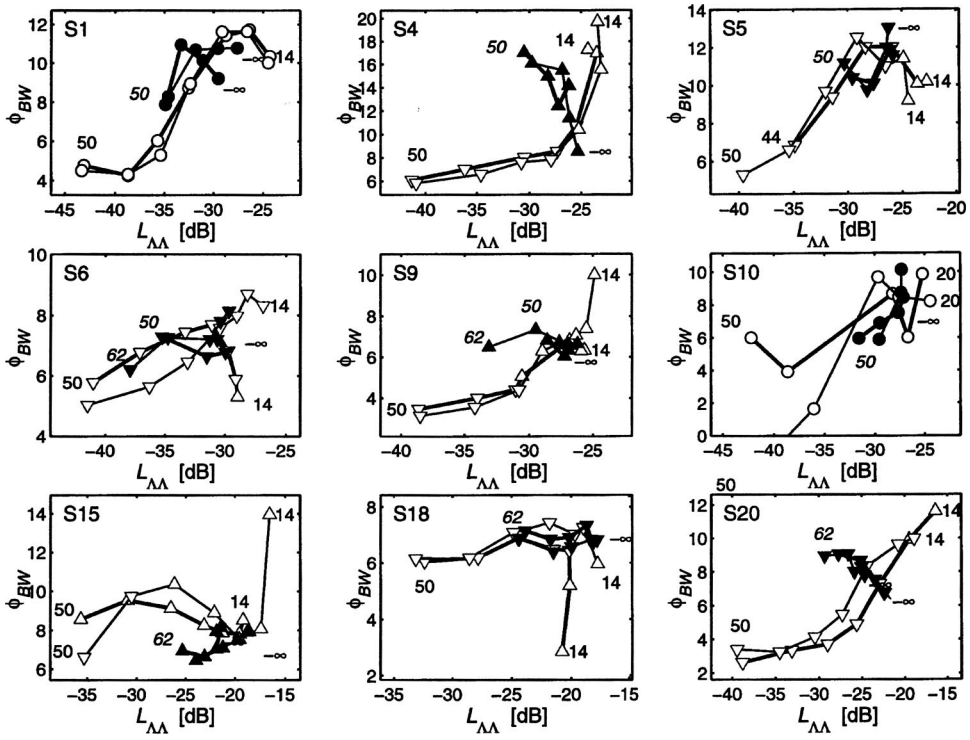


FIG. 6. Variation of the ϕ bandwidth under self- and two-tone suppression for nine ears. The ϕ bandwidth, which is one of four parameters obtained from the fit to the SFOAE frequency sweep, represents a measure of the width of the SFOAE ϕ-spectrum. In Zweig and Shera's theory it is related to the width of the TW peak (Zweig and Shera, 1995; Lineton and Lutman, 2003a). See Fig. 5 for the key to symbols.

alistically high, suggesting that the models underpredict the wavelength of the TW (see also the discussion by Talmadge *et al.*, 2000). However, there is much better agreement between modeled and measured results on the relative changes in ϕ_C under suppression. For example, in the self-suppression experiment, the normalized SFOAE level typically fell by around 18 dB, leading to a reduction in ϕ_C from about 17 to about 10: a reduction of around 40%. In the cochlear model results in Fig. 1, a similar fall in SFOAE level is accompanied by a reduction of around 50% in the value of ϕ_C.

The measured pattern of changes in ϕ_{BW} are less clear than those of ϕ_C, though in some cases a y-shaped pattern is again seen (subjects 1, 4, 9, and 20). This greater variability between subjects is expected, since the cochlear modeling studies showed that estimates of ϕ_{BW} based on a single realization of $\Lambda(\eta)$ are less reliable than those of ϕ_C (Lineton and Lutman, 2003a, Fig. 5, 2003b). Thus, the ϕ_{BW} parameter

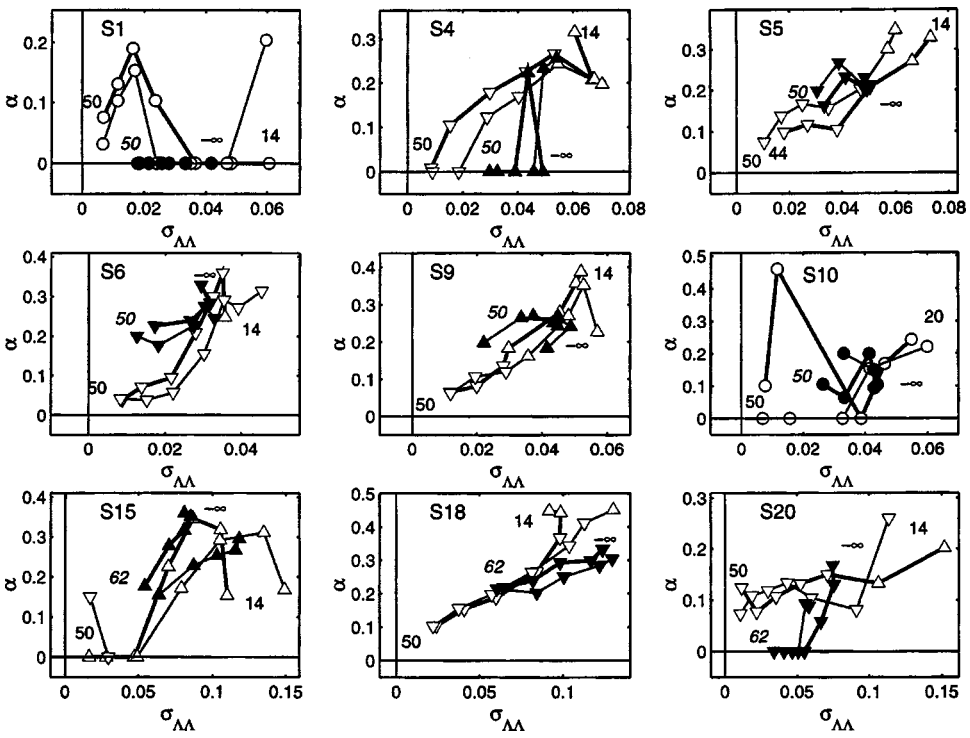


FIG. 7. Variation of α under self- and two-tone suppression for nine ears. The quantity α, which is one of four parameters obtained from the fit to the SFOAE frequency sweep, represents a measure of the influence on the SFOAE of multiple reflections within the cochlea. See Fig. 5 for the key to symbols.

is less useful as a measure of spectral period than is ϕ_C.

The validity of comparing the measured results of the high-side two-tone suppression experiment, where the suppressor-to-probe frequency ratio was 1.3, with the simulated results where the ratio was 1.1, may be questioned. The frequency ratio of 1.1 was chosen for the model simulations because higher values did not produce the same degree of suppression of the SFOAE level as is seen in the measured data. Although this indicates that there are some inaccuracies in the cochlear model, these models may still qualitatively simulate the mechanism that gives rise to the measured effects seen in the two-tone suppression experiment. Further modeling and experimental investigations of the effect of the frequency ratio would help to resolve this issue.

In a theoretical analysis, Talmadge *et al.* (2000) identifies a nonlinear cochlear mechanism of SFOAE generation, unconnected with BM roughness, which in self-suppression gives rise to a dramatic increase in the SFOAE period (corresponding to a reduction in ϕ_C) as stimulus level was increased. However, there is no indication in the measured data of any influence of this mechanism. The reduction of ϕ_C with stimulus level seen in the measured results is much smaller and occurs much more smoothly than that seen in the model results of Talmadge *et al.* (2000, Fig. 4). Note also that this nonlinear generation mechanism was specifically excluded from the cochlear model results reported in the companion paper (Lineton and Lutman, 2003a). Therefore, the distinctive y-shaped variation of ϕ_C seen in the model results arose purely from the roughness-based mechanism of SFOAE generation. The similarity of the measured y-shaped variation to the model results suggests that these too were dominated by the roughness-based mechanism. One possible reasons for the absence of any significant effect of the purely nonlinear mechanism on the measured results was the stimulus level used in the measurements, which was limited to 50 dB SPL.

B. Qualitative effect of suppression on the SFOAE frequency spectrum

It is interesting to consider how the changes in period that are indicated by shifts in the ϕ spectra manifest themselves in the SFOAE frequency spectra shown in Figs. 2 and 3. Despite the change in ϕ_C of around 40%, visual inspection of the ripple patterns in panels (a), (d), and (e) does not reveal any obvious change in the frequency interval between adjacent peaks. The explanation for this is that ϕ_C, as obtained from the four-parameter model, represents a weighted mean of the ϕ values from a number of different components which make up the ripple pattern. The reduction in ϕ_C arises from a reduction in the amplitude of those components with a high ϕ value relative to those with a low ϕ value. In contrast, the frequency interval between the ripples tends to be dominated by a single component. This interpretation is illustrated by the change in the amplitude of the peaks seen in the raw ϕ spectra in Fig. 4(a), for the self-suppression experiment.

It might be expected that the slopes of the phase curves shown in Figs. 2(c) and 3(c) would be proportional to ϕ_C. However, this expectation is only partly borne out by the measured results. Although in Fig. 2(c) there is an overall reduction in the total change in the phase of the SFOAE as stimulus level increases, these reductions across level occur in jumps of whole numbers of cycles, indicating that a whole cycle has been ''lost.'' Where no jumps occur in the frequency interval, for example from 1800 to 2800 Hz in panel (c), the phase spectra run roughly parallel. This is because the slope of the phase curve is closely related to the interval between zero crossings, which is only weakly related to the ϕ_C value obtained by the four-parameter fit to the data.

These observations illustrate the difficulty in detecting changes in spectral period and may explain why some authors failed to report any such effects (Kemp and Chum, 1980a; Zwicker and Schloth, 1984; Zwicker, 1990; Dallmayr, 1987). These issues are discussed more fully in the two companion papers (Lineton and Lutman, 2003a, 2003b).

C. The effect of SOAEs and multiple reflections

The S/EOAE rating for each frequency sweep is indicated in Figs. 5–7. In Fig. 5, the y-shaped pattern appears to be unaffected by the presence of SOAEs. For example, a y-shaped pattern is seen for subjects 1 and 10 who have no SOAEs; for subjects 5, 6, and 20, for whom the majority of frequency sweeps has low S/EOAE ratings; and for subjects 4, 9, and 15, for whom a majority of frequency sweeps has high S/EOAE ratings. According to current theories, SOAEs, like SFOAEs, involve apical reflections of the TW from inhomogeneities on the BM. However, in the case of SOAEs, the product of the apical and basal reflection coefficients is unity at the SOAE frequency leading to a self-sustaining OAE (Talmadge and Tubis, 1993; Zweig and Shera, 1995). The explanation for the insensitivity of the y-shaped pattern to the presence of SOAEs may be the similarity of the two types of emission. However, since SOAEs were excluded from the models presented in Lineton and Lutman (2003a), further modeling is required to test this hypothesis.

Since SOAEs arise from multiple reflections of the TW, it is expected that ears with strong SOAEs will in general show higher values of α. This expectation is borne out by the measurements of α shown in Fig. 7. For example, subjects 1 and 10, who have no SOAEs, show values of α below 0.25 (except for one suspect point for subject 10), while the remaining subjects, whose frequency sweeps have either high or low S/EOAE ratings, show values of α above 0.3.

It is also expected that α will be linearly related to $\sigma_{\Lambda\Lambda}$ for a given ear, since both quantities are roughly proportional to the apical cochlear reflectance as discussed in the companion paper on parametric fitting (Lineton and Lutman, 2003b), which reduces during suppression. In general this expected linear relationship is seen in Fig. 7. Where, in Fig. 7, there is a significant deviation from expectations, this may be due to the relatively high variability of the estimates of α. Such deviations were also found in modeling studies (Lineton and Lutman, 2003a, 2003b). This is especially marked at low values of $\sigma_{\Lambda\Lambda}$ and α (e.g., subjects 1, 4, and 10). Note, however, that accurate estimates of α are not necessarily required in order to get accurate estimates of ϕ_C. This is particularly true for low values of α (<0.1), where the influence on the estimate of ϕ_C is negligible. In such cases, the effects

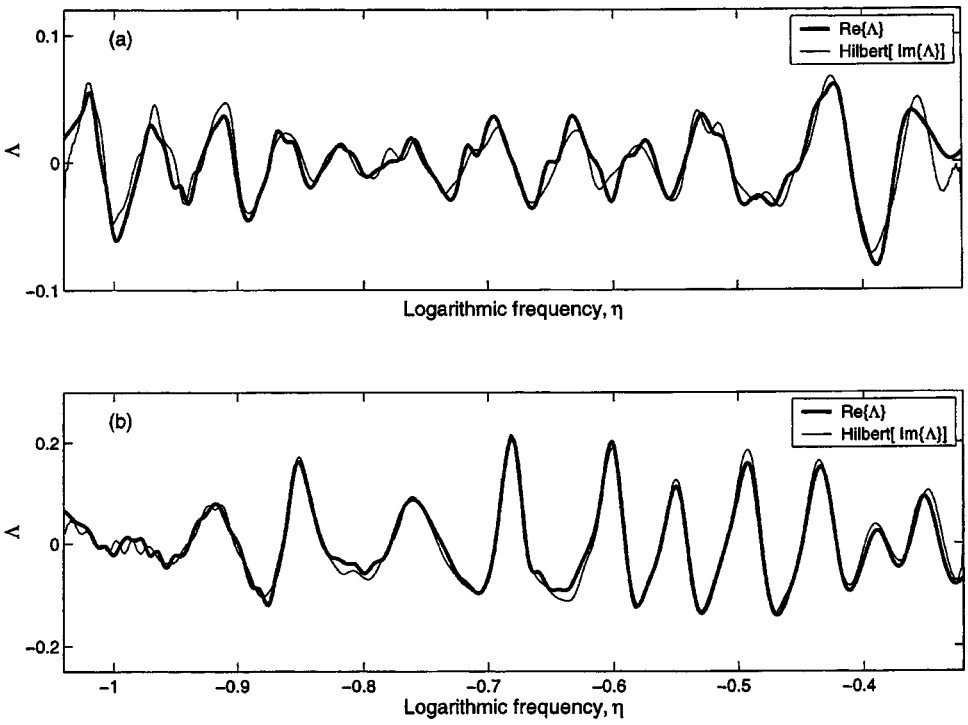


FIG. 8. Analytic properties of measured SFOAE frequency sweeps from (a) ear S1 and (b) ear S20. The thick line represents Re{Λ}; the thin line represents the Hilbert transform of Im{Λ}.

of multiple reflections on the SFOAE spectrum may be safely neglected.

The effect of a nonzero estimate of α on the shape of the fitted ϕ spectrum is to introduce additional peaks located at $\phi = n\phi_C$, where n is a whole number. This can be seen in Fig. 4(b), where several of the ϕ spectra show a second peak (or a points of inflection) at $\phi = 2\phi_C$.

D. Implications of the results

The fact that a reduction of SFOAE can be accompanied by either an increase or a decrease in spectral period may have implications for the interpretation of the various relationship between spectral period and mild cochlear hearing loss reported recently (Don *et al.*, 1998; Avan *et al.*, 2000; Lucertini *et al.*, 2002; Sisto and Moleti, 2002). It may be the case that the detailed pattern of the damage to the cochlea effects both the size and direction of any resulting changes in spectral period.

V. CONCLUSIONS

Measurements of SFOAE spectral period in normal-hearing adults show characteristic changes due to two forms of ipsilateral acoustic suppression: self-suppression and high-side two-tone suppression. During self-suppression, the spectral period was found to increase as SFOAE amplitude was suppressed. In contrast, during high-side two-tone suppression, the spectral period decreased as the amplitude was suppressed.

Both these measured results are in close agreement with the results from cochlear models presented in the companion paper (Lineton and Lutman, 2003a). The measured results thus lend strong support to the distributed roughness theory of SFOAE generation proposed by Shera and Zweig (1993a), and to its corollary that the SFOAE period is determined predominantly by the shape of the TW, and can therefore be altered by suppression. The measured results are inconsistent with predictions based on the hypothesis of a corrugated cochlea (Strube, 1989) that the SFOAE period is independent of TW shape.

The self- and two-tone suppression measurements support the theoretical finding (Lineton and Lutman, 2003a) that different patterns of cochlear amplifier disruption along the BM can lead to quite different changes in SFOAE spectral period.

An important implication of these results is that the spectral period of the SFOAE contains information regarding the shape of the TW. The results also have implications for the interpretation of data on spectral period obtained from ears with mild cochlear hearing loss.

ACKNOWLEDGMENTS

We wish to thank Torsten Marquardt, Amanda Hall, Sarosh Kapadia, Phil Nelson, and Bob Allen for their assistance and for many stimulating discussions. Christopher Shera and an anonymous reviewer also provided useful comments. The research was funded by the University of Southampton.

APPENDIX: USING THE ANALYTIC PROPERTY OF Λ TO OBTAIN AN IMPROVED ESTIMATE OF THE AUTOCOVARIANCE FUNCTION

Rather than the direct estimate of the autocovariance function, $\mathbf{C}_{\Lambda\Lambda}^{(\mathrm{raw})}$, in Eq. (4), a better estimate was achieved by making use of the *a priori* knowledge that $\Lambda(\eta)$ exhibits analyticity while the noise does not. This property constrains the real and imaginary parts of $\Lambda(\eta)$ to be related by the Hilbert transform. The justification for assuming this property comes both from the predictions of Zweig and Shera's theory (1995) and from the measurements themselves, where it was found that in cases where estimates of the signal-to-

noise ratio were high (>10 dB), $\Lambda(\eta)$ exhibited approximate analyticity (as illustrated by data from two subjects in Fig. 8). Analyticity has also been verified in a related measure of the SFOAE spectrum at low stimulus levels by Shera and Zweig (1993b).

To make use of the analytic property, the following function is first calculated:

$$\mathbf{C}^{(\mathrm{raw})}_{\Lambda_{\mathrm{Re}}\Lambda_{\mathrm{Im}}}(m) \equiv \frac{2}{N}\sum_{n=0}^{N-m-1} \Lambda_{\mathrm{Re}}(n)\tilde{\Lambda}_{\mathrm{Im}}(n+m)$$

$$m=0,1,2,...,N-1; \quad n=0,1,2,...,N-1;$$

$$N=512, \tag{A1}$$

where $\Lambda_{\mathrm{Re}}(n)$ is the real part of $\Lambda(n)$, and $\tilde{\Lambda}_{\mathrm{Im}}(n)$ is the Hilbert transform of the imaginary part of $\Lambda(n)$. The Hilbert transform, denoted by $\sim$, is given by

$$\tilde{\Lambda}_{\mathrm{Im}}(\eta) \equiv \mathbb{F}[\mathbb{F}^{-1}\{\Lambda_{\mathrm{Im}}(\eta)\}G(\phi)]$$

$$G(\phi)=\begin{cases} -i & \phi<0 \\ i & \phi>0 \\ 0 & \phi=0, \end{cases} \tag{A2}$$

where $\mathbb{F}$ and $\mathbb{F}^{-1}$ denote the Fourier transform and its inverse. The desired autocorrelation function can then be estimated using

$$\mathbf{C}^{(\mathrm{raw})}_{\Lambda\Lambda}(m) \equiv \mathbf{C}^{(\mathrm{raw})}_{\Lambda_{\mathrm{Re}}\tilde{\Lambda}_{\mathrm{Im}}}(m) - i\tilde{\mathbf{C}}^{(\mathrm{raw})}_{\Lambda_{\mathrm{Re}}\tilde{\Lambda}_{\mathrm{Im}}}(m). \tag{A3}$$

It can be shown that, in the absence of measurement noise, and on the assumption that $\Lambda(\eta)$ is an analytic signal, Eqs. (4) and (A3) yield identical values of $\mathbf{C}^{(\mathrm{raw})}_{\Lambda\Lambda}$. However, in the presence of noise, Eq. (A3) gives an improved estimate.

Avan, P., Wit, H. P., Guitton, M., Mom, T., and Bonfils, P. (**2000**). "On the spectral periodicity of transient-evoked otoacoustic emissions from normal and damaged cochleas," J. Acoust. Soc. Am. **108**, 1117–1127.

Brass, D., and Kemp, D. T. (**1993**). "Suppression of stimulus frequency otoacoustic emissions," J. Acoust. Soc. Am. **93**, 920–939.

Brown, A. M., Williams, P. M., and Gaskill, S. A. (**1993**). "The effect of aspirin on cochlear mechanical tuning," J. Acoust. Soc. Am. **93**, 3298–3307.

Collet, L., Veuillet, E., Moulin, A., Morlet, T., Giraud, A. L., Michyl, C., and Chéry-Croze, S. (**1994**). "Contralateral auditory stimulation and otoacoustic emissions: A review of basic data in humans," Br. J. Audiol. **28**, 213–218.

Dallmayr, C. (**1987**). "Stationary and dynamical properties of simultaneous evoked otoacoustic emissions (SEOAE)," Acustica **63**, 223–255.

Don, M., Ponton, C. W., Eggermont, J. J., and Kwong, B. (**1998**). "The effects of sensory hearing loss on cochlear filter times estimated from auditory brainstem response latencies," J. Acoust. Soc. Am. **104**, 2280–2289.

Giraud, A. L., Perrin, E., Chéry-Croze, S., Chays, A., and Collet, L. (**1996**). "Contralateral acoustic stimulation induces a phase advance in evoked otoacoustic emissions in humans," Hear. Res. **94**, 54–62.

Kanis, L. J., and de Boer, E. (**1993**). "Self-suppression in a locally active nonlinear model of the cochlea: A quasilinear approach," J. Acoust. Soc. Am. **94**, 3199–3206.

Kanis, L. J., and de Boer, E. (**1994**). "Two-tone suppression in a locally active nonlinear model of the cochlea," J. Acoust. Soc. Am. **96**, 2156–2165.

Karlsson, K. K., Berninger, E., and Alván, G. (**1991**). "The effect of quinine on psychoacoustical tuning curves, stapedius reflexes and evoked otoacoustic emissions in healthy volunteers," Scand. Audiol. **20**, 83–90.

Kemp, D. T. (**1986**). "Otoacoustic emissions, travelling waves and cochlear mechanisms," Hear. Res. **22**, 95–104.

Kemp, D. T., and Brown, A. M. (**1983**). "A comparison of mechanical nonlinearities in the cochleae of man and gerbil from ear canal measurements," in *Hearing—Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer, Berlin), pp. 82–88.

Kemp, D. T., and Chum, R. A. (**1980a**). "Observations on the generation mechanism of stimulus frequency acoustic emissions—two tone suppression," in *Psychophysical, Physiological and Behavioural Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen, *Proceedings of the 5th International Symposium on Hearing* (Delft University Press, Delft, The Netherlands), pp. 34–42.

Kemp, D. T., and Chum, R. A. (**1980b**). "Properties of the generator of stimulated acoustic emissions," Hear. Res. **2**, 213–232.

Kemp, D. T., Brass, D. N., and Souter, M. (**1990**). "Observations on simultaneous SFOAE and DPOAE generation and suppression," in *The Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthew, M. A. Ruggero, and C. R. Steele (Springer, Berlin), pp. 202–209.

Lind, O. (**1994**). "Contralateral suppression of TEOAE. Attempts to find a latency," Br. J. Audiol. **28**, 219–225.

Lineton, B., and Lutman, M. E. (**2003a**). "Modeling the effect of suppression on the periodicity of stimulus frequency otoacoustic emissions," J. Acoust. Soc. Am. **114**, 859–820.

Lineton, B., and Lutman, M. E. (**2003b**). "A parametric model of the spectral periodicity of stimulus frequency otoacoustic emissions," J. Acoust. Soc. Am. **114**, 883–895.

Long, G. R., and Tubis, A. (**1988**). "Modification of spontaneous and evoked otoacoustic emissions and associated psychoacoustic microstructure by aspirin consumption," J. Acoust. Soc. Am. **84**, 1343–1353.

Lucertini, M., Moleti, A., and Sisto, R. (**2002**). "On the detection of early cochlear damage by otoacoustic emission analysis," J. Acoust. Soc. Am. **111**, 972–978.

Norton, S. J., and Neely, S. T. (**1987**). "Tone-burst-evoked otoacoustic emissions from normal-hearing subjects," J. Acoust. Soc. Am. **81**, 1860–1872.

Shera, C. A., and Zweig, G. (**1993a**). "Order from chaos: Resolving the paradox of periodicity in evoked otoacoustic emissions," in *Biophysics of Hair-Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 54–63.

Shera, C. A., and Zweig, G. (**1993b**). "Noninvasive measurement of the cochlear traveling-wave ratio," J. Acoust. Soc. Am. **93**, 3333–3352.

Sisto, R., and Moleti, A. (**2002**). "On the frequency dependence of otoacoustic emission latency in hypoacoustic and normal ears," J. Acoust. Soc. Am. **111**, 297–308.

Strube, H. W. (**1989**). "Evoked otoacoustic emissions as cochlear Bragg reflections," Hear. Res. **38**, 35–46.

Sutton, G. J. (**1985**). "Suppression effects in the spectrum of evoked otoacoustic emissions," Acustica **58**, 57–63.

Talmadge, C. L., and Tubis, A. (**1993**), "On modelling the connection between spontaneous and evoked otoacoustic emissions," in *Biophysics of Hair-Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, S. M. van Netten (World Scientific, Singapore), pp. 25–32.

Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). "Modeling otoacoustic emission and hearing threshold fine structures," J. Acoust. Soc. Am. **104**, 1517–1543.

Talmadge, C. L., Tubis, A., Long, G. R., and Tong, C. (**2000**). "Modeling the combined effects of basilar membrane nonlinearity and roughness on stimulus frequency otoacoustic emission fine structures," J. Acoust. Soc. Am. **108**, 2911–2932.

Tavartkiladze, G. A., Frolenkov, G. I., Kruglov, A. V., and Artamasov, S. V. (**1994**). "Ipsilateral suppression effects of transiently evoked otoacoustic emission," Br. J. Audiol. **28**, 193–204.

Tognola, G., Grandori, F., and Ravazzani, P. (**1997**). "Time-frequency distributions of click-evoked otoacoustic emissions," Hear. Res. **106**, 112–122.

Wilson, J. P. (**1980**). "Evidence for a cochlear origin for acoustic re-emissions, threshold fine structure and tonal tinnitus," Hear. Res. **2**, 233–252.

Zweig, G., and Shera, C. A. (**1995**). "The origin of periodicity in the spectrum of evoked otoacoustic emissions," J. Acoust. Soc. Am. **98**, 2018–2047.

Zwicker, E. (**1990**). "On the frequency separation of simultaneously evoked otoacoustic emissions, consecutive extrema and its relation to cochlear traveling waves," J. Acoust. Soc. Am. **88**, 1639–1641.

Zwicker, E., and Schloth, E. (**1984**). "Interrelation of different otoacoustic emissions," J. Acoust. Soc. Am. **75**, 1148–1154.

Zwicker, E., and Wesel, J. (**1990**). "The effect of 'addition' in suppression of delayed evoked oto-acoustic emissions and in masking," Acustica **70**, 189–196.

A parametric model of the spectral periodicity of stimulus frequency otoacoustic emissions

Ben Lineton[a] and Mark E. Lutman
Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom

(Received 10 November 2002; revised 18 April 2003; accepted 23 April 2003)

A model for estimating the spectral period of stimulus frequency otoacoustic emissions (SFOAEs) is presented. The model characterizes the frequency spectrum of an SFOAE in terms of four parameters which can be directly related to cochlear mechanical quantities featuring in the theory of SFOAE generation proposed by Zweig and Shera [J. Acoust. Soc. Am. **98**, 2018–2047 (1995)]. The results of applying the parametric model to SFOAEs generated by cochlear models suggest that it gives a sensitive measure of spectral period. It is concluded that the parametric model may be a useful tool for detecting small changes in cochlear function using SFOAE measurements. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1582176]

PACS numbers: 43.64.Bt, 43.64.Jb, 43.64.Kc [BLM]

I. INTRODUCTION

Stimulus frequency otoacoustic emissions (SFOAEs) measured in humans exhibit a degree of regularity in the interpeak spacing of their spectra[1] (Kemp and Chum, 1980; Wilson, 1980; Zwicker and Schloth, 1984; Dallmayr, 1987; Zwicker, 1990; Lonsbury-Martin *et al.*, 1990; Zweig and Shera, 1995) such that the frequency interval between adjacent peaks is roughly 1/15 of the center frequency (Zweig and Shera, 1995; Talmadge *et al.*, 1998). This regularity, termed *spectral periodicity*, appears as an approximately constant interpeak spacing when the SFOAE spectrum is plotted against the logarithm of stimulus frequency (Zweig and Shera, 1995). This interpeak spacing, measured as an interval of the logarithm of frequency, is termed here the *spectral period*.

Current theories of the generation mechanisms of SFOAEs (Shera and Zweig, 1993a; Zweig and Shera, 1995) imply a close relationship between the sharpness of the peak of the cochlear traveling wave (TW) and the spectral period of the resulting SFOAE. This raises the possibility of investigating the properties of the TW by examining the spectral period of the SFOAE. However, the regularity of the measured SFOAE spectrum is far from perfect—an observation explained in the theory by the involvement of randomly distributed cochlear inhomogeneities. Consequently, the precise value of the period that is obtained for any given spectrum depends to some extent on the choice of signal-processing technique.

Two commonly used methods of calculating spectral period employ general signal-processing techniques, without reference to any underlying theory of SFOAE generation. In the first (e.g., Zwicker and Schloth, 1984; Dallmayr, 1987; Lonsbury-Martin *et al.*, 1990) the peak-to-peak frequency interval is extracted for each discernible spectral ripple (regardless of amplitude of the peak) and from these an average spectral period across the SFOAE spectrum is derived. In the second (e.g., Kemp and Chum, 1980; Wilson, 1980; Kemp and Brown, 1983; Kemp, 1986; Zweig and Shera, 1995), the spectral period is derived from the group delay, obtained from the slope of the phase spectrum. However, neither of these techniques was designed to maximize the correspondence between the calculated spectral period and the underlying TW characteristics.

In contrast, Zweig and Shera (1995, p. 2035) discuss theory-based methods. They show that the problem of estimating the TW characteristics from measured SFOAEs is similar to that encountered in estimating the spectra of random processes, and they outline three approaches to this problem: spectral smoothing, ensemble averaging (in this case across ears), and parametric modeling of the TW characteristics. In this third method, one of the TW parameters is directly related to spectral period.

The aim of the current paper is to develop a signal-processing technique for accurately estimating the spectral period of a single SFOAE spectrum, by adopting a parametric approach similar to that proposed by Zweig and Shera (1995). This technique, called here the four-parameter model, follows Burg (1978a, 1978b) in attempting to improve spectral resolution by maximizing the use of all available data. Unlike the parametric method proposed by Zweig and Shera (1995, p. 2035), the four-parameter model takes into account multiple reflections in the cochlea. The four-parameter model is used in two companion papers (Lineton and Lutman, 2003a, 2003b) to quantify the spectral period of both modeled and measured SFOAEs.

II. THEORY

A. Theoretical basis for the four-parameter model

The four-parameter model is based on models of three mechanical processes: the process of reflection of the forward TW by inhomogeneities distributed randomly along the basilar membrane (BM) (Shera and Zweig, 1993a; Zweig and Shera, 1995), the reflection of the backward TW at the

[a]Currently at: MRC Institute of Hearing Research, Royal South Hants Hospital, Southampton SO14 OYG, United Kingdom. Electronic mail: b.lineton@soton.ac.uk

stapes, and the transmission of acoustical power from the TW back out of the cochlea to the ear canal. The first of these processes can be characterized by the cochlear reflectance. The second and third are characterized by a two-port network model (Shera and Zweig, 1993b).

1. The cochlear reflectance

Zweig and Shera (1995) present a theory of coherent reflection filtering which accounts for the generation of SFOAEs with spectral periodicity. They also present a simplified phenomenological model based on a linear, scaling-symmetric cochlea, which yields a simple expression for the cochlear reflectance. This simplified model has been adopted as the basis for the four-parameter model. The form of the resulting cochlear reflectance is derived below.

In the simplified model, the natural frequency of the BM, $\omega_c(x)$, varies with cochlear place, x (measured from the base), according to the equation

$$\omega_C(x) = \omega_{\mathrm{Ref}} e^{-(x-x_{\mathrm{Ref}})/l}, \tag{1}$$

where ω_{Ref} is an arbitrary reference frequency (chosen here as $2\pi\times 1$ kHz), x_{Ref} is the location of the corresponding reference point, and l is the place-frequency mapping length. The BM response of the model is represented by the TW function, **T**, defined by

$$\mathbf{T}(\Theta) \equiv \frac{\mathbf{v_{BM}}(x,\omega)}{\mathbf{u_{St}}(\omega)}, \tag{2}$$

where $\mathbf{v_{BM}}$ is the BM velocity, $\mathbf{u_{St}}$ is the stapes velocity, ω is the stimulus frequency, and the independent variable, Θ, is defined as

$$\Theta(x,\omega) \equiv -\ln\left[\frac{\omega}{\omega_{\mathrm{Ref}} e^{-(x-x_{\mathrm{Ref}})/l}}\right]. \tag{3}$$

Note that throughout this paper, bold typeface is used to denote complex quantities. As a consequence of scaling symmetry, the right-hand side of Eq. (2), which is a function of both ω and x, can be expressed as a function of the single variable, Θ, on the left-hand side of the equation. This means, for example, that increasing the stimulus frequency causes a simple basalward shift in the pattern of excitation along the BM.

It is convenient to define the nondimensional spatial position, χ, and the logarithmic stimulus frequency, η, as

$$\chi \equiv (x-x_{\mathrm{Ref}})/l \qquad \eta \equiv -\ln(\omega/\omega_{\mathrm{Ref}}), \tag{4}$$

such that

$$\Theta(x,\omega) = \eta(\omega) - \chi(x). \tag{5}$$

For compatibility with measurements (Lineton and Lutman, 2003b), ω_{Ref} is taken as $2\pi\times 1$ kHz, such that x_{Ref} equals the location of the 1 kHz place.[2]

In the reflection hypothesis, Kemp (1978) proposed that OAEs originate from the reflection of the forward TW by BM inhomogeneities. In a linear cochlear model, this reflection can be characterized by the cochlear reflectance (defined as the complex ratio of the backward to the forward TWs), evaluated at the stapes (Shera and Zweig, 1993b; Talmadge *et al.*, 1998). By introducing inhomogeneities into a scaling-symmetric cochlea, and making a number of simplifying assumptions, Zweig and Shera (1995) demonstrate that the form of the cochlear reflectance, as a function of stimulus frequency, results directly from spatial filtering of the BM inhomogeneities by the TW. This can be represented by a convolution in the spatial domain

$$\mathbf{R}(\chi) = \rho_{Sc}(\chi) \otimes \mathbf{T}^2(\chi), \tag{6}$$

where **R** is the cochlear reflectance, ρ_{Sc} is called the scattering potential, $\otimes$ denotes convolution, and **T** is the TW function from Eq. (2), but evaluated at $\Theta=\chi$. From Eqs. (2)–(5), the function, $\mathbf{T}(-\chi)$, represents the BM response as a function of nondimensional place, χ, when the stimulus frequency equals the reference frequency. It follows that $\mathbf{T}(\chi)$ represents the spatially reversed TW. The function $\mathbf{T}^2(\chi)$ can be thought of as the impulse response function of a spatial filter which operates on an ''input signal'' given by the scattering potential. The ''output signal'' of this spatial filtering then equals the apical cochlear reflectance on the left-hand side of Eq. (6). To obtain the cochlear reflectance expressed as a function of logarithmic stimulus frequency, a further substitution of variable is made

$$\mathbf{R}(\eta) = \mathbf{R}(\chi)\big|_{\chi=\eta}. \tag{7}$$

In Zweig and Shera's theory (1995), the function, **T**, takes the form of a bandpass spatial filter, whose spatial center frequency is roughly proportional to the reciprocal of the TW wavelength near the peak of the TW. They hypothesized that the scattering potential, and hence the input signal in Eq. (6), takes the form of spatial broadband noise, arising from random, fine-grained irregularities in the cochlear micromechanics. Consequently, the cochlear reflectance will appear as a narrow-band random function of η. In this theory, the spectral period seen in SFOAEs, which arises as a consequence of the spatial filtering represented in Eq. (6), is determined by the wavelength of the TW near the TW peak.

2. Relating the SFOAE frequency spectrum to the cochlear reflectance

Since the cochlear reflectance is not directly accessible to measurement, the relationship between it and the ear-canal pressure must be derived from considerations of the response of the middle and outer ear, as reported by Kemp (1980) and Shera and Zweig (1993b).

The SFOAE pressure, $\mathbf{p_{SF}}$, can be defined in terms of the change in the ear-canal pressure arising from the reflection of the TW (Shera and Zweig, 1993b)

$$\mathbf{p_{SF}}(\eta) \equiv \mathbf{p_{EC}}(\eta) - \mathbf{p_{EC:R=0}}(\eta), \tag{8}$$

where $\mathbf{p_{EC}}$ is the ear-canal pressure measured for an apical cochlear reflectance of **R**, and $\mathbf{p_{EC:R=0}}$ is the ear-canal pressure that would have been measured if the cochlear reflectance were zero. This latter condition can be approached in OAE measurements either by using a high stimulus level (Kemp and Chum, 1980), or a suppressor tone (Kemp and Brown, 1983; Kemp *et al.*, 1990) to minimize the effect of the cochlear amplifier.

Using a two-port network model of the middle and outer ear, coupled to the cochlear input impedance which is characterized in terms of the cochlear reflectance and nonreflecting impedance, Shera and Zweig (1993b) show that

$$\Lambda = \frac{\mathbf{gR}}{1-\mathbf{rR}} \approx \mathbf{gR}(1+\mathbf{rR}+(\mathbf{rR})^2+\cdots), \tag{9}$$

where Λ is the normalized SFOAE pressure, defined by

$$\Lambda \equiv \frac{\mathbf{p_{SF}}}{\mathbf{p_{EC:R=0}}}. \tag{10}$$

The functions **g** and **r** are related to the characteristics of the middle ear, ear canal, and probe. All terms are complex functions of logarithmic frequency, η. The binomial expansion for Λ shows how the SFOAE can be represented as the sum of multiple reflections within the cochlea. The quantity, **r** (termed the basal reflectance), is the reflectance at the stapes as seen by a backward TW leaving the cochlea, while **R** (the apical cochlear reflectance) is the reflectance due to the BM inhomogeneities encountered by a forward TW. These multiple backward waves sum in the ear canal to give the measured OAE. The magnitude of **R** determines the size of the first reflection, while the product **rR** determines how significant multiple reflections are.

The functions Λ, **g**, and **r** are the same as the functions denoted Δ, p, and q, respectively, in Shera and Zweig (1993b). This change in nomenclature was forced by conflicts elsewhere in this paper and in Lineton and Lutman (2003a, 2003b). Λ was chosen for its resemblance to Δ.

B. A stochastic model of the SFOAE spectrum

To derive a stochastic model of the SFOAE spectrum, the distributed roughness term, $\rho_{Sc}(\chi)$ and the SFOAE frequency spectrum, $\Lambda(\eta)$, must be treated as random processes. For simplicity, it is assumed that $\rho_{Sc}(\chi)$ is a stationary, ergodic, zero-mean, Gaussian, white-noise process in the spatial domain. Since, from standard theory (e.g., Bendat and Piersol, 1966), the linear filtering operation of any stationary, Gaussian random process yields another stationary, Gaussian random process, the cochlear reflectance, $\mathbf{R}(\eta)$, is also stationary and Gaussian. Such a random process is fully characterized by its autocovariance function, or equivalently, by its power spectral density function.

The autocovariance function of the cochlear reflectance is defined by

$$\mathbf{C}_{RR}(\eta') \equiv E[\mathbf{R}^*(\eta)\mathbf{R}(\eta+\eta')], \tag{11}$$

where * denotes the complex conjugate and $E[\ldots]$ denotes the expectation operator. The assumption of stationarity ensures that $\mathbf{C}_{RR}(\eta')$ depends only of the lag-η variable, denoted η', rather than the absolute value of η, and can therefore be calculated from

$$\mathbf{C}_{RR}(\eta') = \lim_{\eta_1\to\infty} \frac{1}{2\eta_1} \int_{-\eta_1}^{\eta_1} \mathbf{R}^*(\eta)\mathbf{R}(\eta+\eta')\,d\eta. \tag{12}$$

Since $\rho_{Sc}(\chi)$ is assumed to be a white-noise process, its autocovariance function is fully characterized by a single parameter: the amplitude of the power spectral density amplitude, ρ_0^2. From this, and by substituting Eq. (6) into Eq. (12), it follows that the autocovariance function of the cochlear reflectance is given by

$$\mathbf{C}_{RR}(\eta') = \lim_{\eta_1\to\infty} \frac{\rho_0^2}{2\eta_1} \int_{-\eta_1}^{\eta_1} [\mathbf{T}^2(\eta)]^* \mathbf{T}^2(\eta+\eta')\,d\eta. \tag{13}$$

The nature of the spectral periodicity in $\mathbf{R}(\eta)$ is more clearly revealed by inverse Fourier transforming Eq. (13), giving

$$S_{RR}(\phi) \equiv \mathbb{F}^{-1}\{\mathbf{C}_{RR}(\eta')\}, \tag{14}$$

where $S_{RR}(\phi)$ is termed the ϕ spectrum of the cochlear reflectance and where the inverse Fourier transform from η to ϕ is defined by

$$\mathbb{F}^{-1}\{\mathbf{A}(\eta)\} \equiv \int_{-\infty}^{\infty} \mathbf{A}(\eta)\exp(i2\pi\phi\eta)\,d\eta. \tag{15}$$

Since $\mathbf{T}^2(\chi)$ is an analytic function (Zweig and Shera, 1995) and $\mathbf{C}_{RR}(-\eta') = \mathbf{C}_{RR}^*(\eta')$, $S_{RR}(\phi)$ is both a right-sided and a purely real function [i.e., $S_{RR}(\phi)=0$ for all $\phi<0$]. The quantity, ϕ, can be interpreted as the envelope delay of a single frequency component, expressed as the number of periods of the carrier frequency (Zweig and Shera, 1995).

From standard theory it can be shown that manipulation of Eqs. (13)–(15) yields

$$S_{RR}(\phi) = \rho_0^2 |\mathbb{F}^{-1}\{\mathbf{T}^2(\eta)\}|^2. \tag{16}$$

Thus, the shape of the ϕ spectrum of the cochlear reflectance is determined purely by the TW function. As Zweig and Shera point out, $S_{RR}(\phi)$ takes the form of a pulse-shaped function which peaks at a location determined by the wavelength of the TW. The location of the peak of $S_{RR}(\phi)$ is termed here the ϕ-center value, denoted ϕ_C, and will be used to characterize the spectral period. This corresponds closely with the quantity denoted $\hat{\varphi}/2\pi$ by Zweig and Shera (1995), and is roughly the reciprocal of the quantity defined as $\Delta f/f$ by previous authors (Zwicker and Schloth, 1984; Dallmayr, 1987). In humans, ϕ_C has a value of about 15 (Zweig and Shera, 1995).

While **R** has the simple stochastic description given by Eq. (16), a stochastic description of Λ is less straightforward for two reasons. First, since the relationship shown in Eq. (9) between Λ and **R** is nonlinear, it follows that Λ is a non-Gaussian random process, and therefore this process cannot be fully characterized by its autocovariance function. Second, the functions $\mathbf{g}(\eta)$ and $\mathbf{r}(\eta)$ lead to deterministic variations in Λ, thereby violating stationarity. Thus, to obtain a stochastic model of Λ, two further simplifying assumptions are required.

First, since the functions $\mathbf{g}(\eta)$ and $\mathbf{r}(\eta)$ vary only slowly with η, it is assumed that they are approximately constant over the measured frequency interval (roughly one octave), such that

$$\mathbf{g}(\eta) = \mathbf{g}_0, \quad \mathbf{r}(\eta) = \mathbf{r}_0. \tag{17}$$

Second, it is further assumed that $\mathbf{r}_0\mathbf{R}(\eta) \ll 1$ for all η over the measured range, such that Eq. (9) can be represented as a truncated power series

$$\Lambda = \frac{\mathbf{g}_0\mathbf{R}}{1-\mathbf{r}_0\mathbf{R}} \approx \mathbf{g}_0\mathbf{R}(1+\mathbf{r}_0\mathbf{R}+\mathbf{r}_0^2\mathbf{R}^2). \tag{18}$$

It is shown in Appendix B that, assuming that Eq. (18) holds, and that $\mathbf{R}$ arises from a Gaussian stationary random process, the following relationship also holds:

$$\mathbf{C}_{\Lambda\Lambda}(\eta') \approx g_0^2\mathbf{C}_{RR}(\eta')[1+2r_0^2\mathbf{C}_{RR}(\eta') + 6r_0^4\mathbf{C}_{RR}^2(\eta')], \tag{19}$$

where $\mathbf{C}_{\Lambda\Lambda}(\eta')$ is the autocovariance function of Λ, defined by

$$\mathbf{C}_{\Lambda\Lambda}(\eta') \equiv \lim_{\eta_1\to\infty} \frac{1}{2\eta_1}\int_{-\eta_1}^{\eta_1} \Lambda^*(\eta)\Lambda(\eta+\eta')d\eta, \tag{20}$$

and

$$g_0^2 \equiv \mathbf{g}_0^*\mathbf{g}_0, \quad r_0^2 \equiv \mathbf{r}_0^*\mathbf{r}_0. \tag{21}$$

The inverse Fourier transform of the autocovariance function yields the ϕ spectrum of Λ

$$S_{\Lambda\Lambda}(\phi) \equiv \mathbb{F}^{-1}\{\mathbf{C}_{\Lambda\Lambda}(\eta')\}. \tag{22}$$

The presence of the nonlinear terms in Eq. (19) complicates the form of $S_{\Lambda\Lambda}(\phi)$. To illustrate this, taking only the first two terms in Eq. (19), it can be shown that the ϕ spectrum of Λ is related to the ϕ spectrum of the $\mathbf{R}$ by

$$S_{\Lambda\Lambda}(\phi) \approx g_0^2 S_{RR}(\phi) + g_0^2 r_0^2[S_{RR}(\phi)\otimes S_{RR}(\phi)], \tag{23}$$

where $\otimes$ denotes convolution in ϕ. Higher-order terms in Eq. (19) lead to multiple convolutions. Thus, $S_{\Lambda\Lambda}(\phi)$ comprises a series of pulse-shaped functions, the first of which has the same shape as $S_{RR}(\phi)$, and is centered at ϕ_C.

C. The four-parameter model of the SFOAE spectrum

The problem now addressed is how to obtain estimates of ϕ_C from real SFOAE measurements. On the assumption that Eq. (19) holds, the problem becomes one of estimating the autocovariance function from only a finite section of the signal, $\Lambda(\eta)$. This is analogous to the problem of spectral estimation from a finite signal. A common method of spectral estimation involves splitting the signal up into a number of segments, Fourier transforming each segment, and then averaging the power spectra across segments (e.g., Randall, 1987). However, an alternative approach known as parametric spectral estimation is to assume that the section of measured signal that is available has arisen from a white-noise input to a certain class of filter whose coefficients are initially unknown (e.g., Burg, 1978a, 1978b). The problem then becomes one of estimating the filter coefficients. Relative to methods based on Fourier transformation, such parametric methods can lead to improved spectral resolution. There are three reasons for this. First, *a priori* knowledge of the class of filter is utilized. Second, unlike Fourier methods, there is no need to distort the data using a tapering window: all the available data points are treated equally. Third, unlike Fourier methods, where the application of the window amounts to the assumption that unavailable data points are zero, no assumptions are made about unavailable data (i.e., data outside the measured interval).

A parametric method is presented here which fits the stochastic model of the signal in Eq. (20) using four parameters. In this method, it is first assumed that the reflectance is given by

$$\mathbf{R}(\eta) = \mathbf{h_{TW}}(\eta)\otimes w(\eta), \tag{24}$$

where $w(\eta)$ is an unknown stationary, Gaussian white-noise signal, and $\mathbf{h_{TW}}(\eta)$ is the impulse response function of second-order Butterworth bandpass filter. Equation (24) is an idealization of Eq. (6) in which the square of the TW function, $\mathbf{T}^2(\chi)$, is replaced by a known class of filter, the scattering potential is replaced with an ideal white-noise process, and the independent variable χ is replaced with η (by invoking scaling symmetry).

Replacing $\mathbf{T}^2(\chi)$ and $\boldsymbol{\rho}_{\mathbf{Sc}}(\chi)$ in Eqs. (12) and (13) with $\mathbf{h_{TW}}(\eta)$ and $w(\eta)$ gives

$$\mathbf{C_{RR}}(\eta') = S_{ww}\mathbf{C_{hh}}(\eta'), \tag{25}$$

where S_{ww} is the power spectral density of $w(\eta)$, and $\mathbf{C_{hh}}$ is a function defined as

$$\mathbf{C}_{hh}(\eta') \equiv \int_{-\infty}^{\infty} \mathbf{h}_{\mathbf{TW}}^*(\eta)\mathbf{h_{TW}}(\eta+\eta')d\eta. \tag{26}$$

To eliminate the unknown quantity, S_{ww}, Eq. (25) is normalized to give

$$\mathbf{K}_{RR}(\eta') = \mathbf{K}_{hh}(\eta'), \tag{27}$$

where $\mathbf{K}$ is known as the autocorrelation function, obtained by dividing the autocovariance function by the mean-square value. Thus, for the left-hand side of Eq. (25)

$$\mathbf{K}_{RR}(\eta') \equiv C_{RR}(\eta')/\sigma_{RR}^2, \tag{28}$$

where σ_{RR}^2 is the mean-square value of the reflectance, given by

$$\sigma_{RR}^2 \equiv \mathbf{C}_{RR}(\eta')\big|_{\eta'=0}. \tag{29}$$

Unlike the general parametric spectral estimation method (Burg, 1978a, 1978b), where a very general form of digital filter is assumed with many unknown filter coefficients, it was decided here to assume that the filter, $\mathbf{h_{TW}}(\eta)$, can be approximated by a one-sided version of a four-pole Butterworth bandpass filter. This has only two free parameters: the center frequency and bandwidth. From consideration of Eqs. (6) and (16) it can be seen that the center frequency is identical to ϕ_C, the reciprocal of spectral period. The bandwidth of the filter is denoted ϕ_{BW}.

The shape of the filter was chosen somewhat arbitrarily, but appears to give a reasonable fit to both measured and simulated SFOAEs (Sec. III). The final estimate of ϕ_C is relatively insensitive to the precise shape of the filter.

The transfer function of a standard (two-sided) second-order Butterworth bandpass filter is represented in the ϕ domain by

$$\mathbf{H_{Bworth}}(\phi) = \frac{1}{\left(\frac{\phi_C^2-\phi^2}{i\phi\phi_{BW}}\right)^2 + \sqrt{2}\left(\frac{\phi_C^2-\phi^2}{i\phi\phi_{BW}}\right)+1}. \tag{30}$$

The Fourier transform of the one-sided version of this yields the impulse response function

$$\mathbf{h_{TW}}(\eta)=\mathbb{F}\{\mathbf{H_{Bworth}}(\phi)U(\phi)\}, \tag{31}$$

where $U(\phi)$ is the unit step function.

Substituting Eq. (28) into Eq. (19) then allows $\mathbf{C}_{\Lambda\Lambda}(\eta')$, to be related to $\mathbf{K_{RR}}(\eta')$ by

$$\mathbf{C}_{\Lambda\Lambda}(\eta')\approx\beta^2\mathbf{K}_{RR}(\eta')[1+2\alpha^2\mathbf{K}_{RR}(\eta') +6\alpha^4\mathbf{K}_{RR}(\eta')], \tag{32}$$

where the parameters α and β are given by

$$\alpha^2\equiv r_0^2\sigma_{RR}^2, \quad \beta^2\equiv g_0^2\sigma_{RR}^2. \tag{33}$$

The parameter α thus characterizes the effective value of the product $|\mathbf{r}(\eta)\mathbf{R}(\eta)|$, taken across the available range of η. From Eq. (32) evaluated at $\eta'=0$, the parameter β is related to the rms amplitude of Λ via

$$\beta^2\approx\frac{\sigma_{\Lambda\Lambda}^2}{1+2\alpha^2+6\alpha^4}. \tag{34}$$

Thus, given estimates of four independent parameters: ϕ_C, ϕ_{BW}, α, and $\sigma_{\Lambda\Lambda}$, an estimate of the autocovariance function of Λ can be calculated from Eqs. (26)–(34). The function obtained in this way from estimates of the four parameters is termed the *fitted* autocovariance function, and denoted $\mathbf{C}_{\Lambda\Lambda}^{(\mathrm{fit})}(\eta')$.

Given $\Lambda(\eta)$ over the interval η_1 to η_2, these four parameters are estimated from the following procedure. First, the autocovariance function of Λ is estimated from

$$\mathbf{C}_{\Lambda\Lambda}^{(\mathrm{raw})}(\eta')\equiv\frac{1}{\eta_2-\eta_1}\int_{\eta_1}^{\eta_2-\eta'}\Lambda^*(\eta)\Lambda(\eta+\eta')d\eta \tag{35}$$

for $0<\eta'<\eta_2-\eta_1$, where the form of the denominator has been chosen to yield the commonly used biased estimator (Jenkins and Watts, 1968). The superscript *raw* here indicates that the function has been estimated from a single measured section of the Λ spectrum.

The parameter, $\sigma_{\Lambda\Lambda}$, can be estimated directly by evaluating $\mathbf{C}_{\Lambda\Lambda}^{(\mathrm{raw})}(\eta')$ at $\eta'=0$. The remaining three parameters ϕ_C, ϕ_{BW}, and α, are then estimated iteratively as follows. From initially arbitrary estimates of these three parameters, an initial estimate of $\mathbf{C}_{\Lambda\Lambda}^{(\mathrm{fit})}(\eta')$ is calculated via Eqs. (26) to (34). The three parameters are then altered iteratively to minimize the mean-square error defined as

$$\varepsilon^2(\alpha,\phi_{BW},\phi_C)\equiv\frac{1}{\sigma_{\Lambda\Lambda}^2\eta'_{\mathrm{Lim}}}\int_0^{\eta'_{\mathrm{Lim}}}|\mathbf{C}_{\Lambda\Lambda}^{(\mathrm{raw})}(\eta') -\mathbf{C}_{\Lambda\Lambda}^{(\mathrm{fit})}(\eta')|^2d\eta'. \tag{36}$$

Rather than use all available nonzero values of the $\mathbf{C}_{\Lambda\Lambda}^{(\mathrm{raw})}(\eta')$, the integral in Eq. (36) is calculated up to a limiting value of η', denoted η'_{Lim}. The reason for doing this is that, given the assumed form of the bandpass filter, only a limited section of this function is required in order to completely specify the filter. This is addressed further in Appendix C, where the relationship between the filter parameters and the $\mathbf{C}_{\Lambda\Lambda}^{(\mathrm{fit})}(\eta')$ is derived. A second reason for limiting the integration is that the estimate of the $\mathbf{C}_{\Lambda\Lambda}(\eta')$ given by $\mathbf{C}_{\Lambda\Lambda}^{(\mathrm{raw})}(\eta')$ becomes increasingly unreliable as η' increases towards $\eta_2-\eta_1$. In fact, it was found that the choice of η'_{Lim} does not greatly influence the values of the four parameters obtained provided it lies between 0.1 and 0.5 of $\eta_2-\eta_1$.

Inverse Fourier transforming the raw and fitted autocovariance functions, as in Eq. (14), yields estimates of the ϕ spectrum, termed the *raw* and *fitted* ϕ spectra, respectively.

III. APPLICATION OF FOUR-PARAMETER MODEL

A. Application of the four-parameter model to simulated SFOAEs

1. Ensemble averaged data

The applicability of the four-parameter model to simulated SFOAEs obtained from a number of cochlear mechanical models, described in Lineton and Lutman, 2003a, has been assessed. Though these cochlear models share the same theoretical elements as those that underpin the four-parameter model presented above, they violate several of the simplifying assumptions. First, the effective shape of the spatial filter is not specified as a Butterworth filter, but is rather more complicated. Second, the models are not constrained to exhibit perfect scaling symmetry. Finally, the dynamics of the middle and outer ears in these models must be represented by Eq. (9) including the functions $\mathbf{g}(\eta)$ and $\mathbf{r}(\eta)$ rather than the simplified version in Eq. (18), thus leading to a further deviation from stationarity. As a consequence of these violations, the four-parameter model will not provide a perfect model of the stochastic process underlying $\Lambda(\eta)$.

To assess the fit between the parametric model and the cochlear model, first a good estimate of the autocovariance function, $\mathbf{C}_{\Lambda\Lambda}$, of the cochlear models must be obtained. This was achieved by using a random number generator to produce a number of different realizations of the scattering impedance, then running the cochlear model for each realization over the interval η_1 to η_2, and thereby generating a

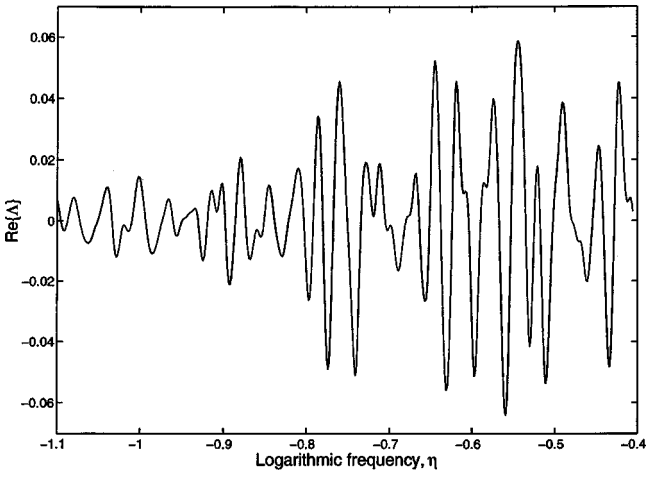


FIG. 1. Simulated SFOAE frequency sweep from a cochlear model. The model, reported by Lineton and Lutman (2003a), is based on Zweig and Shera's theory (Zweig and Shera, 1995) with a random distribution of inhomogeneities along the basilar membrane. A single realization of $\Lambda(\eta)$ is presented. (Adapted from Lineton and Lutman, 2003a, Fig. 2.)

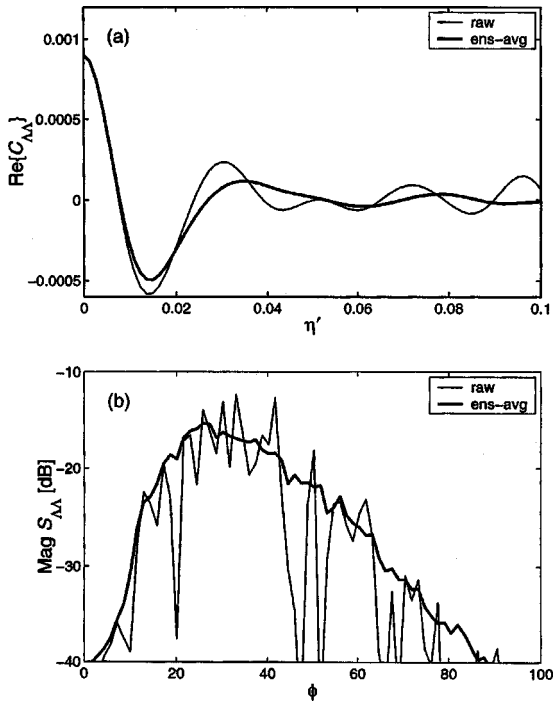


FIG. 2. Raw and ensemble-averaged functions obtained from a cochlear model. In (a) the real part of the autocovariance function is plotted against the logarithmic frequency lag variable. The raw autocovariance function (thin line) is obtained from the single realization of the normalized SFOAE spectrum, $\Lambda(\eta)$, shown in Fig. 1. The ensemble-averaged autocovariance function (thick line) is obtained by averaging across an ensemble of 32 realizations of the raw autocovariance function. In (b) the raw (thin line) and ensemble-averaged (thick line) ϕ spectra, defined as the Fourier transform of the corresponding autocovariance functions, are plotted. (Adapted from Lineton and Lutman, 2003a, Figs. 3 and 4.)

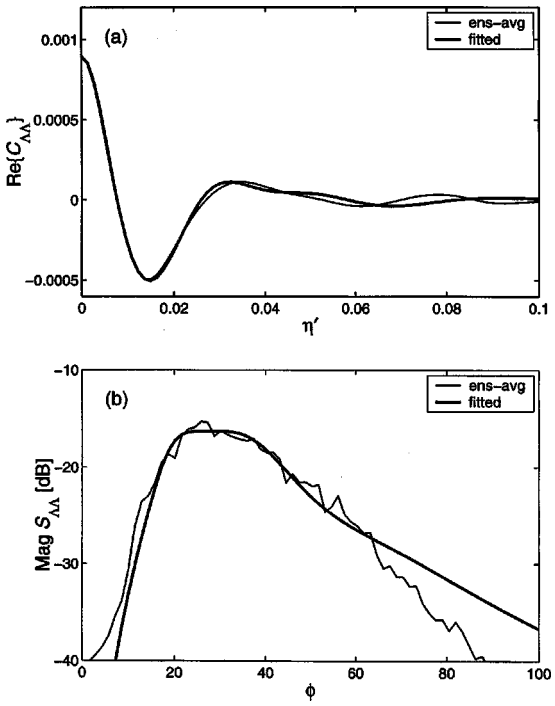


FIG. 3. Ensemble-averaged and fitted functions obtained from a cochlear model. In (a) the real part of the autocovariance function is plotted against the logarithmic frequency lag variable. The ensemble-averaged autocovariance function (thin line) is replotted from Fig. 2 for comparison. The fitted autocovariance function (thick line) is obtained by applying the four-parameter model to the ensemble-averaged autocovariance function. In (b) the ensemble-averaged (thin line) and fitted (thick line) ϕ spectra, defined as the Fourier transform of the corresponding autocovariance functions, are plotted. (Adapted from Lineton and Lutman, 2003a, Figs. 3 and 4.)

number of realizations of the SFOAE spectra, $\Lambda(\eta)$. A single realization of $\Lambda(\eta)$ is illustrated in Fig. 1. For each realization of $\Lambda(\eta)$, $\mathbf{C}_{\Lambda\Lambda}^{(\mathrm{raw})}(\eta')$ is calculated from Eq. (35). Averaging these functions across the ensemble of realizations then yields a reliable estimate of the true autocovariance function, $\mathbf{C}_{\Lambda\Lambda}$, for the cochlear model. This improved estimate is termed the ensemble averaged autocovariance function, $\langle \mathbf{C}_{\Lambda\Lambda}^{(\mathrm{raw})}(\eta') \rangle_{\mathrm{ensemble}}$. [Note that strictly, since the signal is nonstationary, the autocorrelation function has two independent variables: the absolute value of η and the lag value, η', as in the right-hand side of Eq. (11). Use of Eq. (35) to calculate a raw autocovariance function depending only on η' yields a value of the true two-dimensional function averaged over the range η_1 to η_2.] An example of the raw and ensemble averaged autocorrelation functions and the corresponding ϕ spectra is shown in Fig. 2.

The four-parameter model was then applied to $\langle \mathbf{C}_{\Lambda\Lambda}^{(\mathrm{raw})}(\eta') \rangle_{\mathrm{ensemble}}$ to give the fitted ensemble autocovariance function, $\mathbf{C}_{\Lambda\Lambda}^{(\mathrm{ens\text{-}fit})}(\eta')$. Comparing $\mathbf{C}_{\Lambda\Lambda}^{(\mathrm{ens\text{-}fit})}(\eta')$ and $\langle \mathbf{C}_{\Lambda\Lambda}^{(\mathrm{raw})}(\eta') \rangle_{\mathrm{ensemble}}$ in Fig. 3 shows that the four-parameter model provides a good approximation to the random process underlying the generation of $\Lambda(\eta)$ by the cochlear models in this example.

The four-parameter model has also been found to provide a good approximation to a number of other cochlear models with different cochlear amplifier models. An example of the results for a cochlear model with high reflectances both basally and apically is illustrated in Fig. 4. Panel (b) clearly shows the second peak arising from multiple reflections, and leading to a high value of α in the four-parameter model.

From consideration of Eq. (18), it can be seen as $|\mathbf{R}|$ is progressively reduced (for example by suppression) that multiple reflections become less important and Λ becomes dominated by the first reflection. Consequently [as is verified by Eq. (34)], according to the parametric model, α is approximately proportional to $\sigma_{\Lambda\Lambda}$ for small values of α. This expected relationship is seen in the four-parameter fit to data from cochlear models in which $|\mathbf{R}|$ has been progressively reduced by attenuating the cochlear amplifier gain (Fig. 5).

2. Single realizations

As well as being applicable to ensemble averaged data, it is also important that the four-parameter model can be usefully applied to a single realization of Λ. To assess its performance as an estimator for single realizations for a given cochlear model, the four-parameter model was applied to each realization of Λ in a large ensemble to yield the sampling distributions of each of its four estimated parameters. The results (Table I) demonstrate that estimates of ϕ_C and $\sigma_{\Lambda\Lambda}$ are much more reliable than those of ϕ_{BW} and α. Estimates of α for cochlear model 1 are particularly poor. This is because the variance of α does not decrease with the mean value, and thus when the true value of α is quite low

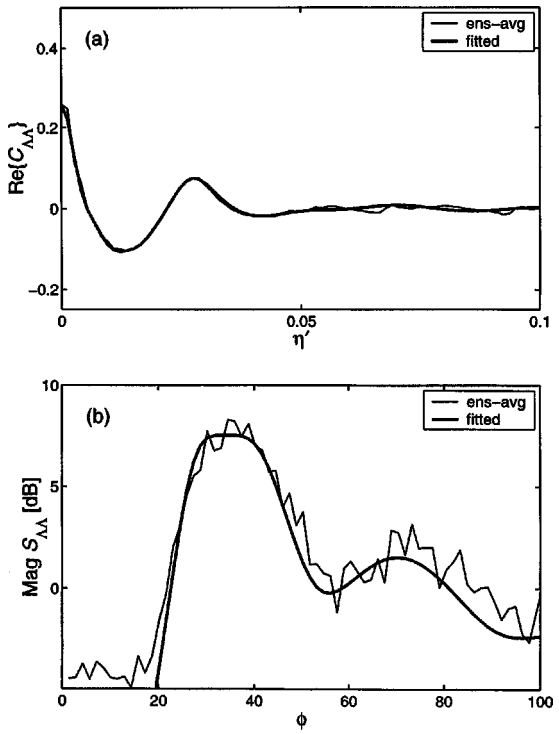


FIG. 4. Ensemble-averaged and fitted functions obtained from a cochlear model showing significant multiple reflections of the TW. To exaggerate the effect of multiple reflections, the cochlear model was specified such that both the basal and apical reflectances were large. These were achieved by setting a large basal impedance and a large scattering impedance, respectively. In (a) the real part of the autocovariance function is plotted against the logarithmic frequency lag variable. The ensemble-averaged autocovariance function (thin line) is obtained by averaging across an ensemble of 32 realizations of the raw autocovariance function. The fitted autocovariance function (thick line) is obtained by applying the four-parameter model to the ensemble-averaged autocovariance function. In (b) the ensemble-averaged (thin line) and fitted (thick line) ϕ spectra, defined as the Fourier transform of the corresponding autocovariance functions, are plotted. The effect of multiple reflections is to cause the second peak clearly seen in the fitted ϕ spectrum.

(0.13 in this case), the relative error increases. Note that for values of $\alpha<0.1$, the effect of α on the other three parameters is negligible.

The performance of the four-parameter model for simulated SFOAEs was also tested in the presence of artificial additive white noise, where it was found that the estimate of ϕ_C changed by less than 1% for signal-to-noise ratios of 0 dB and greater.

B. Comparison of the four-parameter model with other estimators

Three different methods for estimating the spectral period of SFOAEs are discussed here. The performance of these methods has been assessed on the assumption that the SFOAE spectrum arises from the random process described by Eqs. (6) and (9). When averaged across a large ensemble of SFOAE spectra, all three methods, together with the four-parameter method, yield very similar results. However, their sampling distributions differ considerably, with the four-parameter model showing the smallest rms error in its estimate.

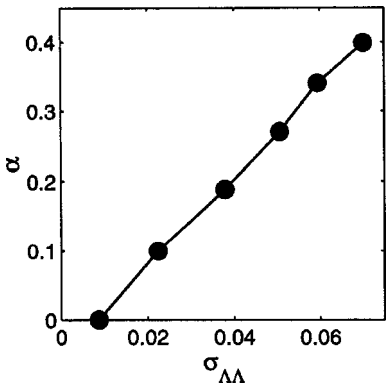


FIG. 5. Variation of α with $\sigma_{\Lambda\Lambda}$ for Zweig–Talmadge cochlear amplifier variant model described in Lineton and Lutman (2003a) with cochlear amplifier attenuations of 0, −0.1, −0.2, −0.4, −0.8, −1.6 dB. Parameters were estimated from the ensemble-averaged autocovariance function obtained from 256 realizations.

In the first method, the spectral period is defined by the frequency spacing between adjacent peaks in the SFOAE spectrum (Kemp and Chum, 1980; Wilson, 1980; Dallmayr, 1987; Zwicker and Schloth, 1984; Zwicker, 1990; Lonsbury-Martin *et al.*, 1990). The drawback with this method is that all frequency intervals are given the same weight, regardless of the ripple amplitude. The observed ripple spacing of a single realization then largely depends on the spectral period of a single dominant component and is little affected by the full distribution of spectral period components present in the signal. Thus, despite changes in the true spectral period, the SFOAE may show no significant change in ripple spacing (see, for example, the experimental data in Lineton and Lutman, 2003a, Fig. 2).

TABLE I. The sampling distributions of the estimates from the four-parameter model and the Fourier transform method for SFOAEs from two cochlear models. Sampling distributions were obtained from an ensemble of 256 realizations.

		Fourier[a]	Four-parameter model			
		ϕ_C[b]	ϕ_C	ϕ_{BW}[c]	α[c]	$\sigma_{\Lambda\Lambda}$[c]
Model 1[d]	True value	27.4	27.4	26.2	0.130	0.0291
	Bias error (%)[e]	9	1	−18	35	−6
	Random error (%)[f]	24	14	33	110	14
	rms error (%)[g]	26	14	38	115	15
Model 2[d]	True value	34.8	34.8	19.9	0.399	0.0701
	Bias error (%)	4	1	−14	−19	−7
	Random error (%)	14	8	29	18	15
	rms error (%)	15	8	32	25	17

[a]In the Fourier transform method, ϕ_C is estimated from the location of the peak of $\mathbb{F}^{-1}\{\Lambda(\eta)w(\eta)\}$, where $w(\eta)$ is a Hanning window.
[b]The true value of ϕ_C was obtained from the location of the peak of $\mathbb{F}^{-1}\{\langle \mathbf{C}_{\Lambda\Lambda}^{(\mathrm{raw})}(\eta')\rangle_{\mathrm{ensemble}}\}$.
[c]The true values of ϕ_{BW}, α, and $\sigma_{\Lambda\Lambda}$ were obtained by applying the four-parameter model to $\langle \mathbf{C}_{\Lambda\Lambda}^{(\mathrm{raw})}(\eta')\rangle_{\mathrm{ensemble}}$.
[d]Models 1 and 2 are the linear baseline and the Zweig–Talmadge cochlear amplifier variant models described in Lineton and Lutman (2003a).
[e]Bias error is defined by the mean of the sampling distribution minus the true value.
[f]Random error is defined as the standard deviation of the sampling distribution.
[g](rms error)^2=(bias error)^2+(random error)^2. All errors are expressed as a percentage of the true value (to nearest 1%).

A second method is to base an estimate of spectral period on the slope of the phase of the SFOAE (e.g., Kemp and Chum, 1980; Wilson, 1980; Kemp and Brown, 1983; Talmadge *et al.*, 2000, Fig. 4). This is analogous to calculating the group delay of the SFOAE measured in cycles rather than units of time. On average, ϕ_C is related to the slope of the phase of $\Lambda(\eta)$ by

$$-\frac{1}{2\pi}\frac{d\theta}{d\eta}\approx\phi_C, \tag{37}$$

where $\theta(\eta)=\arg[\Lambda(\eta)]$. Like the ripple-spacing method, this method is relatively insensitive to changes in TW wavelength. This is illustrated in Fig. 6, where a change in the TW wavelength is induced in a cochlear model by altering the cochlear amplifier gain. The change in spectral period, which is clearly indicated by the ensemble-averaged results, is also detected by the four-parameter fit to a single realization in Fig. 6(b). However, the SFOAE phase spectrum in Fig. 6(a) shows no discernible change for the same single realization. A similar situation is seen in the measured data, where changes detected by the four-parameter model fail to show . up in the phase spectrum [Lineton and Lutman, 2003b, Fig3(c)]. A further serious drawback is that $\arg[\Lambda(\eta)]$ becomes undefined for regions of the SFOAE where the signal falls below the noise floor, thereby complicating the calculation of an average slope.

A third method is to take the location of the peak of the of the raw ϕ spectrum (e.g., Lutman and Deeks, 1999). From the standard theory of random processes (e.g., Bendat and Piersol, 1966), the raw spectrum shows strong idiosyncratic fine structure and thus gives a poor estimate of the true spectrum. This is illustrated in Fig 2(b), where the fine structure leads to considerable ambiguity in the location of the peak raw ϕ spectra. A comparison of the sampling distribution of this method with that of the four-parameter model is given in Table I.

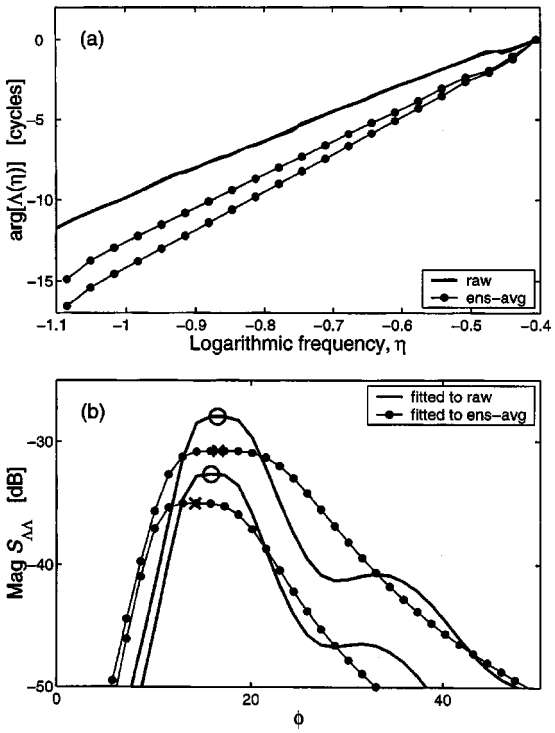


FIG. 6. Effect of altering the TW wavelength on (a) the slope of the SFOAE phase spectrum and (b) the fitted ϕ spectrum obtained from a cochlear model. Two different values of the cochlear amplifier gain were used to obtain two different TW wavelengths. At both values of the gain, an ensemble of 32 realizations of the scattering impedance was used to generate an ensemble of SFOAE spectra. In (a) the ensemble-averaged SFOAE phase spectra (solid line with filled circles) is shown for both values of gain. The change in slope clearly indicates the change in spectral period [Eq. (37)]. Also shown for both values of gain are the phase spectra obtained from a single realization of Λ (thick solid). These are overlaid, but indistinguishable, indicating the difficulty in estimating changes in spectral period from changes in the slope of the phase spectrum. In (b) for each value of gain, the ϕ spectrum obtained from the four-parameter fit to the ensemble-averaged autocovariance function is shown (solid line with filled circles) together with the ensemble estimate of ϕ_C (indicated by a cross). The ϕ spectra obtained from the same single realizations as in (b) are also shown (thick solid) together with the estimates of ϕ_C (indicated by an open circle). Though the change in ϕ_C is underestimated from the single realization in (b), it is still discernible.

C. The four-parameter model applied to measured data

Real ears are likely to deviate even further from the assumptions underlying the four-parameter model than do the cochlear models assessed in Sec. III A. For example, measurements show that spectral period reduces with stimulus frequency (Zweig and Shera, 1995), indicating a deviation from scaling symmetry. Assessing the importance of such deviations for the estimation of spectral period by the four-parameter model is a complex task. Unlike the results from cochlear models, it is not possible to obtain an ensemble-averaged autocovariance function for a given ear. (Averaging data across several ears is not the same as averaging data for which only the scattering impedance varied from one realization to the next.) Thus, there is no known value of spectral period against which the estimate of the four-parameter model can be judged, making it difficult to quantify the discrepancy between the model and experimental data. One approach would be to examine the distribution of the error term in Eq. (36) for measured SFOAEs from a large number of ears under a variety of stimulus conditions. While such a comprehensive comparison has not been attempted here, the error term has been examined from several ears under a range of stimulus levels and found to be of a similar magnitude to that expected from the cochlear models. A typical example of the similarity between the raw and fitted ϕ spectra for measured SFOAE data obtained from Lineton and Lutman (2003b) is shown in Fig. 7.

The relationship between α and $\sigma_{\Lambda\Lambda}$ has also been examined for measured data. As with the results from cochlear models (Fig. 5), the measured data generally show the expected decrease in α as $\sigma_{\Lambda\Lambda}$ is decreased (Lineton and Lutman, 2003b, Fig. 7), at least for values of α above about 0.1. This suggests that α may correctly account for the some of the effects of multiple reflections. Further evidence for the validity of the estimate of α comes from data measured in ears showing strong spontaneous OAEs, for which estimates of α are generally elevated (Lineton and Lutman, 2003b, Fig. 7).

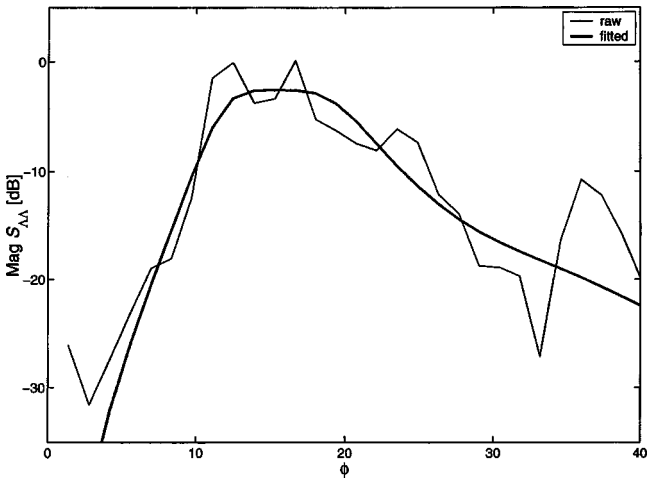


FIG. 7. Raw and fitted ϕ spectra obtained from SFOAE measurements in an ear. The raw (thin line) and fitted (thick line) ϕ spectra are obtained from the Fourier transforms of the raw and fitted autocovariance functions, respectively. The raw autocovariance function was obtained from a frequency sweep over 1376 and 2816 Hz at a nominal earphone level of 20 dB SPL. The fitted autocovariance function was obtained from the four-parameter fit to the raw function. (Adapted from Lineton and Lutman, 2003b, Fig. 4.)

IV. DISCUSSION

It has been demonstrated that the four-parameter model provides a useful characterization of the spectral period of simulated SFOAEs obtained from cochlear models based on Zweig and Shera's theory. Furthermore, the model was generally found to be more sensitive to changes in spectral period than other methods of estimation applied to single realizations of the SFOAE.

The limitations of the model should also be noted. Perhaps the least satisfactory aspect of the model is its treatment of multiple reflections. The assumption that the quantities **g** and **r** in Eq. (9) are independent of frequency allows multiple reflections to be characterized by a single parameter: α. Though the estimates of α were found to be reliable when obtained from ensemble-averaged data, estimates based on a single realization were found to be very variable. In fact, in the majority of cases (whether from simulated or measured SFOAEs), a simpler three-parameter model in which $\alpha \equiv 0$ yields very similar estimates of ϕ_C and ϕ_{BW}. However, modeling results suggest that for those models where $\alpha > 0.15$, the inclusion of α leads to marked improvements in the estimates of ϕ_C and ϕ_{BW}, thereby justifying the additional complexity.

This treatment of multiple reflections has the advantage that the number of free parameters in the model is kept to a minimum. However, this degree of simplification may not always be required. It may instead be possible to examine each spectral ripple in turn to estimate the degree to which multiple reflections influence the SFOAE in that spectral region. Whether this would lead to significant improvements in the estimates of ϕ_C is still uncertain.

It may be possible to apply the general approach outlined here to other OAEs. For example, several studies have investigated the use of measured properties of transient evoked OAEs as indicators of cochlear status (Avan *et al.*, 2000; Lucertini *et al.*, 2002; Sisto and Moleti, 2002). A parametric method of dealing with the inherently random nature of reflection-based OAEs may lead to improved estimators for these purposes.

V. CONCLUSIONS

A stochastic model of SFOAE spectra has been developed based on the theory of SFOAE generation proposed by Zweig and Shera (1995). The model has four parameters, the most important of which quantifies the spectral period of the SFOAE. Application of the model to simulated SFOAEs (Lineton and Lutman, 2003a) suggests that the model gives improved estimates of spectral period compared to previously reported methods. This is particularly important for extracting information regarding the properties of the cochlear TW based on a single measured SFOAE spectrum. The model has been applied to measured SFOAE spectra from a single ear (Lineton and Lutman, 2003b), where it appears to be sensitive to small changes in spectral period brought about by SFOAE suppression. It is therefore a potentially useful tool for investigating changes in TW properties.

ACKNOWLEDGMENTS

We wish to thank Paul White for his assistance. Christopher Shera and an anonymous reviewer also provided useful comments.

APPENDIX A: FREQUENTLY USED SYMBOLS AND THEIR MEANING

The following symbols also appear with the same meaning in the two companion papers by Lineton and Lutman (2003a, 2003b). Where applicable, the number of the defining equation is given. Also where applicable, the corresponding notation used by Shera and Zweig (1993b) or Zweig and Shera (1995), is included. The double-headed arrow ($\leftrightarrow$) indicates a pair of Fourier-conjugate variables. Bold typeface denotes complex quantities.

Operators

$E[\,]$	expectation
$\mathbb{F}\{\,\}$	Fourier transform [Eq. (15)]
$\langle\,\rangle_{\text{ensemble}}$	ensemble average
$\otimes$	convolution
$*$	complex conjugation

Quantities

$\mathbf{C}_{RR}(\eta')$	autocovariance function of **R** [Eq. (11)]
$\mathbf{C}_{\Lambda\Lambda}(\eta')$	autocovariance function of Λ [Eq. (20)]
$\mathbf{C}_{\Lambda\Lambda}^{(\text{fit})}(\eta')$	estimate of $\mathbf{C}_{\Lambda\Lambda}(\eta')$ obtained by the four-parameter fit to $\mathbf{C}_{\Lambda\Lambda}^{(\text{raw})}(\eta')$
$\mathbf{C}_{\Lambda\Lambda}^{(\text{raw})}(\eta')$	estimate of $\mathbf{C}_{\Lambda\Lambda}(\eta')$ obtained from a single realization of Λ [Eq. (35)]
$\mathbf{C}_{\Lambda\Lambda}^{(\text{ens-fit})}(\eta')$	estimate of $\mathbf{C}_{\Lambda\Lambda}(\eta')$ obtained by the four-parameter fit to $\langle \mathbf{C}_{\Lambda\Lambda}^{(\text{raw})}(\eta')\rangle_{\text{ensemble}}$
f	frequency ($t \leftrightarrow f$)

$\mathbf{g}(\eta)$ function characterizing the roundtrip transmission through the middle-ear [denoted q by Zweig and Shera (1995)]

$\mathbf{g}_0$ effective value of $\mathbf{g}$ across a given frequency interval

$\mathbf{h}_{\mathrm{TW}}(\eta)$ approximation to $\mathbf{T}^2$ by a one-sided second-order Butterworth filter [Eq. (31)]

$\mathbf{K}_{RR}(\eta')$ autocorrelation function of $\mathbf{R}$ defined as the normalized autocovariance function [Eq. (28)]

$\mathbf{p_{EC}}(\eta)$ ear-canal pressure

$\mathbf{p_{EC:R=0}}(\eta)$ ear-canal pressure in the absence of inhomogeneities

$\mathbf{p_{SF}}(\eta)$ component of ear-canal pressure due to the presence of inhomogeneities [Eq. (8)]

$\mathbf{r}(\eta)$ middle-ear reflection coefficient for backward TW [also denoted r by Zweig and Shera (1995)]

$\mathbf{r}_0$ effective value of $\mathbf{r}$ across a given frequency interval

$\mathbf{R}(\chi)$ cochlear reflectance (the ratio of backward to forward pressure TW at the stapes) [also denoted R by Zweig and Shera (1995)]

$S_{RR}(\phi)$ the ϕ spectrum of $\mathbf{R}$ defined as the inverse Fourier transform of $\mathbf{C_{RR}}(\eta')$ [Eq. (14)]

$S_{\Lambda\Lambda}(\phi)$ the ϕ spectrum of Λ defined as the inverse Fourier transform of $\mathbf{C}_{\Lambda\Lambda}(\eta')$

$S_{\Lambda\Lambda}^{(\mathrm{fit})}(\phi)$ the fitted ϕ spectrum of Λ defined as the inverse Fourier transform of $\mathbf{C}_{\Lambda\Lambda}^{(\mathrm{fit})}(\eta')$

$S_{\Lambda\Lambda}^{(\mathrm{raw})}(\phi)$ the raw ϕ spectrum of Λ defined as the inverse Fourier transform of $\mathbf{C}_{\Lambda\Lambda}^{(\mathrm{raw})}(\eta')$

$S_{\Lambda\Lambda}^{(\mathrm{ens\text{-}fit})}(\phi)$ the ensemble-fitted ϕ spectrum of Λ defined as the inverse Fourier transform of $\mathbf{C}_{\Lambda\Lambda}^{(\mathrm{ens\text{-}fit})}(\eta')$

t time ($t \leftrightarrow f$)

$\mathbf{T}$ (Θ) at fixed frequency, the forward-traveling cochlear velocity wave, normalized to stapes velocity [Eq. (2)] [denoted T by Zweig and Shera (1995)]

$\mathbf{u_{St}}(\omega)$ the stapes velocity

$\mathbf{v_{BM}}(x,\omega)$ the BM velocity

$w(\eta)$ a Gaussian white-noise signal

x distance along the BM from the stapes

x_{Ref} the distance from the stapes to the point whose natural frequency equals ω_{Ref} ($2\pi\times 1$ kHz)

α a parameter of the four-parameter model characterizing the component of Λ due to multiple reflections of the TW [Eq. (33)]

β a value characterizing the component of Λ due to the first TW reflection [Eq. (33)]

ε a measure of the discrepancy between the fitted and raw autocovariance functions [Eq. (36)]

η logarithmic frequency ($\phi \leftrightarrow \eta$) [Eq. (4)] [differs by a constant from the quantity denoted $\hat{\chi}$ by Zweig and Shera (1995)]

η' logarithmic frequency lag ($\phi \leftrightarrow \eta'$)

Θ dimensionless scaling variable [Eq. (3)] [denoted $\hat{\chi}-\chi$ by Zweig and Shera (1995)]

$\Lambda(\eta)$ ratio of $\mathbf{p_{SF}}(\eta)$ to $\mathbf{p_{EC:R=0}}(\eta)$ [Eq. (10)] [similar to the quantity denoted Δ by Shera and Zweig (1993b)]

χ dimensionless distance [Eq. (4)] [differs by a constant from the quantity denoted χ by Zweig and Shera (1995)]

$\rho_{Sc}(\chi)$ scattering potential [denoted ρ by Zweig and Shera (1995)]

σ_{RR} rms value of $\mathbf{R}$ [Eq. (29)]

$\sigma_{\Lambda\Lambda}$ rms value of Λ

ϕ variable conjugate to η' ($\phi \leftrightarrow \eta'$) [similar to the quantity denoted $\varphi/2\pi$ by Zweig and Shera (1995)]

ϕ_C the location of the peak of $S_{\Lambda\Lambda}(\phi)$ [similar to the quantity denoted $\hat{\varphi}/2\pi$ by Zweig and Shera (1995)]

ϕ_{BW} the width of the peak of $S_{\Lambda\Lambda}(\phi)$

ω angular frequency $2\pi f(t \leftrightarrow \omega/2\pi)$

$\omega_C(x)$ the natural frequency of the BM at point at x [Eq. (1)]

ω_{C_0} the natural frequency of the BM at the base

ω_{Ref} the reference natural frequency$=2\pi\times 1$ kHz

APPENDIX B: DERIVATION OF THE STOCHASTIC MODEL OF THE SFOAE SPECTRUM

In this Appendix, the steps leading from the nonlinear relationship between Λ and $\mathbf{R}$ in Eq. (18) and its stochastic form Eq. (19) are presented.

Consider a signal $\mathbf{x}(n)$ which arises from a random process that is Gaussian, stationary, ergodic, analytic, and of zero mean, and which undergoes the following nonlinear transformation to give the signal, $\mathbf{y}(n)$:

$$\mathbf{y}(n)=\frac{\mathbf{g}_0\mathbf{x}(n)}{1-\mathbf{r}_0\mathbf{x}(n)}. \tag{B1}$$

For small values of the product $|\mathbf{r}_0\mathbf{x}(\mathbf{n})|$ [i.e., $|\mathbf{r}_0\mathbf{x}(n)|<1$], Eq. (B1) can be expanded using the binomial theorem

$$\mathbf{y}(n)\approx\mathbf{g}_0\mathbf{x}(n)[1+\mathbf{r}_0\mathbf{x}(n)+\mathbf{r}_0^2\mathbf{x}^2(n)+\mathbf{r}_0^3\mathbf{x}^3(n)\cdots]. \tag{B2}$$

Note that there is an inconsistency in this approach. The binomial expansion in Eq. (B2) is only valid when $\mathbf{r}_0\mathbf{x}(n)<1$. However, since $\mathbf{x}(n)$ is assumed to be Gaussian, it can take any value from $-\infty$ to $+\infty$. Thus, in theory, the binomial expansion becomes inapplicable for a small proportion of the theoretical signal. In practice, this is not usually a problem, since the physical signal, $\mathbf{x}(n)$, is of course not truly unbounded, and thus not truly Gaussian.

Truncating to the first two terms in Eq. (B2) gives

$$\mathbf{y}(n)\approx\mathbf{A}\mathbf{x}(n)+\mathbf{B}\mathbf{x}^2(n),$$
$$\mathbf{A}\equiv\mathbf{g}_0, \quad \mathbf{B}\equiv\mathbf{g}_0\mathbf{r}_0. \tag{B3}$$

Additional terms may be included, if desired, but the algebra becomes lengthy, and thus for brevity, only these two terms are considered here.

The autocovariance function of $\mathbf{y}(n)$ is then given by

$$\mathbf{C}_{yy}(m)\equiv E[\mathbf{y}(n)^*\mathbf{y}(n+m)]$$
$$\equiv\mathbf{C}_1(m)+\mathbf{C}_2(m)+\mathbf{C}_3(m)+\mathbf{C}_4(m), \tag{B4}$$

where the four terms, obtained by substituting $\mathbf{y}(n)$ from Eq. (B3) in Eq. (B4), are

$$\mathbf{C}_1(m) \equiv \mathbf{A}^2 E[\mathbf{x}(n)^* \mathbf{x}(n+m)], \tag{B5}$$

$$\mathbf{C}_2(m) \equiv \mathbf{AB} E[\mathbf{x}^2(n)^* \mathbf{x}(n+m)], \tag{B6}$$

$$\mathbf{C}_3(m) \equiv \mathbf{AB} E[\mathbf{x}(n)^* \mathbf{x}^2(n+m)], \tag{B7}$$

$$\mathbf{C}_4(m) \equiv \mathbf{B}^2 E[\mathbf{x}^2(n)^* \mathbf{x}^2(n+m)]. \tag{B8}$$

The term $\mathbf{C}_1(m)$ in Eq. (B5) is simply proportional to the autocovariance function of $\mathbf{x}(n)$

$$\mathbf{C}_1(m) = \mathbf{A}^2 \mathbf{C}_{xx}(m). \tag{B9}$$

The terms $\mathbf{C}_2(m)$ and $\mathbf{C}_3(m)$ can be evaluated by noting that they involve the expected value of a product of three Gaussian random variables. It can be shown that, for any three jointly Gaussian random variables, X, Y, and Z of zero mean, the following holds (Deutsch, 1965):

$$E[XYZ] = 0. \tag{B10}$$

Therefore, it follows that:

$$\mathbf{C}_2(m) = \mathbf{C}_3(m) = 0, \quad \text{for all } m. \tag{B11}$$

The term $\mathbf{C}_4(m)$ involves the expected value of a product of four Gaussian random variables. For any four jointly Gaussian random variables, W, X, Y, and Z of zero mean, the expectation of their product is given by Eq. (B12) (Deutsch, 1965)

$$E[WXYZ] = E[WX]E[YZ] + E[WY]E[XZ] + E[WZ]E[XY]. \tag{B12}$$

From this, the expression for $\mathbf{C}_4(m)$ becomes

$$\begin{aligned}\mathbf{C}_2(m) = {} & \mathbf{B}^2 E[\mathbf{x}(n)^* \mathbf{x}(n)^*] E[\mathbf{x}(n+m)\mathbf{x}(n+m)] \\ & + \mathbf{B}^2 E[\mathbf{x}(n)^* \mathbf{x}(n+m)] E[\mathbf{x}(n)^* \mathbf{x}(n+m)] \\ & + \mathbf{B}^2 E[\mathbf{x}(n)^* \mathbf{x}(n+m)] E[\mathbf{x}(n)^* \mathbf{x}(n+m)].\end{aligned} \tag{B13}$$

From the requirement that $\mathbf{x}(n)$ is analytic, it follows that:

$$E[\mathbf{x}(n)^2] = E[\mathbf{x}(n)^* \mathbf{x}(n)^*] = E[\mathbf{x}(n+m)\mathbf{x}(n+m)] = 0. \tag{B14}$$

Thus, Eq. (B13) reduces to

$$\begin{aligned}\mathbf{C}_2(m) &= 2\mathbf{B}^2 E[\mathbf{x}(n)^* \mathbf{x}(n+m)] E[\mathbf{x}(n)^* \mathbf{x}(n+m)] \\ &= 2\mathbf{B}^2 \mathbf{C}_{xx}^2(m).\end{aligned} \tag{B15}$$

Therefore, substituting Eqs. (B9), (B11), and (B15) into (B4) gives

$$\mathbf{C}_{yy}(m) = \mathbf{A}^2 \mathbf{C}_{xx}(m) + 2\mathbf{B}^2 \mathbf{C}_{xx}^2(m), \tag{B16}$$

which gives the first two terms in Eq. (19). The third term in Eq. (19) arises from a similar derivation to that of Eq. (B16), but retaining the cubic term in Eq. (B2).

APPENDIX C: RELATING THE AUTOCOVARIANCE FUNCTION TO THE FILTER PARAMETERS

The terms in $\mathbf{C}_{RR}(\eta')$, Eq. (27), can be calculated directly from the two parameters, ϕ_C and ϕ_{BW} characterizing the bandpass filter. This relationship is derived below for the general case of filtered white noise. Rather than working in the η and ϕ domains, it is convenient in the following derivation to replace η with the discrete time variable, n, and ϕ by the frequency variable, ω. This allows the use of familiar filtering terminology and notation, without leading to confusion.

Consider a discrete time signal arising from Gaussian stationary white noise passed through a bandpass filter such that

$$x(n) = h(n) \otimes w(n) = \sum_{m=-\infty}^{\infty} h(m) w(n-m), \tag{C1}$$

where $x(n)$ is the filter output sequence, $w(n)$ is the (white-noise) input sequence, and $h(n)$ is the unit impulse response function of the bandpass filter. A four-pole Butterworth bandpass filter can be implemented as an autoregressive (AR), moving-average (MA) filter with five MA and four AR filter coefficients. Thus, Eq. (C1) is then represented as

$$x(n) = \sum_{j=0}^{4} a_j w(n-j) - \sum_{k=1}^{4} b_k x(n-k), \tag{C2}$$

where the a_j and b_k's are the five MA and four AR filter coefficients, respectively.

These nine ARMA coefficients are determined by only three independent parameters: the sampling rate, f_s, and the desired upper and lower cutoff frequencies ω_{dU} and ω_{dL} of the Butterworth filter. These relationships are presented in Eq. (C3)

$$\begin{aligned}
&a_0 = a_4 = \omega_{BW}^2 / D, \quad a_2 = -2a_0, \\
&a_1 = a_3 = 0, \\
&b_1 = (-4 - 2\sqrt{2}\omega_{BW} + 4\omega_c^4 + 2\sqrt{2}\omega_{BW}\omega_c^2)/D, \\
&b_2 = (6 - 2\omega_{BW}^2 - 4\omega_c^2 + 6\omega_c^4)/D, \\
&b_3 = (-4 + 2\sqrt{2}\omega_{BW} + 4\omega_c^4 - 2\sqrt{2}\omega_{BW}\omega_c^2)/D, \\
&b_4 = (1 - \sqrt{2}\omega_{BW} + \omega_{BW}^2 + 2\omega_c^2 + \omega_c^4 - \sqrt{2}\omega_{BW}\omega_c^2)/D,
\end{aligned} \tag{C3}$$

where D, ω_c, and ω_{BW} are given by

$$\begin{aligned}
&D = 1 + \sqrt{2}\omega_{BW} + \omega_{BW}^2 + 2\omega_c^2 + \omega_c^4 + \sqrt{2}\omega_{BW}\omega_c^2, \\
&\omega_c = \sqrt{\omega_U \omega_L}, \quad \omega_{BW} = \omega_U - \omega_L, \\
&\omega_{U,L} = \tan(\omega_{dU,dL}/2f_s),
\end{aligned} \tag{C4}$$

and where ω_U and ω_L are the dimensionless prewarped cutoff frequencies of the Butterworth filter, and ω_c and ω_{BW} are the dimensionless center frequency and bandwidth.

The unit impulse response function, $h(n)$, is related to the ARMA coefficients by

$$h(n) = \begin{cases} 0 & \text{for } n \leq 0 \\ \sum_{j=0}^{4} a_j \delta(n-j) - \sum_{k=1}^{4} b_k h(n-k) & \text{for } n > 0 \end{cases}, \tag{C5}$$

where $\delta(n)$ is the unit impulse, defined by

$$\delta(n) \equiv \begin{cases} 1 & \text{for } n=0 \\ 0 & \text{otherwise} \end{cases}. \tag{C6}$$

From Eq. (C2), the autocovariance function of the signal, $x(n)$, as defined in Eq. (11) becomes

$$\mathbf{C}_{xx}(m) = \sum_{j=0}^{4} a_j \mathbf{C}_{wx}(j-m) - \sum_{k=1}^{4} b_k \mathbf{C}_{xx}(m-k), \tag{C7}$$

where $\mathbf{C}_{wx}(m)$ is the cross-covariance function between $x(n)$ and $w(n)$ defined as

$$\mathbf{C}_{wx}(m) \equiv E[w(n)^* x(n+m)]. \tag{C8}$$

The function in Eq. (C8) can be expressed in terms of the impulse response function by substitution from Eq. (C1) to give

$$\begin{aligned} \mathbf{C}_{wx}(m) &= E\left[w(n)^* \sum_{p=-\infty}^{\infty} h(p) w(n+m-p) \right] \\ &= \sum_{p=-\infty}^{\infty} h(p) E[w(n)^* w(n+m-p)] \\ &= \sum_{p=-\infty}^{\infty} h(p)\, \delta(p-m) S_{ww} = S_{ww} h(m). \end{aligned} \tag{C9}$$

Using the result in Eq. (C9) together with the causality property of $h(m)$ from Eq. (C4) yields the following expansion for the autocovariance function in Eq. (C7):

$$\mathbf{C}_{xx}(m) = \begin{cases} S_{ww} \sum_{j=0}^{4} a_j h(j) - \sum_{k=1}^{4} b_k \mathbf{C}_{xx}(k) & \text{for } m=0 \\ S_{ww} \sum_{j=m}^{4} a_j h(j-m) - \sum_{k=m}^{4} b_k \mathbf{C}_{xx}(k-m) & \\ \quad - \sum_{k=1}^{m-1} b_k \mathbf{C}_{xx}(m-k) & \text{for } 1 \leqslant m \leqslant 4 \\ -\sum_{k=1}^{4} b_k \mathbf{C}_{xx}(m-k) & \text{for } m > 4 \end{cases}. \tag{C10}$$

If desired, the right-hand side of Eq. (C10) could be further expanded by replacing the h's with a's and b's by using Eq. (C4), thus yielding equations involving only a's, b's, S_{ww}, and $\mathbf{C}_{xx}(m)$. Furthermore, from Eqs. (C3) and (C4), the a's and b's could be replaced with the two Butterworth filter parameters, ω_C and ω_{BW}. Though the relationship is nonlinear, it is usually possible to invert this system of three equations, such that the three parameters ω_C, ω_{BW}, and S_{ww} can be calculated purely from $\mathbf{C}_{xx}(0)$, $\mathbf{C}_{xx}(1)$, and $\mathbf{C}_{xx}(2)$. Thus, all values of $\mathbf{C}_{xx}(m)$ for $m>3$ are redundant. To perform the inversion, an iterative numerical procedure could be followed. This then illustrates the significance of the variable η'_{Lim} in Eq. (36). To obtain the three parameters from $\mathbf{C}_{xx}(m)$ in this simple example, η'_{Lim} would correspond to $m=3$.

However, this inversion only yields the correct values of the three parameters if all the assumptions in the model are met: the filter must truly be a four-pole Butterworth bandpass filter, and $\mathbf{C}_{xx}(m)$ must be known exactly. In reality, deviations of the cochlea TW from the simple model, the presence of noise, and the finite length of $x(n)$ all cause deviations from these assumptions. From the cochlear model simulations it was found that improved estimates of the model parameters were obtained if η'_{Lim} was somewhat larger than the theoretical value given above, though the exact value of η'_{Lim} was not critical. The effect of this is to perform a least-squares fit to Eq. (C10), rather than performing an exact inversion.

For example, in from the cochlea models it was found that provided η'_{Lim} was set between about 10% and 50% of the available SFOAE signal length, very similar estimates of the four parameters were obtained. Beyond about 50%, the error term contains contributions from estimates of $\mathbf{C}_{xx}(m)$ which become increasingly unreliable, owing to the diminishing interval of the integration in Eq. (35).

[1]In this paper, the SFOAE is defined as the component of the ear-canal pressure due to a backward-traveling wave originating in the cochlea. With this definition, no periodicity is seen in the magnitude spectrum of the SFOAE. Instead, periodicity is seen in both the real and imaginary parts of the frequency spectrum of the SFOAE. Other authors (e.g., Talmadge *et al.*, 2000) characterize SFOAEs using the magnitude spectrum of the ear-canal pressure. This spectrum shows a similar periodicity owing to the pattern of interference between the emission component and the stimulus components of the ear-canal pressure.

[2]This differs from Zweig and Shera (1995), where $x_{\mathrm{Ref}}=0$ and thus ω_{Ref} is the basal natural frequency. This definition could not be used in the measured data, since the basal natural frequency is unknown. As a consequence of this difference, η and χ in Eq. (4) differ from the corresponding quantities denoted $\hat{\chi}$ and χ by Zweig and Shera by an additive constant, while $\mathbf{T}$ is identical in both nomenclatures.

Avan, P., Wit, H. P., Guitton, M., Mom, T., and Bonfils, P. (**2000**). "On the spectral periodicity of transient-evoked otoacoustic emissions from normal and damaged cochleae," J. Acoust. Soc. Am. **108**, 1117–1127.

Bendat, J. S., and Piersol, A. G. (**1966**). *Measurement and Analysis of Random Data* (Wiley, New York).

Burg, J. P. (**1978a**). "Maximum entropy spectral analysis," in *Modern Spectrum Analysis*, edited by D. G. Childers, IEEE Press Selected Reprint Series (Wiley, New York), pp. 34–41.

Burg, J. P. (**1978b**). "A new analysis technique for time series data," in *Modern Spectrum Analysis*, edited by D. G. Childers, IEEE Press Selected Reprint Series (Wiley, New York), pp. 42–48.

Dallmayr, C. (**1987**). "Stationary and dynamical properties of simultaneous evoked otoacoustic emissions (SEOAE)," Acustica **63**, 223–255.

Deutsch, R. (**1965**). *Estimation Theory* (Prentice-Hall, Englewood Cliffs, NJ).

Jenkins, G. M., and Watts, D. G. (**1968**). *Spectral Analysis and Applications* (Holden Day, San Francisco).

Kemp, D. T. (**1978**). "Stimulated acoustic emissions from within the human auditory system," J. Acoust. Soc. Am. **64**, 1386–1391.

Kemp, D. T. (**1980**). "Toward a model for the origin of cochlear echoes," Hear. Res. **2**, 533–548.

Kemp, D. T. (**1986**). "Otoacoustic emissions, traveling waves and cochlear mechanisms," Hear. Res. **22**, 95–104.

Kemp, D. T., Brass, D. N., and Souter, M. (**1990**). "Observations on simultaneous SFOAE and DPOAE generation and suppression," in *The Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthew, M. A. Ruggero, and C. R. Steele (Springer, Berlin), pp. 202–209.

Kemp, D. T., and Brown, A. M. (**1983**). "A comparison of mechanical nonlinearities in the cochleae of man and gerbil from ear canal measurements," in *Hearing—Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer, Berlin), pp. 82–88.

Kemp, D. T., and Chum, R. A. (**1980**). "Observations on the generation mechanism of stimulus frequency acoustic emissions—two tone suppres-

sion,'' in *Psychophysical, Physiological and Behavioural Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen, *Proceedings of the 5th International Symposium on Hearing* (Delft University Press, Delft, The Netherlands), pp. 34–42.

Lineton, B., and Lutman, M. E. (**2003a**). ''Modeling the effect of suppression on the periodicity of stimulus frequency otoacoustic emissions,'' J. Acoust. Soc. Am. **114**, 859–820.

Lineton, B., and Lutman, M. E. (**2003b**). ''The effect of suppression on the periodicity of stimulus frequency otoacoustic emissions: Experimental data,'' J. Acoust. Soc. Am. **114**, 871–882.

Lonsbury-Martin, B. L., Harris, F. P., Stagner, B. B., Hawkins, D. H., and Martin, G. K. (**1990**). ''Distortion product emissions in humans II. Relations to acoustic immitance and stimulus frequency and spontaneous otoacoustic emissions in normally hearing subjects,'' Ann. Otol. Rhinol. Laryngol. Suppl. **147**, 15–29.

Lucertini, M., Moleti, A., and Sisto, R. (**2002**). ''On the detection of early cochlear damage by otoacoustic emission analysis,'' J. Acoust. Soc. Am. **111**, 972–978.

Lutman, M. E., and Deeks, J. (**1999**). ''Correspondence among microstructure patterns observed in otoacoustic emissions and Bekesy,'' Audiology **38**, 263–266.

Randall, R. B. (**1987**). *Frequency Analysis by R. B. Randall*, 2nd ed. (Brüel & Kjaer, Naerum, Denmark).

Shera, C. A., and Zweig, G. (**1993a**). ''Order from chaos: Resolving the paradox of periodicity in evoked otoacoustic emissions,'' in *Biophysics of Hair-Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 54–63.

Shera, C. A., and Zweig, G. (**1993b**). ''Noninvasive measurement of the cochlear traveling-wave ratio,'' J. Acoust. Soc. Am. **93**, 3333–3352.

Sisto, R., and Moleti, A. (**2002**). ''On the frequency dependence of otoacoustic emission latency in hypoacoustic and normal ears,'' J. Acoust. Soc. Am. **111**, 297–308.

Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). ''Modeling otoacoustic emission and hearing threshold fine structures,'' J. Acoust. Soc. Am. **104**, 1517–1543.

Talmadge, C. L., Tubis, A., Long, G. R., and Tong, C. (**2000**). ''Modeling the combined effects of basilar membrane nonlinearity and roughness on stimulus frequency otoacoustic emission fine structures,'' J. Acoust. Soc. Am. **108**, 2911–2932.

Wilson, J. P. (**1980**). ''Evidence for a cochlear origin for acoustic re-emissions, threshold fine-structure and tonal tinnitus,'' Hear. Res. **2**, 233–252.

Zweig, G., and Shera, C. A. (**1995**). ''The origin of periodicity in the spectrum of evoked otoacoustic emissions,'' J. Acoust. Soc. Am. **98**, 2018–2047.

Zwicker, E. (**1990**). ''On the frequency separation of simulataneously evoked otoacoustic emissions, consecutive extrema and its relation to cochlear traveling waves,'' J. Acoust. Soc. Am. **88**, 1639–1641.

Zwicker, E., and Schloth, E. (**1984**). ''Interrelation of different otoacoustic emissions,'' J. Acoust. Soc. Am. **75**, 1148–1154.

The mammalian cochlear map is optimally warped[a)]

Eric L. LePage[b)]
National Acoustic Laboratories, Sydney, Australia (The research division of Australian Hearing), 126 Greville Street, Chatswood, NSW 2067, Australia

(Received 8 July 2002; revised 22 April 2003; accepted 29 April 2003)

The form of the mammalian cochlear frequency-position map has been well described by Greenwood and empirical values found for its coefficients for a number of species. The apical portion of the mammalian map is spatially compressed relative to the base, and this nonuniformity in the representation of frequency is evidently consistent across species. However, an evolutionary reason for this consistency, encompassing critical band behavior with respect to position, is conspicuously missing. Likewise, the length of the cochlea in any mammal, including echolocating species, is related to body size, but attempts to explain the length in terms of frequency limits, range, or resolution have no general explanation. New insight stems from a hypothesis in which the map curvature may be appreciated as an adaptation for *optimal* frequency resolution over the auditory range. It is demonstrated numerically that the mammalian curve may be considered a member of a family of curves which vary in their degree of warp. The ''warp factor'' found to be common across mammals is an optimal trade-off between four conflicting constraints: (1) enhancing high-frequency resolution; (2) setting a lower bound on loss of existing low-frequency resolution; (3) minimizing map nonuniformity; and (4) keeping the whole map smooth, thereby avoiding reflections. [DOI: 10.1121/1.1587150]

PACS numbers: 43.64.Bt, 43.64.Jb, 43.64.Kc [BLM]

I. INTRODUCTION

The form of the cochlear tonotopic map is very basic to the whole of auditory science, being considered at any stage in experimental and modeling investigations through to clinical work, cochlear evolution (Fay and Popper, 2000), and development (Echteler *et al.*, 1989), its extension to echolocating capabilities (Ketten, 2000), even to the fitting of cochlear implants (Skinner *et al.*, 2002).

The application which has triggered this investigation has to do with making sense of otoacoustic emissions, sounds which are emitted from the ear due to activity of the rows of outer hair cells (OHC) stretched along the length of the cochlear partition. In particular, the clinical usefulness of otoacoustic emissions could be greatly enhanced if the emission signal could be deconvolved in such a way as to reveal the extent of regions of dead or poorly functioning OHC. Since the development of the otoacoustic emission technique one should define at least two maps; a reception map and an emission map, which may be the same map for theoretical reasons (Shera and Guinan, 1999), but this needs to be more clearly defined for clinical use. It is not yet known how precisely the map for sounds emitted via reverse propagation overlays the map of forward-propagated sounds input to the ear and also how much emissions are a product of activity at any place or how much they depend upon secondary or multiple reflections. Fundamental to efforts towards a reconciliation is the need for a clear understanding of the form of the cochlear map relating characteristic frequency versus position along the cochlear partition. En route to such a reconciliation, this paper revisits the receptive map for additional clues as to why the map has its particular form, effectively, low frequencies detected at the apex are spatially compressed relative to high frequencies detected at the base.

The generalized form of the mammalian reception cochlear map which is described by the Greenwood relation

$$f = \mathbf{A} \cdot (10^{\mathbf{a}x} - \mathbf{k}), \tag{1}$$

where f is frequency (Hz), x is the distance from the apex of the cochlea (helicotrema end), **A**, **a**, and **k** are coefficients (Greenwood, 1961, 1990, 1996). **a** is the gradient of high-frequency end of the map, i.e., the coefficient of the derivative evaluated at the highest frequency, **A** is a constant which shifts the curve as a whole along the log-frequency axis, and **k** is constant, which introduces curvature into the frequency-position function so as to fit low-frequency data. The gradient **a** may be expressed equivalently in two ways depending upon whether x is expressed in physical distance (mm) or is normalized to the range [0, 1] ([0, 100]%) to deal with variations in length of the cochlear partition which occur between individuals, between species (Bohne and Carr, 1979). The maps of specialized cochleae, such as those from echolocating species, deviate from this relationship and are treated separately below.

Table I shows values for these coefficients extracted from the publications listed and other discussions (Greenwood, personal communication). If x is measured in millimeters, then we denote that $\mathbf{a}'$ is the gradient (oct/mm) of the log frequency-position map at the high-frequency end and is a constant which for humans is equal to 0.06 mm^{-1}. The parameter **a** is the normalized value obtained multiplying $\mathbf{a}'$ by the cochlear length L (mm). The table is listed in roughly

[a)]This material has not been submitted elsewhere. A poster of this material was presented at the conference ''Biophysics of the Cochlea: Molecules to Models,'' Titisee, Germany, 27 July–1 August 2002.

[b)]Electronic mail: Eric.LePage@nal.gov.au

TABLE I. Values of Greenwood model parameters **A**, **a**, and **k** extracted primarily from the references listed. The parameter **a**′ are the non-normalized values obtained from dividing **a** by the cochlear length **L** (mm). The last for house mouse are recent estimates by Greenwood (personal communication) based upon the data of Ou *et al.* (2000). The last three columns indicate estimations of end frequencies F_{apex} and F_{base}, and the number of octaves reception subtended between the F_{apex} and F_{base}. Note the values of F_{apex} are zero for the cases where $\mathbf{k}=1$, or the values indicated in parentheses if the warp theory is valid and $\mathbf{k}\sim 0.85$.

Name	A	a	k	L (mm)	Reference	a' (mm^{-1})	F_{apex} (Hz)	F_{base} (Hz)	#Oct
Elephant	81	1.8	1	60	Greenwood (1961)	0.030	0 (12)	5 111	−8.7
Cow	52.6	2.1	1	38.3	Greenwood (1961)	0.055	0 (8)	6 622	−9.7
Human	165.4	2.1	1	35	Greenwood (1990)	0.060	0 (20)	20 772	−10
Macaque	360	2.1	0.85	25.6	Greenwood (1990)	0.082	54	45 321	9.7
Domestic cat	456	2.1	0.8	25	Liberman (1982)	0.084	91	57 407	9.3
Guinea pig	350	2.1	0.85	18.5	Greenwood (1990)	0.114	53	44 062	9.7
Chinchilla	163.5	2.1	0.85	18.4	Greenwood (1990)	0.114	25	20 583	9.7
Rat	7613.3	0.928	1	8.03	Müller (1991)	0.116	0 (1142)	64 502	−5.8
Opossum	5821	1	0.564	6.4	Müller *et al.* (1993)	0.156	2538	58 210	4.5
Gerbil	398	2.2	0.631	12.1	Müller (1996)	0.182	147	63 079	8.7
Mouse	7130	0.99	1	6.8	Greenwood (pers.)	0.300	0 (1070)	105 262	−6.6

descending order of size of the mammal beginning with available data for elephant and ending with recent estimates of parameters for the house mouse by Greenwood based upon recent data (Ou *et al.*, 2000). The last three columns indicate estimations of end frequencies F_{apex} and F_{base}, and the number of octaves subtended between F_{apex} and F_{base}. Note the values of F_{apex} are either zero for the cases where **k** is designated equal to 1, or the values indicated in parentheses.

In the table there seems to be a pattern across all mammals for values of both **a** and **k** to cluster. However, there appears to be two forms of exception. First, for most primates plus chinchillas $\mathbf{a}\sim 2.1$, while rats and opossum notably stand apart at $\mathbf{a}\sim 1.0$. Second, the values of **k** seem to cluster between 0.5 and 1. Further examination allows three subclusters to be considered, *viz.*, values of $\mathbf{k}\sim 1$, $\mathbf{k}\sim 0.85$, and $\mathbf{k}\sim 0.5$–0.6 (Müller, 1996). Values close to unity tend to be for larger mammals where the data is obtained behaviorally (Heffner and Heffner, 1982), while the values around 0.85 and below tend to be applied to the more precise morphological determinations such as those obtained with dye-injection techniques (Liberman, 1982; Müller, 1991; Müller *et al.*, 1993; Müller, 1996). Echolocating mammals are a special case by virtue that they have multiple map regions, and are treated separately below. For species which have relatively small frequency spans, how the scale is represented (linear or log) is probably not so important. While Greenwood's analyses of the form of the map are extensive, a new reason is provided to show that for species whose range spans >8 octaves, the plotting of the spatial dimension versus the logarithm of frequency has particular advantages for understanding form.

Figure 1(A) shows, overlaid, the resulting maps for the species plotted versus the logarithm of frequency for the empirically determined values of **A**, **a**, and **k** in Table I. In each case the map referred to is the empirically found distance along the cochlear partition (ordinate values) versus the estimated best incoming frequency for any location (abscissa). As might be expected from the clustering of values mentioned above, the curves are mostly very similar in shape. The human and chinchilla curves almost overlay, as do

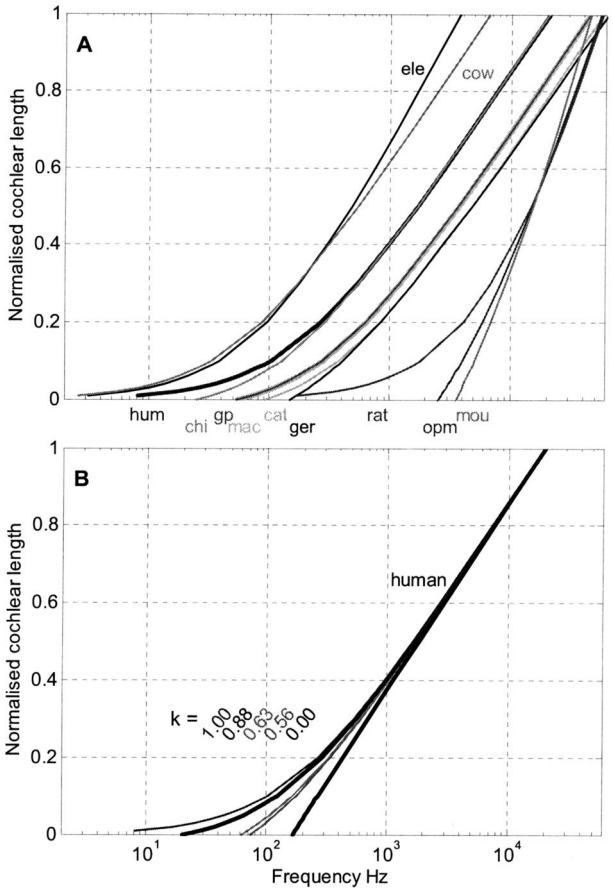


FIG. 1. Panels (A) and (B) show log frequency-position maps according to published values of the three coefficients of Eq. (1), for normalized distance from the apex towards the base. The top panel (A) shows the similarity of 8 maps from elephant (ele), cow, human (hum), chinchilla (chi), guinea pig (gp), macaque (mac), cat, Mongolian gerbil (ger), while the highest frequency maps, for rat, opossum (opm), and mouse (mou), which have shorter cochleae, have different basic gradients. The bottom panel (B) shows for the human case, the hypothetical effect of varying the third of the three parameters **k** over the values indicated in sequence. The higher the value of **k**, the bigger the correction. However, the value of unity quoted several times in Table I incurs the problem that the most apical region of the map is undefined. Any uncertainty in this value therefore generates considerable uncertainty as to the frequency associated with the most apical place.

guinea pig and macaque, and domestic cat and Mongolian gerbil. The notable exceptions in shape occur with the smaller mammals, opossum, rat, and mouse; plotting them on the same axes as the larger animals may lead to misinterpretation when the primary difference seems to be that the cochlea is shorter in these species. The parameter **a** has its biggest effect on the slope of the curves; the higher the value the more extensive the frequency range of the cochlea.

Figure 1(B) illustrates the hypothetical effect of varying **k** for the human case, a feature which is the focus of this paper. The straight dashed lines represent the case of $\mathbf{k}=0$. The value of **a** (2.1 for human) plots as the number of decades spanned between the ends of the map for the $\mathbf{k}=0$ line. For each species five different curves are shown for the values of **k** (0, 0.56, 0.63, 0.89, and 1.0) as indicated by the legend, chosen purely to coincide with the values quoted in Table I. As **k** increases, the curvature increases so that the spatial representation of frequencies at the apex is compressed relative to the basal end (Greenwood, 1996). A common alternate view is to regard the high-frequency representation as greatly expanded relative to the apex. The question of which view to take becomes interesting in the context of cochlear evolution, but to make progress it is necessary to consider how precisely the available data define the low-frequency end of the curve.

Plotted versus log frequency, the high-frequency end of the map is almost straight. For the purposes of curve fitting to data, changing the value of **k** has little effect at the high-frequency end, which is why it is generally taken to be straight (Robles and Ruggero, 2001). Values of $\mathbf{k}>0$ provide a better fit of the model specifically for low-frequency data points. The case of $\mathbf{k}=1.0$ is special. The latter leaves the map undefined at zero distance from the apex because it only reaches the abscissa at zero frequency. At first this seems a minor concern. Table I shows several species as having a value of $\mathbf{k}=1$, and presumably this is due to the sparsity or difficulty of obtaining low-frequency points with precision (Greenwood, 1996). While the high-frequency end of the map will be highly defined by high-frequency points, the same is not true of the low-frequency end, where sensitivity decreases more gradually as frequency is lowered depending upon the size of the helicotrema (Dallos, 1970). In this sense, the form of Eq. (1) is ''open ended.''

The problem is that such values of **k** are basically unsatisfactory because, although this end of the frequency range is important in behavioral terms, slightly differing values close to unity introduce huge variation of which frequency relates to which position. Indeed, while Table I quotes a value of unity, Fig. 1(B) shows that the most satisfactory value of **k** for human is probably closer to 0.9 because this almost coincides with the empirically determined value of 20 Hz at which sensitivity drops. While the data may not support values of $\mathbf{k}<1$, similar theoretical concerns exist, therefore, for every instance of values of unity in Table I.

We are concerned here with how the form of the map came about in terms of evolution, and what general trends may exist, but may be obscured by the variations in the wealth of data for which the equation has been fitted. One of the significant studies on the evolution of hearing is from Masterton *et al.* (1969). They note that the acquisition of high-frequency hearing in mammals is associated with the development of the system of three ossicles in the middle ear, a feature which was apparently present in early small mammals. One of the points of their article is not just the acquisition of high-frequency hearing in mammals, but indeed the apparent ''loss'' of high-frequency reception as the size of the species increased [see again Fig. 1(A)]. They postulate that since larger mammals were not so strongly dependent upon interaural level differences to achieve localization, localization could be achieved by interaural time differences as the head size grew larger. They suppose that loss of high-frequency sensitivity is a later development as the size of mammals increased.

We are concerned here about an earlier stage of evolution, around the Jurassic period, in which the acquisition of high frequencies apparently first took place. Fossils from that period have revealed the *Morganucodon* had an only partly coiled cochlea (like reptiles and birds) but which had early mammalian features—the characteristic beginnings of middle-ear ossicle formation (Kermack, 1989).

We are also concerned about what kinds of pressures led the basilar membrane to elongate, why the length is strongly species dependent, why the increase in length exceeds that expected on the basis of increase in high frequency (Manley, 2000), and ultimately what factors determine cochlear length, a question which remains unanswered (Popper and Fay, 1997; Ketten, 2000). In turn, we are trying to reconcile the ''open-ended'' form of Eq. (1) with the evolutionary pressures which led mammals to acquire high-frequency sensitivity (Popper and Fay, 1997). The considerations shed some light on the Masterton hypothesis, *viz.*, how larger mammals came to lose the highest frequencies.

Our quest for reconciliation leads directly to a paradox. It is a challenging paradox because we generally agree that the equation is a faithful model of most frequency-position data in generalist cochleae, certainly for high-frequency data tested against it. On the one hand, we conceive of high-frequency capability as having been ''added on'' to augment low-frequency reception in the archetypal cochleae of lower vertebrates, e.g., reptiles and turtles which hear up to about 1 kHz (Crawford and Fettiplace, 1980). On the other hand, the equation, in effect, treats the low-frequency region as having been added on to the primary first term, which defines the curve for high frequencies. So, if the high-frequency end of the cochlea is ''newer'' in evolutionary terms, how is it that the first term is decidedly the backbone of the relation? That is, how can we have a faithful model of the data in which the ''older'' end of the map (the archetypal low-frequency detection region) is handled by the second term, which is depicted in Fig. 1(B) essentially as an end correction?

One might suspect instead that during the process of selective adaptation to provide for high-frequency reception, the low-frequency end would have remained essentially invariant, while adaptations occurred to increase basilar-membrane stiffness at the high-frequency end, of the kinds which became extreme in the case of the odontocetes (dophins and porpoises) and bats (Ketten, 2000). This ques-

tion divides into two alternate possibilities. We know that Eq. (1) does not cover such specialist cases, except maybe in piecewise fashion, but overall the equation, despite its minimal set of parameters needed for comparing species, sheds no light on how those parameters became set for any species. Is it possible that there exists a better overall characterization of mammalian frequency-position data?

One possible resolution of the paradox is that high-frequency reception has not simply been added on, in response to selective pressure, but has instead required a fundamental change to premammalian cochlear mechanics. This idea is already embodied in the generally accepted traveling-wave theory (von Bekesy, 1960), but other views exist (Fay and Popper, 2000). An alternate resolution of the paradox may already be contained in the above-mentioned clustering of values of the coefficients seen in Table I. However, it is first necessary to disallow values of **k** equal to unity on the grounds (1) that it is unacceptable for a model to leave the position of the very lowest frequencies undefined, and (2) the more precise (dye-injection) map data (Liberman, 1982; Müller, 1996) set the values of $\mathbf{k}<1$. If, for example $\mathbf{k}\sim0.85$, then two of the three coefficients (**a** and **k**) would be appear virtually constant, at least across eutherian mammals. Therefore, Eq. (1), simple as it is, may not necessarily be the simplest model because it suggests redundancy.

While treating Eq. (1) as a precise characterization of mammalian data, it would seem that the accuracy of the points is not weighted in any way by their thresholds which, by definition, rise at the frequency limits of reception.

A. Hypothesis

The existing data permit the recasting of Eq. (1) with different boundary conditions. Taking into account the rises in hearing threshold at the frequency limits which define the frequency span, we may replace the open-ended Greenwood model with a ''closed-end'' model. We achieve this by accepting the end frequencies to be those which are empirically determined. While there is still some uncertainty as to what the low-frequencies limits are because of the slow rise in thresholds, the principle can still be illustrated for the human case by taking the low-frequency limit to be 20 Hz and the high-frequency limit to be 20 kHz.

This hypothesis gives rise to a series of corollaries which can be tested in the same operation: (1) The map is nonuniform in the sense that it compresses the tonotopic representation towards the low-frequency end. (2) The resulting ''warp'' or smooth sag in the map is advantageous for mammals, i.e., the departure from a uniform map, which comes about by specific tapering of the mechanical properties (most probably stiffness) of the cochlear partition. (3) Evolutionary pressures (maybe over millions of years) have determined the degree of warp. (4) The ''*warp factor* $\mathbf{k}'$'', i.e., the degree of sag or ''warp'' in the map, is essentially common across mammals. (5) This warp factor has important consequences for all processes in the ascending auditory pathway which depend upon the resulting tonotopicity, such as resolution of separate frequency components. (6) The warp factor found is an optimal value in the sense that it is the best trade-off across mammals between conflicting conditions and seeks to fulfill the need for a simpler, more global basis (Fay, 1992) to ''establish a minimal and/or optimal set of parameters for comparing species.''

TABLE II. Values of **A** and **a** are uniquely determined by **k** once the end frequencies are set (for human taken as 20 Hz and 20 kHz), establishing the family (Fig. 4).

k	A	a
0.000	20.00	3.000
0.500	40.00	2.699
0.700	66.67	2.478
0.800	100.00	2.303
0.879	165.43	2.086
0.92	250	1.908
0.95	400	1.707
0.980	1000.00	1.322
0.990	2000.00	1.041
0.995	4000.00	0.778

What follows is a development and testing of the hypothesis and its corollaries, and it will be seen that there is a more generalized way of regarding the map which has widespread implications. The first is that this approach achieves an escape from the severely limiting concept that the basic curve is almost straight. We recognize instead that, plotted versus log frequency the map has the property that it is curved throughout its length, and for good reason. A second is that uncertainties which may arise in the determination of map estimates, e.g., Ou *et al.* (2000), may be partly due to the fact that the Greenwood model is open ended, particularly at the low-frequency end, so that there is no stronger basis for testing the quality of any particular set of data, e.g., whether **A** has the value 3000 or 7900, than being totally at the mercy of sources of experimental error. A third application is to the area of genetic correlates of cochlear development, in particular to answering the question as to how the map becomes established during development (Echteler *et al.*, 1989; Cantos *et al.*, 2000; Brigande *et al.*, 2000).

II. METHOD

The numerical exploration here examines the notion that a better way to think of the curvature of the mammalian frequency-place map is that it is warped. The advantages of the warp concept might be achieved more elegantly in other mathematical terms (Shera *et al.*, 2002), maybe based upon cochlear mechanical model parameters. However, it is not necessary to hypothesize a basis different from Eq. (1) to achieve a general appreciation of the concept and its applications. To the contrary, retaining the form of the Greenwood model with new boundary conditions has one big advantage. This exercise requires considering the map for the human case so that the variables in Eq. (1) are fixed while considering what happens as we allow the three coefficients **A**, **a**, and **k** to vary. Once the end limits are fixed, changing the value of **k** has the very different effect of generating a family of curves with different degrees of warp. Table II shows how various degrees of warp can be achieved within the terms of the basic relation [Eq. (1)]. This is because each value of **k** uniquely sets the values of **A** and **a** as seen in Table II, except for the condition $\mathbf{k}=1$, which generates a singularity.

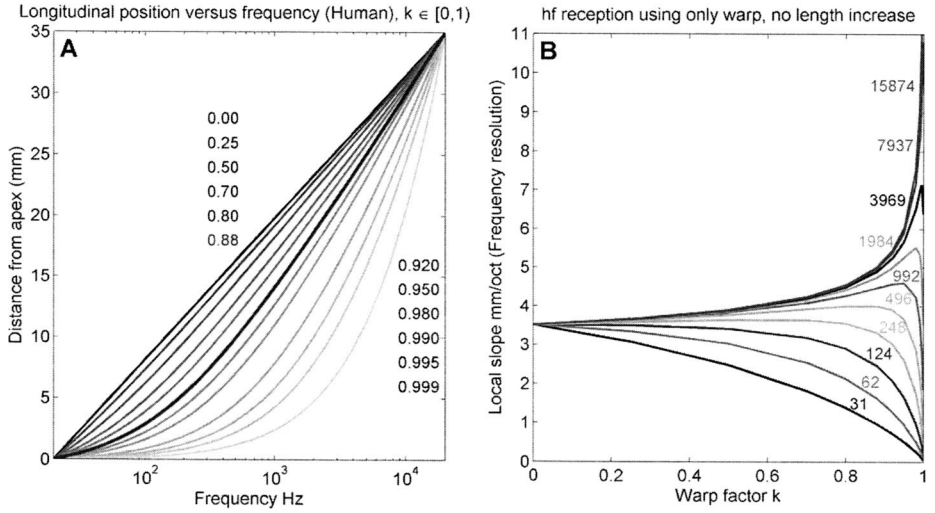


FIG. 2. Panels (A) and (B) show the effects of fixing both end coordinates of the frequency-position map, rather than allowing them to float in the data-fitting process. The justification for doing so is that it is important that the ends of the map be more or less defined by the decreases in hearing sensitivity. The new boundary conditions force the other two coefficients of Eq. (1) (**A** and **a**) to have values defined by the given values of **k** shown in the inset which generates a family of curves [panel (A)], each with a different degree of warp. The thick black line (**k**=0, **a**=3.0) in the top panel (A) is unwarped (undistorted). The most common allowed value of **k**=0.88 in the highlighted curve is thus seen to be a member of the family, each of which has different local values of frequency resolution (the gradient). As **k** rises towards unity, the frequency representation as a whole moves towards high frequencies. Panel (B) reprocesses the curves of panel (A) to show for representative frequencies, an octave apart, how the frequency resolution (mm/oct) changes with the warp factor (**k**). Very high values of frequency resolution may be achieved for high values of **k** without the need to elongate the cochlea. Conversely, the effect of changing **k** on low-frequency resolution is a steady decline.

The new constraints tie the value of **a** to that of **k**, so changing **k** changes the value of the high-frequency slope as well, unlike the original family of curves [Fig. 1(B)].

By setting the end limits to coincide with the limits of hearing sensitivity, we eliminate the possibility of having a very poorly defined lower-frequency limit. The ''add-on correction factor'' **k** is thus transformed into the generalized warp factor $\mathbf{k}'$ ($\mathbf{k}=\mathbf{k}'$). For the purposes of this exercise it suffices to assume, simultaneously, that the corresponding spatial limits are indeed the physical ends of the cochlear partition, i.e., for the human case, distances from the apex of 0 and 35 mm (normalized distances 0 and 1). An important by-product of not reformulating Eq. (1) is that any tentative conclusions about the near-constancy of **k** found across mammals carries across to the degree of warp ($\mathbf{k}'$).

III. RESULTS

Figure 2(A) shows for the human case the effect of varying warp by using the values of **k** in Table II. The order of the curves is reversed from Fig. 1(B), *viz.*, the **k**=0 case still gives rise to a uniform map throughout the length of the cochlea (3.5 mm/oct), but it is now located at the top and spans the 3 decades (**a**=3). The advantages of this approach become clear within the context that frequency resolution is directly proportional to the gradient of the frequency-place map (Greenwood, 1990). As the map becomes increasingly warped, the steeper is the local gradient at any cochlear position and the greater will be the number of independent resonators/channels serving any frequency range, giving rise to greater parallel processing capabilities (Manley, 2000). The salient feature of the left panel, therefore, is that different degrees of warp (12 values from 0 to 0.999 are shown) represent different degrees of trade-off between high- and low-frequency resolution. Figure 2(B) displays the derivatives of the family of curves for octave steps throughout the human frequency range, increasing upward from midrange of the lowest octave (20 to 40 Hz). As the warp increases ($\mathbf{k}'\rightarrow 1$), there is an accompanying steady gradual decline in the low-frequency resolution, whereas high values of resolution may be achieved at high frequencies, but only as $\mathbf{k}'>0.95$. In summary, fixing the lower-frequency limit eliminates the zero-frequency possibility as $\mathbf{k}'\rightarrow$. Now, the map becomes increasingly warped instead, leading to very high values of high-frequency resolution.

It is helpful at this stage, but not essential, to suppose that the hypothesized warping seen in the mammalian cochlear map was not present in archetypal cochleae. This is reasonable considering the morphology of lower vertebrates (Wever, 1978; Fay, 1992). It is likely the case that they are too short for the form of the map to be questioned beyond whether it is simply logarithmic or linear.

A. Conclusion 1

For lower species to have evolved high-frequency hearing, this did not, in the first instance, require elongation of the cochlea. It only required arranging specialized tapering of the cochlear mass, stiffness, and damping to bend the map appropriately to whatever is the behavioral requirement.

Figure 2(B) clearly shows that spatial separation of frequencies within an octave, by substantial fractions of a millimeter, is achievable through warping, without cochlear elongation being essential. The question of what factors led to elongation of the cochlea may be logically separated from the factors leading to map warp and is addressed below. Increasing the warp from zero initially produces a slow change in the map gradients. This slow change with $\mathbf{k}'$ continues for

low frequencies, but as it rises above 0.9 there is a rapid rise in the local gradient for the high-frequency (basal) end. Figure 2(B) shows, however, this apparent advantage comes at a price. Actually, there are two obvious costs: (1) sacrifice of low-frequency resolution, which may already have been the indispensable product of substantial selective advantage, and (2) departure from a tonotopically uniform map (e.g., the constant mm/oct of the straight line at the top of the family of curves). Greenwood actually comments on the center region of the cochlear map with uniform gradient (Greenwood, 1997). The clue which flows from Fig. 2(A) is that one can see how the spatial extent of this central region varies with warp. For the case $\mathbf{k}'=0$ we have 10 octaves of uniform map, but for $\mathbf{k}'=0.999$ we only have about 1 octave of map with the original gradient. We can define this region retaining the uniform original gradient as having a certain ''bandwidth.'' It turns out that it is a very insightful exercise to compute this bandwidth as a function of warp.

We normally think of bandwidth as a concept of a bandpass filter in which the low-frequency and high-frequency slopes are very different in the passband than in the stop bands on either side. To make progress we generalize this notion of bandwidth so that the breakpoints are established where the local gradient differs from the uniform gradient by a set difference.

The three panels of Fig. 3 show how the ''breakpoint'' frequencies are calculated after establishing limits of gradient similarity. Panel (A) shows the derivatives of the curves in Fig. 2(A). Each member of the family has a region where the gradient (mm/oct) crosses the gradient of the unwarped map (horizontal line). Panel (B) is introduced because it is clear that the resulting departure of local gradient is asymmetric about the gradient of the top unwarped member of the family. If we want to compare the effects of warp upon bandwidth, our defined low-frequency and high-frequency limits of similarity should ideally be symmetric about the gradient of the central region. Hence, the logarithm of the gradient is obtained. A criterion for similar positive and negative thresholds can then be determined (here, the gradient threshold is chosen to be ± 0.14, seen as the two horizontal dashed lines above 0 difference). The resulting dependence of the bandwidth upon map warp is shown in Fig. 3(C), where each of the log gradient curves crosses the gradient threshold limits. The higher the degree of warp, the greater the curvature of the family member and the smaller is the bandwidth of the ''unwarped'' (subthreshold) region of the map, as suggested in Fig. 2(A). Note that both the center frequency and the bandwidth vary systematically with the degree of warp.

B. Conclusion 2

For the value of warp apparent in most generalist mammalian cochleae, for which the map is precisely determined, the bandwidth seems to loosely coincide with the bandwidth of frequencies (*viz.*, ~100 Hz to 3 kHz), particularly implicated in neural phase-locking phenomena (Brugge *et al.*, 1969) and periodicity pitch determination important for speech (Sachs, 1984) and music (Tramo *et al.*, 2001).

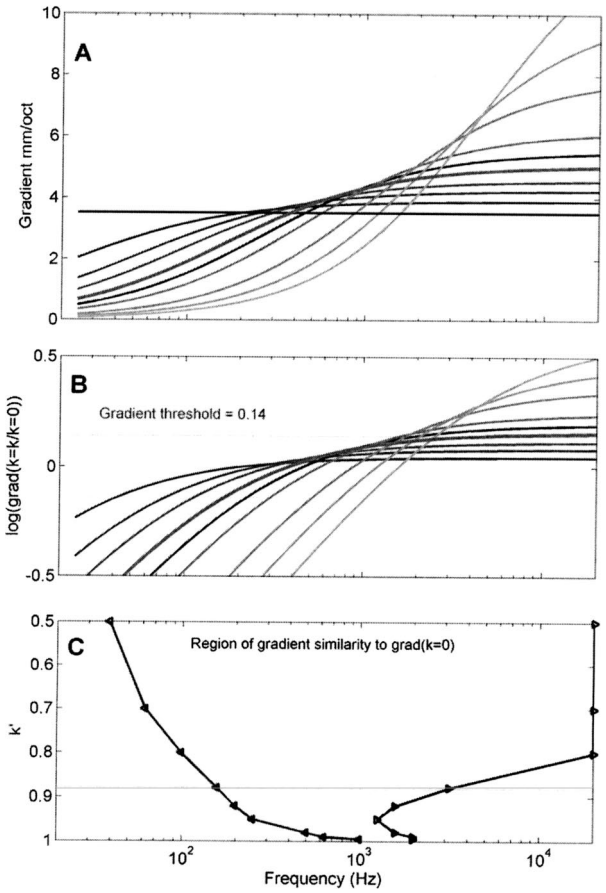


FIG. 3. This figure shows that the central region of each member of the family in Fig. 2(A) has a slope close to that of the unwarped map ($\mathbf{k}=0$) and that its bandwidth varies systematically with the warp factor **k**. Each member of the family in (A) represents the gradient of the basic family which crosses the gradient of the uniform map (horizontal straight line). Panel (B) allows a method of determining the bandwidth of this region by establishing a threshold for gradient change. Panel (C) shows the frequency range between the left and right triangles which applies for each value of **k** less than the threshold value [± 0.14 in panel (B)]. This article was triggered by the realization that, although this bandwidth is seen here to vary markedly with $\mathbf{k}'$, the common value of $\mathbf{k}'\sim 0.88$ results in a bandwidth loosely coinciding with frequencies associated with neural phase locking (important for speech and music), supporting the notion that the common value of **k** is preset according to a best compromise situation.

IV. TESTING THE NOTION OF "OPTIMAL WARP"

If there is any merit to the general notion that the mammalian tonotopy is warped, it should be possible to show, on the basis of actual map data, that there exists some form of compromise between conflicting low- and high-frequency signal-processing requirements, or more general needs. Into this category falls the need for long-term stability of the tonotopic organization and, maybe in turn, protection of the cochlea against acoustic overload. This argument follows from the notion that damage will likely occur first at points of discontinuity of basilar-membrane impedance, or its derivative.

The last of the corollaries above is that the near invariance of **k** and **a** across species is the result of an optimal trade-off for mammals. The nature of the trade-off is therefore of interest. It is clear from Fig. 2 that raising high-frequency resolution by means of warping of the map, with-

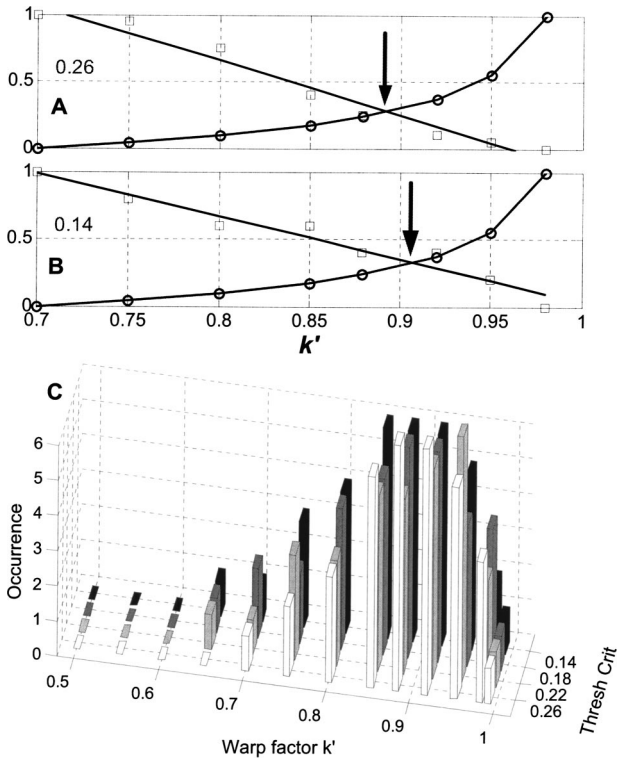


FIG. 4. If there is any merit to the generalized notion of the mammalian cochlear map being warped, it should be possible to show that there exists some form of compromise between conflicting low- and high-frequency signal-processing requirements. We hypothesize that the near invariance of **k** and **a** across species is the result of an optimal trade-off for mammals. Panel (A) shows two curves. The open squares fitted with the straight line shows the bandwidth of the apical unwarped region [from Fig. 3(C)] normalized to the interval [0,1], plotted versus value of **k**, for the criterion value 0.26. The curve defined by the circles shows the normalized frequency resolution [from Fig. 2(B)]. High values of both curves mean good map uniformity in the central region and improved high-frequency resolution, respectively. Since these two conditions vary inversely with rise in **k**, the place where they cross (arrow) represents the best trade-off. This value sits around a value of **k** of 0.9. Varying the threshold criterion of setting bandwidth of the central region does not change the condition of optimal warp [panel (B)]. Panel (C) extends the test for optimal trade-off further to include all threshold criteria and all appropriate ranges of the unwarped map gradient. The resulting histogram shows that the optimal value of **k** in the interval (0.85, 0.95) is robust. The result suggests that the mammalian cochlear map for all unspecialized ears is optimally warped for best high- and midfrequency resolution characteristics independent of length.

out simultaneously changing the length of the cochlea leads to loss of low-frequency resolution and a higher degree of curvature of the map in the central region (Fig. 3).

Figure 4 shows that the common value of $k \sim 0.85$ may be no accident. It is straightforward to plot both the bandwidth in octaves of this central region defined in Fig. 3 along with the high-frequency slope [Fig. 2(B)] versus the warp factor $\mathbf{k}'$. In panels (A) and (B), the curve with the open circles represents high-frequency resolution as a function of $\mathbf{k}'$. The straight line fitted to the open squares shows how the limit of map uniformity varies with $\mathbf{k}'$. Both of these curves are normalized to vary in the interval [0, 1] in order to directly compare them. The two upper panels represent two different thresholds (0.14 and 0.26) for ''gradient-departure'' levels setting the ''bandpass'' breakpoints. In each panel the $\mathbf{k}'$ value where the lines intersect represents the warp factor condition for the *best trade-off between two conflicting conditions: maximizing the high-frequency resolution and maximizing the "bandwidth" of the uniformly mapped region*. No matter which tolerance is considered, the optimal values of $\mathbf{k}' \sim 0.9$. Since the reference gradient for this comparison is somewhat arbitrary [3.5 mm/oct; see Fig. 3(A)], it has further been shown that changing the value of the reference gradient for the comparison shifts the band limits but does not change the estimate of the optimal value of **k**, which remains close to 0.9. When all the likely combinations of reference gradient and gradient departure are tested grouped together, the result is the series of histograms shown in Fig. 4(C). Not only is there a strong clustering of values of optimal trade-off in the range $\mathbf{k}'$ in the interval (0.85 to 0.95), there is a steep decline for higher values of $\mathbf{k}'$ and no incidences of values of $\mathbf{k}' < 0.65$.

A. Conclusion 3

There appears to be an optimal trade-off between two conflicting conditions: increasing high-frequency resolution and limiting the decline of the bandwidth of the central most-uniform-map region. Indeed, this conclusion appears to independently parallel the condition of ''cochlear compromise'' (Zweig *et al.*, 1976) between two conflicting requirements—maximizing both high-frequency resolution and avoidance of reflections of waves along the basilar membrane.

B. Conclusion 4

Subject to caveats, the optimal value of $\mathbf{k}'$ coincides with that value of **k** found to be most common for mammals. The correspondence is not perfect but is indeed remarkable. This has been achieved using the Greenwood relation, which guarantees that the map is smooth.

C. Conclusion 5

Mammals have arrived at a representation of frequency versus position along the cochlear partition which is an optimal trade-off between conflicting requirements. These requirements are: (1) high-frequency reception and resolution is enhanced; (2) the map remains smooth; and (3) the center region is maximally uniform (in terms of the longest possible length of near-equal spacing for adjacent octaves).

Returning to Fig. 1, it is possible now to view the shape of the map in a new context. In particular, the near-common value of the high-frequency slope finally has an explanation in terms of evolutionary forces. No reason has ever been supplied as to why the high-frequency slope should have any value in particular for any species, or why the frequency limits are not tied to body size, as is the frequency range (Ketten, 2000).

V. EVOLUTIONARY PRESSURE TO ELONGATE THE COCHLEA

If high-frequency resolution could be achieved by warping the map, why did mammals develop long cochleae, and what sets the length in any species?

Figure 5 illustrates that this manifestation of cochlear compromise provides an explanation. The warp ensures that the high-frequency slope (versus log frequency) may rise,

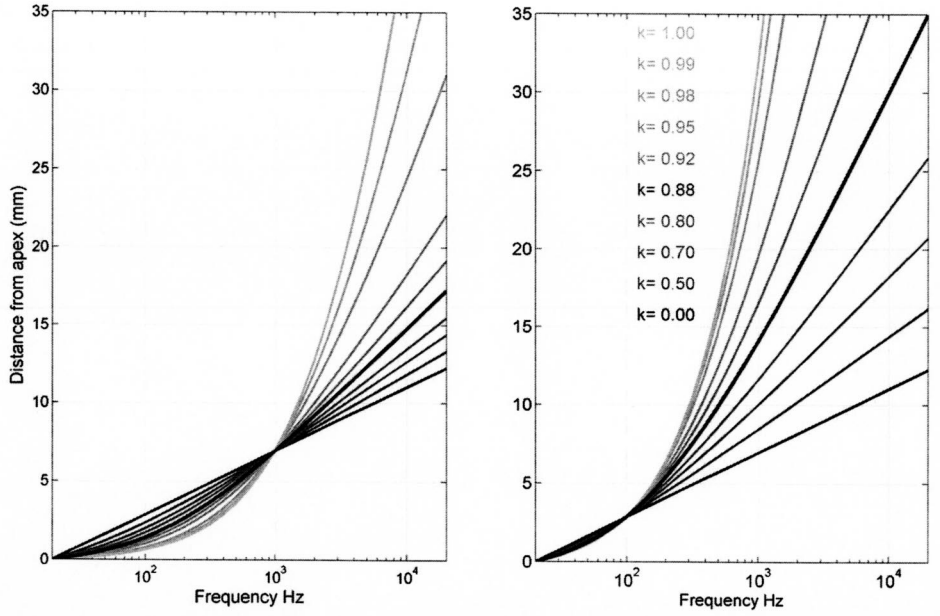


FIG. 5. Panels (A) and (B) demonstrate evolutionary pressure to elongate the cochlea. If warping the cochlear map can increase high-frequency resolution, why do cochleae vary in length? Panel (A) shows that if an archetypal short cochlea gains high-frequency resolution by evolving warp, it simultaneously loses low-frequency resolution [see again Fig. 2(B)]. If the existing low-frequency resolution has species advantage, selective pressure will act so as to offset that disadvantage. Panel (B) shows that simply elongating the cochlea can limit the loss of low-frequency resolution. The existence of the warp means that the whole cochlea does not need to elongate to the extent which would provide whatever required resolution; space saving can be achieved by warping the map as well, reducing the final extension.

but that at low frequencies the slope must drop correspondingly [Fig. 2(B)]. Taking a hypothetical example, the left panel (A) shows how, for a short cochlea (e.g., bird, reptile, even echidna (Ladhams and Pickles, 1996) handling a frequency range up to 10 kHz (say 4 to 7 mm in length) might tend to become warped to transduce higher frequencies. The warping process leads to reduced slope of the low-frequency range (the family of curves below the straight line, $\mathbf{k}'=0$). This is a poor trade-off which does not constitute evolutionary advancement if existing low-frequency resolution is lost. The right panel (B) shows that if the warp is accompanied by elongation of the cochlea to exactly counteract the decrease in low-frequency resolution, loss of low-frequency performance can be avoided. However, the smooth map constraint requires that the whole cochlea is elongated in proportion. Figure 5(B) shows the result. The loss of low-frequency resolution is essentially avoided and the map elongates to (in the case of human) 35 mm long (the curve for $\mathbf{k}\sim0.88$).

A. Conclusion 6

The *lower* the (required) receptive frequency range, the *longer* must be the resulting cochlea. It is therefore simplistic to regard the archetypal cochlea as having elongated to provide high-frequency reception or even a given number of octaves frequency range. Economy dictates that limited increase in frequency resolution can be achieved in the first place purely by warping the map. Elongation occurred because even the optimally warped cochlea could not meet evolutionary demands (e.g., if the bandwidth of the uniform central region is too narrow, or much higher frequency channel separation is required for noisy environments).

B. Conclusion 7

Warping is an indispensable requirement for minimizing space—providing maximally compact cochleae by minimizing the length increase necessary. In other words, evolutionary pressure would first seek to extract the best compromise over the desired frequency range and ''top up'' the frequency resolution by elongating the cochlea, but only so long as to maintain optimal trade-off.

C. Conclusion 8

There exists a frequency span limitation for mammals with generalist cochleae (i.e., echolocating animals excluded), which is about 10 octaves, and this appears to be a hard limit. The reason stems directly from Greenwood's discovery that $\mathbf{a}\sim2.1$ is not exceeded by any species in which the map is smooth and free of kinks (discontinuous derivatives). The reason the limit is hard stems directly from the animal data. The octave range limit cannot be exceeded by further extending (and coiling) the cochlea, because 2.1 (maximum 2.2) is the value pertaining to the *normalized unity-length* cochlea. As observed by Ketten, odontocetes' hearing may cover 12 octaves but this comes at the price of introducing a discontinuity into the map, or into the map gradient, resulting in specialized anisomorphic cochleae.

VI. DISCUSSION

The form of the frequency-position map has its origin in the shape of the critical band curve (Fletcher, 1940). The history of attempts to match frequency or pitch versus place has given rise to many estimations maps including early work (Koenig, 1949; Fant, 1973). The map has been extensively characterized and modeled (Zwicker and Terhardt, 1980; Traunmüller, 1990). With the exception of the first two early attempts, the modern estimates tend to be very similar up to about 3 kHz with the long-considered characterization (Greenwood, 1990) used as the starting point here. The warp hypothesis suggests by contrast that critical bands are established as a result of the optimal trade-off between conflicting requirements.

None of this exploration calls into question the validity of Greenwood's formulation for adequately characterizing the frequency-position map across mammals for the bulk of

the available data. Moreover, none of the formulation loses any of its validity in approach by fixing the boundary conditions on the equation. To the contrary, this formulation adds credibility to the equation by virtue that the end-frequency limits are tied to where the sensitivity rises or by the corresponding physical limits of the cochlear partition. However, while the warp hypothesis should be tested against original morphological data, this task is beyond the scope here.

What is remarkable in light of Masterton's article is that the degree of warp for all the larger species listed (Table I) is about the same since the value of coefficient **a** is about the same; indeed, there is a ceiling of maximum value of $\mathbf{a}=2.15$. It follows from the constraints that the increased sensitivity at lower frequencies in larger mammals is necessarily accompanied by a loss of high frequencies, unless high-frequency specializations such as basilar-membrane stiffening structures are introduced. It is tempting to speculate that the evolutionary process was necessarily constrained to weight the importance of low- and high-frequency information and, in humans, it chose to weight the low-frequency information as more important.

Two things follow from the evident 10-octave limitation. The first is that the notion of a grand basilar membrane (Ketten, 2000) from which a 10-octave segment is selected for any species has new merit. However, the notion of adjusting the taper is something which doesn't necessarily require long-term evolutionary forces. The optimal trade-off between frequency resolution in the various frequency zones is something which might even be dynamic—the result of OHC activity. Second, Masterton's hypothesis of the chronological loss of high frequencies might probably be relaxed. His speculation is that humans lost very high-frequency reception because as the head size grew the interaural time differences sufficed so that high frequencies were no longer needed. The alternate hypothesis here follows from the notion that the 10-octave limit is imposed by the need for a smooth map. Making more use of low frequencies necessitates loss of very high frequencies if the propagation of the traveling wave must be free of irregularities. It is not that large mammals with very low-frequency (infrasound) reception (e.g., great whales (Ketten, 1994) and elephants (Heffner and Heffner, 1982; O'Connell *et al.*, 1997) could not find a good use for higher frequencies and therefore ''lost'' high frequencies over time. Indeed, the odontocetes did get around the problem by introducing stiffening structures, such as the secondary outer-body lamina (Ketten, 2000). This exercise has illustrated that the form of the map likely has a specific origin which relates to forces which presumably came into play as the advantages of high-frequency reception evolved and which although sought has not had any extensive explanation before now.

A. Very highly warped maps and echolocating mammals

Considering the case of echolocating mammals (bats and cetaceans), it would appear that the constraints to provide a very wide bandwidth and very high-frequency resolution cannot be met while retaining a smooth map. Certainly some bats have regions receptive to usual audio frequencies, but they possess a ''fovea'' in which the gradient is up to 60 mm/oct (Kössl and Vater, 1985). Ketten (2000) comments that while there are many sophisticated mathematical models of cochlear function, few are aimed at understanding species-specific variations. It may be consistent with our notion of warp that the price of higher resolution than the ceiling imposed by an $\mathbf{a}=2.1$ is a distinctly kinked and frequently discontinuous map (Henson and Henson, 1979; Vater and Lenoir, 1992) and not just the lack of need of intermediate frequencies as is generally presumed. We have demonstrated that situations may arise such that the advantages of maintaining optimal warp are outweighed by the advantages of rapid echolocation. Very high degrees of map curvature are conceivable in the limit, such as a map where $\mathbf{k}'=1$, resulting in a map which has a square corner, zero midregion bandwidth, zero low-frequency resolution, and infinite high-frequency resolution. Practically such a map can only be approached, and only to the extent that very rapid changes in cochlear mechanical parameters, e.g., mass, stiffness or damping can be achieved. Indeed, mammals with specialized cochleae do exist to provide very much higher high-frequency resolution but only at the cost when the map becomes discontinuous, or its gradient becomes discontinuous—a sufficient condition to establish reflections along the basilar membrane.

B. Cochleae of large mammals

The more pressing the need to accommodate low frequencies, the longer the length of the cochlea. The rapid rise in the required length with increase in value of **k** provides a further reason (other than space limitations) as to why the mammalian cochlea cannot attain such lengths as to achieve arbitrary increases in high-frequency performance. Accordingly, in early animals which have made good use of its existing low-frequency resolution, but needed high-frequency capability, the cochlea as a whole has increased length in proportion to the degradation of low-frequency resolution. The cases of human, cow, and elephant (Table I) are receptive to very low frequencies and so the optimal warp hypothesis will ensure that the cochleae are proportionally longer (up to 60 mm). Whales which are sensitive to infrasound will predictably have longer cochleae again. Indeed, Ketten quotes 71 mm in the case of the blue whale. This may not necessarily be because the animal's cochlea scales to body mass, but because of the reception of very low ocean-transmitted frequencies and the need for uniform mapping in the middle of the hearing range.

C. Gender dependence

While convincing counterarguments may be advanced, it is at least qualitatively consistent with the above argument for elongation that cochleae of human males are 15% longer than the cochleae of females (Sato *et al.*, 1991)—the fundamental speech frequency is lower. Until now no explanation has been offered.

D. Cochleae of very small mammals including rodents

The value of **k** given for Mongolian gerbil is 0.63 and 0.56 for opossum. This lower value could possibly be due to a bad estimate, e.g., due to few data points in the low-frequency region, or it may be a fair estimate. The question arises, is there any room for the value to be consistent with the optimum-warp hypothesis? The value of 0.63 therefore is at the low end of the values of **k** seen in Fig. 4(C), but not inconsistent with the hypothesis.

In the cases of mice, rats, and opossum, the values of **a** are apparently more serious departures to be reconciled with the generalized model. Some of the departure may be explained by the different methods (normal or edge frequency) used to determine distance (Ou *et al.*, 2000). In these species notably the cochlea is much shorter and the number of octaves of audibility lower. As we saw in Table I, mice, rats, and opossum appear to have an **a** value of about 1.0 rather than the value of 2.1. Their estimated cochlear lengths are shorter and their frequency spans are only ~7 octaves. These cases are qualitatively consistent with the warp theory. Masterton hypothesized that they did not possess low-frequency reception and the resulting frequency span required was therefore less than the maximum possible. So, for the sake of economy, their cochleae only extended to the extent that the low-frequency characteristics they possessed were counteracted to meet their high-frequency requirement. It is consistent that the warp is less and the cochleae are shorter.

This exercise has also demonstrated that local increase in frequency selectivity can be achieved without length increase, simply by adjusting the tapering of mechanical parameters to warp the map. The significantly increased frequency selectivity in humans (Shera *et al.*, 2002) is directly predictable from the notion that it is basically tied to spatial constraints (Ehret, 1978; LePage, 1987).

VII. SUMMARY

The form of the Greenwood relation for mammalian tonotopicity initially appears paradoxical by virtue that the equation seems to cater to low-frequency curvature by applying a ''correction'' at what seems to be the ''wrong'' end of the map in evolutionary terms. Without loss of precision for its fit to many data, the paradox is resolved by changing the boundary conditions to fix the end-frequency limits in accordance with behavioral threshold data. This leads directly to the hypothesis that a better way to think about the form of the map is that it is warped. That is, its local gradient varies systematically with position from apex to base. In so doing it shows that the Greenwood relation actually represents a near-symmetric trend so that the correction thus applied to the low-frequency end actually applies neatly to the high-frequency end as well.

The hypothesis provides added meaning in terms that this warp actually evolved in the process of mammalian adaptation to high-frequency reception. Second, once the form of the map is questioned from an evolutionary context, the notion of warp becomes attractive, particularly when it is shown that the common value of **k** coincides with the value which represents a state of warp for optimal trade-off. Third, ''optimal warp'' carries the added implication that the curve is smooth and that the length of the central region is maximally close to being uniform, for as long a distance along the cochlear partition as possible. Fourth, the concept of warp is seen as an important space-saving mechanism in the process of acquiring cochleae which transduce a wide range of frequencies. The notion of optimal warp leaves many aspects to explore, particularly the aspect that some measure of the warp may be due to cochlear activity, accounting for discrepancies in the passive stiffness range necessary for the full range of audio frequencies. The ultimate significance of this warp theory is at least fourfold. First, it provides a natural, encapsulated idea to reconcile with map data derived from various species. Second, it encompasses the ''chicken-and-egg'' problem as to whether critical band placement occurred as a result, i.e., subsequent to, the evolutionary optimization of the map or vice versa. Third, mathematical models may benefit from the process to arrive independently at the form of the map—rather than presuming it at the outset in order to set segmental parameters. Last, it may ultimately yield an added basis for regarding data produced by genetic studies of cochlear development, *viz.*, which end of the cochlea has extended (1) during the evolution process and (2) during maturation.

ACKNOWLEDGMENTS

I thank Dr. Donald Greenwood for his challenging discussions and critique of the first draft of this article; also Professor N. H. Fletcher and Dr. H. Dillon for their comments on later versions, Dr. N. Murray for proofreading the final version, and J. Seymour for assistance with graphics. I am particularly indebted to the reviewers for their extensive constructive comments.

Bohne, B. A., and Carr, C. D. (**1979**). ''Location of structurally similar areas in chinchilla cochleas of different lengths,'' J. Acoust. Soc. Am. **66**, 411–414.

Brigande, J. V., Kiernan, A. E., Gao, X., Iten, L. E., and Fekete, D. M. (**2000**). ''Molecular genetics of pattern formation in the inner ear: Do compartment boundaries play a role?'' Proc. Natl. Acad. Sci. U.S.A. **97**, 11700–11706.

Brugge, J. F., Anderson, D. J., Hind, J. E., and Rose, J. E. (**1969**). ''Time structure of discharges in single auditory nerve fibers of the squirrel monkey in response to complex periodic sounds,'' J. Neurophysiol. **32**, 386–401.

Cantos, R., Cole, L. K., Acampora, D., Simeone, A., and Wu, D. K. (**2000**). ''Patterning of the mammalian cochlea,'' Proc. Natl. Acad. Sci. U.S.A. **97**, 11707–11713.

Crawford, A. C., and Fettiplace, R. (**1980**). ''The frequency selectivity of auditory nerve fibres and hair cells in the cochlea of the turtle,'' J. Physiol. (Lond.) **306**, 79–125.

Dallos, P. (**1970**). ''Low-frequency auditory characteristics: Species dependence,'' J. Acoust. Soc. Am. **48**, 489–499.

Echteler, S. M., Arjmand, E., and Dallos, P. (**1989**). ''Developmental alterations in the frequency map of the mammalian cochlea,'' Nature (London) **341**, 147–149.

Ehret, G. (**1978**). ''Stiffness gradient along the basilar membrane as a basis for spatial frequency analysis within the cochlea,'' J. Acoust. Soc. Am. **64**, 1723–1726.

Fant, G. (**1973**). *Speech Sounds and Features* (MIT Press, Cambridge, MA).

Fay, R. R., and Popper, A. N. (**2000**). ''Evolution of hearing in vertebrates: The inner ears and processing,'' Hear. Res. **149**, 1–10.

Fay, R. (**1992**). ''Structure and function in sound discrimination among vertebrates,'' in *The Evolutionary Biology of Hearing*, edited by D. Webster, R. Fay, and A. N. Popper (Springer, New York), pp. 229–267.
Fletcher, H. (**1940**). ''Auditory patterns,'' Rev. Mod. Phys. **12**, 47–65.
Greenwood, D. D. (**1961**). ''Critical bandwidth and the frequency coordinates of the basilar membrane,'' J. Acoust. Soc. Am. **33**, 1344–1356.
Greenwood, D. D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.
Greenwood, D. D. (**1996**). ''Comparing octaves, frequency ranges, and cochlear-map curvature across species,'' Hear. Res. **94**, 157–162.
Greenwood, D. D. (**1997**). ''The Mel Scale's disqualifying bias and a consistency of pitch-difference equisections in 1956 with equal cochlear distances and equal frequency ratios,'' Hear. Res. **103**, 199–224.
Heffner, R. S., and Heffner, H. E. (**1982**). ''Hearing in the elephant (*Elephas maximus*): Absolute sensitivity, frequency discrimination, and sound localization,'' J. Comp. Physiol. Psychol. **96**, 926–944.
Henson, M. M., and Henson, O. W. (**1979**). ''Some aspects of structural organization in the cochlea of the bat, *Pteronotus parnellii*,'' Scan Electron Microsc. **3**, 975–982.
Kermack, K. (**1989**). ''Palaeontology. Hearing in early mammals,'' Nature (London) **341**, 568–569.
Ketten, D. R. (**1994**). ''Functional analyses of whale ears: Adaptations for underwater hearing,'' IEEE Trans. Underwater Acoust. **1**, 264–270.
Ketten, D. R. (**2000**). *Cetacean Ears*; SHAR Series for Auditory Research (Springer, New York), pp. 43–108.
Koenig, W. (**1949**). ''A new frequency scale for acoustic measurements,'' Bell Lab. Rec. **27**, 299–301.
Kössl, M., and Vater, M. (**1985**). ''The cochlear frequency map of the mustache bat, *Pteronotus parnellii*,'' J. Comp. Physiol. [A] **157**, 687–697.
Ladhams, A., and Pickles, J. O. (**1996**). ''Morphology of the monotreme organ of Corti and macula lagena,'' J. Comp. Neurol. **366**, 335–347.
LePage, E. L. (**1987**). ''A spatial template for the shape of tuning curves in the mammalian cochlea,'' J. Acoust. Soc. Am. **82**, 155–164.
Liberman, M. C. (**1982**). ''The cochlear frequency map for the cat: Labeling auditory-nerve fibers of known characteristic frequency,'' J. Acoust. Soc. Am. **72**, 1441–1449.
Manley, G. A. (**2000**). ''Cochlear mechanisms from a phylogenetic viewpoint,'' Proc. Natl. Acad. Sci. USA **97**, 11736–11743.
Masterton, B., Heffner, H., and Ravizza, R. (**1969**). ''The evolution of human hearing,'' J. Acoust. Soc. Am. **45**, 966–985.
Müller, M. (**1991**). ''Frequency representation in the rat cochlea,'' Hear. Res. **51**, 247–254.
Müller, M. (**1996**). ''The cochlear place-frequency map of the adult and developing Mongolian gerbil,'' Hear. Res. **94**, 148–156.
Müller, M., Wess, F. P., and Bruns, V. (**1993**). ''Cochlear place-frequency map in the marsupial *Monodelphis domestica*,'' Hear. Res. **67**, 198–202.
Ou, H. C., Harding, G. W., and Bohne, B. A. (**2000**). ''An anatomically based frequency-place map for the mouse cochlea,'' Hear. Res. **145**, 123–129.
O'Connell, C. E., Arnason, B. T., and Hart, L. A. (**1997**). ''Seismic transmission of elephant vocalizations and movement,'' J. Acoust. Soc. Am. **102**, 3124.
Popper, A. N., and Fay, R. R. (**1997**). ''Evolution of the ear and hearing: Issues and questions,'' Brain Behav. Evol. **50**, 213–221.
Robles, L., and Ruggero, M. A. (**2001**). ''Mechanics of the mammalian cochlea,'' Physiol. Rev. **81**, 1305–1352.
Sachs, M. B. (**1984**). ''Neural coding of complex sounds: Speech,'' Annu. Rev. Physiol. **46**, 261–273.
Sato, H., Sando, I., and Takahashi, H. (**1991**). ''Sexual dimorphism and development of the human cochlea. Computer 3-D measurement,'' Acta Oto-Laryngol. **111**, 1037–1040.
Shera, C. A., and Guinan, Jr., J. J. (**1999**). ''Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs,'' J. Acoust. Soc. Am. **105**, 782–798.
Shera, C. A., Guinan, Jr., J. J., and Oxenham, A. J. (**2002**). ''Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements,'' Proc. Natl. Acad. Sci. USA **99**, 3318–3323.
Skinner, M. W., Ketten, D. R., Holden, L. K., Harding, G. W., Smith, P. G., Gates, G. A., Neely, J. G., Kletzker, G. R., Brunsden, B., and Blocker, B. (**2002**). ''CT-derived estimation of cochlear morphology and electrode array position in relation to word recognition in Nucleus-22 recipients,'' J Assoc Res Otolaryngol **3**, 332–350.
Tramo, M. J., Cariani, P. A., Delgutte, B., and Braida, L. D. (**2001**). ''Neurobiological foundations for the theory of harmony in western tonal music,'' Ann. N.Y. Acad. Sci. **930**, 92–116.
Traunmüller, H. (**1990**). ''Analytical expressions for the tonotopic sensory scale,'' J. Acoust. Soc. Am. **88**, 97–100.
Vater, M., and Lenoir, M. (**1992**). ''Ultrastructure of the horseshoe bat's organ of Corti. I. Scanning electron microscopy,'' J. Comp. Neurol. **318**, 367–379.
von Bekesy, G. (**1960**). *Experiments in Hearing* (McGraw-Hill, New York).
Wever, E. G. (**1978**). *The Reptile Ear: Its Structure and Function* (Princeton University Press, Princeton).
Zweig, G., Lipes, R., and Pierce, J. R. (**1976**). ''The cochlear compromise,'' J. Acoust. Soc. Am. **59**, 975–982.
Zwicker, E., and Terhardt, E. (**1980**). ''Analytical expressions for critical-band rate and critical bandwidth as a function of frequency,'' J. Acoust. Soc. Am. **68**, 1523–1525.

Estimating bone conduction transfer functions using otoacoustic emissions[a)]

David W. Purcell[b)]
Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Ontario M5S 3G9, Canada and Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario M5S 3G8, Canada

Hans Kunov
Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Ontario M5S 3G9, Canada

William Cleghorn
Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario M5S 3G8, Canada

(Received 18 December 2002; revised 19 April 2003; accepted 22 April 2003)

A technique for estimating the nonparametric bone conduction transfer function using distortion product otoacoustic emissions (DPOAEs) is presented. Individual transfer functions were obtained using DPOAEs recorded from a single ear of five normal-hearing adults. Repeatability of the technique was investigated by performing measurements on at least three dates. Functions were reasonably repeatable, and were unique to each individual as expected from subjective measurements. Input force and DPOAE measurements were made for each individual, and a model of the auditory periphery representative of an average person was employed. The technique is objective and requires only passive cooperation, but robust DPOAEs are needed and the measurement time can be onerous for a wide frequency band or fine frequency resolution. With appropriate adjustments to the model of the auditory periphery, the method could be applied with animal models. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1582436]

PACS numbers: 43.64.Jb, 43.64.Bt [BLM]

I. INTRODUCTION

An objective, noninvasive technique is presented to estimate bone conduction signals at the cochlea. Combining these estimates with the measured force administered to the head, an experimental nonparametric bone conduction transfer function can be obtained.

Otoacoustic emissions (OAEs) are acoustic signals measurable in the ear canal of most humans with normal hearing, and can be used as an objective indicator of processes within the cochlea of the inner ear (Kemp, 1978; Probst *et al.*, 1991). There are several classifications of OAEs and of particular interest here is the distortion product OAE (DPOAE) elicited using two pure tones of frequencies f_1 and f_2. The two tones are typically related by $f_2/f_1=1.22$ (Harris *et al.*, 1989). An emission can be recorded at the cubic distortion frequency $f_{dp}=2f_1-f_2$ from most ears with normal hearing. The magnitude of this emission varies nonlinearly with the magnitudes of the f_1 and f_2 stimuli, labeled L_1 and L_2, respectively (Whitehead *et al.*, 1995). Other distortion products are present in human ears, but the $2f_1-f_2$ DPOAE gives the most robust response (Lonsbury-Martin *et al.*, 1990; Probst *et al.*, 1991), and is therefore employed exclusively in the work described here.

Audiometric bone conduction (BC) refers to delivery of an auditory stimulus to the cochlea through application of a mechanical vibration (usually to the head), using a transducer designated as a bone conductor (Békésy, 1954). The basilar membrane responds the same to a given stimulus that arrives through either bone or air transmission (Békésy, 1955; Bárány, 1938; Lowy, 1942; Tonndorf, 1966) with a traveling wave from base to apex. Transmission through bone to the cochlea is complex and includes a number of distinct linear (Stenfelt *et al.*, 2000; Håkansson *et al.*, 1996, 1986) mechanisms, although nonlinear activity of the stapedius muscle, tactile stimulation, and bone conductor distortion may unbalance the growth of loudness between air and BC at higher stimulus levels (Stenfelt and Håkansson, 2002). The two major mechanisms of BC are referred to as inertial and compressional (Bárány, 1938; Kirikae, 1959; Tonndorf, 1966). Inertial BC, believed to be dominant at lower frequencies (Kirikae, 1959), refers to stimulation of the cochlea primarily through relative movement of the ossicles of the middle ear. Compressional BC, dominant at higher frequencies (Kirikae, 1959; Khanna *et al.*, 1976), involves compression and distortion of the bony labyrinth of the cochlea itself (Tonndorf, 1962). A third mechanism of BC, referred to as osseo-tympanic, is the radiation of sound by the ear canal walls. This energy can reach the cochlea via the middle ear as if it were a normal airborne sound. All three mechanisms

[a)]Part of this work was previously presented at the 26th Canadian Medical and Biological Engineering Society Conference in Halifax in October 2000.

[b)]Author to whom correspondence should be addressed; currently affiliated with the Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, Ontario M6A 2E1, Canada. Electronic mail: d.purcell@utoronto.ca

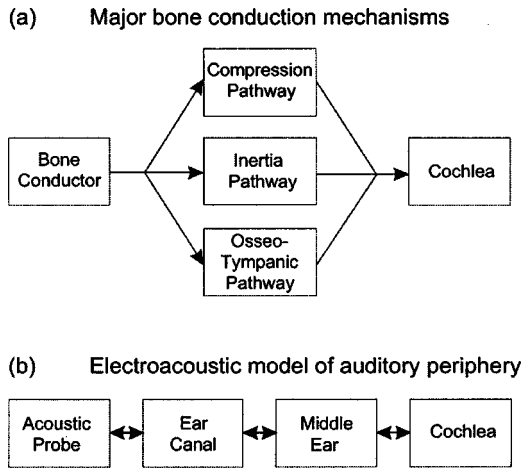


FIG. 1. Schematic of the major BC pathways used to deliver auditory stimuli from the transducer to the cochlea (a). System overview of the electroacoustic model of the human auditory periphery implemented from models in the literature (b).

deliver BC signals that interfere at the cochlea to produce a net stimulus [see Fig. 1(a)].

Extensive research has been performed on the mechanical vibration characteristics of the skull, using dry skulls, cadavers, or individuals with titanium implants for hearing aids (Håkansson *et al.*, 1994, 1986). These studies quantify some characteristics of the BC pathway, but they cannot illustrate the total transfer of energy to the fluid of the cochlea itself, as the response measurements are from the skull alone (Stenfelt *et al.*, 2000). It is desirable to know how a BC stimulus is transformed from the surface of the head to the cochlea. This can be represented as a nonparametric transfer function (TF), where the estimated signal at the cochlea is divided by the measured signal applied to the head in the frequency domain. Large intersubject differences are anticipated in the BC TF from knowledge gained through subjective measurements (Dirks, 1964; Weston *et al.*, 1967); however the individual BC TF would be valuable for basic auditory research that could lead to improvements in clinical diagnosis and design of assistive devices.

BC and OAEs have separate histories of use in clinical audiology and auditory research. A few researchers (Rossi *et al.*, 1988; Collet *et al.*, 1989; Purcell *et al.*, 1998) have reported preliminary studies where emissions were stimulated by BC, but their union can be further exploited. It is not possible to directly measure the BC signal that reaches the cochlea noninvasively. The method described here uses DPOAEs and models of the auditory periphery to estimate the magnitude and phase of applied sinusoidal BC signals at the cochlea. The technique is noninvasive, objective, and estimates the total BC signal acting on the cochlea *in vivo*.

II. METHODS

To evaluate the technique, BC TFs were calculated for the frequency band 2 to 6 kHz in 100-Hz increments for five normal-hearing participants (four male, thresholds <20 dB HL for audiometric frequencies 1–8 kHz), aged 22–27 years, on at least three different dates. An otoscope was employed to inspect the ear canal and eardrum of each individual, but no measurements of acoustic impedance or air–bone gap were performed. On each date, the acoustic probe was sealed in the ear canal using the same rubber tip, and the bone conductor was placed in a similar position on the ipsilateral mastoid; however, the exact position and orientation were necessarily unique for each session. Measurements using air conduction (AC) were always made first to allow time for the bone conductor to settle, in case there was any significant initial movement of fluids and tissue under the point of application (Flottorp and Solberg, 1976). Total test duration was 40 to 60 min. Some individuals found the bone conductor became uncomfortable, and measurements were then terminated at a reduced upper frequency limit (usually 5 kHz).

DPOAEs were measured in the ear canal distal to the tympanic membrane using an acoustic probe, constructed by Poul Madsen Medical Devices, that contained a miniature response microphone and two stimulus speakers (transducers were manufactured by Knowles Electronics). The acoustic probe was always employed for delivery of the f_1 stimulus (typically at 40 to 50 dB SPL; the actual value was determined on an individual basis as discussed below). The probe was used to produce f_2 (typically at 20 to 55 dB SPL) only during AC measurements. The bone conductor was a piezo electric device (Purcell *et al.*, 1998, 1999) driving through an impedance head (integrated force and acceleration sensors manufactured by Brüel & Kjær, Type 8001, contact area 143 mm^2), and held in place by a nonadjustable spring steel headband. The headband provided a nominal application force on the mastoid process of 5 N, but the actual force varied with individual head size. The bone conductor with coupled impedance head was used to deliver the f_2 tone ipsilateral to the acoustic probe during BC measurements (typically at 10 to 45 dB HL).

For each individual measurement, the acoustic and (if applicable) force signals were digitized with 16-bit resolution at a rate of 22 050 Hz. Steady-state data were recorded for near 1.5 s and subdivided into 16 segments (containing an integer number of stimulus and response cycles) that were averaged in the time domain. The response spectrum was calculated using a fast Fourier transform (FFT) of the averaged data (window length 92.9 ms or 2048 samples). In the frequency domain, the mean acoustic noise floor (NF) and NF standard deviation were estimated by averaging values in three frequency bins, of resolution 10.8 Hz, above and below the f_{dp} bin. An acoustic signal in frequency bin f_{dp} was considered statistically different from noise, and therefore a valid DPOAE, if it met two criteria. These were that the magnitude in bin f_{dp} must be at least 3 dB above the estimated NF and greater in magnitude than the NF plus two standard deviations. Magnitude and phase information were stored for subsequent analysis for the acoustic stimuli (f_2 and f_1) and DPOAE (f_{dp}), and the mass compensated force of the bone conductor at frequency f_2.

A. Analysis

An estimate of the net BC signal at the cochlea is required to approximate the experimental BC TF from the surface of the head to the cochlea. The BC signal that will be

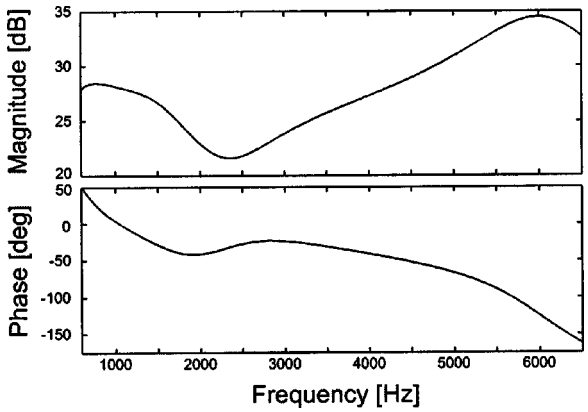


FIG. 2. Transfer function showing the pressure magnitude and phase responses from the acoustic probe tip to the base of the cochlea determined using the electroacoustic model of the human auditory periphery. The upper magnitude plot shows the gain, and the lower phase plot shows the phase shift across the ear canal and middle ear systems.

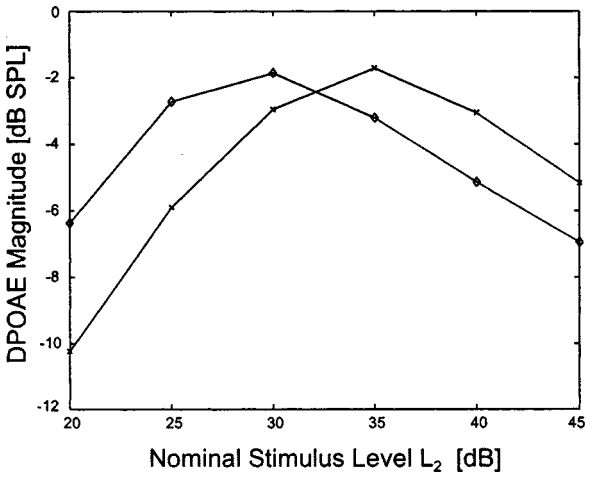


FIG. 3. Examples of AC (diamonds) and BC (symbol x) IOgrams generated by varying stimulus magnitude L_2 while all other stimulus parameters were fixed ($f_2 = 3704$ Hz, $f_1 = 3036$ Hz, $L_1 = 50$ dB SPL). The nominal AC level L_2 was calibrated in dB SPL, whereas the bone conductor was nominally calibrated for the experiment in force units with an arbitrary force reference. This force reference was chosen such that a single L_2 index scale could be used for both AC and BC (i.e., a given L_2 elicited similar level DPOAEs, and was behaviorally of similar level, through AC and BC for a typical participant). At the cochlea for this particular individual, the two modalities deliver stimulus f_2 at different levels for a given value of index L_2, resulting in DPOAEs of different magnitude. The bone conductor can be calibrated to an equivalent dB SPL measured in the ear canal for this individual by aligning the two IOgrams.

estimated ''at the cochlea'' will subsequently be defined as the pressure at the base of the cochlea, or the signal applied at the input to the cochlear model described below. This is an appropriate choice, because the basilar membrane will respond identically with a traveling wave to this pressure, regardless of whether it is generated by an AC signal from the middle ear, or the interference of BC signals arriving through several pathways.

An electrical analog model of the human peripheral auditory system was employed to estimate the magnitude and phase of steady-state sinusoidal acoustic signals at points in the ear where they cannot be measured noninvasively (if it is possible at all). Each subsection of the model shown in Fig. 1(b) was implemented from the literature, and detailed circuits and element values are given in the appendix. Knowles Electronics provided an electrical analog for the miniature microphone used in the acoustic probe. Sound was delivered to this microphone from the probe tip through a small diameter tube (Zuercher *et al.*, 1988). A three-section transmission line was used to simulate the occluded ear canal using an average occluded length of 13.9 mm (Gilman and Dirks, 1986), diameter of 7.48 mm, and temperature of 34.4 °C (Gardner and Hawley, 1972). The middle ear model was originally based on the work of Zwislocki (1962), and was extended by Giguère and Woodland (1994) by modifying the stapes and cochlear windows such that a complex cochlea could be added. The basilar membrane was modeled as a cascade of up to 640 shunt second-order resonant circuits (Giguère and Woodland, 1994), with element values that varied from the base to the apex such that the characteristic resonant frequency changed similarly to a human cochlea.

Figure 2 shows the model pressure TF from the probe tip to the base of the cochlea. The major resonance of the middle ear occurs near 1 kHz, and the resonance of the ear canal can be seen to dominate the response above 3 kHz. It should be noted that in the BC analysis described below, only the probe model and results shown in Fig. 2 were explicitly employed. The ear canal and middle ear model results of Fig. 2 were only used in a forward transmission direction, for estimating the filter applied to a signal at stimulus frequency f_2 from the probe tip to the base of the cochlea. The technique described here does not require an explicit knowledge of the reverse middle ear TF. The reverse middle ear TF affects the measured AC and BC DPOAEs, at a given frequency f_{dp}, identically.

1. Estimating BC magnitude at the cochlea

The magnitude of the BC signal at the cochlea can be estimated using a BC DPOAE calibration technique (Purcell *et al.*, 1999), and the model of the ear canal and middle ear described above. A DPOAE magnitude input/output function, or IOgram (Whitehead *et al.*, 1995), can be obtained by holding stimulus level L_1 fixed, and plotting the DPOAE magnitude as L_2 is varied. When the stimuli are delivered exclusively through AC, the IOgram will have a shape dictated by the magnitude of the emissions elicited with the particular stimulus magnitudes L_1 and L_2 (see Fig. 3). A second IOgram can be obtained, with the same L_1 delivered through AC, but using BC to deliver f_2. The shape of the second IOgram will be similar to the first as in Fig. 3, but the two curves will most likely be shifted with respect to the index of L_2. This is because the nominal calibration of the bone conductor may not be correct (relative to AC) for this individual at this frequency f_2. In other words, the output force of the bone conductor may be set to a certain nominal level L_2, but the actual signal level at the individual cochlea is different than during AC for the same index of L_2.

Using the assumption that emissions of similar magnitude are obtained with stimuli that are of similar magnitude at the place of generation, the BC IOgram may be aligned with the one obtained using AC. The magnitude of an indi-

vidual DPOAE is determined by a nonunique combination of L_1 and L_2, but by using IOgrams a sufficient map of the input/output space is obtained such that the optimal shift required to align the AC and BC IOgrams is unique. Each desired value of BC L_2 can be mapped to an equivalent value of AC L_2 (equivalent as far as eliciting emissions is concerned). This mapping is essentially an objective correction, with respect to AC, of the nominal calibration of the bone conductor at this f_2 frequency for this individual (Purcell *et al.*, 1999).

In the illustrative data of Fig. 3, the BC IOgram can be aligned with the AC curve with a shift left of about 5 dB. For example, when the nominal index value of BC $L_2 = 35$ dB, a DPOAE was elicited with a magnitude of about -2 dB SPL. A similar magnitude emission was obtained when the index value of AC $L_2 = 30$ dB. In the implementation here, the optimal alignment of the two IOgrams was found by minimizing the square error between the curves, as the BC IOgram was shifted in 0.5-dB increments through the L_2 index space. For each participant, and frequency band, the actual value of L_1, and the range of L_2, could be adjusted. This was done to obtain the most number of valid DPOAEs that were significantly different from the background noise, and to give the IOgrams features that make the alignment between the two curves obvious and unique (as in Fig. 3). The absolute value of L_1 used in a given set of IOgram measurements does not affect the outcome (if sufficient DPOAEs are obtained) because it is the relationship between AC and BC DPOAEs that is important (Purcell *et al.*, 1999).

Since the actual experimental values of AC L_2 can be measured at the probe tip, this mapping between BC and AC can give a sound pressure value in the ear canal that elicits an emission of the same magnitude as a given nominal value of BC L_2 index. Using these sound pressure values at the probe tip and the model results shown in Fig. 2, an equivalent pressure at the base of the cochlea can be derived for each BC measurement.

2. Estimating BC phase at the cochlea

Estimation of the phase of the BC signal at the base of the cochlea also relies on models of the acoustic probe, ear canal, and middle ear. While BC phase can be estimated using a cochlear model, an alternative method is presented here that avoids explicit use of a general cochlear model by employing AC phase measurements from the individual under examination.

The pressure phase of the f_1 and f_{dp} tones (ϕ_1 and ϕ_{dp}) can be obtained at the electrical output of the microphone during BC measurements. The net phase of the f_2 tone (ϕ_2) at the cochlea delivered by the bone conductor is unknown and not directly measurable. To estimate ϕ_2 at the base of the cochlea, it is necessary to employ theory on the generation of the DPOAE. There are two sites along the basilar membrane that contribute to the DPOAE for a given set of stimuli (Talmadge *et al.*, 1999; Shera and Guinan, 1999; Knight and Kemp, 2000). The first is the characteristic place for the frequency f_2, which is where the stimuli f_1 and f_2 interact most prevalently. The second is the characteristic place for the frequency f_{dp}, and it may dominate the overall DPOAE response at stimulus frequency ratios of $f_2/f_1 < 1.1$ (Knight and Kemp, 2000; Shera and Guinan, 1999). At larger stimulus frequency ratios, such as $f_2/f_1 = 1.22$ used here, the f_2 place is believed to be the major site of DPOAE generation. The analysis will be simplified by assuming the f_2 place dominates the generation of the DPOAE, and that the emission component from the f_{dp} place does not significantly influence emission phase.

The cubic DPOAE inherits the phase of its stimuli at its generation site, the f_2 place, according to the same relationship that determines its frequency: $\phi_{dp} = 2\phi_1 - \phi_2$ (O'Mahoney and Kemp, 1995; Bowman *et al.*, 1997; Whitehead *et al.*, 1996). For the argument here, it will be assumed that this equation calculates the basilar membrane volume velocity phase at frequency f_{dp} due to the emission, from the basilar membrane displacement phases of the stimuli at frequencies f_1 and f_2. Giguère and Woodland (1994; Strube, 1986) note that the outer hair cell feedback loop, presumed responsible for the DPOAE here, probably includes a stage of differentiation. At any particular place on the basilar membrane, the volume velocity is related to the pressure across the membrane by the point impedance at that place. These values are not explicitly calculated in the analysis, and a different form would simply change the lumped terms in the development below. For all subsequent discussion, the symbols ϕ_{dp}, ϕ_1, and ϕ_2 will again refer to the phases of the pressures at frequencies f_{dp}, f_1, and f_2, respectively.

During IOgram measurements, DPOAEs are elicited using both AC and BC for every frequency of interest. For a single AC IOgram, L_2 is varied such that there are sufficient AC DPOAEs to compare in magnitude with those obtained through BC. For the AC DPOAE data, ϕ_2, ϕ_1, and ϕ_{dp} are all recorded in the ear canal, and are all valid since the steady-state signal in the ear canal may only travel to and from the cochlea via the middle ear. The phase of the AC DPOAE is available at the electrical output of the microphone, and could also be calculated as

$$\phi_{dp}^{mic} = 2\beta_1 - \beta_2 - \gamma_{dp} - \beta_{dp}, \tag{1}$$

$$\beta_1 = (\phi_1^{mic} + a_1 + b_1 + c_1 + \gamma_1 - \pi/2), \tag{2}$$

$$\beta_2 = (\phi_2^{mic} + a_2 + b_2 + c_2 + \gamma_2 - \pi/2), \tag{3}$$

$$\beta_{dp} = (c_{dp} + b_{dp} + a_{dp}). \tag{4}$$

In the above equations, a_i represents the phase shift between the electrical output of the microphone and pressure at the probe tip, b_i is the phase shift between the probe tip and the base of the cochlea, and c_i is the shift between the base of the cochlea and the f_2 place. The phase angle between volume velocity and pressure is represented by γ_i. For all these phase shifts, the subscript denotes the particular frequency of interest (i.e., γ_2 for f_2, γ_1 for f_1, and γ_{dp} for f_{dp}). The phase shift introduced by integration from basilar membrane volume velocity to displacement is $-\pi/2$. Terms β_1 and β_2 in Eq. (1) represent the basilar membrane displacement phases of f_1 and f_2 at the f_2 place, respectively. Together they give the volume velocity phase of the DPOAE on the basilar membrane at its presumed site of generation. The third term in Eq. (1), γ_{dp}, converts this phase to the pressure

variable at the f_2 place. Finally, term β_{dp} shifts this DPOAE pressure phase back to the electrical output of the microphone (in this case, terms c_{dp} and b_{dp} represent reverse transmission phase shifts on the basilar membrane, and across the middle ear and ear canal, respectively).

In this analysis, without employing quantitative values from the model, all the terms a_i, b_i, c_i, and γ_i are unknown. If these unknowns are grouped together, a new term Θ may be defined to represent their summation:

$$\Theta = \phi_{dp}^{mic} + \phi_2^{mic} - 2\phi_1^{mic}. \quad (5)$$

The value Θ is a function of stimulus frequencies, f_2 and f_1, and stimulus magnitudes, L_2 and L_1. For each IOgram comparison between AC and BC modalities, the stimulus frequencies and magnitude L_1 are held fixed. During AC measurements of a single IOgram, Θ will vary due to the change in outer hair cell activity with increasing magnitude L_2.

For BC measurements, only ϕ_1 and ϕ_{dp} can be completely measured in the ear canal. At the f_2 place, ϕ_2 is a result of the net summation of BC components arriving through different pathways. The cochlea does not, however, react differently to the BC stimulus than it would to an equivalent AC signal. Through IOgram comparison, the magnitude of the BC tone has been determined as equivalent to a certain acoustic sound pressure at the probe tip. This specification of an equivalent AC L_2 allows the correct index of Θ to be chosen for the BC case since f_1, f_2, and L_1 are fixed. With the correct Θ available from AC for a given BC measurement, Eq. (5) can be rearranged to obtain the equivalent total microphone phase ϕ_2 during BC:

$$\phi_2^{BC\ mic\ equivalent} = 2\phi_1^{mic} - \phi_{dp}^{mic} + \Theta. \quad (6)$$

Phases ϕ_1 and ϕ_{dp} in Eq. (6) are measured from the microphone in the canal during BC. The calculated ϕ_2 phase value represents what would be digitized from the microphone output during BC for frequency f_2, if all of the BC signal were actually measurable in the ear canal. Employing the ear canal and middle ear models (Fig. 2), this equivalent BC ϕ_2 may then be shifted to the base of the cochlea (known forward model shifts $a_2 + b_2$) for calculation of the total BC TF.

3. Separating the BC signal into ear canal, and middle and inner ear contributions

Once the estimated BC signal at the base of the cochlea is available, the nonparametric TF can be calculated. The experimental TF is found by dividing the complex signal at the cochlea by the complex force signal measured on the surface of the head. This BC TF reflects the multiple contributions made to the net BC signal at the cochlea. During BC, the microphone recorded the sound pressure of the f_2 tone radiated by the canal walls [the osseo-tympanic pathway in Fig. 1(a)]. This sound pressure is not significantly influenced by tympanic membrane radiation of other BC components into the ear canal (Khanna *et al.*, 1976). It will therefore be assumed to accurately represent the energy radiated at stimulus frequency f_2 by the canal walls during BC. This energy crosses the middle ear as a normal AC sound would, and contributes to the net signal at the cochlea.

Once again using the ear canal and middle ear models, the measured ear canal f_2 tone during BC can be transposed to the base of the cochlea. With the ear canal BC contribution known, the TF can be separated into ear canal and middle/inner ear components. Although the ear canal BC contribution is transmitted to the cochlea via the middle ear apparatus, it is distinct from the signal generated by ossicular inertia. Through linear superposition, the middle ear conveys both signals to the cochlea.

The middle/inner ear components are largely due to ossicular inertia and labyrinth compression, but they may also include fluid inertia in the cochlea (Bárány, 1938) and other effects. It is not possible to separate these contributions here; however, the acronym OILC, which derives its name from the major components of ossicular inertia and labyrinth compression, will be used to refer to the net BC signal due to the contributions of the middle and inner ear. The inertial middle ear component contributes most significantly to BC below 2 kHz, and may have less impact for the higher frequencies used here. There is some disagreement in the literature about the role of middle ear inertia in BC at higher frequencies, and this will be discussed further below.

III. RESULTS

Figures 4 and 5 show one example TF from each participant displayed as a Bode plot. In Fig. 4, data from three individuals is shown for f_2 values from 2 to 6 kHz. In Fig. 4(a), both the magnitude and phase plots show that the total BC response for this individual was dominated by the OILC component and that the ear canal component contributed only marginally. The individual in Fig. 4(b) had a more complex TF. Between 2 and 2.6 kHz the OILC and ear canal components were near equal in magnitude and 180° out of phase. At the cochlea the two signals destructively interfered to reduce the total BC response. The most remarkable feature occurs in the OILC response near 4.9 kHz. There is a large null of almost 20 dB in the magnitude plot and an associated phase change of +234° in the phase graph. Figure 4(c) shows regions where the two BC components interfere destructively (near 2.3 and 5.8 kHz) and constructively (near 3 kHz).

Figure 5 shows data from two individuals for f_2 values from 2 to 5 kHz. These individuals found the bone conductor uncomfortable towards the end of the test period, so the band from 5 to 6 kHz was omitted. Figure 5(a) shows an individual whose BC TF was dominated by the OILC component between 3 and 4 kHz, but there is a sharp null in the contribution at 2.3 kHz. In Fig. 5(b) the ear canal component interfered with the OILC contribution for most of the frequency band.

Figure 6 shows data from one individual on three different test dates to demonstrate the repeatability of the measurements [same subject as Fig. 4(c)]. On the last date shown (dotted line), a hardware problem prematurely ended data collection before the band 5 to 6 kHz could be tested.

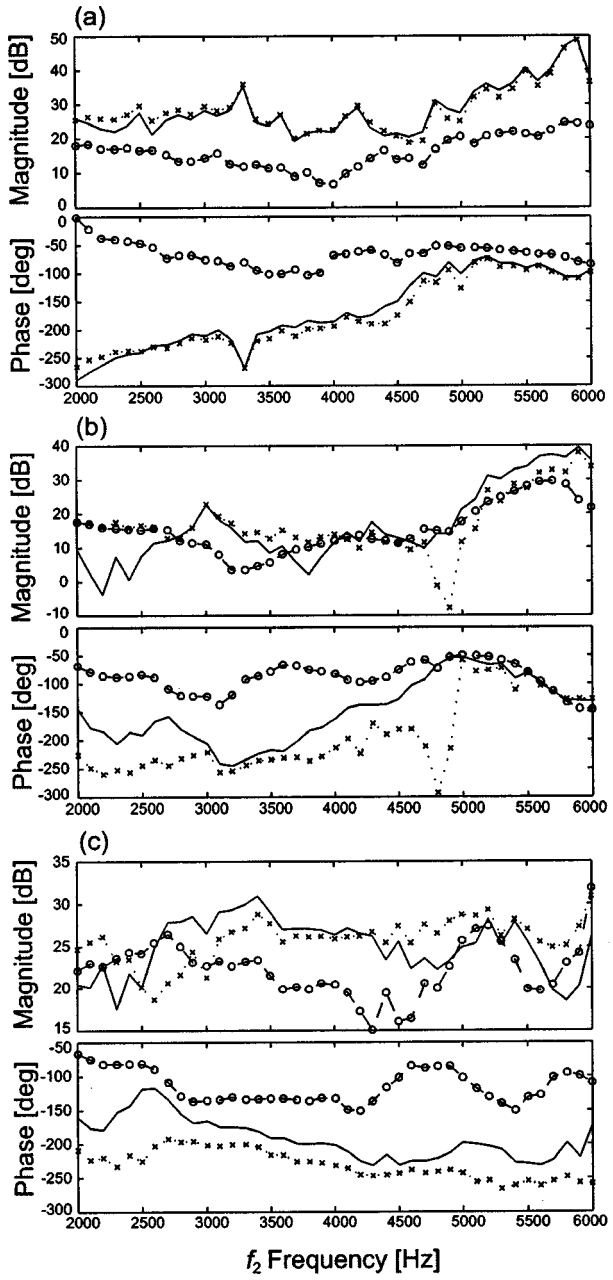


FIG. 4. Bone conduction transfer functions determined for three individuals from 2 to 6 kHz. Data is presented as a Bode plot for each individual with the magnitude response or system gain in dB [$20 \log_{10}$ (output/input) magnitude, where the input was the force applied to the head in μN and the output was the estimated BC pressure signal at the cochlea in μPa] plotted above the unwrapped phase response or phase shift across the system in degrees (output–input phase). The solid line, without symbols, indicates the total BC response. The dash line, with open circles, denotes the ear canal component, and the dot line, with the symbol x, marks the net contribution from the middle and inner ear constituents (OILC).

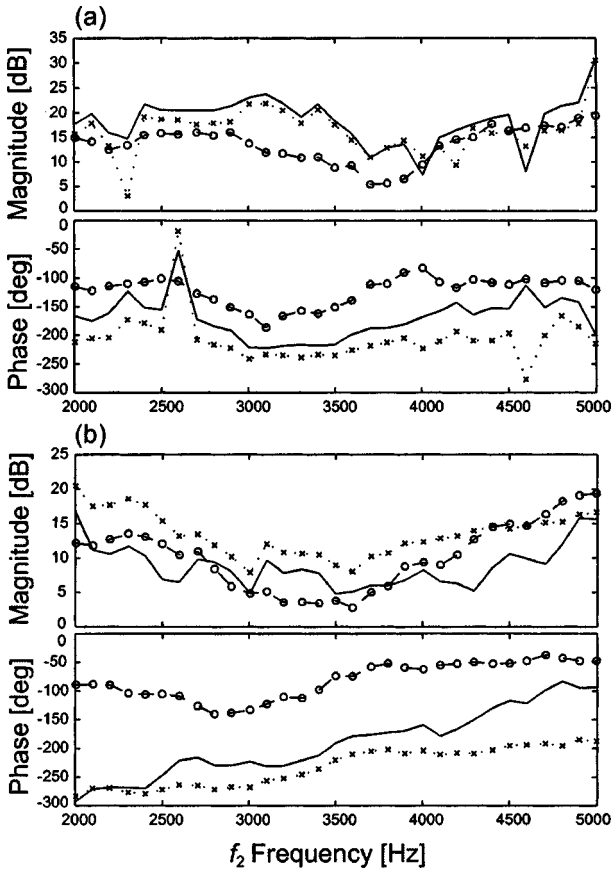


FIG. 5. Bone conduction transfer functions determined for two individuals from 2 to 5 kHz. Symbols are as described in the legend of Fig. 4.

IV. DISCUSSION

A. Nulls in the response

While Fig. 4(b) (near 4.9 kHz) was the best example of a valley in the magnitude response of the OILC component for the five participants, other smaller response dips were not uncommon. If the Θ value and ear canal estimate were reasonably accurate for each individual, then it is possible that these valleys in the magnitude response represent actual nulls in the OILC component. In some cases these nulls were associated with relatively large phase changes, but the complexity of the system, or the coarse frequency resolution (100 Hz), may be responsible for their absence or obfuscation in other instances.

There are a number of mechanisms that could be responsible for a dramatic reduction in the OILC component near specific frequencies. As noted in its definition, OILC represents all contributions to the total BC response, with the exception of the ear canal component. These are presumed to be most predominately inertia of the ossicles and distortion of the cochlea (Khanna *et al.*, 1976). In the measurements described here, it was not feasible to separate these factors. It is possible that at certain frequencies, contributions from the middle and inner ear destructively interfere to form a null in the OILC response. The major resonance of the middle ear occurs at roughly 1 kHz, but some researchers consider that the influence of ossicular inertia may extend to the higher frequencies used here (Stenfelt *et al.*, 2000).

If the middle ear does play a less significant role in BC at these frequencies (Khanna *et al.*, 1976), it is possible that the null is a reduction of the inner ear compressional contribution due to an antiresonance of the skull. Estimation of the properties of a lightly damped system with well-separated resonant frequencies can be accomplished using a wealth of techniques in the literature. Unfortunately, the system under study here is known to have closely spaced resonant frequencies and is heavily damped, which makes analysis relatively

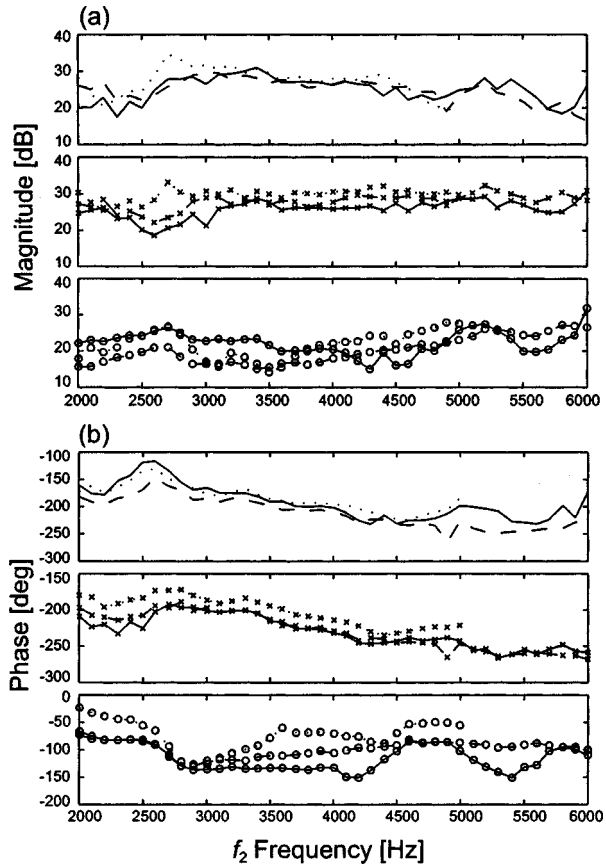


FIG. 6. Bone conduction transfer functions from a single individual [person from Fig. 4(c)] measured on three different dates. This figure is divided into three upper panel magnitude plots (a), and three lower panel phase plots (b). Each subplot shows either the total response, the middle and inner ear component (OILC), or the ear canal component, across three test dates for the individual. This facilitates comparison of the various response constituents on the different test dates. As with Figs. 4 and 5, the total response lines have no symbols, the OILC component is marked with the symbol x, and the ear canal contribution is indicated with open circles. In this figure, however, line style is used to indicate the test order. A solid line, dash line, and dot line demarcate the first, second, and third test dates, respectively.

difficult. Through direct measurements on skin penetrating titanium implants, Håkansson *et al.* (1994) found that skull resonances were unlikely to greatly affect bone-conducted sound due to relative damping coefficients as high as 8.6%. They did conclude, however, that antiresonances may play a significant role. Their transcranial accelerance (acceleration/force) functions show notable dips of 20+ dB in the response from 0.5 to 10 kHz. These measurements were made across the skull *in vivo* from one titanium implant to another, and were not evident in point accelerance measurements. The antiresonances may be due to the measurement point being on a nodal line, or destructive interference of vibrations reaching the contralateral implant through different pathways across the skull apex and base. Measured skull acceleration is not analogous with the compressional BC signal acting on the cochlea (Stenfelt *et al.*, 2000). It is, however, operated on by an unknown function that transforms it into the compressional BC signal. Therefore, an accelerance antiresonance is most likely seen by the cochlea as a dramatic reduction in the compressional BC signal. The TFs determined here are neither point nor transcranial measurements, and the question of whether such antiresonances may appear at the cochlea ipsilateral to the bone conductor is addressed by Stenfelt *et al.* (2000).

Using two sets of three orthogonal accelerometers, Stenfelt *et al.* (2000) were able to investigate the vibration of the temporal bone that encases the cochlea, for various application positions of the bone conductor. The accelerometer assemblies were each cemented on the inside of the temporal bone of a dry skull as near as possible to the corresponding cochlea (one assembly per cochlea). This arrangement allowed them to measure both the ipsi and contralateral response of the bone encasing the cochlea, to a vibration applied to the (external) mastoid portion of the temporal bone. Their findings indicated that up to 10 kHz on the ipsilateral side, the acceleration of the bony labyrinth in the excitation direction is essentially in phase with the excitation. The magnitude response in this direction showed small dips of about 10 dB at several frequencies between 2 and 6 kHz. Acceleration in the two directions perpendicular to the excitation showed much deeper nulls of 20+ dB on the ipsilateral side. If the BC compressional response is a function operating on the summation of the three orthogonal accelerations, then antiresonances may persist at the ipsilateral cochlea, and affect bone conducted sound. The other possibility considered was that the cochlea responds most significantly in the direction of the applied excitation. If this is the case *in vivo*, then the ipsilateral cochlea does not experience significant compressional BC stimulation and the total BC response is dominated by inertial components of the middle and inner ear (Stenfelt *et al.*, 2000).

B. Intrasubject repeatability of transfer functions between sessions

As shown in Fig. 6, there was certainly some variability across dates, but the measurements were in general reasonably supportive of one another. In Fig. 4(b), the largest null in the OILC curve at 4.9 kHz was not found on subsequent dates, but in the magnitude plot of Fig. 5(a), the null at 2.3 kHz was repeated during the second test. Response zeroes can be influenced by transducer placement (whereas system resonant frequencies are not).

On each test date, participants indicated that they were not experiencing any respiratory congestion, but no objective assessment of middle ear impedance was performed. Behavioral hearing thresholds were also only determined on one test date. Unknown changes in middle ear status between dates would add to the variability of the calculated TFs.

It is known from subjective measurements, such as loudness balancing, AC/BC thresholds, and cancellation, that there is significant variability in the magnitude of a BC tone at the cochlea, when the signal is applied to the mastoid portion of the temporal bone (Dirks, 1964; Weston *et al.*, 1967). Cancellation measurements also demonstrate that the phase of the BC tone can change rapidly with movement of the head relative to the bone conductor. Even on the forehead where measurements are more repeatable than the mastoid, variations in position of a few mm can lead to large changes at some frequencies. Khanna *et al.* (1976) found that the sound pressure measured in the ear canal varied by as much

as 25 dB when the vibrator position was varied on the forehead. The inability to replace the bone conductor in exactly the same position and orientation may explain some of the magnitude and phase changes found between experimental dates (Fig. 6). Application force below 5.4 N with a standard bone conductor can also lead to increased variability (Dirks, 1994). Since the headband was nonadjustable, the application force was lower for some subjects depending on head size.

Additionally, the estimate of the BC magnitude at the cochlea is dependent on both AC and BC measurements of DPOAE magnitude. As discussed in Purcell *et al.* (1999), there is some variability of the estimate at each frequency due to the technique itself (typically <3 dB when the transducers are not moved). Variation in the length of the ear canal with different insertions of the acoustic probe would also affect the results, predominately at frequencies of 4 kHz and higher. Insertion depth should, however, have been reasonably consistent as the same acoustic probe tip was used for every test with each individual.

Spontaneous OAEs (SOAEs) could have a significant impact on individual DPOAE measurements during AC and BC, if they were of similar frequency and of significant magnitude relative to the elicited DPOAEs. When employing this BC TF technique, it is prudent to check for the influence of SOAEs. This can be accomplished either through direct measurement of SOAE activity with no stimulus present, or by examining the DPOAE phase responses for each IOgram. During the measurement of an IOgram, multiple DPOAE measurements are made with all stimulus parameters fixed, except L_2. The hardware system employed here produced the stimulus tones (and therefore the elicited DPOAE) with phases that were unsynchronized from measurement to measurement. If SOAEs were significantly affecting the DPOAEs in the canal for a given IOgram, they would be expected to add constructively for some measurements and destructively for others. This would lead to large variations in DPOAE phase in a given IOgram, and these were not observed here. After IOgram comparison, the BC DPOAE with the best signal-to-noise ratio was used to calculate the BC TF. Similar TFs could be obtained with the second or third most robust BC DPOAE, as determined by the signal-to-noise ratio. It is therefore unlikely that SOAEs played a significant role in the results reported here.

C. Comparison of intersubject transfer functions

Large intersubject differences are known to exist when BC is employed. These are believed to be due to the complexity of the BC modality which derives its characteristics from the singular geometry, configuration, dynamics, and mechanical properties of every individual's skin, subcutaneous tissue, middle ear, and skull. Investigations of the resonant frequencies (Håkansson *et al.*, 1994) and mode shapes (Khalil *et al.*, 1979) of the human skull have found that each skull is unique. Intersubject variation of skin/skull mechanical impedance is also quite large (Flottorp and Solberg, 1976; Håkansson *et al.*, 1986). Additionally, it is difficult to predict an individual's response to BC. There is no clear relationship between the size of an individual's head and its

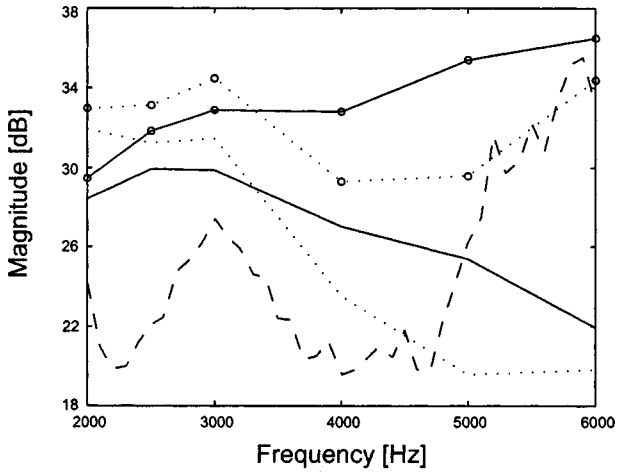


FIG. 7. Illustrative BC magnitude response function (output/input magnitude) determined using standard reference thresholds for AC and BC. In the calculation, the input signal was the reference force threshold for BC on the surface of the head. The output was the signal estimated at the cochlea by scaling the reference AC sound pressure in the ear canal to a pressure at the base of the cochlea. The solid line is for the occluded-ear simulator specified by the ISO, and the dot line is for the Zwislocki coupler. The dot and solid lines with open circles denote the AC reference was scaled by both the ear canal and middle ear models. The dot and solid lines without symbols indicate the air reference was scaled by the middle ear model alone. The dash line is the average response of the five participants.

lowest resonant frequency (Håkansson *et al.*, 1994). The thickness of the skin and subcutaneous tissue layers do not have a direct correlation with the transmission of vibration from the surface of the head to the skull (Mylanus *et al.*, 1994). There are also large variations (SD 6.9 dB on the mastoid) in the threshold of hearing through BC for young otologically normal people (Richter and Brinkmann, 1981; Haughton and Pardoe, 1981).

Each individual's TF was distinctive and contained unique features, but they all shared some similarities. A comparison can be made between mean population standards and the results here by employing the assumption that thresholds through AC and BC should be equal in the absence of a conductive impairment (Hood, 1979). Figure 7 shows magnitude TF plots where the reference equivalent threshold sound pressure by AC (ISO, 1994a) has been transposed across the ear canal and middle ear, or just the middle ear, to the base of the cochlea, and is then divided by the reference equivalent threshold force level for BC (ISO, 1994b). Curves calculated using the AC threshold as measured in a Zwislocki coupler (Wilber *et al.*, 1988) are also shown. The standard sound pressure is given for calibration of audiometric insert earphones (Etymotic Research ER-3A) in a standard occluded-ear simulator (IEC 711 or Zwislocki coupler) with a microphone where the eardrum would be; therefore the effect of the ear canal would already be included. The BC thresholds are also specified for a standard contact area and force, but this exercise demonstrates that the magnitude TFs found for the participants here are not unreasonable. Comparing the curves in Fig. 7 with the magnitude plots in Figs. 4–6, it can be seen that they share gross absolute dB level and in some cases general trends. The average magnitude response of the participants is also plotted for reference in Fig. 7.

D. Sensitivity to model values and other issues

In the calculation of BC TFs discussed here, models of the peripheral auditory system and acoustic probe have been extensively employed. The model of the probe can be specified based on knowledge of its construction, and can be tested experimentally. The model of the ear is far more complex and has been developed over decades by experts in the field. As one moves inwards from the ear canal, the practicality of altering the model to fit a given individual for the purposes of experimental data analysis diminishes.

The model was employed for two main purposes. First, electrical signals digitized from the microphone output had to be corrected, such that the actual magnitude and phase of the acoustic pressure signal sampled at the probe tip were available. Second, the cascaded models of the ear canal and middle ear were employed to estimate the magnitude and phase of the sinusoidal stimulus f_2 acting at the base of the cochlea, from measurements or estimates at the probe tip.

The model ear canal was responsible for elevating the high-frequency end of the magnitude plots in Figs. 4–7. The results could be made more specific to the individual by measuring the canal diameter, and estimating the probe insertion depth either optically or acoustically (Zemplenyi *et al.*, 1985; Chan and Geisler, 1990). Alternatively, a probe tube microphone near the eardrum might be employed if its frequency response characteristics were suitable for the frequency band of interest. Modification of the middle ear model to fit an individual could be possible through measurements of acoustic impedance, or a DPOAE IOgram technique such as developed by Keefe (2002). Changes in the middle ear model would most affect the lowest f_2 values of interest here (those closest to the major forward transmission resonance of the middle ear at about 1 kHz).

For transient evoked OAEs (TEOAEs), Rossi *et al.* (1988) found on average that the peak-to-peak amplitude of the emission measured in the time domain grew 4.6 dB less when elicited by AC than by BC, as the stimulus click level was increased from 10 to 20 dB above the threshold for detecting TEOAEs. As the stimulus level was increased from 20 to 30 dB, the average AC TEOAE grew 0.7 dB more than the BC TEOAE. Also, the threshold for detecting the BC TEOAE was on average 10 dB SL as compared to 0 dB SL for AC. They attributed the differences to a nonlinear competitive inhibition of the click stimuli arriving through multiple pathways during BC. OAEs are sensitive to small increments in stimulus level, and the difference in TEOAE growth between AC and BC (between 10 and 20 dB) could also have been caused by nonlinear behavior of the RadioEar B-71 bone conductor used in the study. Their B-71 was calibrated using input voltage tables, and no direct measurement of the output force of the vibrator was made. If the nonlinear competitive inhibition proposed by Rossi *et al.* (1988) was the cause of their findings, it may be specific to the use of transient stimuli in evoking the TEOAE. Pure tone studies using steady-state stimuli show the basilar membrane responds the same to both AC and BC (Lowy, 1942; Békésy, 1955; Khanna *et al.*, 1976). The outer hair cells on the basilar membrane, presumed responsible for the DPOAEs, should not be able to determine the transmission modality of the stimulus from their frame of reference. Such systematic differences between AC and BC DPOAE growth have not been observed in the present study, or in previous DPOAE data (Purcell *et al.*, 1999, 1998).

V. CONCLUDING REMARKS

Stenfelt *et al.* (2000) provide a helpful summary that puts the BC problem in context. They express that the BC phenomena remain elusive to fully define because of their complexity. This complexity has denied an analytical solution, and the nature of the cochlea makes direct measurements impossible in humans. Therefore ''...either a subjective measurement of the human skull or objective measurements of a cadaver, dry skull, or an animal have been used to link the response of an applied vibration of the skull to that of the cochlea'' (Stenfelt *et al.*, 2000). In an earlier paper, Håkansson *et al.* (1986) identified the need to ''design and perform measurements which objectively determine the vibration transmission through the skin to the bone as well as through the bone to the cochlea.'' They also note the difficulty of making such measurements *in vivo* (Håkansson *et al.*, 1994).

The technique described here may help fulfill some of the requirements specified above. It is objective and estimates the BC TF *in vivo* and noninvasively. The contributions of the middle and inner ear can be separated from the total response but not from each other. The method may be effective at determining frequencies where a cancellation occurs between the middle and inner ear components, or a nodal line passes near the cochlea. Application of the bone conductor to the forehead could minimize the contribution of the middle ear, and hence give more of the compressional constituent.

This technique may be useful in research applications where it is desirable to know the BC TF for the individual under test. Compilation of values across a large number of participants could yield new general information about the mechanisms of BC to improve clinical diagnosis and design of assistive devices, and help advise when surgical intervention would be fruitful. With an appropriate substitution of ear models, the method could be applied to an animal model for investigating fundaments of audition. The low-pass filtering characteristics of the middle ear may make BC an attractive modality for stimulating high frequency OAEs. Another possible application is to use Θ values for tuning the parameters of cochlear and middle ear models to better fit the individual.

A number of limitations exist, the foremost of which is the need for robust DPOAEs in the frequency range of interest. This precludes the use of the technique on most individuals with a mild to severe sensorineural hearing loss. If the signal-to-noise ratio of an individual's DPOAEs is too low, the phase measurements may be grossly contaminated with noise. The presence of a strong spontaneous OAE at the frequency of interest could also influence the measurement, although this effect was not observed here. The lower frequency limit of the technique is determined by the lowest frequency at which the DPOAE can be reliably recorded. This may allow f_2 to go as low as 1 kHz, but noise becomes more of a problem at these frequencies. The high-frequency

limit of the technique is determined by the ability to measure DPOAEs, the frequency response of the bone conductor, and the complexity of the ear models employed. At higher frequencies and smaller wavelengths, the ear models must be more sophisticated, and a practical limit at this time may be 8 kHz. While the frequency resolution employed here was relatively coarse, this is not an explicit limitation of the technique. The resolution may be made as fine as the allowable experimental duration permits.

ACKNOWLEDGMENTS

This work was supported by the Natural Science and Engineering Research Council (Canada), Grant No. OGP0004316.

APPENDIX: ELECTRICAL ANALOG MODEL

This appendix gives some details of the electrical analog elements used in the model of the auditory periphery and acoustic probe. All values were taken directly from, or calculated from equations in, the cited literature. The symbols R_i, C_i, and L_i refer to ideal resistors, capacitors, and inductors, respectively. The subscripts denote their specific placement in the model.

1. Small diameter acoustic tube network

Seven network segments (representing 4 mm each) were used to model the small diameter acoustic probe tube (total length 28 mm) that connects the microphone to the ear canal in the acoustic probe. Identical element values were employed for each segment. Figure 8(a) shows a single segment where terminal V_{out} was connected to the next probe tube segment's V_{in}, or for the final segment, to the acoustic input terminals of the microphone model provided by Knowles Electronics. The V_{in} terminal of the first segment was connected to the ear canal terminal V_{PT}. Using the inside diameter of 1.3 mm and ear canal temperature of 34.4 °C, the frequency-dependent values were calculated according to Warren (1999; Zuercher *et al.*, 1988). Example values for sinusoids of 2 kHz were $R_S = 38.3\ \Omega$, $L_S = 39.7$ mH, $R_P = 692\ k\Omega$, and $C_p = 3.84$ nF.

2. Ear canal network

A network of the ear canal is shown in Fig. 8(b). For the average occluded canal length of 13.9 mm, average diameter of 7.48 mm, and temperature of 34.4 °C, the model values were calculated according to Gardner and Hawley (1972) as $L_1 = L_4 = 0.606$ mH, $L_2 = L_3 = 1.212$ mH, and $C_1 = C_2 = C_3 = 0.143\ \mu$F. An ideal voltage source was placed across the V_{PT} terminal to represent an acoustic pressure source at the probe tip. The V_{TM} terminal was attached to the middle ear at the tympanic membrane.

3. Middle ear network

Values for the middle ear network shown in Fig. 8(c) were taken directly from Giguère and Woodland (1994, 1992) and Lutman and Martin (1979). Terminal V_{TM} represents the tympanic membrane and was connected to the ear canal terminal of the same name. The base of the cochlea is represented by terminal U_0, which is where the cochlear model was attached. In the following list of values, they are grouped according to their intended function in the model.

Middle ear air cavities: $L_a = 14$ mH, $C_p = 5.1\ \mu$F, $R_a = 10\ \Omega$, $R_m = 390\ \Omega$, $C_t = 0.35\ \mu$F

Pars flaccida: $R_{d1} = 200\ \Omega$, $R_{d2} = 220\ \Omega$, $R_{d3} = 5900\ \Omega$, $C_{d1} = 0.8\ \mu$F, $C_{d2} = 0.4\ \mu$F, $C_{d3} = 0.2\ \mu$F, $L_d = 15$ mH

Malleus and incus: $C_0 = 1.4\ \mu$F, $L_0 = 40$ mH, $R_0 = 70\ \Omega$

Incus/stapedial joint: $C_s = 0.25\ \mu$F, $R_s = 3000\ \Omega$

Stapedius muscle (inactive): $C_{st} = 100\ \mu$F

Cochlear windows and annular ligaments: $C_c = 0.6\ \mu$F, $R_{al} = 100\ \Omega$

Ideal transformer representing effective mechanoacoustic impedance transformation across the network: 1:r = 30

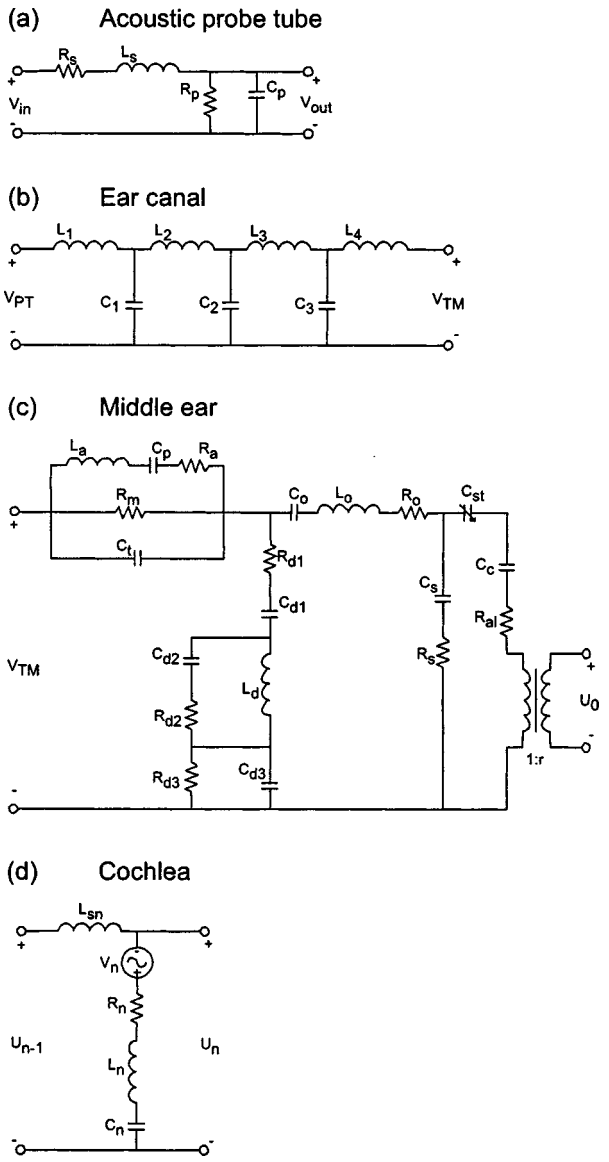


FIG. 8. Electroacoustic model circuits for the acoustic probe tube, and auditory periphery. Ideal resistors, inductors, and capacitors are indicated by R, L, and C, respectively.

4. Segmented cochlea network

The cochlea was modeled as a series of second-order shunt circuits with segments shown in Fig. 8(d) (the first segment attached to the middle ear at terminal U_0). Device values change from base to apex (segment 0 to N), such that the characteristic resonant frequency of each segment falls according to the map of Greenwood (1990). Relationships in Giguère and Woodland (1994, 1992) were employed to determine the appropriate values for each segment. A C program was written to calculate these and subsequently write them to file as a node list for the electrical circuit simulation package. Inductor L_{sn} represents the acoustic mass of the fluids in the scalae at any particular segment. Elements R_n, L_n, and C_n represent the point impedance of the basilar membrane at that place. The network is terminated by an inductor to represent the acoustic mass of the fluids from the final modeled segment N to the apex of the cochlea at the helicotrema. An active outer hair cell in each segment is represented by the nonlinear voltage source V_n.

The complex cochlea was implemented because some investigations (not presented here) were performed where BC phase was estimated using the general cochlear model instead of the individual Θ term. A TF similar to that shown in Fig. 2, from the probe tip to the base of the cochlea, can be obtained using the simpler RLC cochlear termination presented in Lutman and Martin (1979).

Bárány, E. (**1938**). ''A contribution to the physiology of bone conduction,'' Acta Oto-Laryngol., Suppl. **26**, 84–113, 130–167.

von Békésy, G. (**1954**). ''Note on the definition of the term: hearing by bone conduction,'' J. Acoust. Soc. Am. **26**, 106–107.

von Békésy, G. (**1955**). ''Paradoxical direction of wave travel along the cochlear partition,'' J. Acoust. Soc. Am. **27**, 137–145.

Bowman, D. M., Brown, D. K., Eggermont, J. J., and Kimberley, B. P. (**1997**). ''The effect of sound intensity on f1-sweep and f2-sweep distortion product otoacoustic emissions phase delay estimates in human adults,'' J. Acoust. Soc. Am. **101**, 1550–1559.

Chan, J. C., and Geisler, C. D. (**1990**). ''Estimation of eardrum acoustic pressure and of ear canal length from remote points in the canal,'' J. Acoust. Soc. Am. **87**, 1237–1247.

Collet, L., Chanal, J. M., Hellal, H., Gartner, M., and Morgon, A. (**1989**). ''Validity of bone conduction stimulated ABR, MLR and otoacoustic emissions,'' Scand. Audiol. **18**, 43–46.

Dirks, D. D. (**1964**). ''Factors related to bone conduction reliability,'' Arch. Otolaryngol. **79**, 551–558.

Dirks, D. D. (**1994**). ''Bone-Conduction Threshold Testing,'' in *Handbook of Clinical Audiology*, edited by J. Katz (Williams and Wilkins, Baltimore), pp. 132–146.

Flottorp, G., and Solberg, S. (**1976**). ''Mechanical impedance of human headbones (forehead and mastoid portion of the temporal bone) measured under ISO/IEC conditions,'' J. Acoust. Soc. Am. **59**, 899–906.

Gardner, M. B., and Hawley, M. S. (**1972**). ''Network representation of the external ear,'' J. Acoust. Soc. Am. **52**, 1620–1628.

Giguère, C., and Woodland, P. C. (**1992**). ''Network representation of the middle and inner ear in a composite model of the auditory periphery,'' Proc. Inst. Acoust. **14**, 305–312.

Giguère, C., and Woodland, P. C. (**1994**). ''A computational model of the auditory periphery for speech and hearing research. I. Ascending path,'' J. Acoust. Soc. Am. **95**, 331–342.

Gilman, S., and Dirks, D. D. (**1986**). ''Acoustics of ear canal measurement of eardrum SPL in simulators,'' J. Acoust. Soc. Am. **80**, 783–793.

Greenwood, D. D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.

Håkansson, B., Carlsson, P., and Tjellström, A. (**1986**). ''The mechanical point impedance of the human head, with and without skin penetration,'' J. Acoust. Soc. Am. **80**, 1065–1075.

Håkansson, B., Brandt, A., Carlsson, P., and Tjellström, A. (**1994**). ''Resonance frequencies of the human skull *in vivo*,'' J. Acoust. Soc. Am. **95**, 1474–1481.

Håkansson, B., Carlsson, P., Brandt, A., and Stenfelt, S. (**1996**). ''Linearity of sound transmission through the human skull *in vivo*,'' J. Acoust. Soc. Am. **99**, 2239–2243.

Harris, F. P., Lonsbury-Martin, B. L., Stagner, B. B., Coats, A. C., and Martin, G. K. (**1989**). ''Acoustic distortion products in humans: systematic changes in amplitudes as a function of f2/f1 ratio,'' J. Acoust. Soc. Am. **85**, 220–229.

Haughton, P. M., and Pardoe, K. (**1981**). ''Normal pure tone thresholds for hearing by bone conduction,'' Br. J. Audiol. **15**, 113–121.

Hood, J. D. (**1979**). ''Clinical implications in calibration requirements in bone conduction standardisation,'' Audiology **18**, 36–42.

International Standards Organization (**1994a**). ''Acoustics—Reference zero for the calibration of audiometric equipment—Part 2: Reference equivalent threshold sound pressure levels for pure tones and insert earphones,'' ISO 389-2 (International Organization for Standardization, Geneva), pp. 1–7.

International Standards Organization (**1994b**). ''Acoustics—Reference zero for the calibration of audiometric equipment—Part 3: Reference equivalent threshold force levels for pure tones and bone vibrators,'' ISO 389-3 (International Organization for Standardization, Geneva), pp. 1–10.

Keefe, D. H. (**2002**). ''Spectral shapes of forward and reverse transfer functions between ear canal and cochlea estimated using DPOAE input/output functions,'' J. Acoust. Soc. Am. **111**, 249–260.

Kemp, D. T. (**1978**). ''Stimulated acoustic emissions from within the human auditory system,'' J. Acoust. Soc. Am. **64**, 1386–1391.

Khalil, T. B., Viano, D. C., and Smith, D. L. (**1979**). ''Experimental analysis of the vibrational characteristics of the human skull,'' J. Sound Vib. **63**, 351–376.

Khanna, S. M., Tonndorf, J., and Queller, J. E. (**1976**). ''Mechanical parameters of hearing by bone conduction,'' J. Acoust. Soc. Am. **60**, 139–154.

Kirikae, I. (**1959**). ''An experimental study on the fundamental mechanism of bone conduction,'' Acta Oto-Laryngol., Suppl. **145**, 97–107.

Knight, R. D., and Kemp, D. T. (**2000**). ''Indications of different distortion product otoacoustic emission mechanisms from a detailed f1, f2 area study,'' J. Acoust. Soc. Am. **107**, 457–473.

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. Suppl. **147**, 3–14.

Lowy, K. (**1942**). ''Cancellation of the electrical cochlear response with air- and bone-conducted sound,'' J. Acoust. Soc. Am. **13**, 156–158.

Lutman, M. E., and Martin, A. M. (**1979**). ''Development of an electroacoustic analogue model of the middle ear and acoustic reflex,'' J. Sound Vib. **64**, 133–157.

Mylanus, E. A., Snik, A. F., and Cremers, C. W. (**1994**). ''Influence of the thickness of the skin and subcutaneous tissue covering the mastoid on bone-conduction thresholds obtained transcutaneously versus percutaneously,'' Scand. Audiol. **23**, 201–203.

O'Mahoney, C. F., and Kemp, D. T. (**1995**). ''Distortion product otoacoustic emission delay measurement in human ears,'' J. Acoust. Soc. Am. **97**, 3721–3735.

Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). ''A review of otoacoustic emissions,'' J. Acoust. Soc. Am. **89**, 2027–2067.

Purcell, D., Kunov, H., and Cleghorn, W. (**1999**). ''Objective calibration of bone conductors using otoacoustic emissions,'' Ear Hear. **20**, 375–392.

Purcell, D., Kunov, H., Madsen, P., and Cleghorn, W. (**1998**). ''Distortion product otoacoustic emissions stimulated through bone conduction,'' Ear Hear. **19**, 362–370.

Richter, U., and Brinkmann, K. (**1981**). ''Threshold of hearing by bone conduction. A contribution to international standardization,'' Scand. Audiol. **10**, 235–237.

Rossi, G., Solero, P., Rolando, M., and Olina, M. (**1988**). ''Delayed otoacoustic emissions evoked by bone-conduction stimulation: experimental data on their origin, characteristics and transfer to the external ear in man,'' Scand. Audiol. Suppl. **29**, 1–24.

Shera, C. A., and Guinan, Jr., J. J. (**1999**). ''Evoked otoacoustic emissions arise by two fundamentally different mechanisms: a taxonomy for mammalian OAEs,'' J. Acoust. Soc. Am. **105**, 782–798.

Stenfelt, S., and Håkansson, B. (**2002**). ''Air versus bone conduction: an equal loudness investigation,'' Hear. Res. **167**, 1–12.

Stenfelt, S., Håkansson, B., and Tjellström, A. (**2000**). ''Vibration charac-

teristics of bone conducted sound *in vitro*,'' J. Acoust. Soc. Am. **107**, 422–431.
Strube, H. W. (**1986**). ''The shape of the nonlinearity generating the combination tone 2f1-f2,'' J. Acoust. Soc. Am. **79**, 1511–1518.
Talmadge, C. L., Long, G. R., Tubis, A., and Dhar, S. (**1999**). ''Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **105**, 275–292.
Tonndorf, J. (**1962**). ''Compressional bone conduction in cochlear models,'' J. Acoust. Soc. Am. **34**, 1127–1131.
Tonndorf, J. (**1966**). ''Bone conduction. Studies in experimental animals,'' Acta Oto-Laryngol., Suppl. **213**, 105–123.
Warren, D. M. (**1999**). ''Small Acoustic Tubes Revisited, Advanced Engineering Report 076B,'' Report No. 10531-3, Release 6.1, Electrical Analogs for Knowles Electronics, Inc. Transducers, edited by J. LoPresti and E. Carlson (Knowles Electronics, Inc., Itasca, IL), pp. 18–19.
Weston, P. B., Gengel, R. W., and Hirsh, I. J. (**1967**). ''Effects of vibrator types and their placement on bone-conduction threshold measurements,'' J. Acoust. Soc. Am. **41**, 788–792.
Whitehead, M. L., Stagner, B. B., Martin, G. K., and Lonsbury-Martin, B. L. (**1996**). ''Visualization of the onset of distortion-product otoacoustic emissions, and measurement of their latency,'' J. Acoust. Soc. Am. **100**, 1663–1679.
Whitehead, M. L., Stagner, B. B., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995**). ''Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears. II. Asymmetry in L1,L2 space,'' J. Acoust. Soc. Am. **97**, 2359–2377.
Wilber, L. A., Kruger, B., and Killion, M. C. (**1988**). ''Reference thresholds for the ER-3A insert earphone,'' J. Acoust. Soc. Am. **83**, 669–676.
Zemplenyi, J., Gilman, S., and Dirks, D. (**1985**). ''Optical method for measurement of ear canal length,'' J. Acoust. Soc. Am. **78**, 2146–2148.
Zuercher, J. C., Carlson, E. V., and Killion, M. C. (**1988**). ''Small acoustic tubes: New approximations including isothermal and viscous effects,'' J. Acoust. Soc. Am. **83**, 1653–1660.
Zwislocki, J. (**1962**). ''Analysis of the middle-ear function. Part I: Input impedance,'' J. Acoust. Soc. Am. **34**, 1514–1523.

Ipsilateral distortion product otoacoustic emission ($2f1-f2$) suppression in children with sensorineural hearing loss

Carolina Abdala[a] and Tracy S. Fitzgerald[b]
Children's Auditory Research and Evaluation Center, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

(Received 14 December 2002; revised 30 April 2003; accepted 6 May 2003)

Distortion product otoacoustic emission (DPOAE) ipsilateral suppression has been applied to study cochlear function and maturation in laboratory animals and humans. Although DPOAE suppression appears to be sensitive to regions of specialized cochlear function and to cochlear immaturity, it is not known whether it reflects permanent cochlear damage, i.e., sensorineural hearing loss (SNHL), in a reliable and systematic manner in humans. Eight school-aged children with mild-moderate SNHL and 20 normal-hearing children served as subjects in this study. DPOAE ($2f1-f2$) suppression data were collected at four $f2$ frequencies (1500, 3000, 4000, and 6000 Hz) using moderate-level primary tones. Features of the DPOAE iso-suppression tuning curves and suppression growth were analyzed for both subject groups. Results show that DPOAE suppression tuning curves from hearing-impaired subjects can be reliably recorded. DPOAE suppression tuning curves were generally normal in appearance and shape for six out of eight hearing-impaired subjects but showed subtle abnormalities in at least one feature. There was not one single trend or pattern of abnormality that characterized all hearing-impaired subjects. The most prominent patterns of abnormality included: broadened tuning, elevated tip, and downward shift of tip frequency. The unique patterns of atypical DPOAE suppression in subjects with similar audiograms may suggest different patterns of underlying sensory cell damage. This speculation warrants further investigation.
 [DOI: 10.1121/1.1587147]

PACS numbers: 43.64.Jb, 43.64.Kc [BLM]

I. INTRODUCTION

Distortion product otoacoustic emission (DPOAE) iso-suppression tuning curves have been studied extensively in laboratory animals (Brown and Kemp, 1984; Martin *et al.*, 1987), human adults (Harris, Probst, and Xu, 1992; Kummer, Jannsen, and Arnold, 1995; Abdala, 2001a, b) and most recently, in human neonates and children (Abdala *et al.*, 1996; Abdala, 1998, 2001a, b). They have been used to define the generation site of the DPOAE, to study maturation of cochlear function and tuning, and to characterize cochlear nonlinearity. *DPOAE suppression tuning curves* (STCs) are recorded by presenting an ipsilateral suppressor tone (*fs*) simultaneously with the two primary tones ($f1$ and $f2$). The suppressor tone is increased in level until a criterion reduction in DPOAE amplitude is achieved. A family of suppressor tones centered around the target $f2$ frequency is presented to generate a DPOAE STC. The level of each suppressor tone required to reduce DPOAE amplitude by a criterion amount, typically 6 dB, is plotted as a function of suppressor frequency. In addition to generating a DPOAE STC, this process results in a series of *suppression growth functions*, each function representing the unique pattern and rate of amplitude reduction produced by that particular suppressor tone.

DPOAE suppression tuning curves from the normal cochlea have a sharp tip, are narrow in width, show a steep high-frequency flank, and a shallower low-frequency flank. The tip of the DPOAE STC is generally centered around the $f2$ frequency, the primary generation site for the DPOAE in the normally functioning cochlea (Martin *et al.*, 1987). DPOAE STCs are typically designated by their corresponding $f2$ frequency.

To some extent, DPOAE STCs behave like typical tuning functions in the auditory system. They become narrower with increases in $f2$ frequency (e.g., Abdala *et al.*, 1996) and broadened at higher primary tone levels (e.g., Abdala, 2001a; Kummer *et al.*, 1995). Additionally, they are sensitive enough to reflect cochlear immaturities in premature neonates, as has been shown by the age effects that have been reliably observed using this paradigm (Abdala, 1998, 2001a). In the mustached bat, DPOAE STCs have been shown to accurately reflect regions of specialized function in the cochlea (Frank and Kossl, 1995). However, the complexity of the DPOAE generation process, including multiple DPOAE sources and the recording paradigm itself (presenting a two-tone stimulus and recording a response of cochlear origin from a distant site in the ear canal), makes it unlikely that the DPOAE suppression tuning curve reflects the same level or type of frequency resolution as neural or basilar-membrane tuning curves. In the normal cochlea, DPOAE STCs most likely represent an outline of the primary DPOAE generation site centered around the $f2$ region on the basilar membrane, particularly when the process is conducted using relatively high suppression criteria. This generation site will likely be impacted by cochlear tuning, but also by other factors.

Research with laboratory animals has shown that DPOAE STCs are not affected in the same way as neural or basilar-membrane tuning curves when the cochlea is reversibly damaged by ototoxins (Martin *et al.*, 1998) or by noise

[a]Electronic mail: cabdala@hei.org
[b]Current affiliation: Department of Hearing and Speech Sciences, University of Maryland—College Park, College Park, MD 20742.

(Howard *et al.*, 2002); however, it is not known whether DPOAE STCs accurately reflect permanent cochlear damage and are sensitive to mild-moderate sensorineural hearing loss (SNHL) in humans. Additionally, it is not known whether DPOAE STCs reflect differences in underlying pathology or patterns of damage among subjects with hearing loss; that is, whether two subjects with the same audiometric results will have a different pattern or degree of altered DPOAE suppression tuning.

The objective of this experiment was to provide an initial description of DPOAE suppression in a small group of children with known cochlear damage associated with SNHL. The hypotheses to be tested are: (1) DPOAE suppression tuning curves will reflect cochlear damage and dysfunction in a systematic fashion; this prediction is based on the understanding that DPOAEs effectively reflect cochlear function, output, and integrity; and (2) Unique and variable alterations of DPOAE suppression will be observed in subjects with a similar degree and configuration of hearing loss. This may elucidate different underlying cochlear pathology or patterns of hair-cell loss in subjects with the same audiometric configuration.

II. METHOD

A. Subjects

Hearing-impaired children, rather than adults, were tested in this experiment in an attempt to maximize residual cochlear function. The assumption was that school-aged children had less exposure to high-level noise and ototoxic medications than adults. Additionally, hearing-impaired children have had fewer years wearing hearing aids than their adult counterparts and thus, have had less exposure to amplified sound. By testing children, we were attempting to enhance the likelihood of finding hearing-impaired subjects with DPOAEs that met our signal-to-noise ratio (SNR) criteria.

Eight children, four males and four females, with mild-moderate SNHL of unknown etiology participated in the study. They ranged in age from 10–17 years (mean $=13.06$ years). To obtain this experimental group of eight subjects, 67 hearing-impaired children were initially contacted to participate in this study. Of the 67, 19 did not respond to the initial query, 25 declined for various reasons, and appointments were successfully made with the 23 remaining children. Of these 23 children who were screened for participation, only eight met criteria for inclusion into the study. This yields a rate of 12% for successful recruitment of hearing-impaired children (8/67) and 35% for successful generation of DPOAE STCs in the hearing-impaired children who were screened (8/23). DPOAEs were required to be at least 9 dB above the noise floor at any test frequency to participate in this experiment. Additionally, DPOAE levels were required to be at least 10 dB above the mean system distortion levels for that $f2$ frequency. The hearing-impaired subjects were recruited from the patient population of the Children's Auditory Research and Evaluation (CARE) Center at the House Ear Institute in Los Angeles over an 18-month period.

The hearing thresholds of the eight children who comprised the hearing-impaired group ranged from -5 to 55 dB HL across the audiometric test frequencies. Seven of the eight children had bilateral SNHL and were fit with binaural amplification. The exception was Subject H12, who had normal hearing except for a mild notch at 4000 Hz in the left ear and, therefore, did not use amplification. All hearing-impaired subjects had normal tympanograms (ASHA, 1990) on each day of test. Appendix A provides descriptive information on the hearing-impaired subjects. The audiograms for the test ear of the hearing-impaired subjects are presented in Appendix B. To optimize DPOAE SNR, the ear and $f2$ frequency combination with the largest DPOAE amplitude(s) was chosen for testing. This inability to select the desired test frequency limited experimenter options considerably and thus led to limited data collection at any one $f2$ frequency.

Twenty children, eight males and 12 females, with normal hearing thresholds also participated in this study. These subjects ranged in age from 7–15 years (mean $=11.03$ years). The normal-hearing children were recruited from the Los Angeles area by means of written announcements distributed at elementary schools or by association with employees of the House Ear Institute. Subjects from the normal-hearing group had thresholds of 15 dB HL or better at 250–8000 Hz, word recognition scores of 85% correct or better (PBK or NU-6 50 word lists, depending on age), normal tympanograms (ASHA, 1990), and a negative history for otologic disease. One ear from each subject was chosen for DPOAE testing, resulting in a sample of ten right ears and ten left ears. The number of normal-hearing ears tested at each $f2$ frequency was as follows: six for 1500 Hz, eight for 4000 Hz, and 10 for both 3000 and 6000 Hz.

B. Instrumentation and signal analysis

DPOAE suppression data were collected using customized software developed at the CARE Center, in conjunction with an Ariel DSP16+ signal processing and acquisition board housed in a Compac Prolinea 590 personal computer. The Ariel board was connected to an Etymotic Research ER-10C probe system and to an analog high-pass filter (12 dB/octave, 710-Hz high-pass cutoff). The ER-10C probe contains two output speakers and a low-noise microphone, and was coupled to the ear using a disposable foam eartip. The DSP processor generated the two primary tones and suppressor tone. The lower frequency primary tone ($f1$) was generated by one D/A converter and delivered via one output transducer. The higher frequency primary tone ($f2$) and suppressor tones (fs) were generated by a second D/A converter and output through the second transducer.

The signal at the probe microphone was high-pass filtered and sampled at a rate of 50 kHz with a sweep length of 4096 samples, resulting in a frequency resolution of 12.2 Hz. Twenty-five sweeps of the microphone signal were added and comprised one block for $f2=3000$, 4000, and 6000 Hz. Due to elevated noise at low frequencies, 50 sweeps were added for one block at $f2=1500$ Hz.

C. Data acceptance criteria

The customized DPOAE software used two criteria to determine whether a measured DPOAE level would be ac-

cepted into the grand average: (1) noise measurements for the three frequency bins (12.2 Hz wide) on either side of the $2f1-f2$ frequency had to be less than 0 dB SPL (to ensure appropriate subject state) and (2) the measured DPOAE level had to be at least 3 dB higher than the average noise measured in the same six bins around the distortion product frequency. This 3-dB SNR criterion has been successfully applied in many DPOAE investigations of hearing-impaired, as well as normal-hearing human subjects (Erminy, Avan, and Bonfils, 1998; Lonsbury-Martin *et al.*, 1990; Moulin, Bera, and Collet, 1994). The program attempted up to six blocks of either 25 or 50 sweeps to achieve the two criteria. If both criteria were not met after six blocks, no data were collected at that condition and the next condition was initiated.

D. Calibration

Intermodulation distortion produced by the recording system at $2f1-f2$ was measured for all test conditions with the probe placed in a Zwislocki coupler. The mean level of system distortion was −21 dB SPL. In no case did the level of system distortion exceed −15 dB SPL. The level of system noise ranged between −22 and −27 dB SPL depending on frequency.

At the start of DPOAE testing for each subject, a calibration procedure was conducted on both output transducers with the probe inserted in the subject's ear. Tones of a fixed voltage were presented to the transducers at 250 Hz intervals from 500 to 15 000 Hz and the resulting SPL of these tones recorded in the ear canal. Based on this information, an equalization of output levels was performed for each subject to achieve target stimulus levels across test frequencies. Because the calibration process was conducted at the entrance of the ear canal, it is probable that standing waves in the canal produced misestimations of the actual stimulus level arriving at the tympanic membrane and, subsequently, at the cochlea for $f2=3000$, 4000, and 6000 Hz. However, both hearing-impaired and normal-hearing children were subjected to the same potential error in level estimation; therefore, it should not have affected between-group comparisons in any systematic fashion.

E. Procedure

DPOAE testing was conducted with the child seated comfortably in an easy chair inside a single-walled sound-attenuating booth. Subjects were given the option of watching a video with subtitles or reading quietly during testing. Primary tones were presented at a fixed $f2/f1$ ratio of 1.2 and fixed levels of $L1=65$ and $L2=55$ dB SPL for all DPOAE testing. DPOAE suppression data were collected at $f2$ frequencies of 1500, 3000, 4000, and 6000 Hz. Each DPOAE suppression session lasted approximately 2 h; all DPOAE suppression data were collected during only one session for each subject. However, several subjects required two visits to complete the entire screening protocol and 50% of the subjects returned for an additional visit to provide retest data.

An unsuppressed DPOAE level was measured at the start of data collection and again prior to the presentation of each new suppressor tone. Each suppressor tone was presented simultaneously with the primary tones and its level increased in 2-dB steps over a range of intensities (maximum range was 40–84 dB SPL). Twelve to 15 suppressor tones with frequencies ranging from one octave below to 1/4 octave above $f2$ were presented at intervals of 25 to 150 cents (1 octave = 1200 cents). Table I lists the suppressor frequencies for each of the $f2$ frequencies. Completion of the procedure yielded a series of suppression growth functions, an example of which is given in Fig. 1. The top panel displays the functions obtained using suppressor frequencies below $f2$, while the bottom panel displays the functions obtained using suppressor frequencies above $f2$. To generate iso-suppression tuning curves (STCs), the suppressor level that reduced DPOAE amplitude by 6 dB was determined from the suppression growth function using linear interpolation and was plotted as a function of suppressor frequency. The STC plotted from the data in Fig. 1 is shown in Fig. 2.

TABLE I. The suppressor frequencies (*fs*) presented for each of the four $f2$ frequencies. An asterisk indicates which suppressor frequencies were included in the analysis of suppression growth.

$f2$	1500 Hz	3000 Hz	4000 Hz	6000 Hz
$fs<f2$	850*	1863*	2246*	3044*
	1050*	2091*	2588*	3620*
	1208	2347	2911*	4063*
	1345*	2634*	3268	4560
	1436*	2751	3516*	5119*
	1478	2873*	3662	5350
		2957*	3833*	5582*
			3943	5746
				5914*
$fs>f2$	1522*	3044	4053*	6087
	1566	3133*	4175	6266*
	1612*	3225	4297*	6449
	1684	3367*	4492	6638*
	1784*	3516	4688*	6932*
	1890*	3725*	4895*	7239
		3946*	5188	

F. DPOAE data analysis

The following suppression tuning curve features were quantified and analyzed: (1) Q_{10} value; (2) slope (dB/octave) of the low- and high-frequency sides; (3) tip frequency; and (4) tip level. The Q_{10} value is a measure of STC width that is calculated by dividing the tip frequency by the bandwidth of the tuning curve 10 dB above the tip level. When a Q_{10} value could not be calculated, the STC was assigned a value 3 standard deviations below the normal mean.

Additionally, suppression growth was evaluated for six to eight of the suppressor frequencies in each $f2$ condition. The suppressor frequencies included in this analysis were selected to provide a representative range of frequencies above and below $f2$ and are indicated with an asterisk in Table I. To estimate the slope of suppression growth, a regression line was fit to the most linear portion of each function. Points comprising the linear portion were chosen by visual inspection and were agreed upon by two independent examiners. If an initial plateau in the function was observed,

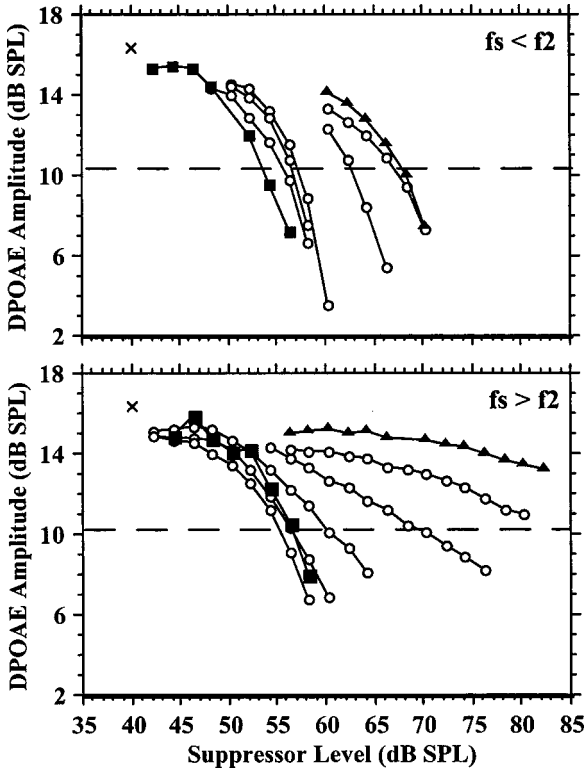


FIG. 1. A series of suppression growth functions recorded from the right ear of a normal-hearing subject. The $f2$ frequency was 3000 Hz. The top panel displays suppression growth for suppressor frequencies (fs) lower than $f2$. The bottom panel displays suppression growth for fs higher than $f2$. Filled triangles designate the fs furthest from $f2$, and filled squares indicate the fs closest to $f2$. "×" represents the unsuppressed DPOAE level and dashed line indicates the level at which 6 dB of suppression was reached. The complete list of suppressor frequencies for each $f2$ is provided in Table I.

those points were not included. An initial plateau consisted of a sequence of continuous points at the start of the function showing a 1-dB or lower decrease in DPOAE amplitude as suppressor level increased. Points at the end of the function that fell below the noise floor were eliminated.

III. RESULTS

A. DPOAE and noise levels

Mean DPOAE and noise levels are presented in Fig. 3 as a function of $f2$ frequency for both the normal-hearing (open symbols) and hearing-impaired (filled symbols) groups. As expected, the mean DPOAE amplitudes for the hearing-impaired subjects were lower than those for the normal-hearing subjects. However, noise levels were similar between the two groups. There was no systematic relationship between subject age and noise levels (data not shown).

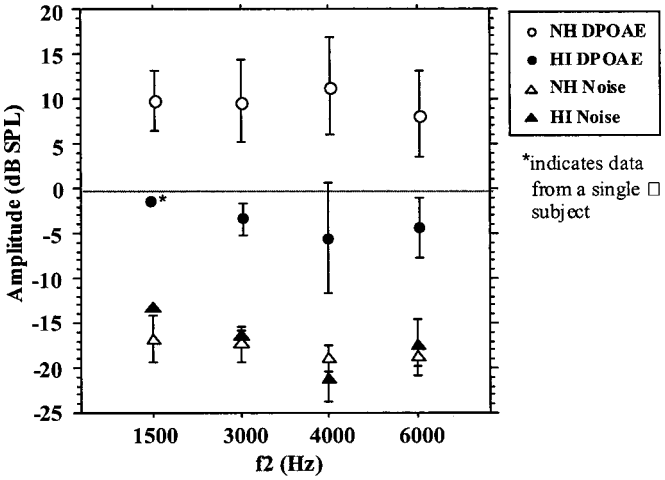


FIG. 3. Mean unsuppressed DPOAE amplitudes and overall noise levels for the normal-hearing (NH) and hearing-impaired (HI) subjects. Error bars indicate ±1 standard deviation.

The data in Fig. 3 show that mean DPOAE amplitude in the hearing-impaired subjects ranged from 0 to −5 dB SPL, whereas the average noise floor was between −14 dB (at $f2=1500$ Hz) and −22 dB SPL (at 4000 Hz). The breakdown of SNR for unsuppressed DPOAEs in the eight hearing-impaired subjects is as follows: H04=9 dB; H08 and H21=11 dB; H09, H14, H15, and H19=14.5 to 15.5 dB; H12=21 dB SNR. These DPOAE SNR values allowed for sufficient dynamic range to successfully conduct 6-dB suppression of DPOAE amplitude during the experimental procedure.

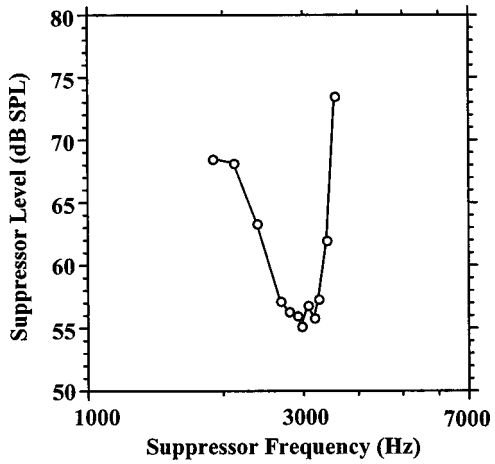


FIG. 2. Example of an STC derived from the suppression growth series in Fig. 1 using an $f2$ frequency of 3000 Hz and a suppression criterion of 6 dB.

B. Suppression tuning curves

The STCs measured from each of the eight hearing-impaired subjects are presented in Fig. 4. Because of the limited number of subjects at any one $f2$ frequency, we were unable to analyze or detect any frequency-related trends in suppression tuning for this group of eight children. Each panel represents data for a single hearing-impaired individual compared to mean data for the normal-hearing group at that $f2$ frequency. The SNR for the data points comprising each DPOAE STC from hearing-impaired subjects was calculated to ensure data quality and reliability. Only 3.8% (4/105) of points comprising the eight STCs collected in hearing-impaired subjects required use of the minimum 3-dB SNR rule, whereas 89% of the STC data points had ⩾5-dB SNR and 42% had ⩾8-dB SNR.

The STC morphology varied among the hearing-impaired subjects. Six of the STCs exhibited typical morphology; that is, a sharp tip, a steep high-frequency side, and a shallower low-frequency flank. On the other end of the spectrum, the STC morphology for subject H04 is grossly abnormal. The STC is flattened with no low-frequency or

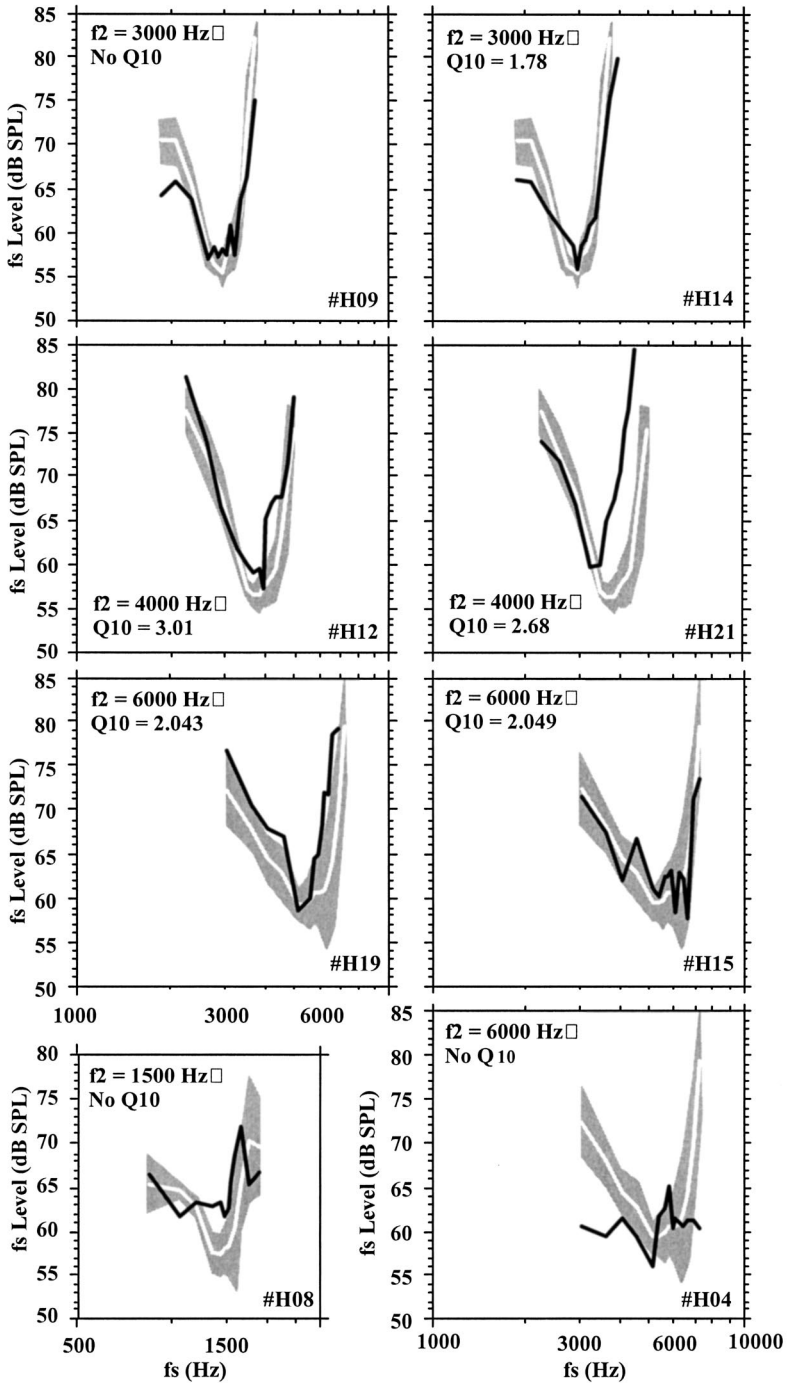


FIG. 4. STCs from the eight hearing-impaired subjects. Each panel displays the STC from one hearing-impaired subject (black line) compared to the normal mean (white line) and ± 1 standard deviation (gray-shaded area). The $f2$ frequency, Q_{10}, and subject number are listed in each panel. "No Q_{10}" indicates that a Q_{10} value could not be calculated.

high-frequency boundaries. Except for subject H04 and H08, all STCs exhibited a typical bandpass characteristic, accompanied by subtle morphological abnormalities, such as the absence of a sharp tip or excessive jitter. In a few cases, the slopes of the STC flanks were more shallow than normal on either the low-frequency side (H08, H14, and H15) or the high-frequency side (H04 and H12).

The Q_{10} values varied across hearing-impaired subjects. Figure 5 shows Q_{10} as a function of $f2$ frequency. Means and ± 1 standard deviation for the normal-hearing group are indicated by open circles and error bars, respectively. The hearing-impaired subjects can be classified into three groups based on Q_{10}: (1) normal Q_{10} values; (2) smaller Q_{10} values, indicating a sharper tuning curve than normal; and (3) larger Q_{10} values, indicating a broader tuning curve than normal. The largest contingent $(n=4)$ demonstrated broad STCs. A Q_{10} could not be calculated for three of these subjects; therefore, this value was assigned, as described in the data analysis section (H04, H08, and H09). The inability to measure Q_{10} was attributable to an elevated tip level, an abnormally shallow low-frequency flank, or a combination of the two. Two subjects showed a sharper than normal STC (H12, H19); however, their Q_{10} values fell just outside of the normal range.

In Fig. 6, STC tip level is plotted as a function of tip frequency for each hearing-impaired individual compared to the normal means (white circles). The shaded areas represent ± 1 standard deviation for the normal-hearing subjects. Only

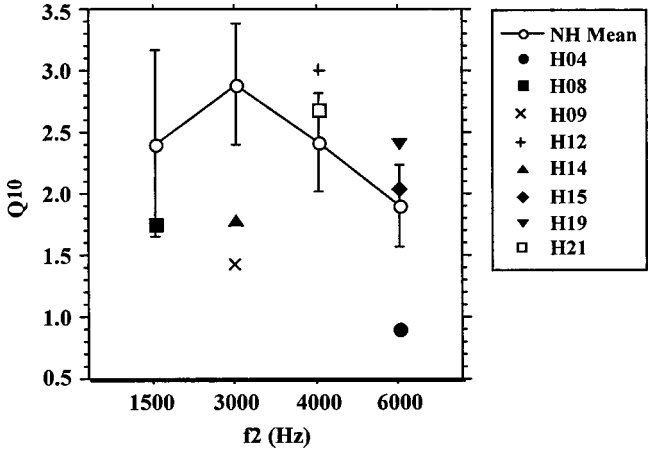


FIG. 5. Individual STC Q_{10} values as a function of $f2$ frequency for the hearing-impaired subjects. Normal-hearing (NH) subject's mean data and variability are presented for comparison. Error bars represent ± 1 standard deviation.

two hearing-impaired children had an STC tip that fell within normal limits on both dimensions. Three subjects had tip levels that were elevated. The most prominent trend, however, was for the STC tip frequency to be shifted to a lower-than-normal frequency in the hearing-impaired subjects.

In summary, six out of eight DPOAE STCs recorded from hearing-impaired subjects retained their shape and bandpass characteristic. Nevertheless, other more subtle features of the tuning curves were atypical in these subjects with hearing loss. No single pattern of STC abnormality was observed across all the data; however, patterns emerged, such as the tendency for STCs to be broader than normal, for tip levels to be somewhat elevated, and for tip frequency to shift slightly to a lower frequency. It is not clear what relationship exists between the STC and audiometric hearing level at the $f2$ frequency. H04 had the second most elevated audiometric threshold at $f2$ (45 dB HL) and showed the most grossly abnormal STC; however, this is only one observation. The two subjects with the least overall hearing loss produced STCs that were excessively narrow (H12, H19). In both cases they have hearing thresholds that fall within the normal

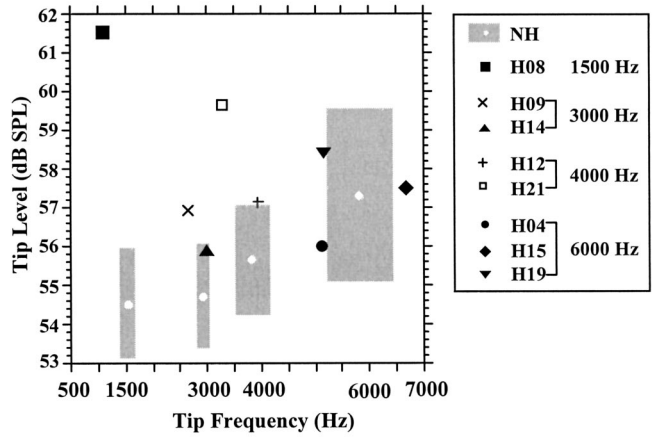


FIG. 6. STC tip level as a function of tip frequency for the hearing-impaired subjects. Normal-hearing (NH) subjects mean data and ± 1 standard deviation are presented for comparison. The $f2$ frequency for the hearing-impaired subjects is indicated in the key.

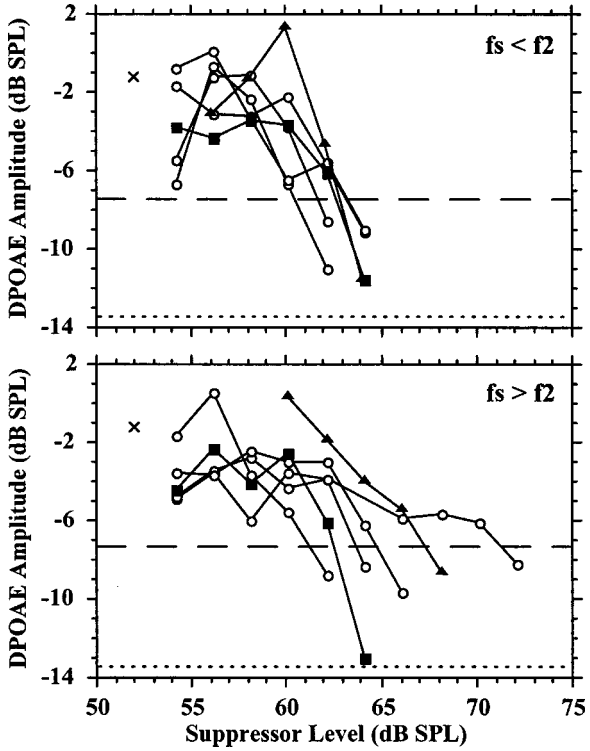


FIG. 7. A series of suppression growth functions recorded from the left ear of a hearing-impaired subject (H08). The $f2$ frequency was 1500 Hz. The top panel displays suppression growth for suppressor frequencies (*fs*) lower than $f2$. The bottom panel displays suppression growth for *fs* higher than $f2$. Filled triangles designate the *fs* furthest from $f2$, and filled squares designate the *fs* closest to $f2$. "×" represents the unsuppressed DPOAE level and the dashed line indicates the level at which 6 dB of suppression was reached. The dotted line near the bottom of each panel indicates the mean noise level across all suppressor frequencies for this subject (−13.4 dB SPL). The complete list of suppressor frequencies for each $f2$ is provided in Table I.

range at frequencies both above and below the $f2$ frequency tested (see Appendix B). A larger database will be necessary to draw firm conclusions about the relationship between audiometric hearing level and DPOAE STC abnormality.

C. Slope of suppression growth functions

Figure 7 is an example of a series of suppression growth functions recorded from the left ear of hearing-impaired subject H08. This figure can be compared to the series from the normal-hearing subject in Fig. 1. The dotted line at the bottom of each panel indicates the mean noise level for this subject across all suppressor frequencies.

Suppression growth functions recorded from ears with normal hearing such as those in Fig. 1 follow a systematic pattern that depends on suppressor frequency (e.g., Kummer *et al.*, 1995; Abdala, 1998). For suppressor frequencies below $f2$ (low-frequency-side suppressors), the slope of suppression growth is steep, approximating linear values (1 dB/dB). Such a pattern is readily apparent in the top panel of Fig. 1. The functions are for the most part equally steep and differ mainly in terms of the level at which suppression begins, with lower frequencies requiring a higher level of the suppressor to achieve a reduction in DPOAE amplitude. However, for suppressor frequencies above $f2$ (high-frequency-side suppressors) illustrated in the bottom panel of Fig. 1, the slope of suppression growth becomes shallower as

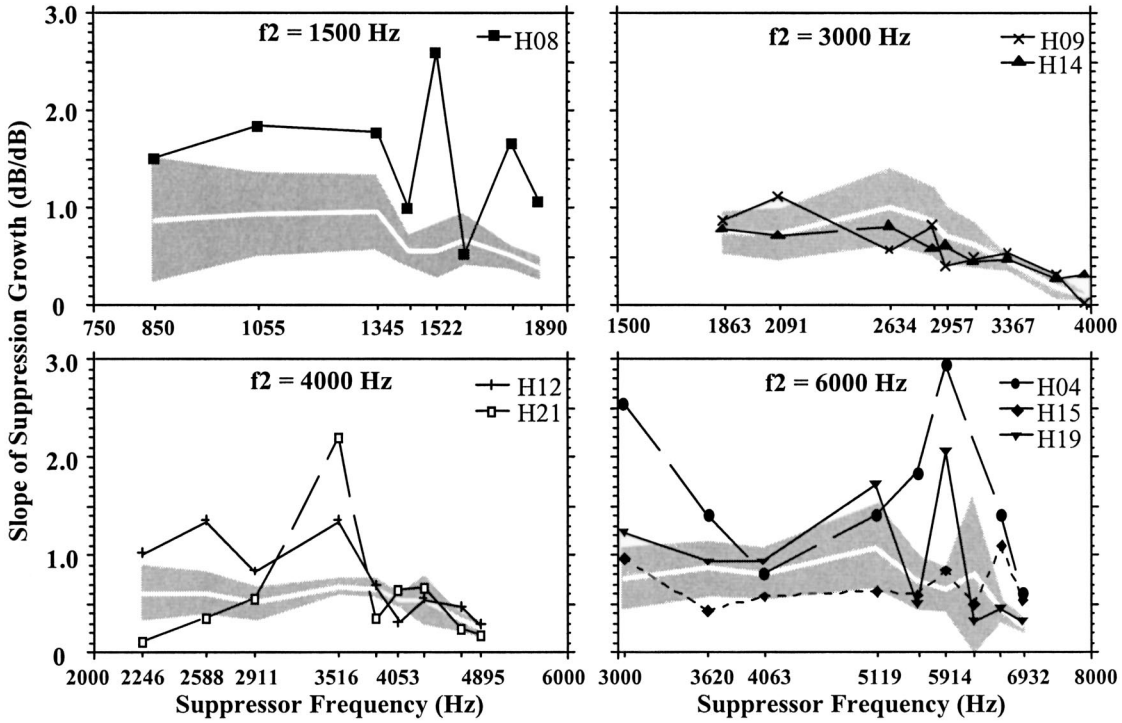


FIG. 8. Slope of suppression growth for a range of suppressor frequencies (see Table I). Each panel presents data for one $f2$ frequency as indicated. Data for individual hearing-impaired ears are plotted using black lines and symbols. The thick white line and shaded region indicate the mean slope and ± 1 standard deviation for the normal-hearing group, respectively.

the suppressor frequency increases. This frequency-dependent pattern of suppression growth is based on normal basilar-membrane nonlinearity and will be considered further in the Discussion section. Data from the majority of the hearing-impaired subjects did not demonstrate this systematic pattern of suppression growth. In several cases, the functions from hearing-impaired ears demonstrated steeper suppression growth compared to the normal-hearing ears. In Fig. 7, steeper suppression growth is particularly noticeable for high-frequency-side suppressors (bottom panel) that are normally shallow and compressive. For this child with hearing loss, the suppression growth remains fairly steep even as suppressor frequency increases above $f2$.

Figure 8 shows the slope of suppression growth as a function of suppressor frequency for each hearing-impaired subject. Each panel represents data for one of the four $f2$ frequencies tested. The typical, asymmetrical pattern of suppression growth for low-frequency-side versus high-frequency-side suppressors described earlier can be observed in the data for the normal-hearing group (white line and gray-shaded area). With the exception of the two subjects tested at an $f2$ frequency of 3000 Hz, the hearing-impaired subjects generally show atypical slope of suppression growth. However, no single pattern is evident across suppressor frequency for the hearing-impaired ears. In some cases, a steeper slope occurred at only a single suppressor frequency. As expected, the two subjects with the most abnormal tuning curves (H04 and H08; see Fig. 4) also had consistently steep suppression growth.

D. Test–retest reliability of STC results

An attempt was made to readminister the experimental protocol to all eight hearing-impaired subjects to evaluate how repeatable the results were in this group; however, four hearing-impaired subjects declined to participate in this retest process due to constraints imposed by academic, familial, and extracurricular demands.

Four hearing-impaired children were tested a second time. The time interval between test session #1 and #2 ranged from 4.5 to 37.5 weeks. Figure 9 presents STCs #1 and #2 for each of the four subjects. STC morphology was consistent between sessions in three of the four subjects. In the case of subject H04, although the two curves do not

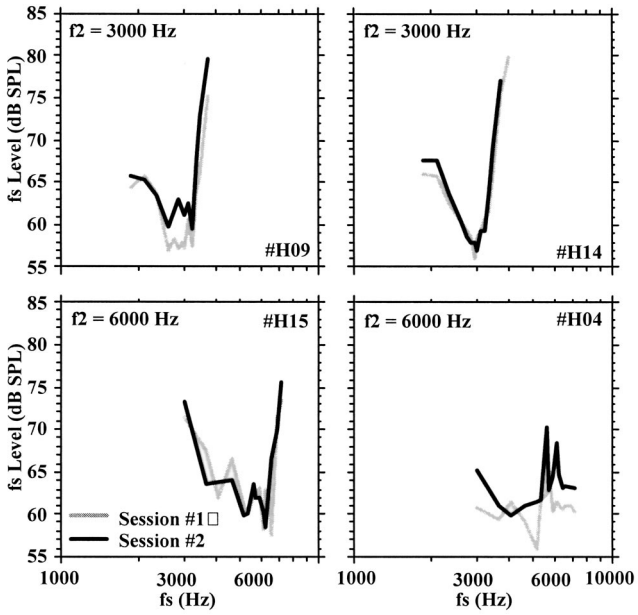


FIG. 9. Test–retest reliability of STCs recorded from four of the hearing-impaired subjects.

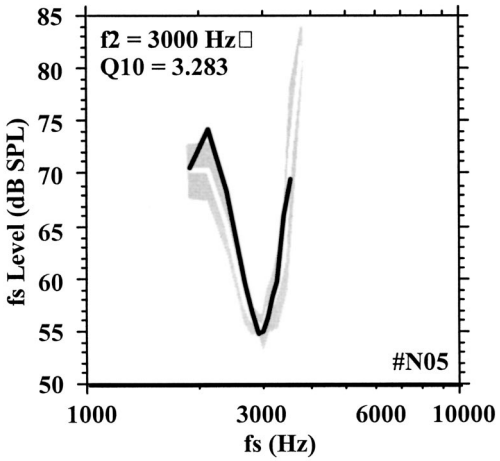


FIG. 10. DPOAE STC ($f2 = 3000$ Hz) recorded from one normal-hearing child with a low-level DPOAE (black line). The white line represents the normal mean and the gray shaded area represents ± 1 standard deviation.

overlap precisely, both curves are grossly abnormal and a Q_{10} value could not be calculated for either STC. Thus, the assigned Q value was the same in each instance. Q_{10} values varied minimally from session #1 to session #2 (change in $Q10 = 0.25-0.54$) in those for whom Q_{10} could be calculated. The category into which an STC was classified based on Q_{10} (i.e., ''normal,'' ''broad,'' ''narrow'') remained the same for all four retest subjects.

IV. DISCUSSION

The purpose of the present study was to characterize ipsilateral DPOAE suppression in eight children with permanent SNHL and to gain further insight into how DPOAE STCs reflect underlying pathology. Studies of DPOAE suppression in humans with permanent hearing loss have not been previously conducted. The findings of the present study indicate that DPOAE STCs retained their normal bandpass characteristic and appearance in six of eight hearing-impaired subjects, while two of these subjects showed grossly abnormal suppression tuning. Subtle abnormalities of DPOAE suppression were observed in all hearing-impaired children, but even these did not follow a single predictable pattern of abnormality as has been reported for other types of tuning functions such as basilar-membrane or neural tuning curves.

One might argue that the low-amplitude, reduced SNR DPOAEs in the hearing-impaired subjects (relative to normal-hearing subjects) could lead to inferior-quality recordings and be responsible for the abnormal DPOAE suppression tuning observed in some of these hearing-impaired children. However, during the course of this experiment, DPOAE suppression data were recorded from several normal-hearing children with low DPOAE amplitude to test this possibility. The data from one such subject are shown in Fig. 10. This subject had an unsuppressed DPOAE of -1.74 dB SPL, noise-floor levels comparable to our hearing-impaired subjects, and completely normal STC morphology and width. It does not appear that a low-level distortion product or reduced SNR is a sufficient condition or can account for abnormal DPOAE suppression tuning.

Another way in which to explore this relationship between DPOAE SNR and abnormal suppression tuning is by scrutinizing data from the two hearing-impaired children with the most grossly abnormal suppression tuning (H04 and H08). DPOAE SNRs from these two subjects are comparable to other hearing-impaired subjects who produced more typical-looking STCs. For example, all but one data point out of the 11 points comprising the STC for subject H08 had $\geqslant 5$ dB SNR; likewise, data points comprising the STC from H04 ranged in SNR from 3.9 to 9.1 dB. The similarity of DPOAE SNR between the two hearing-impaired subjects that produced abnormal STCs and the other hearing-impaired subjects that produced more typical suppression tuning effectively dissociates SNR and abnormal suppression tuning. The data, instead, suggest that the cochlear pathology associated with SNHL in subject H04 and H08 was responsible for altering STC morphology and for disrupting the normal frequency-dependent pattern of suppression growth.

It must be acknowledged that the sensation level of primary tones was not matched between groups. Because of the presence of higher audiometric thresholds, the primary tones were stimulating the cochleae of the hearing-impaired subjects at attenuated sensation levels. This difference could have contributed somewhat to the different DPOAE suppression results in hearing-impaired and normal-hearing subjects. This is a common methodological dilemma when recording psychoacoustic tuning curves in individuals with hearing loss (Moore and Glasberg, 1986). However, in a recent study of DPOAE STCs as a function of primary tone level, tuning curves recorded with low-level primary tones (55–45 dB SPL) were sharper and narrower than tuning curves recorded with higher primary tones (75–65 dB SPL) (Abdala, 2001a). This argues against the sensation level variable as a significant factor in the abnormal suppression reported here from hearing-impaired ears. It is not clear, however, how primary-tone sensation level impacts DPOAE STCs measured from the damaged cochlea. Future studies could record DPOAE suppression data from hearing-impaired subjects at various primary-tone levels to address this question.

A. Suppression tuning curves

Suppression tuning from each of the hearing-impaired ears differed from normal in at least one feature; however, abnormalities were subtle and most STCs recorded from hearing-impaired children retained a normal appearance and a clearly bandpass characteristic. The gross abnormality in DPOAE suppression for two of these subjects may reflect the extent of cochlear damage or a unique and particularly detrimental pattern of sensory damage. In the six other hearing-impaired subjects, the primary patterns of abnormality included broadened tuning, somewhat elevated tip levels, and slightly shifted tip frequencies (to a lower frequency). Given the variability in the findings, analyzing mean data may not be useful with this population and may miss subtle but significant effects of cochlear pathology. It is possible that smaller suppression criteria may enhance the sensitivity of

DPOAE suppression techniques to SNHL in humans. This variable should be further investigated; however, preliminary data recently presented (Gorga *et al.*, 2003) suggest that when a 3-dB suppression criterion was used in a group of hearing-impaired adults, DPOAE STCs were no more sensitive to SNHL and still did not reflect cochlear damage in a ''conventional'' fashion (i.e., as observed with basilar-membrane or neural tuning curves).

DPOAE STC results from the hearing-impaired subjects can be compared with the well-established effects of sensory cell damage on basilar-membrane or neural tuning measurements. Both basilar-membrane and neural tuning curves become broader and demonstrate an elevated tip following cochlear damage (e.g., Harrison and Evans, 1979; Liberman and Dodds, 1984; Ruggero and Rich, 1991; Sewell, 1984). Broader tuning curves with larger Q_{10} values were evident for four of the eight individuals, but two other subjects had normal Q_{10} values and two had sharper than normal STCs. Likewise, although three of the subjects had elevated tip levels, the other five did not.

It is evident that DPOAE STCs have similar features to tuning curves derived from single auditory-nerve fibers or the basilar membrane, suggesting that they might reflect cochlear tuning in a similar manner. However, the data reported here from children with SNHL loss and recent studies of DPOAE STCs measured in laboratory animals (Howard *et al.*, 2002; Martin *et al.*, 1998) have indicated that comparisons between DPOAE STCs and other tuning curves are not straightforward. Martin and colleagues (1998) recorded DPOAE STCs in rabbits before and after administering ethacrynic acid, a loop diuretic known to cause reversible alterations to neural and basilar-membrane tuning curves (e.g., Evans and Klinke, 1982; Ruggero and Rich, 1991; Sewell, 1984). Martin *et al.* did not observe major changes in the width and morphology of the DPOAE STCs during the period when ethacrynic acid is known to have maximum effect, even though DPOAE amplitudes were substantially reduced. They did, however, report decreased tip frequencies and increased tip levels, similar to results from some of the hearing-impaired subjects tested in this investigation.

Howard *et al.* (2002) reported markedly different results when they produced temporary threshold shifts in rabbits using moderate-level noise. They measured DPOAE STCs before and after noise exposure and observed changes in STC morphology, but not in the expected direction. DPOAE STCs became slightly narrower following the noise exposure. They also reported a tendency for STC tip levels to decrease and tip frequencies to increase. These results, as well as those presented here, indicate that changes in DPOAE STCs do not necessarily mirror those of neural and basilar tuning curves following temporary or permanent cochlear damage.

The shift of the STC tip to a lower frequency in the majority of the hearing-impaired ears does match the pattern seen in neural tuning curves following OHC loss. In the case of neural tuning curves, the shift of tip frequency probably reflects the loss of cochlear amplifier activity. The cochlear amplifier is thought to enhance basilar-membrane motion at the region that is just basal to the area of maximum displacement for the stimulus-driven traveling wave (e.g., Neely and Kim, 1986). Loss of this enhancement would result in a loss of the sharply tuned tip of the excitation pattern and a shift in the tip to the place of maximum displacement dictated by passive features of the basilar membrane, which occurs at a lower frequency. In the case of DPOAEs, the STC tip frequency is considered the DPOAE generation site in the normal cochlea and corresponds to the most effective suppressor tone, which is typically centered near $f2$ when 4–6-dB suppression criteria are used (e.g., Kummer *et al.*, 1995; Martin *et al.*, 1987). These STCs would have had slightly different morphology and may have been centered at lower frequencies if lower suppression criteria had been applied (Kummer *et al.*, 1995).

The differences between DPOAE and other tuning curves are not surprising when we consider the differences in the measurement paradigms. DPOAEs STCs are measured using three simultaneous tones, including the two moderate-to-high-level tones used to evoke the DPOAE. In contrast, basilar-membrane and neural tuning curves are recorded using low-level tones. Additionally, the DPOAE measurements involve detection of a low-level acoustic response in the ear canal, distant from its site of origin in the cochlea. With DPOAEs it is likely that the generation site being suppressed is distributed along a larger region of the basilar membrane and possibly consists of multiple sources, compared to the response of a single nerve fiber or a very focal region of the basilar membrane. Six out eight hearing-impaired children retained normal STC morphology and appearance despite documented cochlear damage; therefore, it is clear that DPOAEs STCs do not reflect cochlear tuning in the same way as neural or basilar-membrane tuning curves. It may be that they give a more comprehensive or integrated view of cochlear integrity and one that can be used to monitor changes in function.

B. DPOAE sources and generation

One interpretation of the shifted DPOAE STC tip frequency is that the relative contribution from multiple DPOAE sources in the cochlea changes due to cochlear damage. The DPOAE measured in the ear canal appears to be a combination of energy from more than one source in the cochlea. Results from several studies support the hypothesis that a second source from the DPOAE frequency place, in addition to the primary tone overlap region (i.e., generation site), contributes significantly to the ear canal recording (e.g., Konrad-Martin *et al.*, 2002; Mauermann *et al.*, 1999a, b; Stover, Neely, and Gorga, 1999; Talmadge *et al.*, 1999). The two-source model hypothesizes that distortion is initially generated due to interaction of the two primary tones on the basilar membrane, close to the $f2$ region and then propagates in two directions; basally toward the stapes and apically to the $2f1-f2$ place (i.e., reflection site) to generate a stimulus frequency OAE (SFOAE) (Kim, 1980). This SFOAE is also believed to propagate basally through reverse transduction, into the middle and outer ear where it sums with the primary distortion component. Depending on the phase relationship between the two components, their summation could pro-

duce enhancement or cancellation, resulting in DPOAE fine structure (e.g., Talmadge *et al.*, 1999).

In the present experiment, we are not able to distinguish the relative contribution from the DP place at $2f1-f2$ and the $f2$ region as the generation site, because most subjects had flat or very mildly sloping SNHLs; that is, both sources may have been equally disrupted by the cochlear damage. Other investigators have tested ears with sloping and rising hearing loss to manipulate the contribution of these two sources to the ear canal DPOAE (Konrad-Martin *et al.*, 2002; Mauermann *et al.*, 1999a). Konrad-Martin and colleagues (2002) recently studied inverse fast Fourier transforms of DPOAE filter functions in ears with various audiometric configurations to address this question. Their findings indicate clear support for the two-source model of DPOAE generation in subjects with rising SNHL. These subjects showed an enhanced contribution from the $f2$ region and reduced contribution from the damaged DP site. Mauermann *et al.* (1999a, b) reported the loss of DPOAE fine structure associated with the DP reflection site for this same type of hearing loss. The generally flat audiograms from our subjects do not allow for this kind of analysis, although it has been hypothesized that when equal damage is present at the $f2$ and DP-site frequency (i.e., a flat loss), the relative magnitudes of each of these sources is normal (Konrad-Martin *et al.*, 2002). Thus, it remains uncertain whether the slightly shifted tip frequencies in this study reflect altered patterns of contribution from the two sources.

A second possibility is that the DPOAE generation site, at the $f2$ region, may shift toward the apex in subjects with SNHL. It may be that the point of maximum overlap of traveling waves produced by $f1$ and $f2$ has been altered somewhat due to cochlear damage and its impact on vibratory motion of the basilar membrane. This altered basilar-membrane activity could produce a shift in STC tip frequency by changing the point at which the excitation patterns of the two primaries interact in a nonlinear manner. This hypothesis warrants further study.

C. DPOAE suppression growth

DPOAE iso-suppression tuning curves capture a static point along the suppression growth function to provide a measure of cochlear function. DPOAE suppression growth provides a dynamic measure of function and has been shown to reflect cochlear nonlinearity (Abdala and Chatterjee, 2002). In the normally functioning cochlea, suppression of the $2f1-f2$ DPOAE follows a systematic, frequency-dependent pattern (Abdala and Chatterjee, 2002; Gorga *et al.*, 2002; Harris *et al.*, 1992; Kummer *et al.*, 1995). Suppressor tones lower in frequency than $f2$ produce at least linear, usually expansive suppression of the DPOAE; suppressor tones above the $f2$ frequency produce compressive, shallow suppression growth. This pattern is consistent with direct measurements of basilar-membrane response growth (Rhode and Cooper, 1993), psychoacoustic measures of masking growth (Plack and Oxenham, 1998), and measures of neural rate growth (Delgutte, 1990).

The frequency-dependent pattern of DPOAE suppression growth has its basis in the recognized nonlinear characteristics of basilar-membrane motion and depends upon the known differential effect of a masker or suppressor well below the probe (i.e., the upward spread of masking) compared to a masker/suppressor well above the probe or at probe frequency. In this paradigm, the ''probe'' is considered to be $f2$ because this is the primary generation site of the DPOAE and is the region to be suppressed. For suppressor tones lower than $f2$, the ''tail'' of the excitation profile corresponding to the suppressor tone grows at least linearly with increasing suppressor level, thus overlapping and interfering with the $f2$ site effectively. This linear growth of basilar-membrane motion for low-frequency-side suppressors produces steep growth of DPOAE suppression.

In contrast, suppressor tones higher than $f2$ do not produce effective overlap with the $f2$ site and thus do not produce effective DPOAE suppression even when increased in level. This overlap is often not possible because of the compressive nature of basilar-membrane motion at the region tuned to the high-frequency suppressor tone. Therefore, for high-frequency suppressor tones, it is typical to see shallow, compressive growth of DPOAE suppression.

This normal frequency-dependent pattern of DPOAE suppression growth is not observed in term-born neonates or neonates that were born prematurely but have reached the equivalent of term status (39–41-weeks postconceptional age). Neonates show shallow growth of suppression for suppressors lower in frequency than $f2$ (''low-frequency side'' suppressor tones), suggesting a more focal excitation pattern and less upward spread of masking. It may indicate that the immature cochlea is excessively compressive (Abdala, 2001a, b; Abdala and Chatterjee, 2002). The hearing-impaired subjects tested in this study also show atypical DPOAE suppression growth; however, they have a more variable pattern of disruption.

Three aspects of DPOAE suppression growth are abnormal in these hearing-impaired subjects: (1) The normal frequency-dependent pattern of cochlear nonlinearity has been disrupted; that is, we do not see the normal compressive growth for high-frequency-side suppressors in contrast to the steep growth of suppression expected for low-frequency-side suppressor tones; (2) In six of the eight subjects, the slopes are abnormally steep for at least one suppressor frequency, suggesting that it is generally easier to suppress the DPOAE in hearing-impaired subjects. The steeper slopes may indicate localized regions of pathology. Excessively steep suppression growth may indicate broadened excitation patterns and excessive spread of masking in impaired cochlea; (3) The slopes of the suppression growth obtained in these subjects are irregularly distributed across suppressor frequencies and show marked intersubject variability. The hearing-impaired data show peaks and troughs spaced at erratic intervals in frequency. These three irregularities in suppression growth may suggest abnormalities in cochlear function, possibly related to basilar-membrane nonlinearity and compression.

D. DPOAE suppression and underlying pathology

The findings from this study clearly indicate that different DPOAE suppression results are observed in subjects with

similar audiometric results, that is, similar degree (±5 dB) and configuration of hearing loss. Five of the subjects presented here had relatively flat mild-moderate hearing losses (see Appendix B), yet their STCs varied widely in the extent of abnormality. A striking example can be found by comparing subjects H04 and H21. Both are teenage females with nearly identical bilateral, symmetrical hearing losses. However, their STCs are extremely different. Subject H21 presents with an STC of normal width and morphology, while H04 presents with an STC that is grossly abnormal in morphology. The results may suggest that subject H04's hearing loss stems from more widespread loss of OHCs or a more detrimental pattern of sensory cell loss. A similar comparison can be made between H08's STC, which is grossly abnormal in shape and H14's normal-appearing tuning curve, although both subjects show similar audiometric thresholds. These examples are limited by the inability to compare two subjects with similar audiograms at the same $f2$ frequency.

A pattern observed in only two hearing-impaired subjects was narrower suppression tuning. The two subjects with sharper STCs (H12, H19) also have the least overall hearing loss with some normal-hearing thresholds both above and below the $f2$ frequency. Similar findings of narrow suppression tuning have been reported in premature human infants (e.g., Abdala, 2001a, b) and may reflect a very subtle dysfunction of the cochlear amplifier due to mild OHC damage or, in the case of premature neonates, OHC immaturity. Martin and colleagues (1998) have hypothesized that a narrower STC reflects a DPOAE generation region that is significantly reduced or narrowed. However, it is not clear how hearing loss would create a more focal overlap region between $f1$ and $f2$ traveling waves. It is also possible to produce narrower DPOAE STCs by manipulating stimulus parameters such as $f2/f1$ ratio and primary tone level. This explanation is unlikely since, with the exception of $f2$ frequency, stimulus parameters were the same in all subjects. At present, it is not clear why narrow STCs were observed in these two hearing-impaired subjects and whether this result reflects a unique type of damage to the cochlear amplifier.

The variation of DPOAE suppression results in this relatively small sample of hearing-impaired subjects may well stem from variations in the local patterns of hair-cell survival and consequently, of cochlear amplifier function. Unfortunately, our current understanding of cochlear mechanics does not extend so far as to predict the patterns of suppression in regions with irregularly distributed or partially functioning hair-cell survival. Further study of DPOAE suppression growth patterns in populations with defined etiology of hearing loss may elucidate this relationship. At present, however, it is apparent that the normal frequency-dependent pattern of cochlear nonlinearity as measured by DPOAE suppression growth is disrupted in subjects with cochlear damage.

V. CONCLUSIONS

Results of the present study suggest that DPOAE STCs measure a different aspect of cochlear function and tuning than traditional physiological tuning curves. It is presumed that the cochlear amplification process is disrupted in these subjects with mild-moderate SNHL, because outer hair-cell function is abnormal. Functional outer hair cells are required for normal cochlear amplifier function; therefore, in the absence of normal outer hair cells, phenomena associated with the amplifier, such as sharp tuning and optimal sensitivity, are compromised (Liberman and Dodds, 1984; Ruggero and Rich, 1991). Even so, six out of eight STCs recorded in children with documented cochlear damage (i.e., SNHL) showed generally normal DPOAE STC appearance and shape. Interestingly, however, these six hearing-impaired children showed subtle abnormalities in DPOAE suppression tuning, some related to tuning curve width and others related to shifting tip frequencies and levels. These subtle abnormalities support the first hypothesis and suggest that the DPOAE suppression paradigm reflects cochlear dysfunction in a systematic, although novel and subtle, way. The answer to the second hypothesis is less clear. Although our results suggest that the same audiometric hearing loss may alter DPOAE suppression in varied ways, it is not clear whether this variation reflects different patterns or types of cochlear damage. Further study with a larger database is needed to elucidate the relationship between DPOAE STCs, and underlying patterns or etiologies of hearing loss in both humans and laboratory animals.

ACKNOWLEDGMENTS

The authors would like to thank Leslie Visser-Dumont for her assistance in the preparation of this manuscript. This experiment was funded by the March of Dimes Grant No. 12-FY00-211 and by the House Ear Institute.

APPENDIX A: DESCRIPTIVE INFORMATION ON THE EIGHT HEARING-IMPAIRED PARTICIPANTS

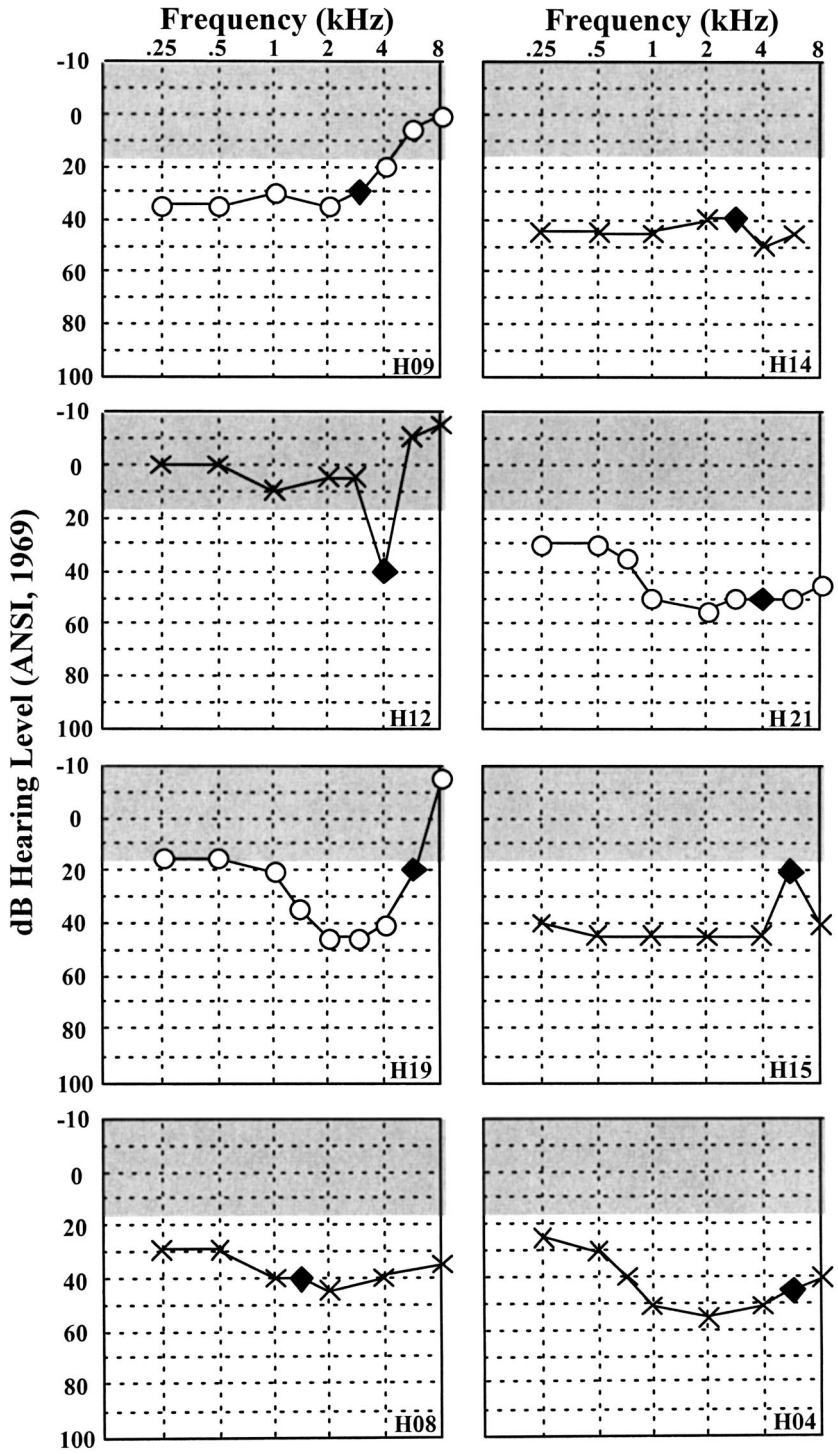
ID #	Age in study (years)	Age at ID[a] (years)	Age 1st amp[b] (years)	Gender	Test ear	$f2$ tested (Hz)	dB HL @$f2$	DPOAE level (dB SPL)
H04	17	4	4	F	L	6000	45	−7.44
H08	12.5	5	5	F	L	1500	40	−1.18
H09	11	6	6.5	M	R	3000	30	−4.41
H12	10	7.5	N/A	M	L	4000	40	−1.00
H14	10.5	5	6	F	L	3000	40	−1.94
H15	14.5	5	6	M	L	6000	20	−4.35
H19	14.5	14	14.5	M	R	6000	20	−0.66
H21	14.5	4	6	F	R	4000	50	−9.70

[a]Age at which the child's hearing loss was first identified.

[b]Age at which the child first began using amplification.

APPENDIX B

Audiograms for the test ear of each of the eight hearing-impaired subjects. Standard audiometric symbols are used: circles=right ear and ''×''s=left ear. The $f2$ frequency tested for a given subject is indicated by the black diamond.



Abdala, C. (**1998**). ''A developmental study of DPOAE ($2f1-f2$) suppression in premature human neonates,'' Hear. Res. **121**, 125–138.

Abdala, C. (**2001a**). ''Maturation of the human cochlear amplifier: Distortion product otoacoustic emission suppression tuning curves recorded at low and high primary tone levels,'' J. Acoust. Soc. Am. **110**, 1465–1476.

Abdala, C. (**2001b**). ''DPOAE suppression tuning: Cochlear immaturity in premature neonates or auditory aging in normal-hearing adults?'' J. Acoust. Soc. Am. **110**, 3155–3162.

Abdala, C., and Chatterjee, M. (**2003**). ''Maturation of DPOAE suppression growth as a measure of cochlear nonlinearity in humans,'' J. Acoust. Soc. Am. (in press).

Abdala, C., Sininger, Y., Ekelid, M., and Zeng, F-G. (**1996**). ''Distortion product otoacoustic emission suppression tuning curves in human adults and neonates,'' Hear. Res. **98**, 38–53.

American Speech-Language-Hearing Association (**1990**). ''Guidelines for screening for hearing impairment and middle-ear disorders,'' ASHA **32** (Suppl 2), 17–24.

Brown, A., and Kemp, D. (**1984**). ''Suppressibility of the $2f1-f2$ stimulated acoustic emissions in gerbil and man,'' Hear. Res. **13**, 29–37.

Delgutte, B. (**1990**). ''Two-tone rate suppression in auditory-nerve fibers: Dependence on suppressor frequency and level,'' Hear. Res. **49**, 225–246.

Erminy, M., Avan, P., and Bonfils, P. (**1998**). ''Characteristics of the acoustic

distortion product $2f2-f1$ from the normal human ear,'' Acta Oto-Laryngol. **118**, 32–36.
Evans, E., and Klinke, R. (**1982**). ''The effects of intracochlear and systemic furosemide on the properties of single cochlear nerve fibres in the cat,'' Physiol. (London) **331**, 409–427.
Frank, G., and Kossl, M. (**1995**). ''The shape of $2f1-f2$ suppression tuning curves reflects basilar membrane specializations in the mustached bat, *Pternonotus parnellii*,'' Hear. Res. **83**, 151–160.
Gorga, M., Neely, S., Dorn, P., and Konrad-Martin, D. (**2002**). ''The use of distortion product otoacoustic emission suppression as an estimate of response growth,'' J. Acoust. Soc. Am. **111**, 271–284.
Gorga, M., Neely, S., Dierking, D., Dorn, P., and Fitzpatrick, D. (**2003**). ''DPOAE suppression in normal-hearing and hearing-impaired human subjects,'' Abstr. Assoc. Res. Otolaryn., Daytona Beach, FL, Vol. **26**, p. 171.
Harris, F., Probst, R., and Xu, L. (**1992**). ''Suppression of the $2f1-f2$ otoacoustic emission in humans,'' Hear. Res. **64**, 133–141.
Harrison, R., and Evans, E. (**1979**). ''Cochlear fibre responses in guinea pigs with well defined cochlear lesions,'' Scand. Audiol. Suppl. **9**, 83–92.
Howard, M., Stagner, B., Lonsbury-Martin, B., and Martin, G. (**2002**). ''Effects of reversible noise exposure on the suppression tuning of rabbit distortion-product otoacoustic emissions,'' J. Acoust. Soc. Am. **111**, 285–296.
Kim, D. (**1980**). ''Cochlear mechanics: Implications of electrophysiological and acoustical observations,'' Hear. Res. **2**, 297–317.
Konrad-Martin, D., Neely, S., Keefe, D., Dorn, P., Cyr, E., and Gorga, M. (**2002**). ''Sources of DPOAEs revealed by suppression experiments, inverse fast Fourier transforms, and SFOAEs in impaired ears,'' J. Acoust. Soc. Am. **111**, 1800–1809.
Kummer, P., Janssen, T., and Arnold, W. (**1995**). ''Suppression tuning characteristics of the $2f1-f2$ distortion product otoacoustic emission in humans,'' J. Acoust. Soc. Am. **98**, 197–210.
Liberman, M., and Dodds, L. (**1984**). ''Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves,'' Hear. Res. **16**, 55–74.
Lonsbury-Martin, B., Harris, F., Hawkins, M., Stagner, B., and Martin, G. (**1990**). ''Distortion product otoacoustic emissions in humans. I. Basic properties in normal-hearing subjects,'' Ann. Otol. Rhinol. Laryngol. **99**, Suppl. 147, 29–44.
Martin, G., Lonsbury-Martin, B., Probst, R., Scheinin, S., and Coats, A. (**1987**). ''Acoustic distortion products in rabbit ear canal. II. Sites of origin revealed by suppression contours and pure-tone exposures,'' Hear. Res. **28**, 191–208.
Martin, G., Jassir, D., Stagner, B., and Lonsbury-Martin, B. (**1998**). ''Effects of loop diuretics on the suppression tuning of distortion product otoacoustic emissions in rabbits,'' J. Acoust. Soc. Am. **104**, 972–983.
Mauermann, M., Uppenkamp, S., van Hengel, P., and Kollmeier, B. (**1999a**). ''Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. I. Fine structure and higher-order DPOAE as a function of the frequency ratio $f2/f1$,'' J. Acoust. Soc. Am. **106**, 3473–3483.
Mauermann, M., Uppenkamp, S., van Hengel, P., and Kollmeier, B. (**1999b**). ''Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. II. Fine structure for different shapes of cochlear hearing loss,'' J. Acoust. Soc. Am. **106**, 3484–3491.
Moore, C., and Glasberg, B. (**1986**). ''Comparisons of frequency selectivity in simultaneous and forward masking for subjects with unilateral cochlear impairments,'' J. Acoust. Soc. Am. **80**, 93–107.
Moulin, A., Bera, J., and Collet, L. (**1994**). ''Distortion product otoacoustic emissions and sensorineural hearing loss,'' Audiology **33**, 305–326.
Neely, S., and Kim, D. (**1986**). ''A model for active elements in cochlear biomechanics,'' J. Acoust. Soc. Am. **79**, 1472–1480.
Plack, C., and Oxenham, A. (**1998**). ''Basilar-membrane nonlinearity and the growth of forward masking,'' J. Acoust. Soc. Am. **103**, 1598–1608.
Rhode, W., and Cooper, N. (**1993**). ''Two-tone suppression and distortion production on the basilar membrane in the hook region of cat and guinea pig cochleae,'' Hear. Res. **66**, 31–45.
Ruggero, M., and Rich, N. (**1991**). ''Application of a commercially manufactured Doppler shift laser velocimeter to the measurement of basilar membrane vibration,'' Hear. Res. **51**, 215–230.
Sewell, W. (**1984**). ''The effects of furosemide on the endocochlear potential and auditory-nerve fiber tuning curves in cats,'' Hear. Res. **14**, 305–314.
Stover, L., Neely, S., and Gorga, M. (**1999**). ''Cochlear generation of intermodulation distortion revealed by DPOAE frequency functions in normal and impaired ears,'' J. Acoust. Soc. Am. **106**, 2669–2678.
Talmadge, C., Long, G., Tubis, A., and Dhar, S. (**1999**). ''Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **105**, 275–292.

Maturation of cochlear nonlinearity as measured by distortion product otoacoustic emission suppression growth in humans

Caroline Abdala[a)]
Children's Auditory Research and Evaluation Center, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

Monita Chatterjee
Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

(Received 11 December 2002; revised 7 May 2003; accepted 9 May 2003)

The growth of distortion product otoacoustic emission (DPOAE) suppression follows a systematic, frequency-dependent pattern. The pattern is consistent with direct measures of basilar-membrane response growth, psychoacoustic measures of masking growth, and measures of neural rate growth. This pattern has its basis in the recognized nonlinear properties of basilar-membrane motion and, as such, the DPOAE suppression growth paradigm can be applied to human neonates to study the maturation of cochlear nonlinearity. The objective of this experiment was to investigate the maturation of human cochlear nonlinearity and define the time course for this maturational process. Normal-hearing adults, children, term-born neonates, and premature neonates, plus a small number of children with sensorineural hearing loss, were included in this experiment. DPOAE suppression growth was measured at two $f2$ frequencies (1500 and 6000 Hz) and three primary tone levels (55–45, 65–55, and 75–65 dB SPL). Slope of DPOAE suppression growth, as well as an asymmetry ratio (to compare slope for suppressor tones below and above $f2$ frequency), were generated. Suppression threshold was also measured in all subjects. Findings indicate that both term-born neonates and premature neonates who have attained term-like age, show non-adult-like DPOAE suppression growth for low-frequency suppressor tones. These age effects are most evident at $f2=6000$ Hz. In neonates, suppression growth is shallower and suppression thresholds are elevated for suppressor tones lower in frequency than $f2$. Additionally, the asymmetry ratio is smaller in neonates, indicating that the typical frequency-dependent pattern of suppression growth is not present. These findings suggest that an immaturity of cochlear nonlinearity persists into the first months of postnatal life. DPOAE suppression growth examined for a small group of hearing-impaired children also showed abnormalities.
[DOI: 10.1121/1.1590973]

PACS numbers: 43.64.Jb, 43.64.Kc [BLM]

I. INTRODUCTION

Distortion product otoacoustic emissions (DPOAEs) can be suppressed by a third tone presented simultaneously with the two stimulating primary tones ($f1$,$f2$). Ipsilateral suppression of the $2f1-f2$ distortion product has been conducted in laboratory animals (Brown and Kemp, 1984; Frank and Kossl, 1995; Koppl and Manley, 1993; Martin *et al.*, 1987), human adults (Abdala *et al.*, 1996; Gorga *et al.*, 2002; Harris *et al.*, 1992; Kummer *et al.*, 1995) and more recently, in human neonates and children (Abdala *et al.*, 1996; Abdala, 1998; Abdala, 2001a,b) to observe and define characteristics of cochlear nonlinearity and function.

The DPOAE iso-suppression tuning curve has also been used as a tool to study maturation of human cochlear function. Recent studies using longitudinally collected data from premature and term-born neonates clearly show that DPOAE suppression tuning remains non-adult-like beyond term birth (Abdala, 2003). DPOAE suppression tuning curves (STC) from term-born neonates, or from prematurely born neonates that have reached term-like status (38–41 weeks postconceptional age), are excessively narrow and steep on the low-frequency flank relative to adult STCs. This age effect is observed at $f2$ frequencies of 1500 Hz and, more prominently, at 6000 Hz.

DPOAE suppression tuning curves provide a useful *snapshot* of cochlear function for a fixed suppression criterion. DPOAE suppression growth and patterns of growth as a function of suppressor frequency, however, provide a more dynamic and comprehensive measure of cochlear function. Suppression of the $2f1-f2$ DPOAE follows a systematic, frequency-dependent pattern (Kummer *et al.*, 1995). In the mature, healthy cochlea, suppressor tones lower in frequency than $f2$ produce at least linear (1 dB/dB) DPOAE suppression growth; suppressor tones above the $f2$ frequency produce compressive, shallow suppression growth (<1 dB/dB). This pattern is consistent with direct measurements of basilar-membrane response growth (Rhode and Cooper, 1993), psychoacoustic measures of masking growth (Hicks

a)Electronic mail: cabdala@hei.org

and Bacon, 1999; Oxenham and Plack, 1998), and measures of neural rate growth (Delgutte, 1990).

The frequency-dependent pattern of DPOAE suppression growth has its basis in the recognized nonlinear characteristics of basilar-membrane motion and depends upon the known differential effect of a masker or suppressor well below the probe (i.e., the upward spread of masking) compared to a masker/suppressor well above the probe or at probe frequency (Oxenham and Plack, 1998). In the DPOAE suppression paradigm, the ''probe'' is considered to be $f2$ because this is the primary generation site of the DPOAE and is the region to be suppressed. For suppressor tones lower than $f2$, growth of suppression is steep because the ''tail'' of the excitation profile corresponding to the suppressor tone grows in a linear fashion with increasing suppressor level, thus overlapping and interfering with the $f2$ site effectively. This linear growth of the basilar-membrane response for low-frequency side suppressors produces a slope of DPOAE suppression growth approximating 1.0 (Abdala, 2001a; Kummer *et al.*, 1995).

In contrast, suppressor tones higher than $f2$ do not produce effective overlap with the $f2$ site and thus do not produce effective suppression even when increased in level. This overlap is not easily achieved and often not possible because of the compressive nature of basilar-membrane motion at the region tuned to the high-frequency suppressor tone. Therefore, for high-frequency suppressors, we observe shallow growth of DPOAE suppression and compressive slope values (<1.0) (Abdala, 2001a,b; Kummer *et al.*, 1995).

The nonlinear behavior of the basilar membrane is clearly represented in the asymmetry of DPOAE suppression growth for tones lower than $f2$ versus tones higher in frequency than $f2$. Because the DPOAE suppression growth paradigm offers evidence of basilar-membrane nonlinearity, it provides an effective assay for studying maturation of this phenomenon in humans. In this experiment, maturation of cochlear nonlinearity was studied by measuring DPOAE suppression growth in several age groups for a range of primary tone levels and $f2$ frequencies. The research objectives were to: (1) investigate the maturation of basilar-membrane nonlinearity in humans and (2) describe the time course for this maturational process.

II. METHODS

A. Subjects

A portion of the subjects described in this report participated in previous studies of DPOAE suppression tuning in our laboratory. However, new subjects were included to fill in experimental conditions and provide adequate subject numbers in each group. The subjects participating in this experiment included: 37 normal-hearing adults, 47 normal-hearing children, 109 premature neonates, 49 term neonates, and 4 children with sensorineural hearing loss. Table I shows their respective ages, genders, and test ear.

The neonatal subjects were born at Women and Children's Hospital, Los Angeles County+University of Southern California Medical Center and tested prior to discharge in the Infant Auditory Research Laboratory, a quiet room away from the neonatal intensive care unit and postpartum suites. All neonates were screened with a 30-dB HL click-evoked ABR prior to inclusion into this study.

TABLE I. Mean subject age and age ranges, gender, and ear.

Group	Mean age (range)	Gender M/F	Test ear R/L
Premature	35 Wks (31–39)	62/47	46/63
Term	39 Wks (37–42)	36/13	30/19
NH Child	11 Yrs (6.5–17)	16/31	17/30
SNHL Child	14 Yrs (12.5–15.3)	2/2	1/3
Adult	26 Yrs (17–35)	11/26	14/23

Premature subjects were born between 24 and 36 weeks gestational age and ranged from 31 to 41 weeks postconceptional age (PCA) at test. Postconceptional age is defined as gestational age at birth plus weeks between birth and test. Average birthweight for premature neonates was 1964 grams and average 5- and 10-min APGAR scores, which are indicators of neonatal health at birth, were 7.3 and 8.54, respectively (10 reflects optimum health). A small subset of the premature neonates ($n=9$) was tested longitudinally from 31–33 through 45 weeks PCA.

Term neonates were born between 37 and 41 weeks of gestation and tested within 84 h of birth. Average birth weight for term infants was 3283 grams, and 5- and 10-min APGARS were 8.0 and 8.96, respectively. None of the neonates included in this study had high-risk factors for hearing loss other than prematurity (Gerkin, 1984). Adult and child subjects had screening tympanometry (226-Hz probe tone) and a complete audiogram conducted prior to testing. Audiometric thresholds were $\leqslant$15 dB HL from 500 to 8000 Hz. The adults and children were tested at the House Ear Institute in a sound-attenuated booth while sitting quietly, reading, or watching a subtitled video movie.

B. Instrumentation and signal analysis

An Ariel DSP16+ signal processing and acquisition board housed within a Compaq Prolinea 590 personal computer with Pentium processor was used to generate stimuli and acquire data. The Ariel board was connected to an Etymotic Research ER-10C probe system and to an analog high-pass filter (12-dB/oct; 710-Hz high-pass cutoff). The ER-10C probe contains two output transducers and a low-noise microphone. The two primary tones and the suppressor tone were generated by the DSP processor. The primary tone at $f1$ was generated by one D/A converter and delivered via one transducer. The primary tone at $f2$ and the suppressor tone (f_s) were produced by the second D/A converter and output through the second transducer.

The signal at the probe microphone was high-pass filtered and sampled at a rate of 50 kHz with a sweep length of 4096 samples, giving frequency resolution of 12.2 Hz. Twenty-five sweeps of the microphone signal were added and comprised one block for $f2=6000$ Hz. Due to elevated noise, 50 sweeps comprised one block at $f2=1500$ Hz.

C. Data acceptance criteria

Acceptance criteria were as follows: (1) Noise measurements for three frequency bins (12.2 Hz wide) on either side of the $2f1-f2$ frequency had to be <0 dB SPL to assure appropriate subject state and (2) The measured DPOAE level had to be at least 5 dB above the average noise measured in the same six bins around the distortion product frequency to be accepted into the grand average.

The program attempted up to six blocks of either 25 or 50 sweeps to achieve the absolute noise criterion of 0 dB SPL and the S/N ratio of 5 dB. If both of these criteria were not met after six attempted blocks, no data were collected and the next condition was initiated. In addition, sweeps were accepted into a block of data only when the estimated rms level in that sweep did not exceed a user-controlled artifact rejection threshold. This level was set for each subject based on observations of baseline activity level determined early in the test session, and modified if necessary during the experiment.

Intermodulation distortion produced by the recording system at $2f1-f2$ was measured with the probe in a Zwislocki coupler for all test conditions. The mean level of distortion was -21 dB SPL and in no case exceeded -17 dB SPL. The recording system noise floor was determined using a similar method with no tones present. The level of system noise floor ranged between -22 and -27 dB SPL depending on frequency.

D. Procedure

An *in situ* calibration procedure was conducted on both output transducers before each subject was tested. Tones of fixed voltage were presented to the transducers at 250 Hz intervals from 500 to 15 000 Hz and the resulting SPL of these tones recorded in each ear canal. Based on this information, an equalization of output levels was performed for each subject to achieve target stimulus levels across test frequencies.

Custom-designed software for the collection of DPOAE suppression data was developed at the House Ear Institute. For a given $f2$ condition (either 1500 or 6000 Hz), an unsuppressed DPOAE was initially recorded. A suppressor tone (f_s) of given frequency was then presented simultaneously with the primary tones, and its level increased in 5-dB steps over a range of intensities from 40 to 85 dB SPL. Between 9 to 15 suppressor tones with frequencies ranging from one octave below to 1/4 octave above $f2$ were presented at intervals between 25–150 cents (one octave=1200 cents). An unsuppressed DPOAE was recorded before each new suppressor tone was presented. DPOAE amplitude was plotted as a function of suppressor level for each suppressor tone, and the slope of this line was calculated to obtain suppression growth rate.

E. Data analysis

DPOAE suppression data were analyzed in the following manner.

(1) For each subject, a graph of DPOAE amplitude as a function of suppressor level was generated for every suppressor tone (f_s) within a given $f2$ condition. *Suppression growth slope* was measured from this function by fitting a regression equation to the line, beginning 1–2 dB below the unsuppressed DPOAE amplitude and using all points with sufficient signal/noise ratio. If there was a plateau present in the initial part of the suppression growth function, it was eliminated in order to represent the monotonic portion of suppression growth. In this way, a slope value was obtained for each suppressor tone, reflecting the rate of suppression growth as level increased. In order to observe age effects on suppression growth slope, an ANOVA (age$\times f_s$) with repeated measures on f_s was conducted at each $f2$ frequency and primary tone level independently.

(2) Once the suppression growth slope values were determined, a slope *asymmetry ratio* (AR) was calculated for each subject. This ratio was a comparison of suppression growth slope for suppressor tones lower than $f2$ (low-frequency side or LFS suppressors) and slope for suppressors tones higher than $f2$ (high-frequency side or HFS suppressors). At $f2=1500$ Hz, the AR was computed from slope measured for $f_s=1050$ Hz and $f_s=1890$ Hz (i.e., LFS slope/HFS slope). At $f2=6000$ Hz, the AR was computed from slope measured for $f_s=5119$ Hz and slope at $f_s=6932$ Hz. The larger the ratio value, the more asymmetrical the DPOAE suppression growth slope across suppressor frequency. To search for age effects on the asymmetry ratio metric, an ANOVA (age$\times$primary tone level) was conducted at each $f2$ frequency independently.

(3) The suppression growth data collected longitudinally in a subset of premature neonates and adult subjects at $f2=6000$ Hz were analyzed in the following manner: A mean *LFS suppression growth slope* (average of suppression growth slope for the two lowest suppressor tones: 3620 and 4063 Hz) was plotted as a function of postconceptional age/test session. Past studies have shown that age effects are not present for HFS suppressor tones, and for this reason, they were excluded from the analysis. An ANOVA (age group$\times$test session) with repeated measures on test session was conducted to test for differences in LFS suppression growth between adult and neonatal groups and for change in LFS suppression growth with increasing postconceptional age.

(4) A DPOAE *suppression threshold* was also calculated as the suppressor level required for a suppressor tone to just initiate suppression; that is, produce a 2-dB reduction in DPOAE amplitude. This threshold value was extrapolated from each suppression growth function individually and plotted as a function of suppressor frequency. Suppression threshold provides a metric of DPOAE suppression that is not redundant with measures of suppression growth measurement. For example, it is possible for two different suppressor tones to cause comparable growth of suppression rates; however, for one tone, suppression may not be initiated until the suppressor reaches 80 dB SPL; for the other tone, suppression might be

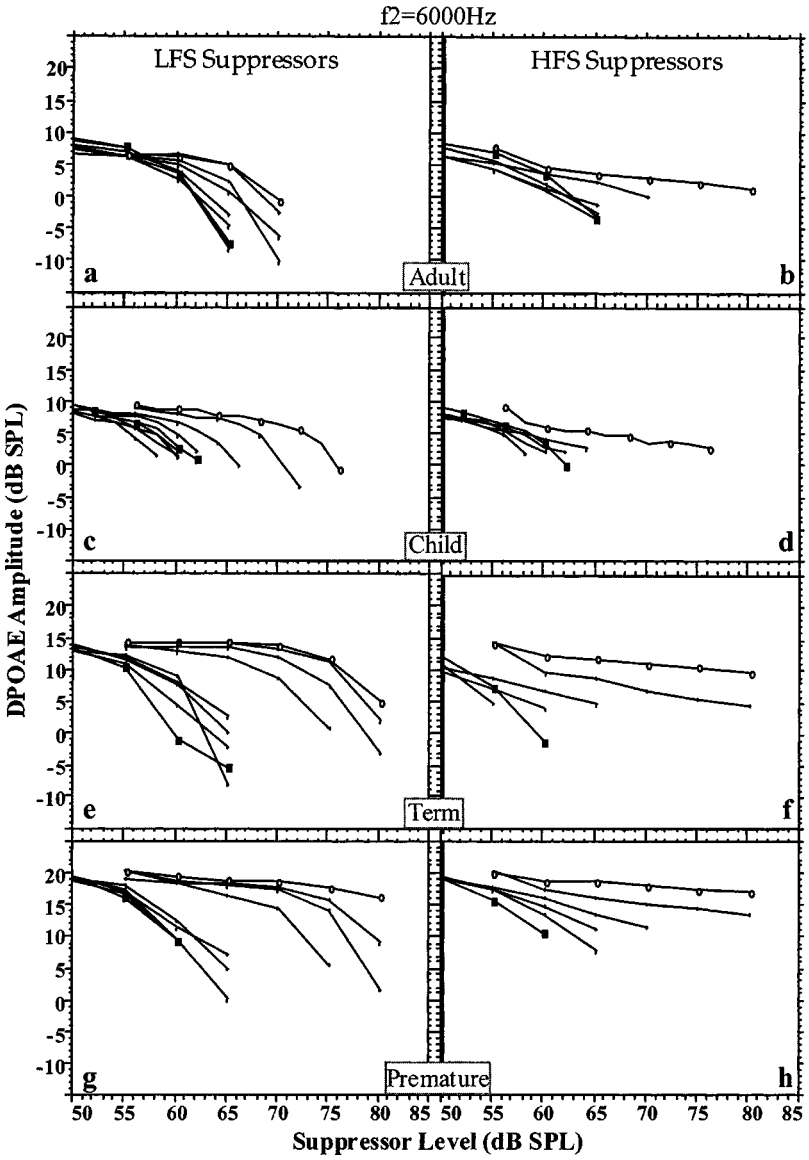


FIG. 1. DPOAE amplitude as a function of suppressor level for a group of suppressor tones centered around 6000 Hz. One representative adult, child, term-born neonate, and premature neonate is shown. Low-frequency side (LFS) suppressors are tones that are lower in frequency than $f2$; HFS suppressors are higher in frequency than $f2$. In each panel, the frequency farthest from f2 is shown as an open circle; the frequency closest to $f2$ is shown by filled squares. Actual suppressor tone frequencies are listed in Appendix A.

initiated at 45 dB SPL. The DPOAE suppression threshold measure distinguishes between these two suppression patterns, whereas the slope measure alone would not. DPOAE threshold measures were displayed as mean values at each $f2$ frequency and primary tone level, for the four age groups (premature, term, child, and adult).

III. RESULTS

A. Individual suppression growth series

Figures 1(a)–(h) show randomly chosen DPOAE suppression growth functions at $f2 = 6000$ Hz for one individual in each age group. The boundary suppressor tones (those nearest and farthest from $f2$ frequency) are designated by filled square and open circle, respectively. The left panels display suppression growth functions for LFS suppressor tones; the right panels display data from HFS suppressor tones. Therefore, the open circles depict the lowest frequency suppressor tone on the left panels and the highest frequency suppressor tones on the right panels. Table II includes the list of suppressor frequencies used in each f2 condition. The adult example [Figs. 1(a) and (b)], represents a typical and expected pattern of suppression growth in the normally functioning, mature cochlea. Low-frequency side-suppressor tones produced steep suppression growth, and HFS suppressor tones produced shallow suppression growth. The normal-hearing child in Figs. 1(c) and (d) showed a similar pattern.

The term neonate [Figs. 1(e) and (f)], like the adult, had very shallow suppression growth for the HFS suppressors, but also showed relatively shallow slope for some of the LFS

TABLE II. Suppressor frequencies (f_s). Suppressor tone frequencies used for $f2 = 1500$-Hz and $f2 = 6000$-Hz conditions. The numbers with asterisks are those values used in the suppression growth slope graphs and analyses.

$f2 = 1500$ Hz		$f2 = 6000$ Hz	
850	1522*	3044	5914*
1050*	1566	3620*	6087
1208	1612*	4063*	6266*
1345*	1684	4560	6449
1436*	1784*	5119*	6638*
1478	1890*	5350	6932*
		5582*	7239
		5746	

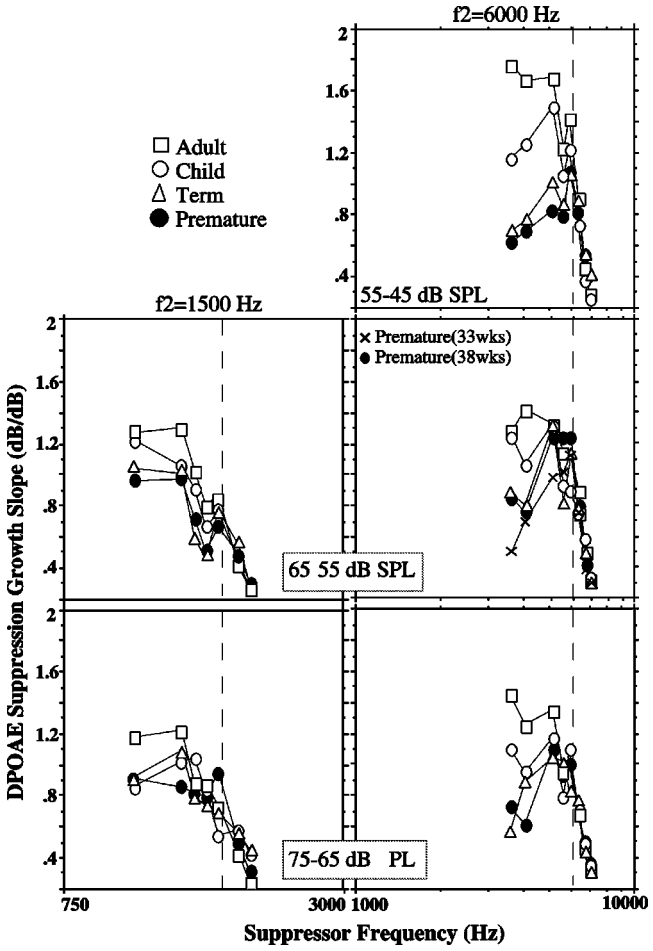


FIG. 2. Mean DPOAE suppression growth slope as a function of suppressor frequency for two $f2$ conditions (1500 and 6000 Hz) and for low-, mid-, and high-level primary tones. The vertical line is placed at the $f2$ frequency. Mean data from four age groups are presented. Actual suppressor tone frequencies are detailed in Appendix A. The standard deviations of the mean slope values are presented in Table III.

suppressor tones. Additionally, the difference in DPOAE suppression threshold between the lowest frequency suppressor tones and the other LFS tones was more marked in the neonate than the adult (left panels). The term neonate shown in Fig. 1 has elevated DPOAE suppression thresholds for the four lowest frequency suppressors (see functions shifted to the right). The premature neonate [Figs. 1(g) and (h)] showed the same elevated suppression threshold for the lowest frequency suppressor tones. This pattern of requiring higher levels to initiate DPOAE amplitude reduction for LFS suppressor tones is unique to neonates and was not observed in the representative adult data. It is consistent with previous reports that neonates have narrower DPOAE suppression tuning curves with steeper low-frequency flanks than adults (Abdala, 2001a). It is also evident in Figs. 1(g) and (h) that LFS suppressor tones produced relatively shallow suppression growth in this premature neonate, rather than the characteristically steep growth (>1.0) typically observed in adults.

B. Mean DPOAE suppression growth slope

Figure 2 shows mean suppression growth slope plotted as a function of suppressor frequency for the four age groups at $f2$ frequencies of 1500 and 6000 Hz and three primary tone levels. In the midlevel condition, data from premature neonates are presented at two postconceptional ages (33 and 38 weeks). DPOAE suppression data could not be collected with low-level primary tone levels for $f2=1500$ Hz because of the higher noise floor associated with this frequency. The dashed vertical line in Fig. 2 is at the $f2$ frequency. The standard deviations at each suppressor frequency are presented in Table III.

The mean suppression growth slope data shown in Fig. 2 quantitatively confirms the anecdotal observations made from the individual suppression growth functions in Fig. 1. The mean adult data show frequency-dependent suppression

TABLE III. Standard deviations for the mean values presented in Fig. 2. Primary tones: LL=55–45 dBSPL, ML=65–55 dB SPL, HL=75–65 dB SPL. Note that an additional category of prematurity is present at 6000 Hz/ML. P=Premature, T=Term, C=Children, A=Adult.

	6000 Hz/LL				6000 Hz/ML					6000 Hz/HL			
f_s	P	T	C	A	P/33	P/38	T	C	A	P	T	C	A
3620	0.282	0.439	0.527	0.321	0.284	0.386	0.557	0.255	0.492	0.489	0.308	0.787	0.540
4063	0.643	0.333	0.602	0.392	0.430	0.312	0.356	0.277	0.324	0.300	0.604	0.305	0.505
5119	0.189	0.421	0.474	0.264	0.300	0.301	0.482	0.254	0.486	0.398	0.421	0.424	0.521
5582	0.369	0.317	0.369	0.193	0.223	0.367	0.281	0.143	0.240	0.291	0.375	0.271	0.201
5914	0.371	0.209	0.254	0.399	0.466	0.366	0.245	0.210	0.283	0.349	0.207	0.372	0.278
6266	0.189	0.204	0.150	0.223	0.099	0.191	0.238	0.183	0.344	0.240	0.236	0.172	0.182
6638	0.081	0.216	0.080	0.208	0.120	0.101	0.184	0.104	0.063	0.210	0.070	0.130	0.116
6932	0.116	0.203	0.060	0.268	0.129	0.227	0.042	0.091	0.052	0.182	0.098	0.071	0.091

	1500 Hz/ML				1500 Hz/HL			
f_s	P	T	C	A	P	T	C	A
1050	0.415	0.414	0.462	0.318	0.507	0.220	0.383	0.466
1345	0.418	0.445	0.294	0.408	0.305	0.556	0.344	0.347
1436	0.448	0.275	0.404	0.427	0.374	0.347	0.423	0.495
1522	0.061	0.097	0.121	0.296	0.256	0.296	0.542	0.419
1612	0.232	0.322	0.424	0.126	0.202	0.347	0.281	0.354
1784	0.085	0.132	0.123	0.116	0.148	0.297	0.041	0.163
1890	0.103	0.053	0.112	0.088	0.138	0.020	0.038	0.092

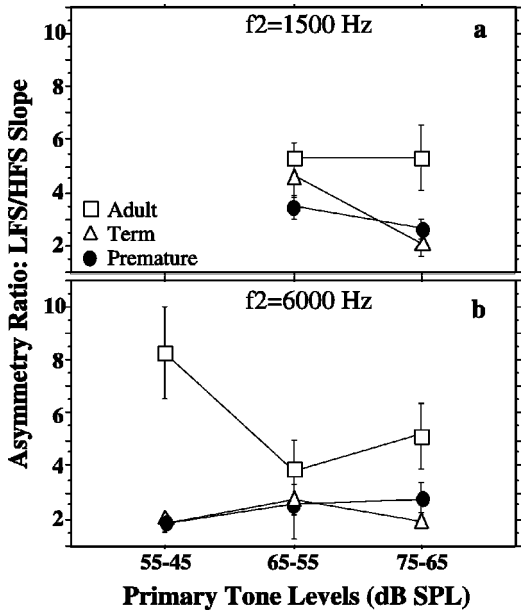


FIG. 3. Mean asymmetry ratio (suppression growth slope for LFS suppressor/slope for HFS suppressor) for three age groups, two $f2$ frequencies and low-, mid-, and high-level primary tones. Error bars represent ± 1 s.d.

growth: HFS suppressors produced compressive rates of suppression growth (0.2–0.8), whereas LFS suppressors produced suppression growth slope values $\geqslant 1.0$. ANOVAs (age$\times f_s$) with repeated measures on f_s, were conducted at each $f2$ frequency and primary tone level separately. Four age groups were included in these analyses: premature neonate, term neonate, child, and adult.

Results at $f2=1500$ Hz indicated a main effect of suppressor frequency at both levels, thus confirming the frequency dependence of DPOAE suppression growth and a main effect of age at 65–55 dB SPL only ($f=6.47$; $p=0.0001$). There was also an interaction between age and suppressor frequency at 75–65 dB SPL ($f=2.36$; $p=0.0078$). From data presented in Fig. 2, it is evident that the interaction effect reflects age differences that are present only for the lowest frequency suppressors, to the left of the dashed vertical line and not for the HFS suppressors.

Results at $f2=6000$ Hz indicated a main effect of suppressor frequency at all levels and a main effect of age for primary tone levels of 55–45 and 65–55 dB SPL only ($f=24.24$; $p<0.0001$ and $f=12.25$; $p<0.0001$, respectively). Interactions between frequency and age were observed at all levels and, again, are consistent with age differences that are present only for the LFS suppressor tones. High-frequency suppression growth showed complete overlap among ages.

Asymmetry ratio (AR) data are shown in Fig. 3. At $f2=1500$ Hz, mean adult ARs were around 5.0 and neonatal values ranged from 2.0 to 4.5. At 6000 Hz, adult AR ranged from 3.7–8.2 while neonatal values were below 3.0. The larger AR values reflect greater asymmetry in suppression growth produced by LFS versus HFS suppressor tones. Factorial ANOVAs (age$\times$primary tone level) were conducted at each $f2$ frequency separately. At 1500 Hz, main effects of age and primary tone level on asymmetry ratio were observed, indicating that adults had larger AR values than neonates ($f=5.39$; $p<0.0087$; $f=14.1$; $p<0.0001$, respectively). There was also an interaction between age and level. At 6000 Hz, there was a main effect of age only ($f=16.92$; $p<0.001$) and an interaction between age and level. The interactions indicate that age effects on the AR were present at all but the mid primary tone levels (see Fig. 3).

C. Longitudinal suppression growth data

Longitudinal suppression growth data were collected on a subset of nine premature neonates and five adults subjects at $f2=6000$ Hz only. Figure 4 shows suppression growth slope as a function of f_s recorded at weekly sessions for two premature neonates and two adults. In this figure, the neonatal test sessions (postconceptional ages) are distinguished by symbol. These data were plotted using a linear frequency scale in order to provide visual detail and to better observe session to session change. The dashed vertical line is at $f2$. The adults showed less session-to-session variability than neonates, and the maxima in adult suppression growth slope ($\cong 1.50$ to 1.75 dB/dB) was always observed at frequencies

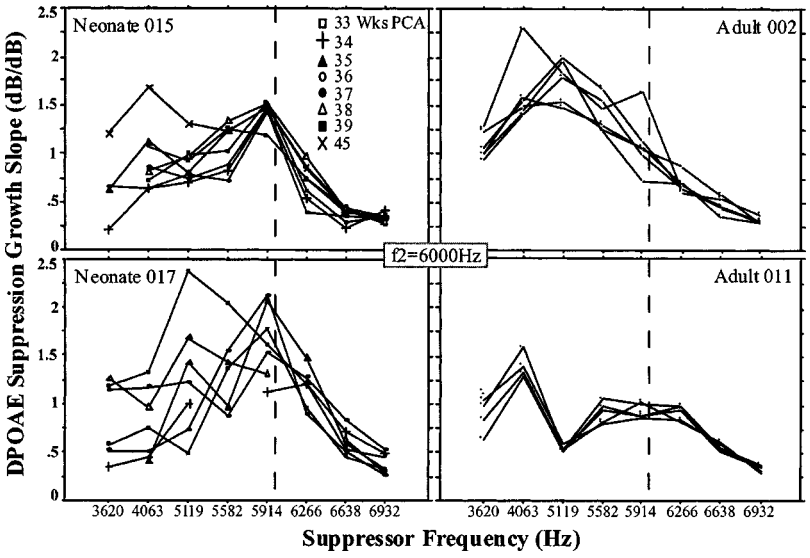


FIG. 4. DPOAE suppression growth slope as a function of suppressor frequency for $f2=6000$ Hz at primary tone levels of 65–55 dB SPL. Each panel displays data from one individual subject tested over repeated weekly sessions. Postconceptional age (PCA) is distinguished by symbol for the two neonatal subjects.

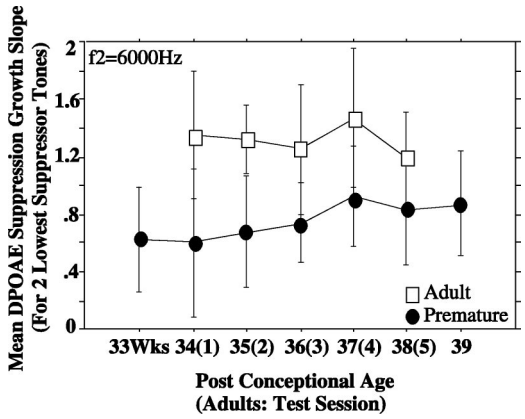


FIG. 5. Mean DPOAE suppression growth slope (for $f2=6000$ Hz) measured for the two lowest frequency suppressor tones, as a function of post-conceptional age/test session. Adult and neonatal data (term and premature neonates combined) are shown. Primary tone levels are 65–55 dB SPL. Error bars represent ± 1 s.d.

lower than $f2$. In contrast, the neonates showed a maxima of suppression growth slope around the $f2$ frequency. This maxima appears to be shifting and approximating adult-like values in these two neonates as they mature during weekly testing; however, as a group, neonates did not have adult-like DPOAE suppression growth for LFS suppressors even after reaching ''term-like'' age (see Fig. 5).

Figure 5 shows mean LFS suppression growth in adult subjects tested weekly for five sessions and premature neonates tested initially at 33 weeks PCA and weekly thereafter. Mean LFS suppression growth slope is plotted as a function of test session for nine neonates and five adult subjects. Mean slope values approximated 1.4 dB/dB in adults. Mean suppression growth slope for neonates remained compressive for all test sessions, showing values <1.0 through 39 weeks PCA. An ANOVA of age group (premature and adult)$\times$test session with repeated measures on test session was conducted on these data and indicated an age effect ($f=9.921$; $p=0.0084$) but no session effect. This suggests that suppression growth for LFS suppressors was significantly shallower in neonates than adults, but that slope values did not change significantly as neonates increased in age from 33 to 39 weeks PCA; that is, even after these premature neonates reached term-like status, they did not have mature DPOAE suppression growth for LFS suppressor tones.

D. Mean suppression threshold

Mean DPOAE suppression threshold as a function of suppressor frequency is presented for each age group and level in Fig. 6. These threshold functions are actually mean DPOAE suppression tuning curves using 2-dB suppression criteria. At $f2=1500$ Hz, there were not marked differences in mean DPOAE threshold among ages at mid primary tone levels but, at high primary tone levels, adults and children showed slight threshold elevation around the tuning curve tip region. At $f2=6000$ Hz, age differences were more evident; neonates had a higher suppression threshold ($\approx 70-80$ dB SPL) for the low-frequency flank of the tuning curves at all primary tone levels, although the effect was more prominent at mid- and high levels. In contrast, mean adult suppression thresholds were at relatively low levels (≈ 65 dB SPL) on the low-frequency flank. Additionally, at mid- and high primary tone levels, adults and children showed threshold elevation around the tuning curve tip.

E. Contribution of the middle ear

Figure 7 addresses the possibility that simple attenuation of primary tone levels through an immature neonatal middle-ear system may account for the age differences observed in DPOAE suppression growth slope. Keefe and colleagues (1993, 1994) have documented functional changes in the conductive system of human neonates over the first year of life. It is possible that these changes contributed to age effects observed in measures of DPOAE suppression. Figure 7 shows suppression growth slope for $f2=6000$ Hz as a function of f_s for neonates and adults. Neonatal data are shown using primary tones at levels that were presented 10 dB higher than adult levels. If the immaturity we have observed (i.e., shallow neonatal suppression growth for LFS suppressors) is due to attenuated primary tone levels arriving at the neonatal cochlea, a simple level compensation should make adult and neonatal suppression growth data comparable. As seen in Fig. 7, this was not the case. A 10-dB level ''advantage'' for neonates did not make their DPOAE suppression growth data look more adult-like.

An immature middle ear may also attenuate *reverse* transmission and consequently reduce the DPOAE level recorded in the ear canal. One could speculate that this attenuation might contribute to the reported age effects in DPOAE suppression. However, data from our laboratory indicate that the initial amplitude of the unsuppressed DPOAE does not determine the pattern of suppression growth or DPOAE STC width. We have found that neonates with uncharacteristically low DPOAE amplitude (i.e., more ''adult-like'' amplitude; <3 dB SPL) do not show more adult-like suppression growth for LFS suppressors. Additionally, statistical correlations between initial unsuppressed DPOAE amplitude and DPOAE STC width (Q10) are insignificant at both 1500 and 6000 Hz (data not shown). These results suggest that the initial unsuppressed amplitude of the DPOAE does not determine suppression growth patterns or width of suppression tuning curves.

Finally, if the immature middle ear alters suppressor tone level in neonates, it is possible that we are measuring different segments of the suppression growth function in each age group. This error could lead to artificial age effects in DPOAE suppression growth slope. The DPOAE suppression paradigm, however, includes an intrinsic control for this possibility. In all subjects, the starting point for measurement of suppression growth slope was 1–2 dB down from their unsuppressed DPOAE. This uniform starting point was not based on any absolute DPOAE level but varied according to each subject's own unsuppressed value. This process served to normalize all suppression growth slope measurements and ensure that the same segment of the suppression growth function was measured, regardless of age group.

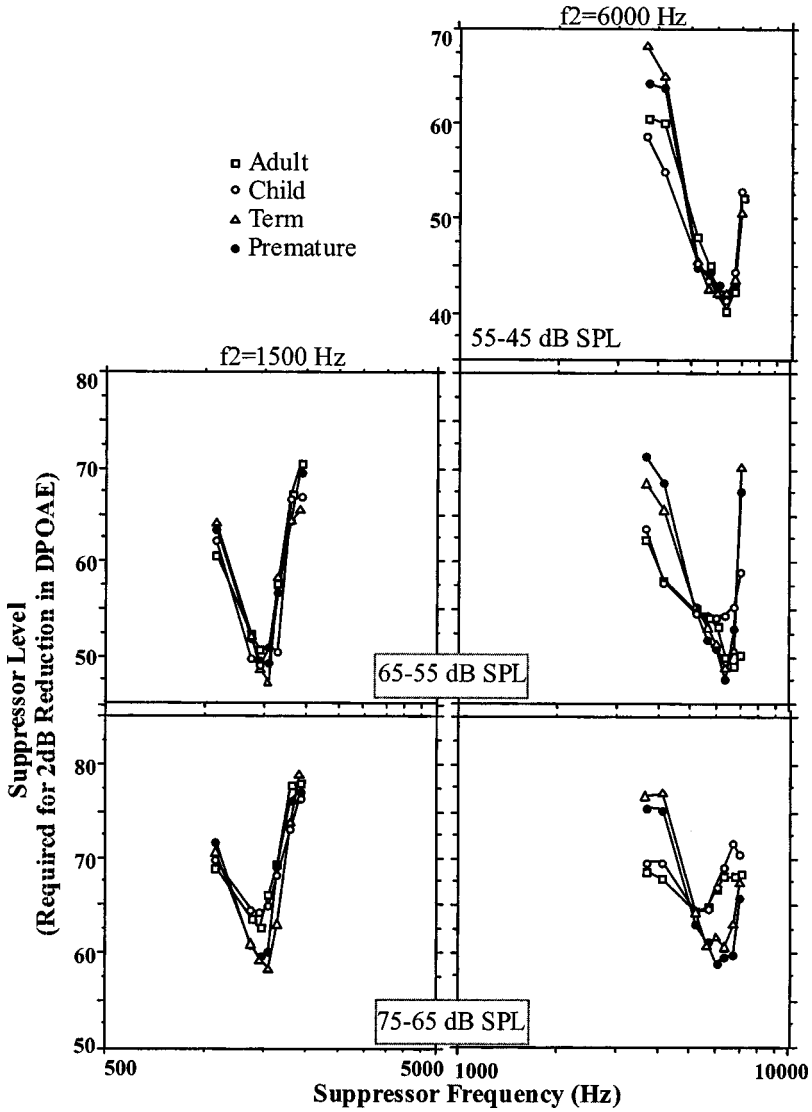


FIG. 6. Mean DPOAE suppression threshold (measured as the suppressor level required to produce a 2-dB reduction in DPOAE amplitude) as a function of suppressor frequency. Data from four age groups are presented at two $f2$ frequencies for low-, mid-, and high primary tone levels.

IV. DISCUSSION

In agreement with earlier studies conducted in humans, the results presented here show that for fixed-frequency primary tones, the rate of DPOAE suppression growth with increasing suppressor level depends strongly on the suppressor frequency and, to a lesser extent, on the level of the primary tones. More importantly, the present experimental results confirm previous work (Abdala, 1998, 2001a) showing a strong effect of developmental age on DPOAE suppression growth. They also extend previous findings to show that age influences the ''asymmetry ratio,'' a ratio between suppression growth on the low-versus high-frequency side of $f2$. Additionally, all the observed differences between the age groups in this study are smaller in the 1500-Hz region than at the 6000-Hz region. This is consistent with the widely held view that the cochlear amplifier is less active in the apex than in the base of the cochlea. Because there may be a less significant impact of the cochlear amplifier in the cochlear apex, DPOAE differences between adults and neonates may be smaller in magnitude and more difficult to detect. At 6000 Hz, in contrast, the amplifier is thought to be highly active and thus its dysfunction or immaturity in neonates may produce more robust age effects.

A. Maturation of cochlear nonlinearity

When the suppressor tone (f_s) is lower in frequency than a fixed $f2$, the tail of the suppressor excitation pattern overlaps with the $f2$ site to suppress the DPOAE. The basilar-membrane response in this tail region is expected to grow linearly with increases in suppressor level. Thus, when $f_s < f2$, normal DPOAE suppression should also grow linearly with increases in suppressor level and produce a slope value approximating 1.0.

When f_s is higher in frequency than a fixed $f2$, we expect a different pattern of interaction. In this case, the overlap between the excitation pattern evoked by the suppressor (f_s) and the $f2$ site is much less. Further, because the region of overlap is in the apical cutoff region of the suppressor's excitation pattern, where the f_s response grows most compressively, DPOAE suppression also grows compressively (<1.0). In this case, the slope of DPOAE suppression is likely to reflect the characteristic frequency (CF) compressive nonlinearity at the f_s place almost exactly.

Thus, we expect to find a change in the slope of DPOAE suppression growth as a function of suppressor frequency in the normally functioning, nonlinear cochlea. This frequency-dependent pattern is observed in adult and children's data as

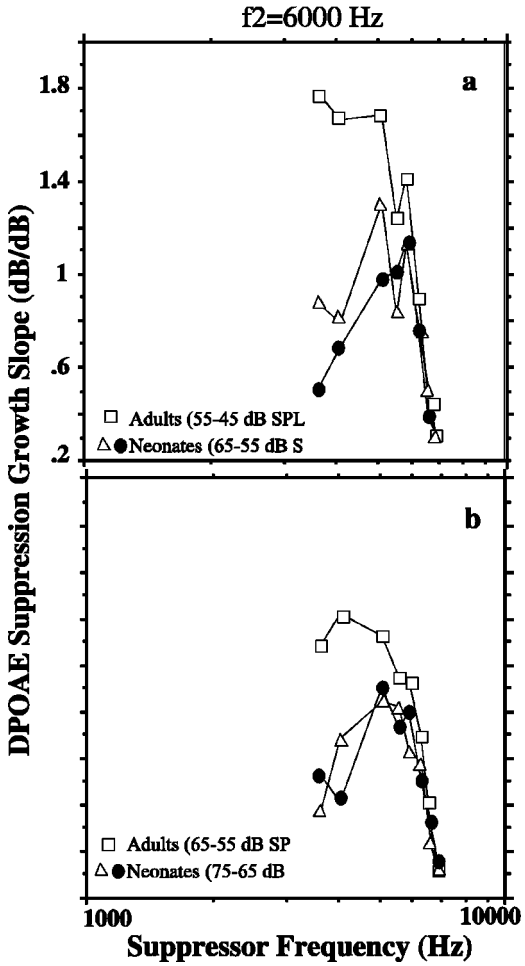


FIG. 7. Mean DPOAE suppression growth slope as a function of suppressor frequency for adults and neonates (triangle=term; filled circle=premature). In both panels, the DPOAE was generated in neonates using 10-dB-higher primary tone levels than adults.

shown in Fig. 2. In fact, the pattern of suppression growth is qualitatively very similar to that obtained by Delgutte (1990) in two-tone suppression experiments on the auditory-nerve response of cats. Other studies of DPOAE suppression in normal-hearing adults have reported a similar dependence of suppression growth rate on suppressor frequency (Abdala, 2001a; Gorga *et al.*, 2002; Kummer *et al.*, 1995). This typical pattern of suppression growth observed in adults and children is not observed in term or premature neonates. Neonates do not show characteristically steep (>1.0) growth of suppression for LFS tones. The DPOAE suppression growth pattern changes systematically (toward adult-like values) as postconceptional age increases. The slopes of suppression growth are similar across age groups for HFS suppressor tones; thus, age effects are dominated by and can be explained completely by suppression growth for LFS suppressor tones.

The systematic frequency-dependent nature of DPOAE suppression growth strongly suggests that it accurately reflects cochlear/basilar-membrane nonlinearity, albeit from a distant source, the ear canal. Given this assumption, excessively shallow growth for suppressor tones below $f2$ in neonates suggests that the tail region of the suppressor's excitation pattern does not increase in a linear fashion with increasing level, in developing cochleae. This finding is relevant, in particular combined with the fact that neonatal DPOAE suppression tuning is sharper and amplitude saturation occurs at elevated levels in neonates when recording DPOAE input/output functions (Abdala, 2000; Abdala, 2001a). Combined, these findings support the hypotheses that the immature cochlea: (1) functions in an active, nonlinear mode for a wider portion of the dynamic range than the mature cochlea, where the active system appears to disengage at high levels; (2) provides excessive sharpness of suppression tuning or an ''overshoot'' of tuning during certain developmental periods; and (3) shows response growth in the neonate that is excessively compressive (<1.0) for tones lower in frequency than CF. A similar period of overshoot has been observed in neonatal gerbils that show excessvie cochlear amplifier gain during certain developmental stages (Mills and Rubel, 1996). We have hypothesized that this developmental period may reflect a cochlear amplifier that is less well regulated in the neonate due to immaturity of the medial efferent system; this system modulates OHC motility and thus, can impact aspects of cochlear amplifier function. Although the source of this effect is not completely known at present, clearly, the results of this study indicate that an immaturity in cochlear nonlinearity is present beyond term birth.

B. Does DPOAE suppression reflect cochlear nonlinearity?

1. The asymmetry ratio (AR)

The results presented here indicate that a key difference between the age groups lies in the asymmetry between the rate of suppression growth for suppressor frequencies above and below $f2$. The ratio between the slope of suppression growth for LFS suppressors and the corresponding slope for HFS suppressors provides an index of this asymmetry. The asymmetry is present in the normal-functioning cochlea and reflects the frequency dependence of basilar-membrane nonlinearity. If the slope of suppression growth reflects cochlear nonlinearity as we hypothesize, then the ratio of the slopes on the apical and basal side of the $f2$ place should be approximately 1.0/0.2=5.0. We note here that the mean slope for adults is in reasonable agreement with this value, ranging from 3.77 to 5.1 at mid- and high primary tone levels and 8.2 at low levels.

Figure 3 shows that the AR values are higher in the adults than premature and term neonates. The lower ratios in the neonates likely reflect an excessively compressive growth of the cochlear response on the basal or ''tail'' side of the peak.

2. Physiological studies

Physiological studies of two-tone suppression on the auditory nerve and basilar membrane have found rates of suppression growth varying from 1 to 3 dB/dB for suppressor frequencies lower than CF. The auditory-nerve data show steeper slopes than the basilar-membrane data (Ruggero *et al.*, 1992; Rhode and Cooper, 1993; Delgutte, 1990). For suppressor frequencies higher than CF, basilar-membrane

measurements show rates of suppression growth varying from 0.28 to 0.48 dB/dB (Ruggero *et al.*, 1992). In the auditory nerve, Delgutte (1990) reported rates of suppression growth ranging from 0.17 to 0.7 dB/dB. Although it is appealing to directly compare the slopes obtained in different experiments, it is important to recognize that in many of the physiological studies cited, suppression is defined as the horizontal shift in the input/output function at CF (i.e., *effective attenuation* of the input by the suppressor). In the present study and in other studies of DPOAE suppression in humans, suppression is defined as a decrement in DPOAE amplitude. The compressive nonlinearity at the source of the DPOAE generation and suppression mechanism is likely to modify the rate of suppression growth relative to that recorded from the basilar membrane using the effective attenuation measure. The intervening hair cell transducer nonlinearity makes the auditory-nerve data even less comparable to the present results.

Rhode and Cooper (1993) and Cooper (1996) measured suppression growth as the rate of change in basilar-membrane output with increasing suppressor level. Because of the similarity between their measurement of suppression growth and DPOAE suppression growth, a comparison between these two slope values is appropriate. For suppressor frequencies lower than CF, these investigators found rates approaching, but not greater than 1.0. This result is similar to the suppression rates obtained with children in the present study for suppressors lower in frequency than $f2$. The adults in our experiments generally showed $\geqslant$1-dB/dB slopes for LFS suppressors. This steep suppression growth in adults may suggest possible dysfunction that produces a greater spread of excitation (consistent with the broader tuning) in the tail region of the suppressor's excitation pattern. The corresponding slopes obtained in the children are much more similar to those obtained in comparable studies of basilar-membrane suppression in animals (Rhode and Cooper, 1993). This may suggest that children are the most ''normal'' hearing of the age groups. As hypothesized earlier (Abdala, 2001b), adults may be showing subtle, subclinical effects of aging and noise exposure that are not revealed in the audiogram.

3. Psychoacoustic studies

Psychophysical experiments in forward masking have long shown steep growth of masking ($>$1.0) for maskers that are lower in frequency than the signal, and fairly linear ($\approx$1.0) slopes for masker frequencies at CF or higher than signal frequency. Oxenham and Plack (1997, 1998) have proposed that growth-of-masking functions reflect basilar-membrane response growth and that this differential growth of masking for a signal at CF and a masker lower than CF is an indication of basilar-membrane nonlinearity. Experiments with naturally existing hearing loss or induced hearing loss support this idea. Hicks and Bacon (1999) showed that growth-of-masking functions can be temporarily and predictably altered by the administration of salicylates to normal-hearing subjects. Salicylates are known to alter cochlear amplifier activity by disrupting outer-hair-cell electromotility (Dieler *et al.*, 1991; McFadden *et al.*, 1984). In contrast to the normal pattern of masking growth for maskers higher versus lower in frequency than CF, the responses from these subjects were the same (linear) for maskers higher and lower than CF and did not exceed 1.0.

Oxenham and Plack (1997) measured growth of masking using low-frequency maskers (comparable to our $f_s < f2$ condition) in hearing-impaired and normal-hearing listeners. As expected, in normal hearers, the growth of masking slopes for low-frequency side maskers were much steeper than growth for a masker at CF where the basilar-membrane response will be compressive. In hearing-impaired adults, however, they found both on-frequency and low-frequency masking conditions produced the same effect, linear growth of masking (1.0). Although these psychoacoustic experiments are fundamentally different from our DPOAE suppression growth measurements, they offer results that are qualitatively similar to the results of DPOAE suppression growth.

The data of Oxenham and Plack suggest growth of masking slope values ranging from 5.88 to 6.25 (corresponding to estimates of compression at CF of 0.16 to 0.17). Note that in their experiments, the low-frequency masker was an octave below the signal, whereas the asymmetry ratio recorded in the present study utilized an LFS slope value closer in frequency to $f2$. Even considering this difference, the ratio values in the present experiment suggest compressive growth at the CF place ranging from 0.285 to 0.117 dB/dB. Given the fundamental differences between psychoacoustic measures and measures of DPOAE suppression, the consistency among results is noteworthy. This consistency supports the contention that both paradigms are probing the same underlying phenomenon, basilar-membrane nonlinearity.

C. DPOAE suppression growth and hearing impairment

Some subjects with mild to moderate hearing loss still produce DPOAEs and thus, given sufficient signal-to-noise ratio, can participate in studies of DPOAE suppression. Preliminary suppression growth data from four mildly hearing-impaired school-aged children are included here for comparison with adult and neonatal suppression (please see Abdala and Fitzgerald, 2003 for detailed report). In cases of mild-moderate SNHL, it is possible that hair-cell survival and/or functionality are irregular and patchy, even when the audiogram shows a flat loss. In such cases, cochlear amplifier activity and cochlear nonlinearity may be compromised and the residual nonlinearity may be distributed unevenly along the cochlea.

Figure 8 shows plots of DPOAE suppression growth slope as a function of suppressor frequency in four hearing-impaired subjects (thin black lines). The normative mean values measured in a group of normal-hearing children are indicated by the thick black lines and $\pm$1 s.d. outlined in gray. The one subject providing suppression growth data at $f2 = 1500$ Hz had an audiometric pure-tone average (@ 500, 1000, and 2000 Hz) of 38 dB HL. The three subjects providing data at $f2 = 6000$ Hz had audiometric pure-tone averages

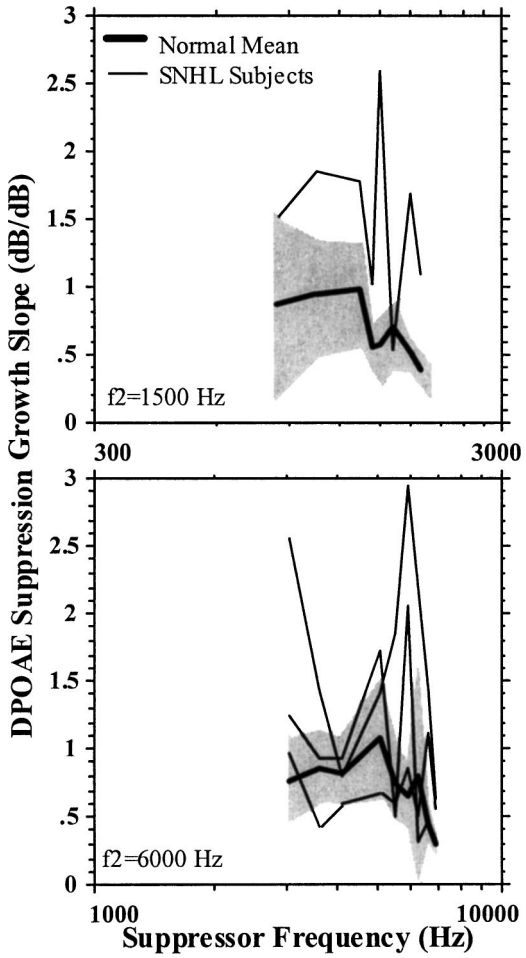


FIG. 8. DPOAE suppression growth slope as a function of suppressor frequency for one hearing-impaired subject at $f2 = 1500$ Hz and three hearing-impaired subjects at $f2 = 6000$ Hz (thin black lines). The thick black line represents the normal mean values for a group of normal-hearing children, and the gray region outlines ± 1 s.d.

of 27, 40, and 45 dB HL; their audiometric thresholds at 6000 Hz were 20, 45, and 25 dB HL.

There are three aspects of DPOAE suppression growth that are abnormal in these hearing-impaired subjects: (1) The growth is not asymmetrical; that is, we do not see the normal compressive growth for HFS suppressors in contrast to the linear growth of suppression expected for LFS suppressor tones. The frequency-dependent pattern of cochlear nonlinearity has been disrupted; (2) In three of the four subjects, the slopes are abnormally steep, for certain suppressor frequencies, suggesting that it is generally easier to suppress the DPOAE. The steeper slopes may indicate localized regions of pathology. Excessively steep suppression growth may indicate broadened excitation patterns and excessive spread of masking in impaired cochleae; (3) The slopes of the suppression growth obtained in these subjects are irregularly distributed across suppressor frequencies and show marked intersubject variability. The hearing-impaired data show peaks and troughs spaced at erratic intervals in frequency.

Unfortunately, our current understanding of cochlear mechanics does not extend so far as to predict the patterns of suppression in regions with irregularly distributed or partially functioning-hair-cell survival. The large variation in this small sample of results from hearing-impaired children may well stem from large or small variations in the local patterns of hair-cell survival/cochlear amplifier function. It is possible that further studies of DPOAE suppression growth patterns in such populations can further elucidate this relationship. These preliminary data, however, suggest that the typical frequency-dependent pattern of cochlear nonlinearity as measured by DPOAE suppression growth may be disrupted in subjects with cochlear damage. It is also apparent that the disruption in DPOAE suppression growth due to cochlear damage is unlike the suppression growth immaturity observed in neonates. This may suggest different underlying mechanisms although this question warrants further investigation.

V. SUMMARY

The results of the present study suggest that immaturity of cochlear nonlinearity persists into the first months of postnatal life. Neonates tested in this experiment show an excessively compressive cochlear response as measured by DPOAE suppression growth for LFS suppressor tones. This immaturity is most marked at $f2 = 6000$ Hz, where we would expect the cochlea to exhibit maximum nonlinearity. Investigation of DPOAE suppression growth patterns into the postnatal period is warranted to fully describe the time course for maturation of cochlear nonlinearity.

The results of this study also suggest that DPOAE suppression growth provides an effective assay of cochlear nonlinearity. The results of this paradigm compare well with other physiological and psychoacoustic measures of this same phenomenon, and DPOAE suppression growth appears to be sensitive to known cochlear damage. Preliminary results indicate that the normal frequency-dependent pattern of cochlear nonlinearity, as measured by DPOAE suppression growth, is disrupted in subjects with cochlear damage. This area also warrants further investigation to detail whether cochlear nonlinearity is disrupted in a similar fashion for cochlear pathology and cochlear immaturity. The preliminary results in hearing-impaired subjects presented here suggest otherwise.

ACKNOWLEDGMENTS

The authors would like to thank Leslie Visser-Dumont and Ellen Ma for their outstanding data collection and management assistance and Tracy Fitzgerald for her work with the hearing-impaired and normal-hearing children. This experiment was funded by a grant from the National Institutes of Health, NIDCD R29 DC03552 and by the House Ear Institute.

Abdala, C., Sininger, Y., Ekelid, M., and Zeng, F-G. (**1996**). "Distortion product otoacoustic emission suppression tuning curves in human adults and neonates," Hear. Res. **98**, 38–53.

Abdala, C. (**1998**). "A developmental study of DPOAE ($2f1-f2$) suppression in premature human neonates," Hear. Res. **121**, 125–138.

Abdala, C. (**2000**). "Distortion product otoacoustic emission ($2f1-f2$) amplitude growth in human adults and neonates," J. Acoust. Soc. Am. **107**, 446–456.

Abdala, C. (**2001a**). ''Maturation of the human cochlear amplifier: Distortion product otoacoustic emission suppression tuning curves recorded at low and high primary tone levels,'' J. Acoust. Soc. Am. **110**, 1465–1476.
Abdala, C. (**2001b**). ''DPOAE suppression tuning: Cochlear immaturity in premature neonates or auditory aging in normal-hearing adults?'' J. Acoust. Soc. Am. **110**, 3155–3162.
Abdala, C. (**2003**). ''A longitudinal study of DPOAE ipsilateral suppression and input/output characteristics in human neonates,'' J. Acoust. Soc. Am. (submitted).
Abdala, C., and Fitzgerald, T. (**2003**). ''Ipsilateral distortion product otoacoustic emission $(2f1-f2)$ suppression in children with sensorineural hearing loss,'' J. Acoust. Soc. Am. (in press).
Brown, A., and Kemp, D. (**1984**). ''Suppressibility of the $2f1-f2$ stimulated acoustic emissions in gerbil and man,'' Hear. Res. **13**, 29–37.
Cooper, N. (**1996**). ''Two-tone suppression in cochlear mechanics,'' J. Acoust. Soc. Am. **99**, 3087–3098.
Delgutte, B. (**1990**). ''Two-tone rate suppression in auditory-nerve fibers: Dependence on suppressor frequency and level,'' Hear. Res. **49**, 225–246.
Dieler, R., Shehata-Dielee, W., and Brownell, W. (**1991**). ''Concomitant salicylate-induced alterations of OHC subsurface cisternae and electromotility,'' J. Neurocytol. **20**, 637–653.
Frank, G., and Kossl, M. (**1995**). ''The shape of $2f1-f2$ suppression tuning curves reflects basilar membrane specializations in the mustached bat, *Pternonotus parnellii*,'' Hear. Res. **83**, 151–160.
Gerkin, K. (**1984**). ''The high-risk register for deafness,'' ASHA **26**, 17–23.
Gorga, M., Neely, S., Dorn, P., and Konrad-Martin, D. (**2002**). ''The use of distortion product otoacoustic emission suppression as an estimate of response growth,'' J. Acoust. Soc. Am. **111**, 271–284.
Harris, F., Probst, R., and Xu, L. (**1992**). ''Suppression of the $2f1-f2$ otoacoustic emission in humans,'' Hear. Res. **64**, 133–141.
Hicks, M., and Bacon, S. (**1999**). ''Effects of aspirin on psychophysical measures of frequency selectivity, two-tone suppression, and growth of masking,'' J. Acoust. Soc. Am. **106**, 1436–1451.
Keefe, D., Bulen, J., Hoberg Arehart, K., and Burns, E. (**1993**). ''Ear-canal impedance and reflection coefficient in human infants and adults,'' J. Acoust. Soc. Am. **94**, 2617–2638.
Keefe, D., Bulen, J., Campbell, S., and Burns, E. (**1994**). ''Pressure transfer function and absorption cross section from the diffuse field to the human infant ear canal,'' J. Acoust. Soc. Am. **95**, 355–371.
Koppl, C., and Manley, G. (**1993**). ''Distortion-product otoacoustic emissions in the bobtail lizard. II. Suppression tuning characteristics,'' J. Acoust. Soc. Am. **93**, 2834–2843.
Kummer, P., Janssen, T., and Arnold, W. (**1995**). ''Suppression tuning characteristics of the $2f1-f2$ distortion product otoacoustic emission in humans,'' J. Acoust. Soc. Am. **98**, 197–210.
McFadden, D., Plattsmier, H., and Pasanen, E. (**1984**). ''Aspirin-induced hearing loss as a model of sensorineural hearing loss,'' Hear. Res. **16**, 251–260.
Martin, G., Lonsbury-Martin, B., Probst, R., Scheinin, S., and Coats, A. (**1987**). ''Acoustic distortion products in rabbit ear canal. II. Sites of origin revealed by suppression contours and pure-tone exposures,'' Hear. Res. **28**, 191–208.
Mills, D., and Rubel, E. (**1996**). ''Development of the cochlear amplifier,'' J. Acoust. Soc. Am. **100**, 428–441.
Oxenham, A., and Plack, C. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.
Oxenham, A., and Plack, C. (**1998**). ''Suppression and the upward spread of masking,'' J. Acoust. Soc. Am. **104**, 3500–3510.
Rhode, W., and Cooper, N. (**1993**). ''Two-tone suppression and distortion production on the basilar membrane in the hook region of cat and guinea pig cochleae,'' Hear. Res. **66**, 31–45.
Ruggero, M., Robles, L., and Rich, N. (**1992**). ''Two-tone suppression in the basilar membrane of the cochlea: Mechanical basis of auditory-nerve rate suppression,'' J. Neurophysiol. **68**, 1087–1099.

Input–output functions for stimulus-frequency otoacoustic emissions in normal-hearing adult ears[a)]

Kim S. Schairer,[b)] Denis Fitzpatrick, and Douglas H. Keefe
Center for Hearing Research, Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 20 December 2002; revised 23 May 2003; accepted 27 May 2003)

Input-output (I/O) functions for stimulus-frequency (SFOAE) and distortion-product (DPOAE) otoacoustic emissions were recorded in 30 normal-hearing adult ears using a nonlinear residual method. SFOAEs were recorded at half octaves from 500–8000 Hz in an $L_1=L_2$ paradigm with $L_2=0$ to 85 dB SPL, and in a paradigm with L_1 fixed and L_2 varied. DPOAEs were elicited with primary levels of Kummer *et al.* [J. Acoust. Soc. Am. **103**, 3431–3444 (1998)] at f_2 frequencies of 2000 and 4000 Hz. Interpretable SFOAE responses were obtained from 1000–6000 Hz in the equal-level paradigm. SFOAE levels were larger than DPOAEs levels, signal-to-noise ratios were smaller, and I/O functions were less compressive. A two-slope model of SFOAE I/O functions predicted the low-level round-trip attenuation, the breakpoint between linearity and compression, and compressive slope. In ear but not coupler recordings, the noise at the SFOAE frequency increased with increasing level (above 60 dB SPL), whereas noise at adjacent frequencies did not. This suggests the existence of a source of signal-dependent noise producing cochlear variability, which is predicted to influence basilar-membrane motion and neural responses. A repeatable pattern of notched SFOAE I/O functions was present in some ears, and explained using a two-source mechanism of SFOAE generation.
[DOI: 10.1121/1.1592799]

PACS numbers: 43.64.Jb, 43.64.Ri, 43.64.Kc [BLM]

I. INTRODUCTION

Evoked otoacoustic emissions (EOAEs) are biological responses to sound that are generated in the inner ear, travel back through the middle ear, and into the external ear canal where they are recorded (Kemp, 1978). In terms of stimulus protocol, EOAEs produced by continuous tonal stimuli fall into two main classes: (1) distortion-product otoacoustic emissions (DPOAE) and (2) stimulus-frequency otoacoustic emissions (SFOAE). DPOAE techniques use two sinusoidal tones of frequencies f_1 and f_2 as stimuli, and measure responses produced at distortion-product (DP) frequencies mf_1+nf_2 (for nonzero integers m and n). SFOAE techniques use one or more sinusoidal tones as the stimulus set, and measure responses produced at the same probe frequency as one of the stimulus tones. DPOAEs represent nonlinear intermodulation distortion generated within the cochlea, and SFOAEs are thought to be a reflected component from the tonotopic site of the probe tone within the cochlea.

A study of EOAE response properties as a function of stimulus frequency and level may have relevance to other physiological and psychoacoustical studies of hearing. In particular, EOAE level as a function of stimulus level, expressed as an input–output (I/O) function, may be an analog to the basilar-membrane (BM) I/O function that is thought to influence such behavioral responses as loudness growth and masking. DPOAE I/O functions in normal-hearing and hearing-impaired human ears have been described in some detail, while studies of SFOAE I/O functions have been based on few subjects and a limited range of stimulus levels.

The research described in this report is concerned with the measurement of SFOAE I/O functions in normal-hearing adult ears over a broad range of stimulus frequencies and levels, comparisons with measurements of DPOAE I/O functions, and relationships between SFOAEs and models that incorporate features of BM mechanics. The remainder of this Introduction reviews literature on: (A) SFOAEs and aspects of DPOAEs relevant to this investigation, and (B) properties of mechanical measurements of BM response growth relevant to the interpretation of EOAE responses.

A. Review of SFOAE and other OAE types

1. Early studies of SFOAEs

If a sine tone is slowly swept in frequency at a fixed voltage level applied to the output transducer, the sound-pressure level (SPL) in the ear canal plotted versus stimulus frequency has fine structure (i.e., sharp, closely spaced peak and valleys) in the response that cannot be attributed to the transducer, the ear canal, or the middle ear (Kemp, 1979; Wilson, 1980). Cochlear reflection locations were considered as sources of the fine structure observed in the SPL spectra, and Kemp (1979) related this SFOAE fine structure to the microstructure in the behavioral audiogram. Zwicker and

[a)]Portions of this work were presented in: D. H. Keefe, K. Schairer, F. Zhao, and J. L. Simmons, ''Measurement and interpretation of stimulus frequency otoacoustic emissions: Input/output functions and transient analyses,'' Presented at the 141st Meeting of the Acoustical Society of America, 4–8 June 2001, Chicago, IL [J. Acoust. Soc. Am. **109**, 2373–2374(A) (2001)].

[b)]Electronic mail: schairerk@boystown.org

Schloth (1984) found that the minima in the fine structure of behavioral thresholds nearly coincided with maxima in the SFOAE fine structure in the same ear.

SFOAE recording methods have taken advantage of the fact that SFOAE level grows less rapidly than the evoking stimulus level. Kemp and Chum (1980) used a vector subtraction technique based on ear-canal responses at high (70–80 dB SPL) and low (40 dB SPL) stimulus levels to separately measure the SFOAE as a nonlinear residual, based on the assumption that the SFOAE was negligible at high levels because it was suppressed by the stimulus. With this vector subtraction method, any distortion in the measurement system led to inaccuracies, and reduced the dynamic range of the SFOAE I/O function. Kemp and Brown (1983) used a difference technique to measure SFOAEs in an analog, swept-tone paradigm, in which the suppressor was 10 dB above the probe-tone level, and 100 Hz above the probe-tone frequency. This technique reduced the effect of distortion because the suppressor remained at moderate levels; however, there was no clear criterion for full suppression of the SFOAE.

2. Digital measurements of SFOAEs

Digital techniques do not necessarily solve the problems of distortion within the measurement transducers, but they do allow more complex signal processing. One way to reduce the amount of distortion associated with SFOAE recordings is to present a four-interval suppression method in which a continuous tone that elicits the SFOAE is suppressed by a pulsed tone (Kemp *et al.*, 1990; Brass and Kemp, 1991, 1993; Souter, 1995). The responses recorded in the ear canal are subtracted such that a ''residual'' contains only a nonlinear portion of the SFOAE response. The two primary stimuli, or tones f_1 and f_2, are presented at levels L_1 and L_2, and may be equal in frequency, in which case the SFOAE is recorded at the frequency of the primaries. Alternatively, SFOAEs may be elicited by primaries that have various f_2/f_1 relationships (Kemp and Chum, 1980; Brass and Kemp, 1991, 1993; Souter, 1995), with one of the primaries acting as a suppressor of the OAE recorded at the frequency of the other primary.

Another digital method to measure SFOAEs used a two-interval suppression method (Guinan, 1990; Dreisbach *et al.*, 1998; Shera and Guinan, 1999), which is a generalization of the analog difference technique. This difference method used a slightly lower or higher frequency tone to suppress the SFOAE recorded at a probe-tone frequency. The spectral difference at the probe frequency between the probe tone alone, and the probe tone plus the suppressor tone, corresponded to the SFOAE. Dreisbach *et al.* (1998) elicited SFOAEs in humans using suppressor tones that were 43 Hz below the probe-tone frequencies of 800 to 14 000 Hz, and were 15–20 dB above the probe-tone level. Shera and Guinan (1999) used a suppressor tone of 15 dB above the probe-tone level of 40 dB SPL, and 43.9 Hz above the probe-tone frequencies of approximately 1000 to 7000 Hz as shown in their Fig. 9. Shera *et al.* (2002) used a similar method. In such SFOAE difference measurements it is unknown whether the SFOAE was partially or fully suppressed by the suppressor tone, but the restrictions to moderate-level probe tones and suppressor tones not more than 20 dB higher in level helped control against system distortion. A three-interval suppression method to measure continuous and gated OAEs (Keefe, 1998) was used in the research reported herein and is described in Sec. II.

Finally, a spectral decomposition method (Shera and Zweig, 1993; Kalluri and Shera, 2001) was based on the rate of phase change as a function of frequency, and was designed to study SFOAE fine structure. The components that varied slowly in phase as a function of frequency were separated by filtering from the components that varied rapidly, and the components that varied rapidly with phase were taken as the SFOAE.

3. Sources of SFOAEs and DPOAEs

SFOAE time delays calculated from the phase slope decrease with increasing frequency, and generally agree with time delays measured using tone-burst responses (Wilson, 1980; Souter, 1995). This is consistent with the generation of SFOAEs near the tonotopic place of the stimulus frequency, so that the SFOAE latencies are equal to the round-trip latency to the tonotopic place. One possible mechanism for generating SFOAEs is the saturating nonlinearity in the BM impedance, which also is invoked for DPOAE generation. The same nonlinearity could act on a single primary and produce a zero-order DP in the sense that the output of a nonlinear function at frequency f would include nonlinear components at the same frequency f (Brass and Kemp, 1993). Two problems with this nonlinear-distortion model have been described: (1) SFOAEs have linear growth at low levels, which suggests a linear mechanism predominates at low levels, and (2) the latencies of SFOAEs predicted from this model are much shorter than are observed (Strube, 1989; Zweig and Shera, 1995; Shera and Guinan, 1999).

To resolve these problems, Zweig and Shera (1995) proposed a source mechanism for SFOAEs (and CEOAEs) based on a linear coherent reflection from distributed inhomogeneities along the BM, which might arise from irregularities in outer hair cell (OHC) patterns. Although SFOAE latencies in gerbil have long latencies at low stimulus levels, consistent with the linear reflection theory, a short-latency SFOAE has been observed in gerbil at higher stimulus levels (Kemp and Brown, 1983). Such a short-latency SFOAE is consistent with the nonlinear-distortion model.

A model of DPOAE generation includes both the linear coherent reflection mechanism and the nonlinear-distortion mechanism with either long or short latencies depending on which source mechanism is dominant (Kalluri and Shera, 2001; Knight and Kemp, 2001). The component near the f_2 place when f_2/f_1 is greater than 1.1 is referred to as the ''wave-fixed'' (Knight and Kemp, 2001), ''low-latency'' (Kemp, 1986), or ''distortion-emission'' (Shera and Guinan, 1999), and it is generated by nonlinearity of the cochlear mechanics. The component from the place of the DPOAE frequency is the ''place-fixed'' (Kemp, 1986; Knight and Kemp, 2001) or ''reflection-source'' emission (Shera and Guinan, 1999), and it is generated by reflection of energy

from a fixed point on the BM. SFOAEs also have been described as a mixture of both place-fixed and wave-fixed sources at higher stimulus levels (Talmadge *et al.*, 2000), so it may be more a matter that: (1) the place-fixed and wave-fixed source regions are closer for SFOAEs at a particular stimulus frequency than for DPOAEs, and (2) the place-fixed region is dominant for SFOAEs at low levels, whereas the wave-fixed region must generate a DPOAE at any level before it can propagate to the place-fixed region and generate a SFOAE to mix with the wave-fixed component.

B. Relationship of BM response growth to behavioral measurements and EOAEs

Direct recordings from the BM (Rhode, 1971; Sellick *et al.*, 1982; Robles *et al.*, 1986; Ruggero *et al.*, 1997; Rhode and Recio, 2000) and derivations from eighth-nerve rate-intensity functions (Yates *et al.*, 1990) in animal models have revealed a nonlinear, compressive growth of BM displacement or velocity with increases in stimulus level when the frequency of the stimulus corresponds to the tonotopic location along the BM at which the response is recorded (i.e., the characteristic frequency, or CF). The BM I/O function has a lower-level linear region, and then a more compressive region (with slope approximately 0.2) above some transition region (Yates *et al.*, 1990; Ruggero *et al.*, 1997). Response growth becomes more linear with damage to the cochlea or death of the experimental animal (Sellick *et al.*, 1982; Robles *et al.*, 1986; Ruggero and Rich, 1991; Ruggero *et al.*, 1997), and when the frequency of the stimulus is above or below CF in a healthy cochlea (Rhode, 1971; Sellick *et al.*, 1982; Robles *et al.*, 1986; Ruggero *et al.*, 1997; Rhode and Recio, 2000). This nonlinear response growth of the BM has been attributed to OHC functioning (Ruggero, 1992; Rhode and Recio, 2000). For an extensive review of this topic, see Robles and Ruggero (2001).

The BM compressive nonlinearity may underlie patterns of results observed in various psychoacoustic paradigms, including intensity difference limens (Yates *et al.*, 1990; Baer *et al.*, 2001), forward masking (Oxenham and Plack, 1997; Plack and Oxenham, 1998; Oxenham and Plack, 2000), and loudness growth (Yates *et al.*, 1990; Moore and Glasberg, 1997; Schlauch *et al.*, 1998). These functions change in listeners with hearing loss, presumably due to loss of the nonlinearity (Moore and Oxenham, 1998). Because direct BM recordings cannot be performed in human listeners, some authors have attempted to use behavioral methods to estimate or derive BM I/O functions (Oxenham and Plack, 1997; Schlauch *et al.*, 1998). However, these behavioral studies can require extended test time and a great deal of attention and cooperation from the listener, and the links between physiology and behavior are indirect.

Another approach may be to use OAE I/O functions as a link between the BM response growth and patterns of results observed in psychoacoustic measurements. If OAE I/O functions are related to BM response growth, they might be used to predict behavioral results that reflect this characteristic. This would be particularly useful for individuals who cannot respond to more demanding behavioral tasks (Buus *et al.*, 2001; Dorn *et al.*, 2001). The first step in this approach is to study the relationship between OAE I/O functions and BM response growth.

DPOAE I/O functions have been reported by several investigators (Lonsbury-Martin *et al.*, 1990; Gorga *et al.*, 1994; Whitehead *et al.*, 1995; Kummer *et al.*, 1998; Withnell and Yates, 1998; Buus *et al.*, 2001; Dorn *et al.*, 2001). Withnell and Yates (1998), for example, recorded DPOAE I/O functions in guinea pigs that were comparable in form to the BM I/O functions reported in the literature. Their DPOAE I/O functions were characterized by an approximately linear slope at low stimulus levels, a breakpoint of 22 to 33 dB SPL, and an upper slope of 0.14 to 0.27 dB/dB. Results from DPOAE I/O function studies in humans demonstrate a compressive, nonlinear growth in ears with normal hearing, similar to BM recordings with stimulus frequencies at CF (e.g., Dorn *et al.*, 2001). Also similar to the BM recordings, ears with hearing loss have more nearly linear DPOAE I/O functions or higher thresholds with steeper growth at moderate stimulus levels (Kummer *et al.*, 1998; Dorn *et al.*, 2001).

SFOAEs may provide a more localized picture of nonlinear response growth because DPOAEs have multiple generator sites, whereas SFOAEs may have a more restricted place of generation (e.g., Brass and Kemp, 1993). Zwicker and Schloth (1984) demonstrated that SFOAE level grows linearly with increasing stimulus level up to approximately 20-dB sensation level in normal-hearing adult ears. Wilson (1980) reported compressive I/O functions at moderate levels in two subjects. Brass and Kemp (1991) reported I/O functions obtained with equal-frequency primaries of 1599 Hz in one subject, with one primary fixed at levels ranging from 40 to 60 dB SPL and the other primary varied by -9 to $+15$ dB around the fixed level. The SFOAE was recorded in the channel of the fixed-level primary as a function of the varied-level primary. Their I/O functions had linear growth at lower levels and more compressive growth at higher levels. Brass and Kemp (1993) later obtained another set of compressive SFOAE I/O functions from three normal-hearing adults.

Existing reports of SFOAE I/O functions have been based on measurements in a sparse number of ears over a modest range of frequencies and levels. While the theoretical issues separating SFOAE and DPOAE mechanisms have been elucidated by SFOAE latency measurements at low levels, and by a large number of DPOAE studies, the problem of SFOAE generation at higher stimulus levels has rarely been systematically studied. The goals of the current study were to: (A) measure SFOAE I/O functions in a population of normal-hearing ears using a wide range of stimulus levels at select frequencies; (B) compare SFOAE I/O functions to DPOAE I/O functions; (C) confirm the biological origin of SFOAEs by accompanying measurements of distortion and noise in a test cavity; (D) characterize the distortion in ears with cochlear implants in which a biological response would not be expected; (E) demonstrate the repeatability of the responses; (F) characterize the difference in noise variance as a function of level between SF and DP conditions; (G) examine novel features found in individual I/O functions; and (H) fit the measured SFOAE I/O function to a two-slope model of SFOAE response growth. Goals A, B, C, and D are ad-

dressed in experiment 1, goals E, F, and G are addressed in experiment 2, and goal H is addressed in the Discussion.

II. THEORY UNDERLYING MEASUREMENTS

The OAE measurements reported in this study were measured using a double-evoked (2E) method, which uses a three-interval paradigm, and an analysis that yields a nonlinear, residual OAE response. In the first interval, a stimulus s_1 is output by a digital-to-analog converter (DAC) channel 1, and DAC channel 2 outputs zeros. The pressure response measured by the microphone in this interval is p_1. In the second interval, a stimulus s_2 is output by DAC channel 2, and DAC channel 1 outputs zeros. The pressure response in this interval is p_2. In the third interval, the stimulus s_1 is output by DAC channel 1, s_2 is output by DAC channel 2, and the pressure response is p_{12}. The 2E OAE waveform response $p_d[n]$ (also termed the distortion waveform) is written as an explicit function of the discrete time step n by

$$p_d[n]=p_1[n]+p_2[n]-p_{12}[n]. \tag{1}$$

The three stimuli satisfy

$$0=s_1[n]+s_2[n]-s_{12}[n]. \tag{2}$$

For present purposes, assume system distortion and noise to be negligible. In the absence of OAEs, the measured stimulus pressure response to $s_1[n]$ is $q_1[n]$, to $s_2[n]$ is $q_2[n]$, and to $s_{12}[n]$ is $q_{12}[n]$. These pressure responses are linear so that

$$0=q_1[n]+q_2[n]-q_{12}[n]. \tag{3}$$

Including OAE contributions and omitting the implied time dependence, the total first-component pressure p_1 includes the sum of the stimulus response and the total OAE $p_{\mathrm{oae}}(q_1)$, i.e., $p_1=q_1+p_{\mathrm{oae}}(q_1)$, in which the OAE pressure varies in general with the level and phase of q_1. The total pressures p_2 and p_{12} have similar forms. It follows that

$$p_d=p_{\mathrm{oae}}(q_1)+p_{\mathrm{oae}}(q_2)-p_{\mathrm{oae}}(q_{12}). \tag{4}$$

If q_1 and q_2 are each a sine tone of the same frequency $f_1=f_2=f$ but different levels L_1 and L_2, the above relates the measured 2E distortion p_d to a linear combination of total SFOAEs. Different SFOAE I/O functions are measured in accordance with different choices of L_1 and L_2. Results have been reported using this technique for CEOAEs, transient-gated SFOAEs and DPOAEs, and continuous SFOAEs (Keefe and Ling, 1998; Keefe *et al.*, 1999; Keefe *et al.*, 2001, 2003; Konrad-Martin *et al.*, 2002; Schairer and Keefe, 2002).

A *fixed-L_1 protocol* has L_1 fixed at a high level on the saturating part of the I/O curve, and varies L_2, under the constraint that L_2 does not exceed L_1. If L_1 is sufficiently large that the SFOAE is approximately fully saturated, then the SFOAE at L_{12}, which is slightly higher than L_1, is also saturated. It follows that the measured distortion approximates the total SFOAE at L_2: $p_d=p_{\mathrm{oae}}(q_2)$. In practical measurements, this prediction was not confirmed because the noise produced by the large L_1 was often high in level, contradicting the assumption of low noise. Some typical results are considered in the Discussion section. This fixed-L_1 protocol has been further explored in a subsequent study in which f_1 and f_2 differ, with f_1 as the suppressor of the SFOAE at f_2 (Schairer and Keefe, 2002).

An *equal-level protocol* is used in most of the results reported herein such that $L_1=L_2$, or $q_1=q_2=q$, and $q_{12}=2q$. Note that the two loudspeakers used in the study to produce the stimulus waveforms q_1 and q_2 were sufficiently well matched in amplitude and phase that q_1 and q_2 were essentially equal whenever their inputs s_1 and s_2 were equal.

It follows from Eq. (4) that the measured distortion in the equal-level protocol is

$$p_d(q)=2p_{\mathrm{oae}}(q)-p_{\mathrm{oae}}(2q). \tag{5}$$

If the SFOAE saturates at L_1, then it also saturates at $L_{12}=L_1+6$ dB, so that $p_d=p_{\mathrm{oae}}(q)$. In the low-level regime for which SFOAEs are described as a linear response, then the equal-level protocol measures two signals that would exactly cancel ($p_d=0$), while the fixed-L_1 paradigm would extract the low-level SFOAE if the noise were sufficiently small.

The coherent reflection emission theory of SFOAE generation was formulated for sufficiently low levels that the SFOAE varies linearly with stimulus level. It is expressed by (Zweig and Shera, 1995)

$$P_{\mathrm{oae}}(Q)=AQ, \tag{6}$$

in which upper-case letters denote spectral variables defined in the frequency domain, e.g., Q is the discrete Fourier transform (DFT) of the waveform q, and P_{oae} is the DFT of p_{oae}. The index corresponding to the center frequency is omitted from all spectral variables. The DFT of each of Eqs. (1)–(5) is well defined, e.g., Eq. (5) in the frequency domain is expressed as $P_d(Q)=2P_{\mathrm{oae}}(Q)-P_{\mathrm{oae}}(2Q)$. The amplitude of the complex attenuation function A in Eq. (6) is the ratio of the SFOAE to the incident pressure amplitudes, and its phase is essentially the round-trip delay from ear canal to cochlear source location near the tonotopic place. It is related to the forward middle-ear (pressure) transmittance T^+, the reverse middle-ear transmittance T^-, the forward middle-ear reflectance R^+, the reverse middle-ear reflectance R^-, and the apical reflectance R that describes the ratio of incoming to outgoing cochlear traveling wave near the basal end of the cochlea by

$$A=M\left(\frac{R}{1-R^-R}\right), \quad M=\left(\frac{T^+T^-}{1+R^+}\right). \tag{7}$$

The forward middle-ear transmittance and reflectance are referenced to a measurement location at the probe microphone in the ear canal. The term M in the equation on the left depends only on ear-canal and middle-ear dynamics, and its numerator is the round-trip transfer function for middle-ear pressure transmission. The term M in the equation on the right depends on cochlear mechanics; its numerator is the apical cochlear reflectance and the denominator describes round-trip internal reflections between tonotopic place and the basal end of the cochlea, which is terminated by the reverse middle-ear reflectance.

An alternative model posits that the SFOAEs are produced by nonlinear distortion within the cochlea (Brass and Kemp, 1993). While this mechanism should vanish in the

low-level regime in which the cochlear response is linear, it is possible that it may contribute to moderate or high-level SFOAEs. There is some evidence for such a two-source model of SFOAE generation in gerbil (Kemp and Brown, 1983), but no evidence was found in human responses. While the reflection emission is differentiated by its long latency from the nonlinear-distortion with its short latency (Strube, 1989), it is useful to regard each component as a complex phasor so that there can be constructive or destructive interference between the components as frequency or level are varied. Varying frequency at a fixed level allows the examination of SFOAE fine structure as a function of level, as modeled by Talmadge *et al.* (2000). Varying level at a fixed frequency allows the examination of SFOAE I/O functions as a function of frequency, which is the approach used in the present study. Although an I/O function would be recovered if various fixed levels were used in methods such as that of Shera and Zweig (1993) that reveal fine structure, the goal of the current study was to focus on characterizing the level dependence of the responses at select frequencies rather than across a large range of closely spaced frequencies. In either approach, it is possible that the reflection emission and nonlinear-distortion mechanisms would exhibit destructive interference as stimulus amplitude is varied. Such destructive interference would produce a local minimum in the I/O function and provide evidence for two sources for SFOAE generation. Any constructive interference would be difficult to differentiate from a single-source mechanism, except perhaps in terms of a stronger variation of latency with stimulus level than might be expected based on the linear coherent reflection emission theory alone.

In experiment 1, I/O functions were measured in the ears of 30 normal-hearing subjects for the condition in which $L_1 = L_2$ and $f_1 = f_2$. SFOAE I/O functions were compared with predictions from the model and with DPOAE I/O functions. The distortion of the system was measured with the probe inserted into a coupler, and in two patients with cochlear implants to evaluate the degree of separation of biological responses from nonbiological distortion. The final protocol was presented to four of the original subjects from the left-ear data collection group in order to examine overall repeatability in experiment 2. Level-dependent ''notches'' in the SFOAE I/O function were observed in some subjects in experiment 1, and the repeatability of these notches was studied in experiment 2 using stimulus levels with finer spacing. It should be noted that there are two types of measured ''noise'' described in the results. The noise calculated in the 2E method is a measure of the variability of the responses. ''On-band'' noise or variability is calculated in the bin in which the response is measured, and ''off-band noise'' is the average in the adjacent frequency bins above and below the on-band frequency. The difference between on- and off-band noise was further investigated in experiment 2. Unless otherwise specified in the text and figure captions, the noise reported is on-band noise.

TABLE I. Summary demographic information.

	Left ears		Right ears	
	Male	Female	Male	Female
N	6	9	7	8
Minimum age	18	18	20	22
Maximum age	28	35	27	35
Mean age	21.83	23.78	23.29	27.5
s.d.	3.60	4.92	2.29	5.21

III. EXPERIMENT 1

A. Subjects

1. Subjects with normal hearing

Thirty adults with normal hearing were recruited from a local university and from the clinical staff in our audiology clinic. Data were collected in 15 left ears and 15 right ears. All subjects were informed of the procedures and purpose of the study and signed a consent form prior to participation. Summary demographic data are displayed in Table I. The range of ages was restricted between 18 and 35 years in order to avoid any decrease in EOAE amplitudes that might be associated with older age groups, particularly at higher stimulus frequencies, at least for DPOAEs (Lonsbury-Martin *et al.*, 1991; Dorn *et al.*, 1998). An attempt was made to balance the numbers of males and females in right and left ear groups; however, there was a reduced representation from male volunteers due to difficulties in recruiting. Hearing was screened at audiometric test frequencies of 250 to 8000 Hz in both ears using conventional procedures and a GSI-16 or GSI-61 clinical audiometer. Subjects were included in the normal-hearing group if their bilateral thresholds were at or better than 15 dB HL at audiometric test frequencies. Middle-ear function was assessed using a GSI-37 Auto Tymp screener or a GSI-33 middle-ear analyzer and a 226-Hz probe tone. Subjects were included in the study if middle-ear functioning of the test ear was normal, as defined in this study by static pressure between -100 and $+50$ daPa, and static admittance between 0.2 and 2.5 mmhos.

Subjects were seated in a comfortable chair in a sound-attenuated booth during OAE testing. They were allowed to remain awake and quietly read, or to sleep during data collection. The time to collect all data, including the hearing screening and tympanogram, was limited to approximately 2 to 2 1/2 h. No prior training was provided for the subjects.

2. Subjects with cochlear implants

Two subjects with cochlear implants were recruited from our clinic population. Subject C01 was a 48-year-old male with cochlear implants in both ears. Both of C01's ears were tested in two test sessions. Subject C02 was a 46-year-old female who had a cochlear implant in her right ear, and wore a hearing aid on her left ear. Subject C02's implanted ear was tested in one test session. For both subjects, the cochlear implant processor was turned off and removed from the test ear during OAE recording. Subject C02 was allowed to keep amplification on the nontest ear during the test. The middle-

ear analysis and criteria were the same as for the normal-hearing group.

B. Stimulus generation and OAE recording

Stimulus generation and data acquisition were under the control of a custom software program. The DAC and analog-to-digital conversions (ADC) were accomplished with a 24-bit Card Deluxe sound card (Digital Audio Labs) using a sample rate of 22 050 Hz. An Etymotic ER-10C low noise probe microphone and receiver system was used with a modification to provide an extended 20-dB range of output to each receiver. Two channels of DAC stimuli were delivered to separate receiver ports of the probe. The probe assembly was coupled to the ear canal using Etymotic foam probe tips. The microphone of the probe assembly recorded the sound-pressure signal in the ear canal, which was high-pass filtered with a cutoff frequency of 450 Hz (Krohn-Hite model 3343, analog-signal filter). The high-pass filter was used to attenuate low-frequency responses that would otherwise degrade the artifact rejection algorithm. The filtered microphone signal was sampled by the ADC.

In this implementation of the 2E OAE method, the elementary stimulus s_1 of buffer length $N=512$ samples and duration 23.2 ms was a sine tone at frequency f_1 and level L_1. The elementary stimulus s_2 of buffer length N was a sine tone at frequency f_2 and level L_2. The total stimulus buffer of buffer length $15*N$ consisted of 5 repetitions of s_1 in DAC channel 1 and zeros in DAC channel 2, followed by 5 repetitions of s_2 in DAC channel 2 and zeros in DAC channel 1, and followed by 5 repetitions each of s_1 in DAC channel 1 and s_2 in DAC channel 2. The corresponding total response buffer consisted of five repetitions each of p_1, p_2, and p_{12}, in which p_{12} was the response to the joint presentation of s_1 and s_2. Such responses were contaminated by transients in the transitions between the sets of repetitions of p_1 and p_2, between sets of p_2 and p_{12}, and between sets of p_{12} and p_1 (the latter from the beginning of the next presentation of the stimulus set). These transients were excluded by discarding the first of the five repetitions of each elementary response. Thus, each response buffer provided four repetitions of each of p_1, p_2, and p_{12} for subsequent analyses (a pilot study that varied the number of discarded buffers is described in the Discussion section, which demonstrates that the above discard rule eliminated significant artifacts associated with transient-evoked OAEs). Successive total response buffers were recorded and temporarily stored for the artifact rejection test.

The artifact rejection test was designed to detect and reject data buffers that were contaminated by intermittent noise, and was based on comparing a previous valid response buffer to the current response buffer using a test of dissimilarity. At the beginning of data acquisition, the software acquired two response buffers before conducting the first artifact rejection test, and the initially recorded response served as the initial ''previous'' buffer. The time-domain test of dissimilarity was based on a threshold value of the maximum absolute difference in the previous and current buffers across all samples in the buffers (Keefe and Ling, 1998). If the current buffer was valid, it became the new previous valid buffer in the next artifact rejection test. After a first fail (at startup or following one or more valid tests), a new current buffer was compared to the previous valid buffer. If artifact rejection failed a second time in succession, the second invalid buffer was used as the new previous buffer for subsequent tests. This allowed the artifact rejection test to adapt to slight changes in physiological status during data acquisition by updating the contents of the previous buffer only after two successive fails.

Each valid total response buffer was successively stored in one of eight storage locations. A total of 32 valid buffers was stored, with 4 time averages per storage location. This provided 32 repetitions overall of p_1, p_2, and p_{12} at each stimulus level of the I/O functions (4 repetitions within each total response buffer in each of 8 storage locations, with each time average of 4 responses). For each of the 32 sets of repetitions, a corresponding 2E waveform p_d was calculated using Eq. (1). A N-sample DFT was calculated for p_1, p_2, p_{12}, and p_d for each repetition. The mean and noise SPLs were calculated in accordance with procedures described n the Appendix. The noise SPL was based on the standard error of the mean of the 32 responses measured in each spectral bin.

C. Stimulus conditions

DPOAEs were elicited using f_2 frequencies of approximately 2000 and 4000 Hz with an f_2/f_1 of 1.21. In subjects with normal hearing, L_2 levels for the DPOAE conditions were presented in descending 5-dB steps from 85 dB SPL (left ear in nine subjects, and right ear in 15 subjects) or 75 dB SPL (left ear in six subjects) down to 0 dB SPL. For L_2 levels of 65 dB SPL and above, L_1 was set equal to L_2. At each L_2 below 65 dB SPL, $L_1=0.4L_2+39$ dB (Kummer *et al.*, 1998). It should be noted that in the Kummer *et al.* (1998) study the lowest L_2 presented was 20 dB SPL, whereas the lowest L_2 presented in the current study was 0 dB SPL. For our purposes, the Kummer *et al.* equation was used to select L_1 levels at each L_2 down to 0 dB SPL. As the stimulus conditions were refined for SFOAE data collection as described below, more time in a test session was available for data collection at the higher L_2 levels (80 and 85 dB SPL) in the DPOAE conditions. The six subjects who were presented with a maximum L_2 of 75 dB SPL in the left ear were tested earlier in the data collection process before the higher L_2 levels were added. For the subjects with cochlear implants, L_2 levels were presented in descending 5-dB steps from 85 to 0 dB SPL, with L_1 as described above.

SFOAEs were elicited with equal-frequency primaries at octave and half-octave frequencies from 500 to 8000 Hz. Left-ear data were collected first in the group of subjects with normal hearing, and during this first half of the data collection process, various protocol changes were made as difficulties with the recording parameters were resolved. For example, poor signal-to-noise levels (SNR) were observed across the I/O function for SFOAEs recorded with primaries of 500 and 8000 Hz, and these frequencies were removed from the protocol. As mentioned in Sec. II, a fixed-L_1 condition was also used in which L_1 was fixed at 75, 70, 65, or 60 dB SPL, and L_2 was varied below the fixed L_1. This

TABLE II. Summary of DPOAE and SFOAE conditions presented.

Condition	Maximum L_2 (dB SPL)	Step size (dB SPL)	N Left	N Right
DP2000	85	5	9	15
DP2000	75	5	6	
DP4000	85	5	9	15
DP4000	75	5	6	
SF500	75	5	6	
SF750	85	5	2	15
SF750	70	10	5	
SF1000	85	5	2	15
SF1000	75	5	10	
SF1500	85	5	2	15
SF1500	70	10	5	
SF2000	85	5	5	15
SF2000	75	5	10	
SF3000	85	5	2	15
SF3000	70	10	5	
SF4000	85	5	5	15
SF4000	75	5	10	
SF6000	85	5	2	15
SF6000	70	10	5	
SF8000	75	5	10	

condition produced high signal-dependent noise levels at all L_2 levels, which resulted in poor SNRs, and it was eventually removed from the protocol on which summary results were based. Thus, the focus of the Results section is on the equal-level primary condition, and data from the fixed-level condition are presented only in Fig. 11. In the $L_1=L_2$ condition, the primaries were presented in descending 5- or 10-dB steps from an initial maximum level (70, 75, or 85 dB SPL) down to 0 dB SPL.

The final protocol consisted of the two DPOAE conditions (with f_2 frequencies of 2000 and 4000 Hz), and the SFOAE $L_1=L_2$ conditions at octave and half-octave frequencies from 750 to 6000 Hz, with levels descending in 5-dB steps from 85 to 0 dB SPL. The final protocol was presented to two of the left ears and all 15 of the right ears in the group with normal hearing, and to both of the subjects with cochlear implants.

D. Coupler measurements

Distortion and noise were measured in a Brüel & Kjær ear simulator type 4157 (also referred to as coupler) for each condition to assess measurement system distortion. Coupler measurements were recorded with the same stimulus generation and recording parameters as for biological measurements.

E. Results

1. Subjects with normal hearing

The numbers of ears (N) for each condition presented to the group with normal hearing are summarized in Table II. Some of the differences in N are due to test-time limitations before the protocol was finalized. The condition names refer to OAE type (DP or SF) and frequency, which is the f_2 frequency for DPOAEs. For example, SF6000 refers to the SFOAE condition with primaries of 6000 Hz, and DP2000 refers to a DPOAE condition with an f_2 of 2000 Hz.

a. Median DPOAE and SFOAE I/O functions. Median DPOAE and SFOAE I/O functions for the right ear are displayed in Fig. 1. The left-ear median responses were similar and are therefore not shown. It should be noted that although SFOAEs were elicited with $L_1=L_2$, responses are described as a function of L_2 in order to be consistent with the description of DPOAEs. No data are plotted at the highest L_2 of 85 dB SPL for the SF6000 condition, because the system was unable to produce this level in the coupler, possibly due to a standing-wave effect. A biological response is clearly separate from the noise level for all conditions, with the largest SNRs at moderate L_2 levels for SFOAEs and at moderate to high L_2 levels for DPOAEs. A small separation of the biological SFOAE response from the distortion in the coupler and noise in ears is identified at the highest L_2 levels (with the exception of the SF1500 and SF2000 conditions), but the SNR is much less than DPOAE SNR at these higher levels. There are two reasons for this difference. First, the distortion observed in the coupler is greater at the higher stimulus levels in the SFOAE conditions than for the higher stimulus levels in the DPOAE conditions. Second, in the SFOAE conditions, the noise levels in ears (but not the coupler) tend to increase with increases in L_2 across all frequencies, and tend to decrease with increases in frequency, particularly at low L_2 levels. A higher L_2 is required for the response to emerge from the noise for the lower (SF750) and higher (SF6000) frequency SFOAEs.

DPOAE functions are lower in level than SFOAEs, and more compressive over the moderate level stimulus range. In general, SFOAEs recorded with the equal-level protocol tend to separate from the noise floor at higher input stimulus levels, have steeper growth, and achieve higher amplitudes than DPOAEs.

b. Individual DPOAE and SFOAE I/O functions. Some observations can be made concerning the I/O functions of individual subjects that are not reflected in the quantitative summary across subjects. Responses from individual subjects R11 and R08 are shown in Figs. 2 and 3, respectively. The first observation concerning subject R11 (Fig. 2) is the substantial SNR (12 dB) for the lowest stimulus level ($L_2=0$ dB SPL) in the DP4000 condition. This may be due to the presence of a spontaneous otoacoustic emission at the f_2 or $2f_1-f_2$ frequency, although this cannot be verified because spontaneous emissions were not recorded. Among the notable differences in the individual responses, there is a clear separation between distortion and noise down to $L_2=5$ dB SPL in the SF3000 condition in Fig. 2, but the response does not separate from the noise floor until $L_2=35$ dB SPL in Fig. 3. The overall level of the SF2000 response is much lower in Fig. 2 than Fig. 3, but the reverse is true for the SF4000 condition. The noise is higher for Fig. 3 in the SF3000 and SF4000 conditions, but this does not account for the smaller response amplitudes and higher thresholds. Finally, a feature that is observed in several SFOAE I/O functions is a level-dependent notch, such as that shown in the SF4000 condition in Fig. 2 (and present to a lesser extent in the SF2000, SF3000, and SF6000 conditions). These notches do not appear at the same frequencies or intensities in all subjects, and not all subjects have notches.

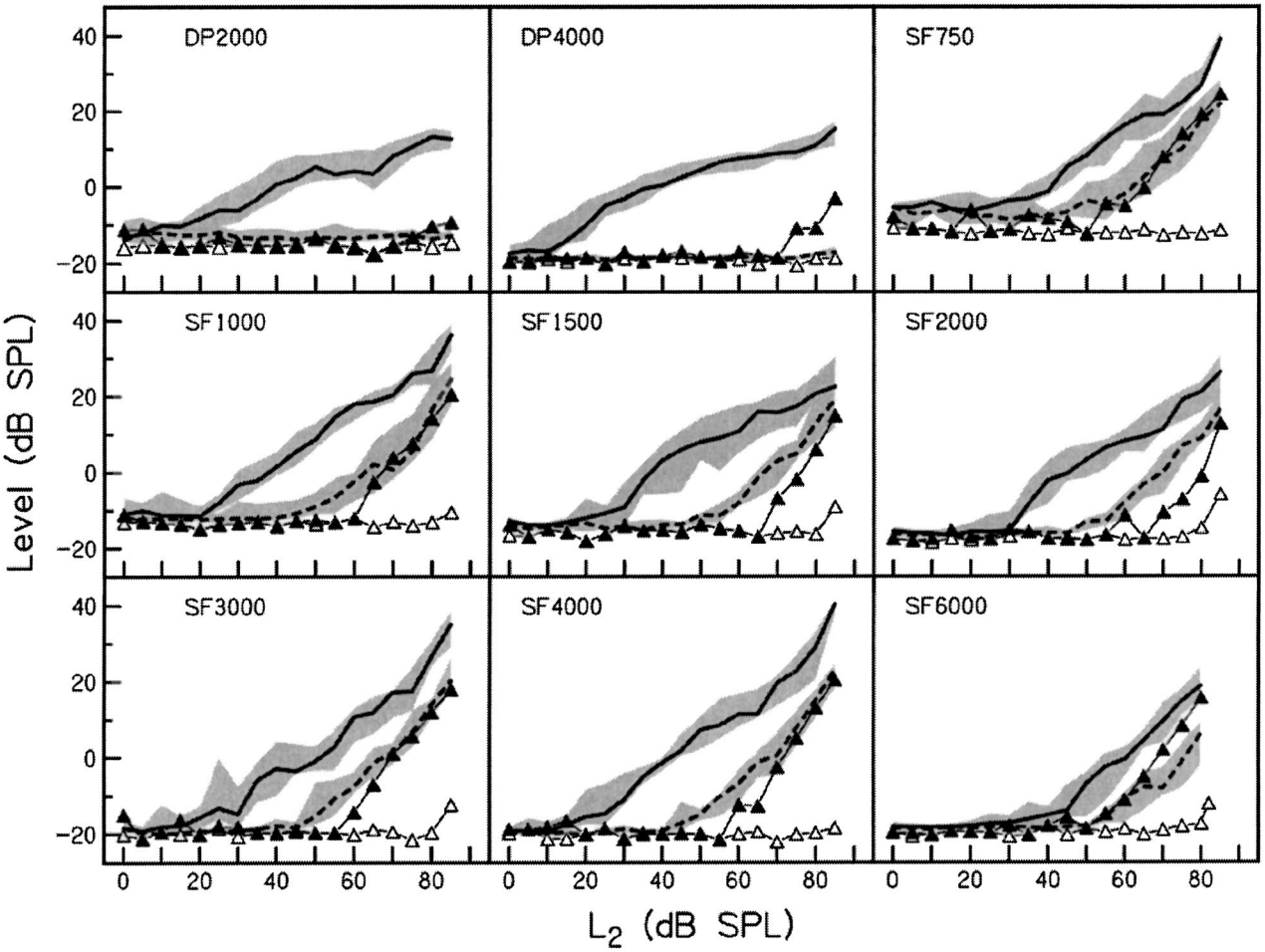


FIG. 1. SPL of median distortion (solid line) and median on-band noise (dashed line) for DPOAE and SFOAE conditions in normal-hearing right ears as a function of stimulus level L_2. The 25th to 75th percentile ranges for distortion and on-band noise SPL are represented by the shaded areas. Coupler distortion and on-band noise SPL are represented by the filled and open triangles, respectively. In comparison to DPOAE I/O functions, SFOAE I/O functions demonstrate: (A) higher thresholds (i.e., the L_2 at which the OAE separates from the noise); (B) steeper growth; (C) an increase in distortion in the coupler with increases in L_2; and (D) an increase in noise in ears (but not in the coupler) with increases in L_2.

The number of subjects who had notches in their I/O functions for each $L_1=L_2$, SFOAE condition was counted for the right-ear data set. A notch was defined as a 3-dB decrease in distortion in comparison to the preceding (i.e., lower) L_2 condition. Only decreases observed in the range including $L_2=40$ dB to 70 dB SPL were counted as notches in order to avoid confusion with the noise floor at low L_2 levels, and with rollover or a decrease in the response level at high L_2 levels. There were only two exceptions for which notches were identified at L_2 levels below 40 dB SPL, because the distortion in both cases at the notch was clearly separated from the noise floor. Four subjects (out of a total of 15 subjects) had notches that met the criteria in the SF750, SF1000, and SF3000 conditions, three subjects in the SF1500 and SF4000 conditions, two subjects in the SF2000 condition, and one subject in the SF6000 condition. The repeatability of notched I/O functions is addressed in experiment 2.

2. Subjects with cochlear implants

Measurements were made in ears with cochlear implants in order to compare with the biological origin of the responses from normal-hearing ears. Data from the ears with cochlear implants are compared in Fig. 4 with the right-ear 25th to 75th percentiles of signal and noise replotted from Fig. 1. The right-ear responses were selected for comparison because they received the same final protocol as the ears with cochlear implants. Note that the range on the ordinate is greater in Fig. 4 to accommodate the larger range of responses.

In the DPOAE conditions (DP2000 and DP4000), there is a wide range of L_2 levels over which there is a separation between the distortion from ears with cochlear implants and 25th to 75th percentile of distortion from ears with normal hearing. The only exception is for subject C01's left ear at the highest L_2 levels for the DP2000 condition. The range of separation is much smaller in the SFOAE conditions. In general, the SFOAE distortion from the ears with cochlear implants tends to fall below the 25th percentile of distortion from ears with normal hearing at moderate L_2 levels, but overlap and in some cases exceed this range at higher L_2 levels. For frequencies of 1500 Hz and above, there is an overlap of the distortion from subject C02's ear and the ears with normal hearing from L_2 levels of approximately 65 dB

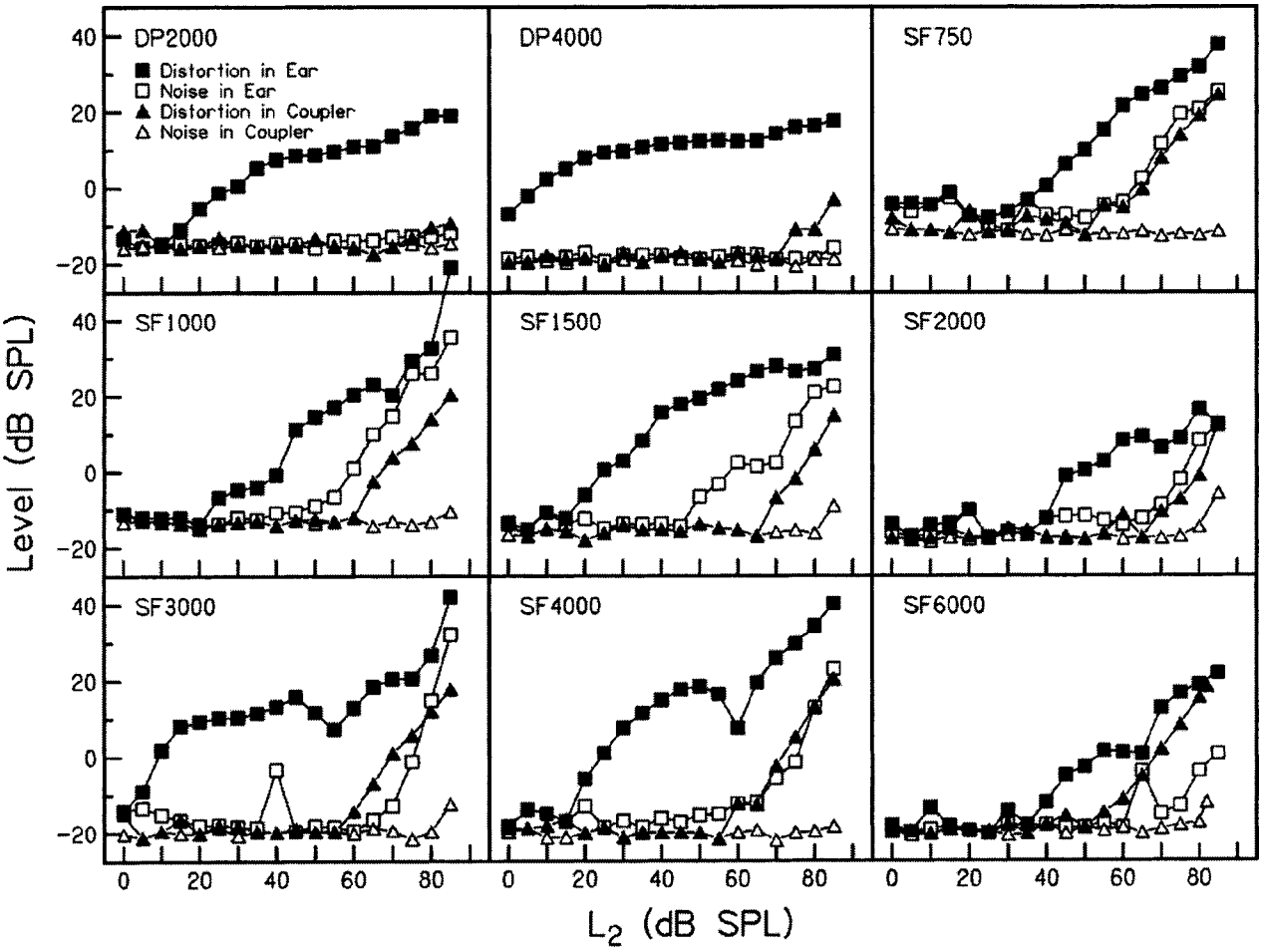


FIG. 2. Individual OAE I/O functions for normal-hearing right ear in subject R11, with distortion and on-band noise recorded in the ear (filled and open squares, respectively) compared to the distortion and on-band noise recorded in the coupler (filled and open triangles, respectively). Individual I/O functions reveal certain features that are not reflected in the median data. For example, this subject has a large SNR at the lowest L_2 in the DP4000 condition, which was rarely observed in other subjects. This could be due to a spontaneous emission at the f_2 or $2f_1-f_2$ frequency, but this cannot be verified in the current data. There also is a level-dependent notch in the SF4000 condition. Notches were often observed in individual I/O functions, but they did not appear at consistent stimulus frequencies or levels across subjects.

SPL and above. In the SF6000 condition, C02's distortion falls within the range of distortion recorded from normal ears over most of the I/O range.

The SFOAE distortion in ears with cochlear implants exceeds the coupler distortion at 1500 and 2000 Hz, and less so at 3000 Hz. This suggests a biological origin for the SFOAE distortion in ears with cochlear implants, despite the fact that there is compelling evidence of pathology in OHC functioning.

IV. EXPERIMENT 2

A. Subjects and stimulus conditions

1. Signal-dependent SFOAE noise and overall repeatability

Four subjects were asked to return and were presented with the final protocol as described in experiment 1. The only difference between the original and repeat tests was that an analog high-pass filter was used for the original test, and a real-time digital high-pass filter was used for the repeat test. The digital filter had a similar transfer function to the analog filter. Digital filtering was performed prior to artifact rejection. The delay between the recording sessions for subjects L04 and L05 was approximately 1 month. There was approximately a 2-month delay for subject L09, and a 3-month delay for subject L03. Because these subjects were among the group of subjects tested early in the process, not all of the conditions originally presented were part of the protocol that was presented during the second visit, and not all of the conditions presented during the second visit were presented during the first visit. For responses that were measured using identical stimulus conditions, Pearson product moment correlations were calculated to compare on-band distortion, or the levels measured in the frequency bin of the OAE, between the two visits. The squared correlation (r^2) was the measure of repeatability.

It was noted earlier that the on-band noise increased as L_2 increased for SFOAEs, but not for DPOAEs. In addition to the test of repeatability, on-band and off-band noise was compared to determine whether the noise increased only in the on-band bin (i.e., in the frequency bin of the OAE), or whether it also increased in the off-band bins. An increase in noise in the off-band bins would be expected if the source of the noise were environmental or physiological, but unrelated

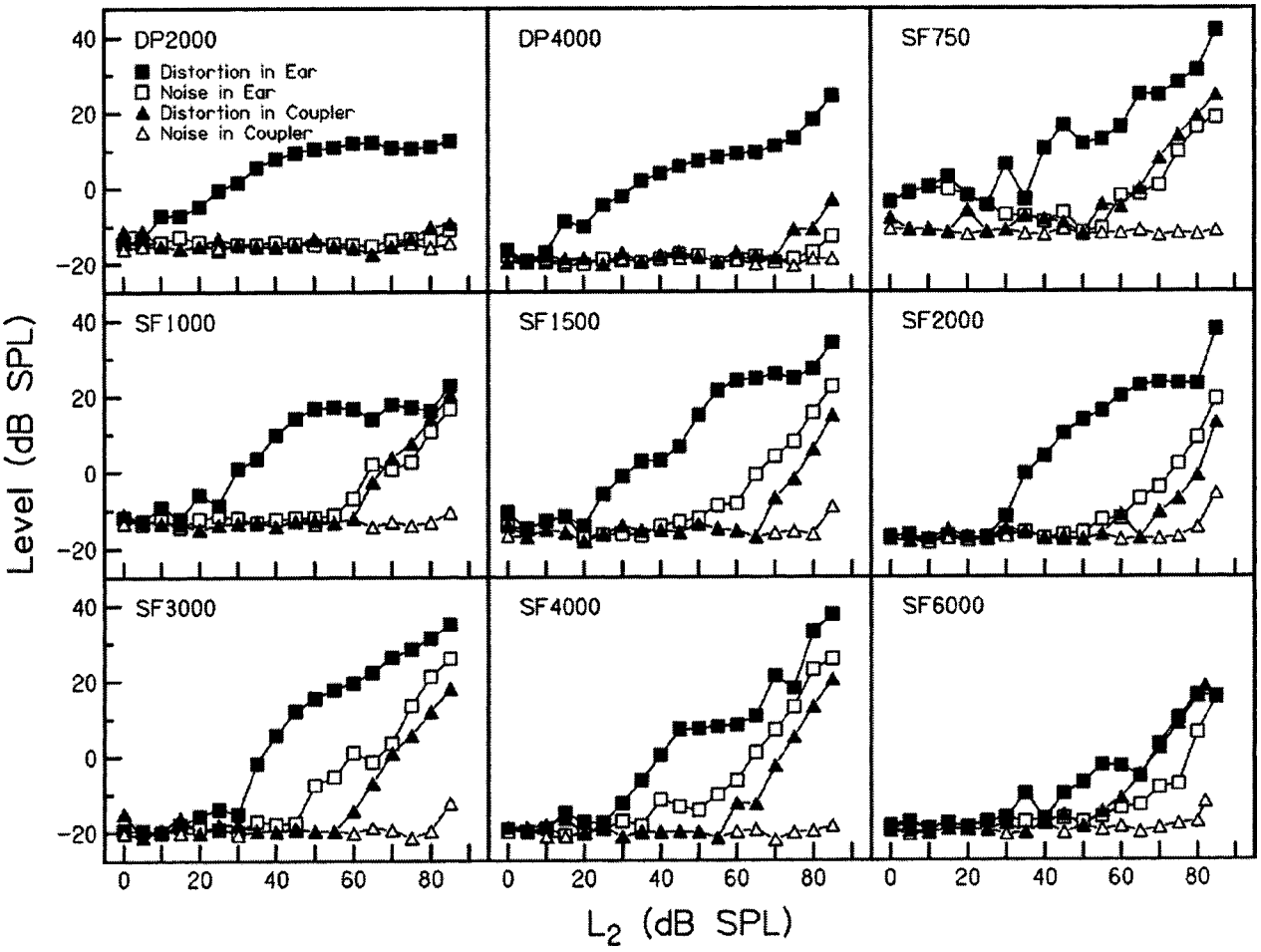


FIG. 3. Individual OAE I/O functions for normal-hearing right ear in subject R08, with data represented as in Fig. 2. A detailed comparison of this subject's responses with subject R11's responses (Fig. 2) is available in the text.

to the presentation of the sound stimuli (e.g., noise generated by the functioning of the circulatory or respiratory systems).

2. Notch repeatability

More detailed data collection around the notches observed in individual SFOAE I/O functions was completed in three of the subjects that had notches in the first recordings. The first recordings were completed using L_2 levels in 5-dB steps. For each subject in each SFOAE condition that demonstrated a notch, L_2 levels were presented in 2-dB steps over a ± 20 dB range around the level at which the notch occurred in the first recordings.

B. Results

1. Signal-dependent SFOAE noise

The variation in noise at and around the signal frequency is evident in the SFOAE spectrum recorded in one ear at 4000 Hz and at equal stimulus levels of 70 dB SPL (Fig. 5). Over a broad frequency range (left panel), the noise SPL based on the standard error of the mean (see the Appendix) is smoothly varying through the approximate midline of the off-band distortion SPL. This off-band distortion has larger fluctuations across frequency than does the off-band noise. The on-band noise (at 4000 Hz) is elevated above the off-band noise, in this case by 13.8 dB. Given that the SFOAE SNR is conventionally defined as the level difference between signal and noise in the same frequency bin, it is clear that this signal-dependent elevation of noise reduces the apparent SNR by 13.8 dB. A definition of noise based on an average of off-frequency bins is a poor estimate of on-band noise associated with SFOAEs.

2. Overall repeatability

Table III lists the r^2 values for the comparison of on-band distortion between the two visits for each subject. The lowest r^2 value is for L03 in the DP2000 condition ($r^2 = 0.75$). The remaining 22 r^2 values are 0.88 or above, with an averaged squared correlation of 0.95. Thus, the SFOAE and DPOAE I/O functions were highly repeatable.

Figure 6 shows a summary of responses from the two visits for L_2 levels of 75 to 5 dB SPL in 10-dB steps. Measures from the first and second visits are represented by filled and open symbols, respectively. On-band distortion (circles), i.e., the OAE signal level, shows good repeatability for the SF and DP conditions. On-band noise (squares) shows good repeatability, with the exception of a sharp increase in noise at $L_2 = 50$ dB SPL for run 2 for subject L05 in the SF4000 condition, and small differences across the moderate to high L_2 range in the SF4000 condition for subjects L03 and L09. Off-band noise (triangles) was highly repeatable and re-

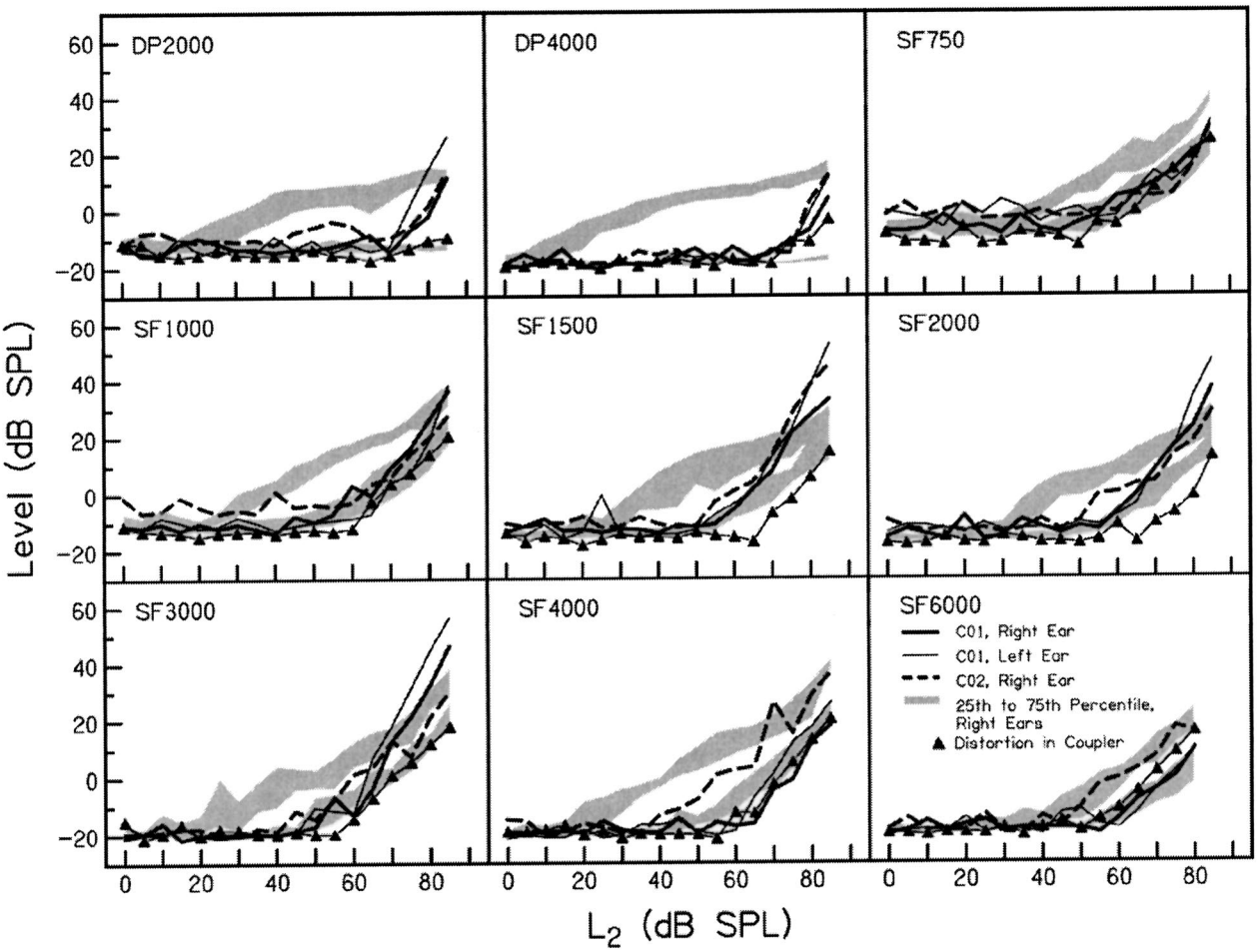


FIG. 4. Individual distortion I/O functions in ears of two subjects with cochlear implants (thick solid, thin solid, and dashed lines) compared to 25th to 75th percentile distortion and on-band noise recorded from the normal-hearing right ear data (shaded areas copied from Fig. 1) and distortion recorded in the coupler (filled triangles). The thin solid line represents distortion from subject C01's left ear, the thick solid line from C01's right ear, and the dashed line from C02's right ear.

mained nearly constant and low in level throughout the I/O function for all SF and DP conditions. This is in contrast to the on-band noise, which increased with increases in L_2 for the SF conditions, but remained constant for the DP conditions.

3. Notch repeatability

Examples of notched I/O functions measured in subject R09 are displayed in Fig. 7. A selected region around the notch area is shown for the original 5-dB step data (squares)

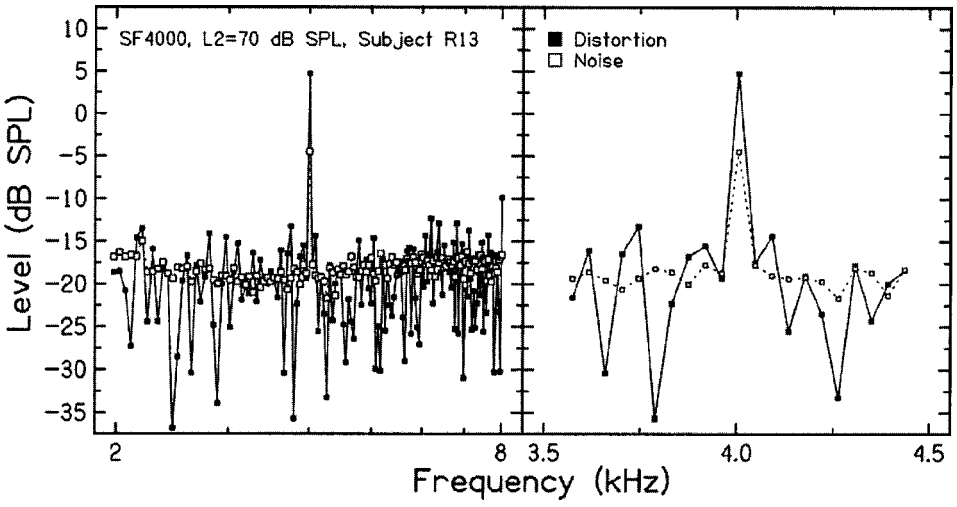


FIG. 5. SFOAE and noise spectra for SF4000 condition at equal primary levels of 70 dB SPL over broad frequency range (left panel) and narrow frequency range (right panel) for subject R13. This figure demonstrates that the on-band noise (i.e., at the frequency at which the response is recorded) is elevated above the off-band noise (i.e., adjacent to the frequency at which the response is recorded). This occurs in SFOAEs, but not in DPOAEs or in responses recorded in the coupler.

TABLE III. r^2 values for repeatability of on-band distortion (OAE) in four subjects.

	Subject			
Condition	L03	L04	L05	L09
DP2000	0.75	0.96	0.88	0.88
DP4000	0.97	0.99	0.94	0.99
SF750				0.99
SF1000			0.97	0.95
SF1500				0.96
SF2000	0.87	0.99	0.89	0.99
SF3000				0.99
SF4000	0.98	0.96	0.98	0.98
SF6000				0.98
SF8000			0.91	

and for the 2-dB step data (triangles) obtained during the same visit. The 2-dB step recordings demonstrate the repeatability of the notch structure, both in the L_2 value at which the notch occurs and the depth of the notch. As expected, the notch is better defined in the smaller-step recording. For the SF1000 condition, the minimum of the notch in the 2-dB step recordings appears to be shifted slightly upwards on the L_2 axis by a smaller amount than the step size in the original data set. For the SF750 condition, the noise level is much higher across the I/O function in the first recording; however, the distortion levels between the two recordings demonstrate good repeatability. Based on the phase of P_d for its waveform defined in Eq. (1), there also is a phase shift in the L_2 region of the level-dependent notch, and the amount of phase

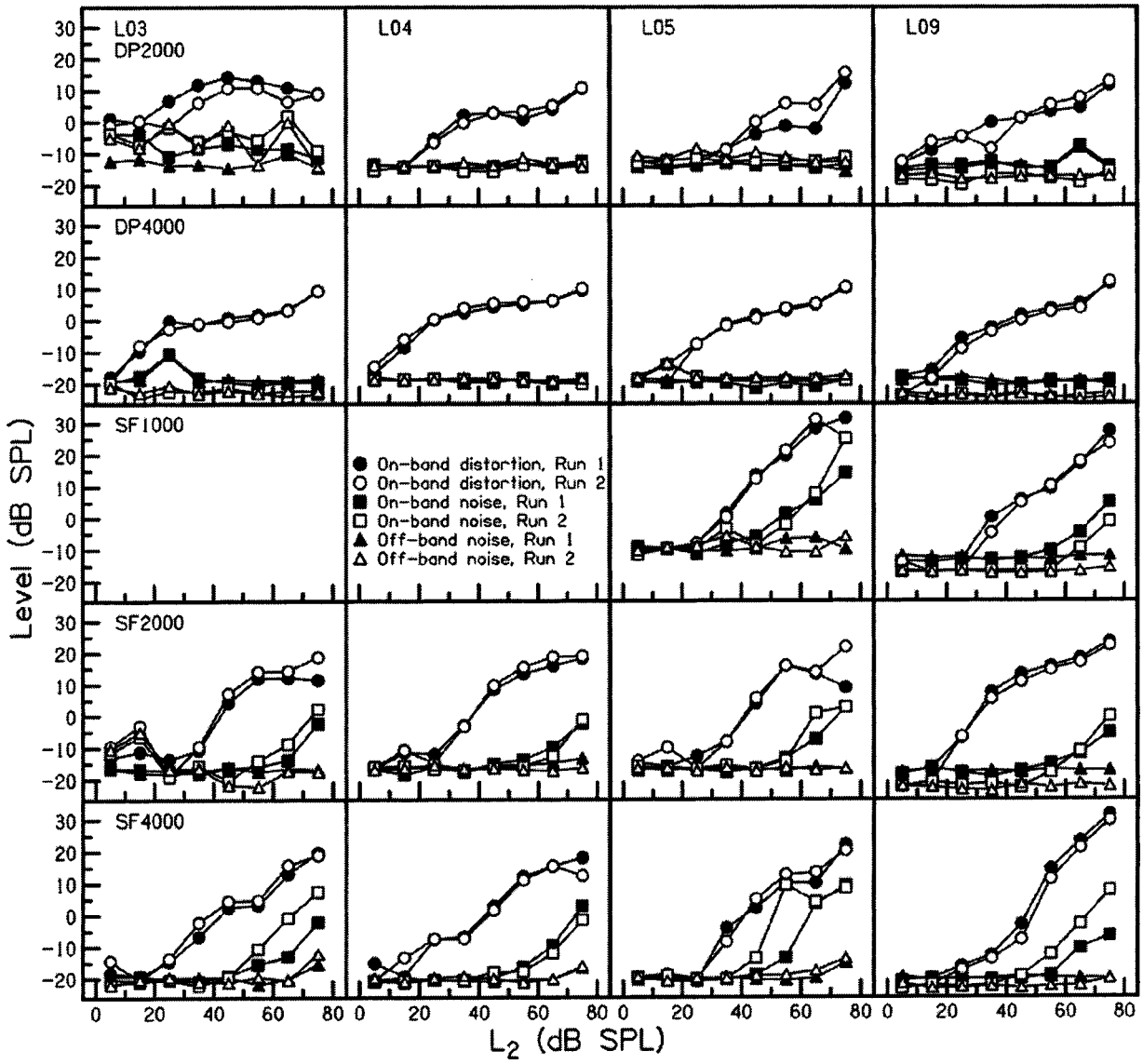


FIG. 6. Repeatability of OAE and noise I/O functions recorded in normal-hearing left ears of subjects from visit 1 (filled symbols) and visit 2 (open symbols). There was at least 1 month between visits. Plotting symbols are: on-band distortion (circles), on-band noise (squares), off-band noise (triangles). On-band noise is the noise SPL at the OAE frequency and off-band noise is the average noise SPL at the two adjacent frequencies above and below the OAE frequency. Responses shown in rows 1–5 are for DP2000, DP4000, SF1000, SF2000, and SF4000 conditions. Responses shown in columns 1–4 are for subjects L03, L04, L05, and L09, respectively.

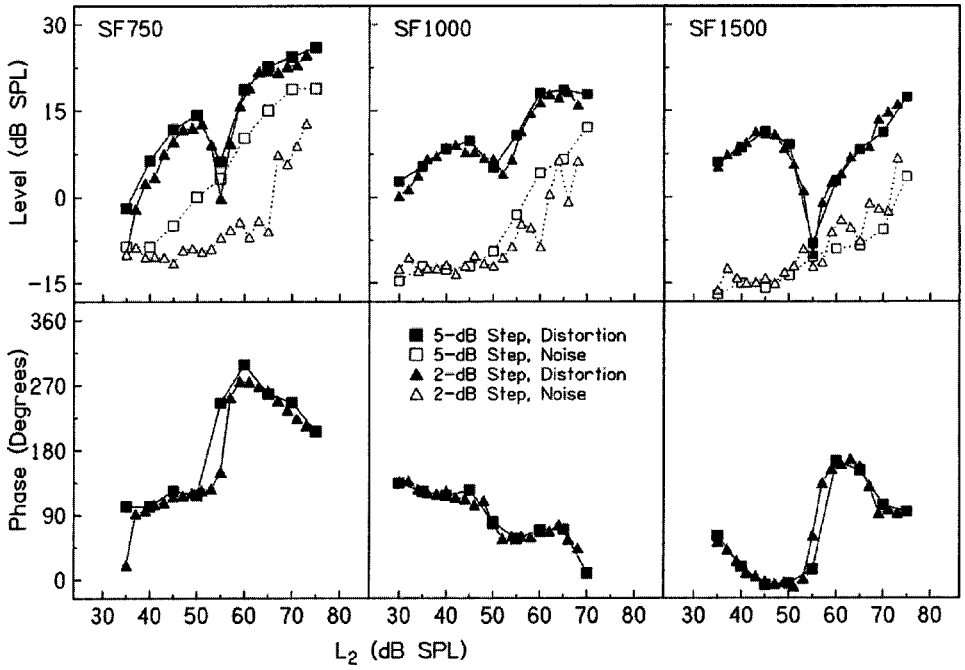


FIG. 7. Repeatability of notched SFOAE I/O functions for SF750, SF1000, and SF1500 conditions in the normal-hearing right ear of subject R09. Responses were obtained during the same visit with L_2's in 5-dB steps (squares), as part of the general protocol, and in 2-dB steps (triangles) around the notch. Level is shown across the three conditions in the top row (filled symbols=distortion, open symbols=on-band noise), and phase is shown in the bottom row. There is a phase shift in the L_2 region of the level-dependent notch, and the amount of phase shift appears to be related to the depth of the notch. For example, there is approximately a 180° phase shift in the SF750 and SF1500 conditions in which there is a deep notch, but less of a phase shift in the SF1000 condition in which the notch is not as deep.

shift appears to be related to the depth of the notch. For example, there is approximately a 180° phase shift in the SF750 and SF1500 conditions in which there is a deep notch, but less of a phase shift in the SF1000 condition in which the notch is not as deep.

V. DISCUSSION

A. General form and comparison of I/O functions in ears with normal hearing

Despite some methodological differences, the general form of the I/O functions for normal-hearing ears was similar to functions reported in the literature for SFOAEs with equal-frequency, equal-level primaries (e.g., Brass and Kemp, 1991, 1993) and DPOAEs using optimized primary level presentation with f_2 frequencies of 2000 and 4000 Hz (e.g., Dorn *et al.*, 2001). Brass and Kemp (1993) reported one of their subjects required 6 months of practice for monitoring probe placement and for maintaining an appropriate state during SFOAE data collection in order to produce smooth functions. However, the responses reported here showed good repeatability, despite the fact that the subjects did not receive instruction or practice in monitoring probe placement.

In the current data, the SF500 and SF750 conditions demonstrated the poorest SNR throughout the I/O range, and responses in the SF8000 condition were contaminated by excessive system distortion as measured in the test cavity through the mid- to high-L_2 range. The SF500 and SF8000 conditions were poor enough that they were excluded from further analyses. For the remaining SFOAE compared to DPOAE I/O functions, the equal-level, equal-frequency primary condition produced SFOAEs that had higher absolute SPLs but smaller SNR at moderate to high L_2 levels (Fig. 8). In the left ear, with the exception of the SF6000 condition, the SFOAE SPL exceeded the DPOAE SPL for stimulus levels of 50 dB SPL and higher (top left panel). However, the corresponding SNRs (bottom left panel) decrease sharply at higher stimulus levels for SFOAEs, while SNRs for DPOAEs remain high. This trend was similar for the right-ear data. The smaller SNR for SFOAEs at high levels was due to increased noise in the ears and system distortion at higher levels in excess of that observed for DPOAEs. SFOAE I/O functions recorded using the equal-level, equal-frequency paradigm were steeper or more nearly linear than DPOAE I/O functions.

For clinically oriented SFOAE recordings intended to assess whether a response lies within a baseline range of responses in normal-functioning ears, the equal-level protocol at stimulus levels near 50 dB SPL may be of particular interest. This is because the SNR in a normal-functioning ear tends to be close to its maximum value at 50 dB SPL across all test frequencies.

The SFOAE and DPOAE I/O functions were fitted using a two-slope model in order to compare compression slopes, the breakpoint pressure at which compression became significant, and overall amplification. The data used in the parameter estimation were the right-ear median data in Fig. 1; however, only responses up to levels of 75 dB SPL were included. This was because some of the SFOAE data at higher levels were influenced by system distortion. The 15 responses comprising each OAE I/O function were specified at stimulus levels in the range 0–75 dB SPL at 5-dB increments. The model related the input pressure spectrum Q and output SFOAE pressure spectrum P_{oae} at each frequency using the form

$$|P_{\text{oae}}(Q)|^2=\left|\frac{A_0 Q}{[1+|Q/Q_c|^{(1-c)/c}]^c}\right|^2+|N_{\text{oae}}|^2. \tag{8}$$

The average measured noise energy averaged over low levels up to 35 dB SPL was $|N_{\text{oae}}|^2$. The three model parameters included the *attenuation* A_0, the *breakpoint pressure* Q_c, and the *compressive slope* c. Each parameter was real. At low stimulus levels ($|Q|\ll|Q_c|$) in the absence of noise, the

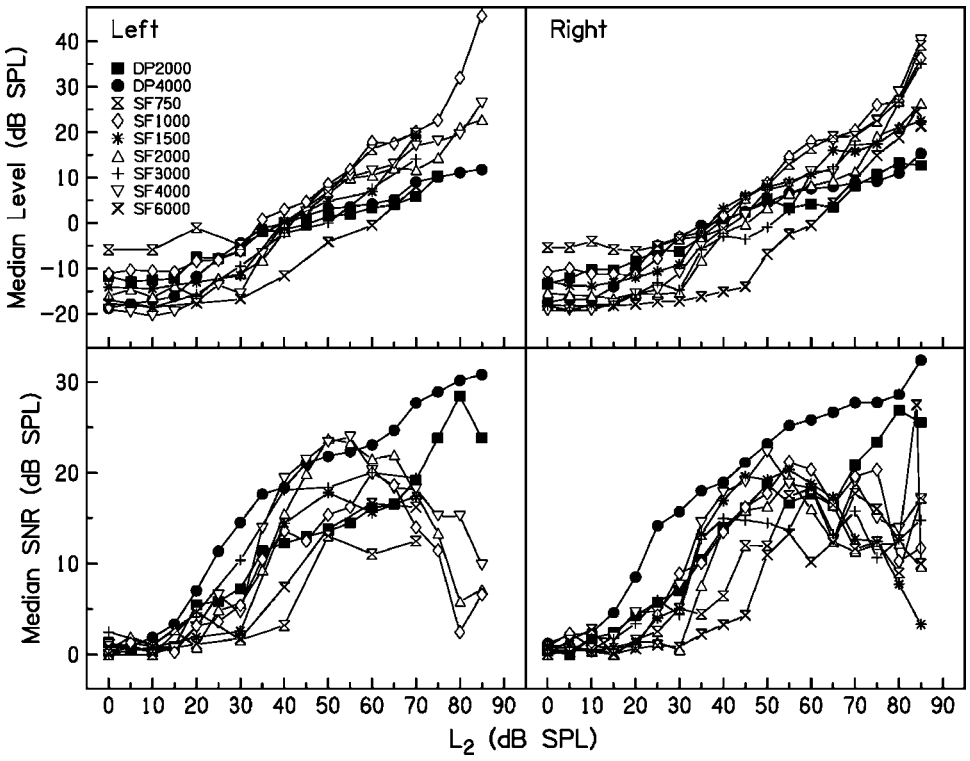


FIG. 8. Median distortion level (top row) and SNR (bottom row) versus L_2 from left ears (left column) and right ears (right column) with OAE condition as the parameter. DPOAE conditions are identified by f_2 frequency and labeled by filled squares and circles, and SFOAE conditions are identified by stimulus frequency and labeled by nonfilled symbols as indicated in the legend. The right-ear responses are the same medians shown in Fig. 1. The N for each condition is listed in Table II for both ears. In general, the equal-level, equal-frequency primary condition produced SFOAEs that had higher absolute SPLs, but smaller SNRs at moderate to high L_2 levels than DPOAEs.

model has linear growth, $|P_{\text{oae}}(Q)| \simeq |A_0 Q|$. This is of the same form as the magnitude of Eq. (6) for the low-level prediction of the coherent reflection emission theory of SFOAEs. Thus, the resulting values of attenuation may be used to estimate the magnitude of the attenuation in Eq. (7). At high stimulus levels ($|Q| \gg |Q_c|$) in the absence of noise, the model has a compressive slope, $|P_{\text{oae}}(Q)| \simeq |(A_0 Q_c^{1-c}) Q^c|$ for $0 \leqslant c < 1$. The compression function in the denominator of the model has been used to fit BM I/O functions (Yates *et al.*, 1990). The breakpoint pressure $|Q| = |Q_c|$ is that pressure amplitude for which the response level (in the absence of noise) is reduced by $6c$ dB from the predicted linear-growth response. As discussed earlier, the SFOAE responses measured using the equal-level protocol eliminate the linear-growth component of the response, so that they are nonzero only at stimulus levels comparable to the level of the breakpoint pressure and above. Thus, the breakpoint pressure is operationally well defined in the measured data.

A somewhat similar model using a low-level linear slope and moderate-level compressive slope has been previously used to fit DPOAE I/O functions (Dorn *et al.*, 2001). However, their compressive slope was in a function that differed slightly from that of Yates *et al.* (1990), on which the present model is based. Dorn *et al.* represented the noise level as a parameter, while it is a measured input value in the present model. In order to fit DPOAE responses at f_2 frequencies of 2000 and 4000 Hz up to $L_2 = 95$ dB SPL, Dorn *et al.* included a high-level steeply sloping component that is omitted in the present analyses, which analyzed data only up to $L_2 = 75$ dB SPL.

The breakpoint pressure plays a different role in interpreting DPOAE and SFOAE responses, because the DPOAE is measured directly as the response to L_{12} while the SFOAE is measured indirectly using Eq. (5). The latter procedure used the measured SFOAE I/O function $|P_d(Q)|$ obtained using the equal-level protocol, and the above model for the total SFOAE magnitude $|P_{\text{oae}}(Q)|$. This model form of $|P_{\text{oae}}(Q)|$ was substituted in Eq. (5) to calculate a predicted $|P_d(Q)|$. The parameters were calculated using a least-squares predictor function `lsqcurvefit` from the MATLAB Optimization Toolkit.

The SFOAE and DPOAE I/O function data and model predictions are shown in Fig. 9 and the estimated model parameters are listed in Table IV. The SFOAE solid lines in Fig. 9 are the predictions of the 2E measurements in noise using the equal-level protocol. The SFOAE model parameters are those for the total SFOAE amplitude $|P_{\text{oae}}(Q)|$, and this predicted function (as SPL) is plotted in dashed lines on the SFOAE panels of Fig. 9. This is the I/O function of the underlying total SFOAE produced by a single sine tone.

The model predictions (solid lines) have adequate agreement with the measured data (symbols). The DPOAE I/O functions are more compressive than the SFOAE I/O functions, and their slopes (approximately 0.3 dB/dB in Table IV) are similar to those reported by Dorn *et al.* (2001). The attenuation level is approximately -30 dB for each of the DPOAE responses, while the DPOAE at 4000 Hz has a breakpoint pressure level that is 5.4 dB higher than that at 2000 Hz.

For SFOAEs, the breakpoint pressure sets the threshold of the onset of nonlinearity in the I/O function, and the breakpoint pressure level is approximately 19 dB higher for the SFOAE at 2000 Hz compared to SFOAEs at 1000 and 4000 Hz (Table IV), i.e., there is an SFOAE minimum at 2000 Hz relative to 1000 and 4000 Hz. The tabulated breakpoint-pressure levels are consistent with the plotted

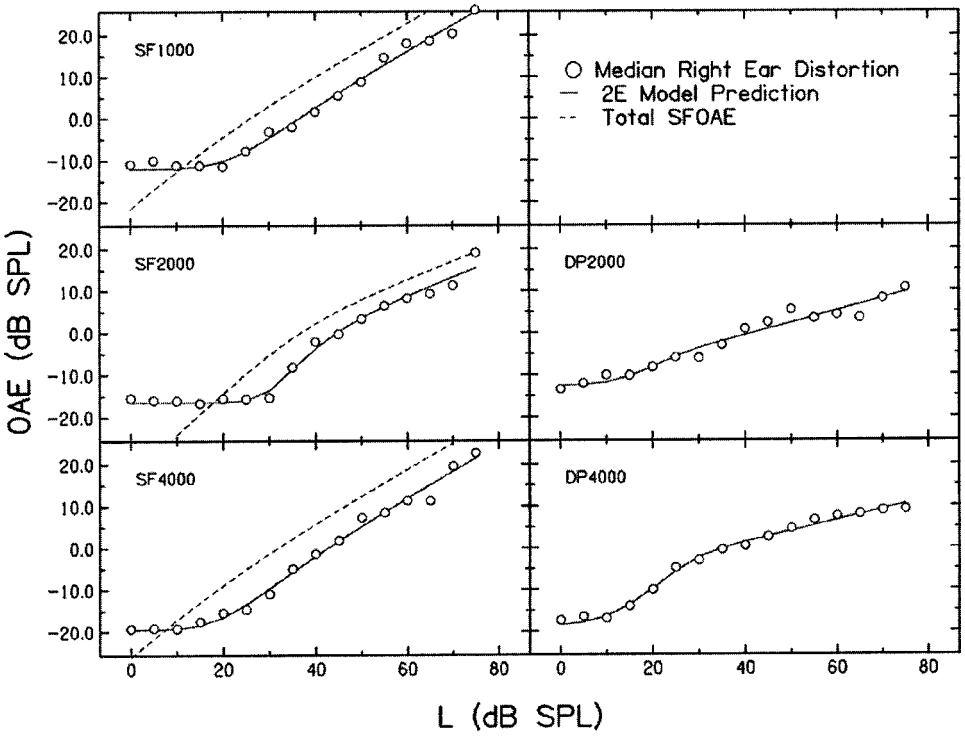


FIG. 9. The SFOAE I/O function measurements and predictions are shown in the left panel as a function of the equal-level parameter L, and the DPOAE I/O function measurements and predictions are shown in the right panel as a function of L_2. The measurements are shown as symbols and the prediction as solid line. The dashed line in each SFOAE panel is the predicted level based on the total SFOAE $P_{oae}(Q)$, which includes the low-level, linear-growth component. Each SFOAE frequency and DPOAE f_2 frequency is displayed in text within its panel.

thresholds of the measured SFOAE I/O functions with respect to their noise levels.

The slopes of the SFOAE I/O function are compressive in the range of 0.44–0.61 dB/dB. These exceed the compressive slopes of BM I/O functions that are in the range 0.20–0.25 (Yates *et al.*, 1990). The BM nonlinearity represents forward propagation only from the ear canal to tonotopic place, whereas the SFOAE nonlinearity from the coherent emission viewpoint represents both forward and reverse propagation between ear canal and tonotopic place, as mediated by a spatial convolution of the traveling wave with the spatial wave numbers associated with distributed inhomogeneities. One possibility is that the round-trip propagation is associated with the approximate doubling of the slope of the SFOAE I/O function relative to that of the BM.

A minimum across frequency at 2000 Hz is also evident in the SFOAE attenuation level (Table IV), which has a depth of 10 to 14 dB compared to neighboring frequencies. The SFOAE attenuation levels may be interpreted in terms of the coherent reflection emission model, which predicts the values in Eq. (7). It is likely that median averaging tends to remove the influence of multiple internal reflections in this prediction. Moreover, the term $1+R^+$ in the denominator of Eq. (7) would not act to produce a minimum at 2000 Hz: it tends towards a minimum in the vicinity of a standing-wave frequency in the ear canal close to 6000 Hz. The remaining part of the predicted SFOAE attenuation, $A \sim T^+T^-R$, is relevant to explaining the minimum at 2000 Hz. This is the product of the forward and reverse middle-ear transmittance, and the cochlear reflectance. In particular, the round-trip middle-ear transfer function may control this minimum.

TABLE IV. Summary of model parameters estimated from measured SFOAE and DPOAE I/O functions. (The frequency of the DPOAE responses is the f_2 frequency.)

Type	Frequency (Hz)	Attenuation level (dB)	Breakpoint pressure (dB SPL)	Compressive slope (dB/dB)
SFOAE	1000	−20.3	17.0	0.61
SFOAE	2000	−34.1	36.8	0.44
SFOAE	4000	−24.8	18.1	0.61
DPOAE	2000	−29.4	23.3	0.30
DPOAE	4000	−30.3	28.7	0.26

DPOAE I/O functions have been used to estimate the spectral shapes of forward and reverse middle-ear transfer functions, and the estimated forward and reverse transfer functions have a minimum at 2000 Hz. Based on Fig. 8 of Keefe (2002), the sums of the relative levels of forward and reverse transfer functions are approximately 0, −15, and 4 dB at 1000, 2000, and 4000 Hz, respectively. Normalizing the SFOAE attenuation levels at 2000 and 4000 Hz to that at 1000 Hz in the data of Table IV, the relative attenuation levels are 0, −13.8, and −4.5 dB at 1000, 2000, and 4000 Hz, respectively. The round-trip middle-ear transfer is likely responsible for this minimum at 2000 Hz. The relative agreement between these predictions of relative attenuation is notable because they are based on unrelated predictive models and unrelated measurements (SFOAEs versus DPOAEs). The SFOAE model provides an absolute prediction in terms of the coherent reflection emission model, whereas the DPOAE model only predicts the relative levels; the SFOAE model makes no assumption concerning the existence of a scaling symmetry in the cochlea, whereas the DPOAE model is based on this assumption. However, the minimum in the attenuation level does not explain the more compressive slope of the SFOAE at 2000 Hz compared to the SFOAE I/O function slopes at 1000 and 4000 Hz.

One property of the present model is linear growth of the SFOAE at low levels, in agreement with the findings of

Zwicker and Schloth (1984) and of Shera and Zweig (1993). This linear growth is also in agreement with the present data, in which the absence of a SFOAE residual at low stimulus levels provided evidence for linear growth. This is the case in the normal-ear responses in Fig. 1 up to stimulus SPLs equal to the breakpoint-pressure levels listed in Table IV. The model in Eq. (8) allows estimation of SFOAE properties in the linear-growth regime and thus a direct measurement of attenuation level, which is the only relevant coefficient at low levels. It would be useful in future research to obtain individual measurements, rather than population estimates, of the model parameters of SFOAE I/O functions, which would generalize the technique to individual ears. It also may be possible to include predictions of the phase of A_0 with which to predict the phase of A. This model may be testable in ears with mild conductive hearing loss that have detectable levels of SFOAEs. A limitation of the current SFOAE model is that it does not describe notched I/O functions that were observed in some subjects. These notched I/O functions are further described below.

B. I/O functions in ears with cochlear implants

The measurements from ears with cochlear implants (Fig. 4) showed unexpectedly large distortion at stimulus levels at and above 60 dB SPL, in excess of the distortion observed in the coupler; in some cases, this distortion was in the range of SFOAE responses in ears with normal hearing. Dorn *et al.* (2001) demonstrated clear separation of DPOAE responses from ears with cochlear implants from ears with normal hearing across the I/O range for f_2 frequencies of 2000 and 4000 Hz (see their Fig. 2). However, they noted that the distortion measured from ears with cochlear implants exceeded their predicted system distortion at $L_1 = L_2 = 85$ dB SPL for an f_2 of 4000 Hz. They suggested that this might have been due to system distortion, or to a biological source. Responses in the present study from ears with cochlear implants began to overlap with the range of responses from ears with normal hearing at $L_2 = 85$ dB SPL, which is consistent with the findings of Dorn *et al.* However, one outlier case is CO1's left ear in the DP2000 condition, which overlapped the range of responses from normal-hearing ears at $L_2 = 75$ dB SPL, and exceeded the range above that L_2.

There are no data in the literature with which to compare the current measurements in ears with cochlear implants for the SFOAE conditions. For SFOAE conditions in the current study, distortion at high L_2 levels occurred in at least one ear with a cochlear implant at all frequencies but 750 Hz. This conclusion is based on whether the distortion in each ear with a cochlear implant exceeded the 25th percentile of the responses in normal-hearing subjects. For the SF6000 condition, measurements from C02's right ear were within the range of responses from normal-hearing ears throughout the I/O function. There are insufficient data to explain why the distortion from ears with cochlear implants exceeded the coupler measurements, or why it exceeded the responses from ears with normal hearing in some conditions. Possibilities include an unknown source of system distortion that has larger effects in ears than in the coupler, or a biological origin to the SFOAE responses in ears with cochlear implants.

Such an origin would not involve the OHC functioning that underlies the generation of SFOAEs at low and moderate levels in a normal-functioning cochlea, because ears with cochlear implants clearly lack normal OHC functioning. However, a reflected wave would be generated at high stimulus levels in an ear without normal OHC functioning, as long as there is some inhomogeneity in the stiffness-gradient properties of the BM associated with the cochlear map. Such an ear would lack the compressive slope in the SFOAEs of ears with normal-functioning OHCs. SFOAE I/O functions for the cochlear-implant ears have a less compressive slope than those for normal-hearing ears (compare Fig. 4 to Fig. 1). Latency measurements would be helpful to sort out whether these SFOAE and high-level DPOAE responses in ears with cochlear implants are produced by middle-ear or cochlear nonlinearity, because a middle-ear or system nonlinearity would have a short latency compared to the normal round-trip cochlear delay. More data are also needed on SFOAE I/O functions in ears with sensorineural hearing loss, but without cochlear implants, in order to better understand SFOAE source mechanisms in ears with cochlear pathology. The existing sparse data from such ears have been difficult to interpret (Konrad-Martin *et al.,* 2002).

C. Repeatability

The r^2 values indicated good repeatability of the on-band distortion for two runs of select conditions for the four returning normal-hearing subjects (Fig. 6). There was also repeatability of the pattern of the elevated on-band noise above the off-band noise in the SFOAEs, a pattern that is absent in DPOAEs. However, the SFOAE distortion I/O functions had better repeatability than the SFOAE on-band noise I/O functions, which suggests that whatever mechanism(s) is producing the elevated on-band noise may be somewhat decoupled from the mechanisms producing the SFOAE distortion. In summary, DPOAE and SFOAE distortion are robust cochlear responses, and the overall patterns of on-band and off-band noise levels differ between DPOAEs and SFOAEs in a repeatable fashion.

D. Signal-dependent noise in SFOAE generation

The on-band noise SPL increased with increasing L_2 (above approximately 60 dB SPL) for SFOAEs, but not for DPOAEs. These patterns were generally observed in the median responses (Fig. 1) as well as in the individual responses (Figs. 2–3). The relative elevation of on-band to off-band noise SPL in the median data was 0 dB for the two DPOAE conditions, while the maximum elevation was approximately 30 dB across SFOAE conditions. A key result in interpreting this effect is that the corresponding on-band noise in the coupler (see Fig. 2 or the same data plotted in Fig. 3) was essentially constant with increasing L_2 up to $L_2 = 80$ dB SPL. At the highest L_2 tested (85 dB SPL), the on-band coupler noise increased across various frequencies by 0–10 dB, but it was still lower than the on-band noise observed in the normal-hearing ears. Thus, the increased signal-dependent SFOAE noise has a biological origin and is a large effect.

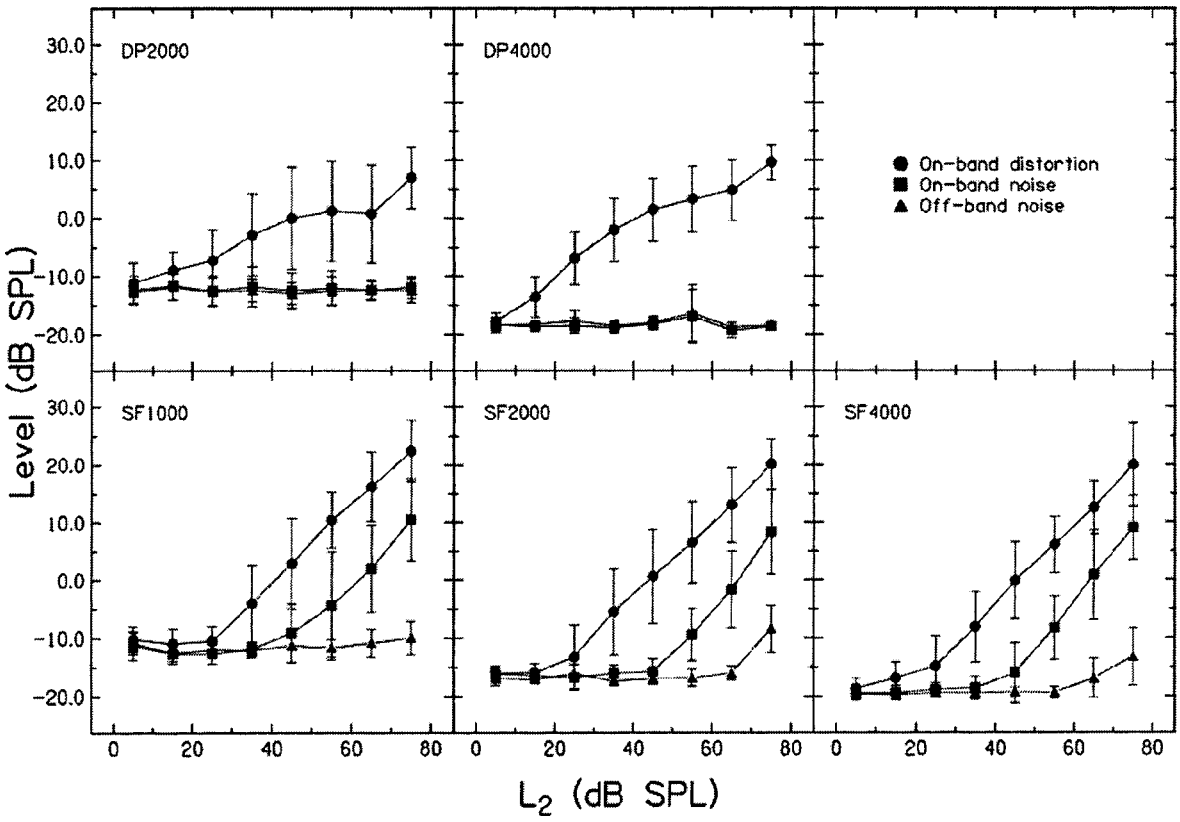


FIG. 10. Means and standard deviations for distortion (circles), on-band noise (squares), and off-band noise (triangles) as a function of L_2 for DPOAEs (top row) and select SFOAEs (bottom row) for a population of nine normal-hearing right ears. On-band and off-band noise levels were essentially constant and equal over all L_2 levels in the DPOAE conditions. Similarly, the SFOAE off-band noise level was nearly constant with L_2 except for a slight growth of approximately 7 dB in SF2000 and SF4000 at the highest L_2 levels. However, the on-band SFOAE noise SPL increased approximately linearly with L_2 for L_2 exceeding 50 dB SPL.

DPOAE and SFOAE results in Fig. 10 show the inter-subject means and standard deviations of the distortion, on-band noise, and off-band noise as functions of L_2 from 5 to 75 dB SPL in 10-dB increments for a subset of nine of 15 normal-hearing subjects with right-ear tests. This subset of subjects was chosen based on limitations of the underlying database; the remaining six right-ear responses were similar to those illustrated. Data at L_2 levels at 80 and 85 dB SPL were excluded because of some uncertainty regarding the presence of variability in system distortion. These mean OAE responses were similar to the median OAE responses in Fig. 1. The DP2000 and DP4000 conditions show that on-band and off-band noise levels were essentially constant and equal over all L_2 levels. The SFOAE off-band-noise level was nearly constant with L_2 for SF1000, SF2000, and SF4000 except for a slight growth of approximately 7 dB in SF2000 and SF4000 at the highest L_2 levels. The on-band noise SPL increased approximately linearly with L_2 for L_2 exceeding 50 dB SPL.

As described in Sec. III C, SFOAEs were measured by a second protocol in which L_1 was fixed at a high level and L_2 was varied from low levels up to 5 dB below L_1. As described in Sec. II, this fixed-L_1 protocol was anticipated to be useful in measuring the total SFOAE I/O function. The results were unexpected insofar as recordings in many subjects were contaminated by high levels of on-band noise, which precluded the ability to measure a detectable SFOAE. A comparison recorded in two subjects is shown in Fig. 11 for SFOAEs at 2000 Hz recorded with the equal-level and the fixed-L_1 protocols. The left panel shows fixed-L_1 responses recorded with $L_1 = 65$ dB SPL, while the right panel shows analogous responses recorded with $L_1 = 70$ dB SPL. The equal-level SFOAE responses for a given subject are identical on left and right panels to facilitate comparison. The expected pattern of results was obtained for the ear of subject L03 (bottom panels), in which the SFOAE for the equal-level protocol had a higher threshold relative to the on-band noise than did the SFOAE for the fixed-L_1 protocol. These results are consistent with the measurement theory described in Sec. II in that the fixed-L_1 protocol allows the measurement of a lower level SFOAE response. This is because the on-band noise is similarly low for both protocols. As expected from the measurement theory, the fixed-L_1 SFOAE I/O function for subject L03 (bottom left panel of Fig. 11) is similar to the total SFOAE I/O function predicted based on the equal-level protocol (the SF2000 dashed-line curve in Fig. 9).

An example of the unexpected class of results is shown for subject L02 (top panels). The SFOAE response measured with the equal-level protocol was similar to that of subject L03, but the SFOAE response measured with the fixed-L_1 protocol had high on-band noise for all L_2 levels. Thus, the SFOAE threshold for the fixed-L_1 protocol was constrained by the noise to be approximately 50 dB SPL when L_1 was

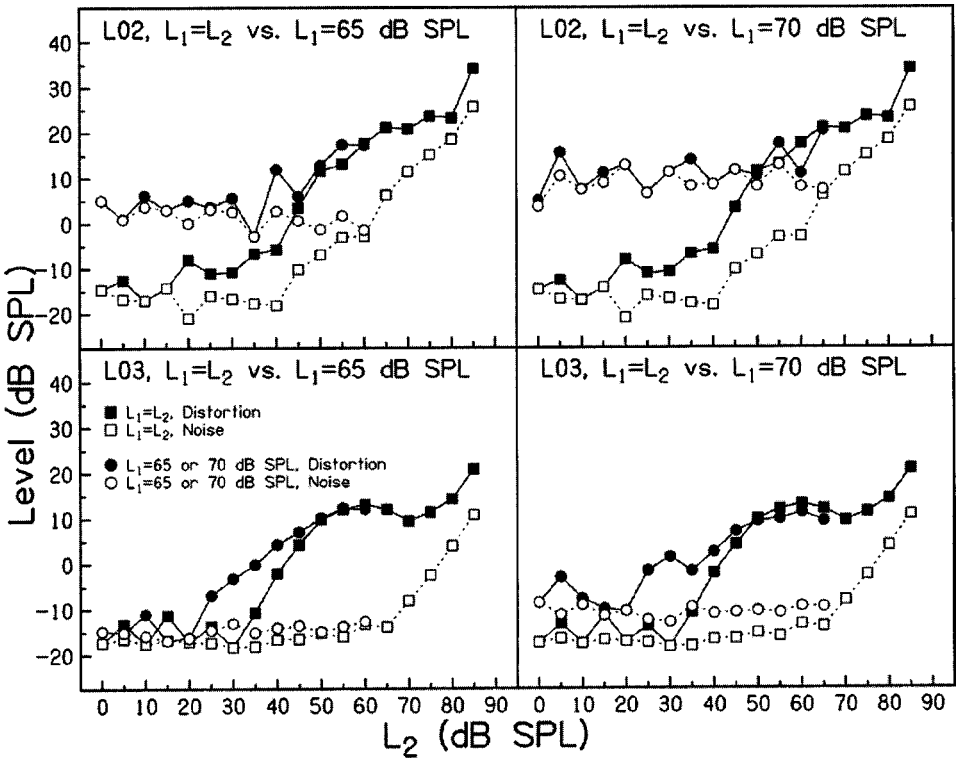


FIG. 11. SFOAE I/O functions at 2000 Hz for subjects L02 (top row) and L03 (bottom row). Each panel shows the SFOAE distortion (filled symbols) and on-band noise (open symbols) for each of the SFOAEs measured with the equal-level protocol (squares) and fixed-L_1 protocol (circles). The left column shows fixed-L_1 results with $L_1=65$ dB SPL and the right columns shows results with $L_1=70$ dB SPL.

fixed at 65 dB SPL (left upper panel), and the threshold was approximately 65 dB SPL when L_1 was fixed at 70 dB SPL. The presentation of the fixed high-level L_1 tone at 65–70 dB SPL created a high, signal-dependent noise that obscured the ability to detect the SFOAE increment produced by the addition of the L_2 tone until L_2 became sufficiently large to boost the joint presentation of L_1 and L_2 above the noise floor. For example, for the conditions that $L_1=70$ dB and $L_2=65$ dB, their joint level was $L_{12}=71.2$ dB SPL, corresponding to an increment of 1.2 dB above the level of the fixed ''pedestal'' tone L_1. Such a level of cochlear variability may be sufficient to influence intensity encoding by auditory-nerve fibers and in behavioral experiments on intensity discrimination, but only in subjects with elevated on-band SFOAE noise. This level of cochlear variability may also be observable in mechanical measurements of BM motion.

The two protocols led to similar results within both subjects between the fixed-L_1 and equal-level protocols when the stimulus levels were similar—e.g., the fixed-L_1 protocol at $L_1=65$ dB (left panels) had its highest L_2 at 60 dB SPL. It would be expected that this SFOAE response should be close to the equal-level responses at $L_1=L_2=60$ or $L_1=L_2=65$ dB SPL. This was the observed pattern.

A first conclusion is that the source of signal-dependent noise lies either within the cochlea or the middle ear, because there is no corresponding off-band noise that can be attributed to environmental or general physiological functioning such as circulation or respiration. It would appear less likely that the middle ear was the source of the on-band SFOAE noise based on the observation that there was no corresponding elevation of on-band noise in DPOAE recordings. Because the DPOAE component was in a distinct frequency bin from either of the primary stimuli, any noise added to the primary stimuli by a source within the middle ear would have also resulted in added noise to the DPOAE that these primaries produced. However, it is possible that on-band DPOAE noise was not observed because it was below the system noise floor of approximately −20 dB SPL. An increase in on-band DPOAE noise relative to off-band noise levels might be revealed in future studies if more averaging were performed to further reduce the off-band noise floor. A more complete theory would need to take into account the influence of BM compression on the relative forward-transmitted levels of the primary signal and its noise, and the relative reverse-transmitted levels of the OAE and its noise, in combination with the noise source properties assumed for the middle ear and/or cochlea. A potential source of signal-dependent noise includes fluctuations in the transmission pathway between ear canal and tonotopic place that is not synchronous with the repetitive signal averaging. This might include fluctuations in the stapedius muscle tension within the middle ear, or fluctuations in the medial olivocochlear (MOC) efferent feedback to the OHCs in the cochlea. A second conclusion is that the noise effects would be transmitted to the cochlea even if the noise source were found to lie within the middle ear.

It also is not likely due to inadequate cancellation of ''ringing'' in the cochlea. The concern is that the single buffer discarded in each of responses to the five repetitions of stimuli s_1, s_2, and s_{12} may have been too short to eliminate cochlear transients (i.e., the ringing) evoked when the stimulus changed from s_1 to s_2, s_2 to s_{12}, or s_{12} to s_1 (as described in Sec. III B). There are two pieces of evidence to support the conclusion that this ringing was negligible. First, Keefe *et al.* (2003) presented data demonstrating an increase in on-band noise relative to off-band noise in transient SFOAE responses measured in normal-hearing ears. A

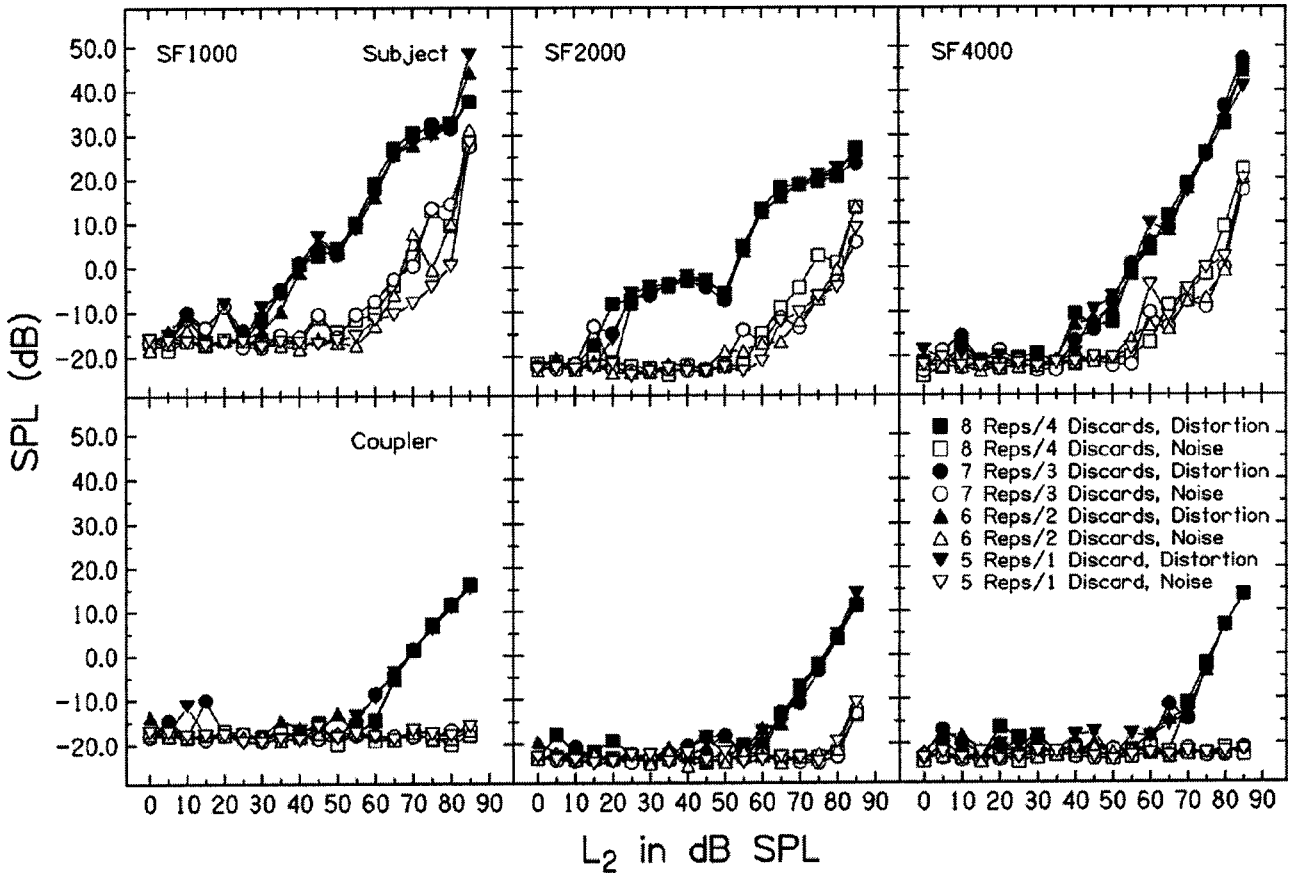


FIG. 12. Distortion (filled symbols) and on-band noise (open symbols) in an example normal-hearing subject (top row) and the coupler (bottom row) at SFOAE frequencies of 1000 Hz (left column), 2000 Hz (middle column), and 4000 Hz (right column) using four different analysis conditions. For example, in the 8 reps/4 discards condition, there were eight repetitions of the stimulus in each interval s_1, s_2, and s_{12}, the first four repetitions were discarded, and responses to the last four repetitions were used in the final average. The purpose is to demonstrate that (a) the noise does not increase as a function of L_2 in the coupler (except at $L_2 = 85$ dB SPL at 2000 Hz), whereas the noise increases as a function of L_2 in the normal-hearing ear, and (b) the noise is not dependent upon the analysis condition for the coupler or the ear. Thus, the level-dependent noise is not likely due to inadequate cancellation of cochlear ringing.

400-ms transient in the stimulus was followed by 600 ms of silence in each of the s_1, s_2, and s_{12} buffers, so that the duration of each was 1000 ms. Because the 600 ms of silence provided sufficient delay for any cochlear ringing to subside, the observed increase in on-band noise was unlikely to have been associated with cochlear ringing. Second, additional responses in this study were obtained in the IEC 711 coupler and in the first author's right ear for the SF1000, SF2000, and SF4000 conditions with four different analysis conditions: (1) eight repetitions of the stimulus per interval s_1, s_2, and s_{12}, with the first four repetitions discarded and the last four repetitions averaged for further analysis; (2) seven repetitions and three of the initial repetitions discarded; (3) six repetitions and two discards; and (4) five repetitions and one discard. The last condition was the same method used in the remainder of this report. In each case, the last four repetitions were used in the final averages. Each repetition had a duration of 23 ms so that in the first condition, the initial 92 ms (i.e., four of eight repetitions) of each response was discarded. If the level-dependent increase in on-band noise observed in normal-hearing ears was due to inadequate cancellation of ringing, then the noise should decrease as the number of repetitions and discards increases. Results in Fig. 12 show that there was no increase in on-band noise as a function of L_2 across analysis conditions in the coupler, except at the highest L_2 in the 2000 Hz condition. There was a level-dependent increase in on-band noise in the subject's ear in every analysis condition, and there did not appear to be a systematic change as a function of analysis condition. Thus, no effects were observed that could be attributed to cochlear ringing.

These findings concerning on-band SFOAE noise suggest that a source of signal-dependent variability exists in cochlear mechanics at relatively high stimulus levels. Variability has apparently not been reported in other reports of OAE measurements or mechanical measurements of BM motion, nor has it been considered in models of cochlear mechanics. There is internal variability associated with auditory-nerve firing, such that the identical input leads to differing sequences of nerve spikes (e.g., Teich and Khanna, 1985), but external mechanical variability in the input to these inner hair cell fibers might also contribute to the total variability. Such a mechanical source of variability would be expected to lead to correlations between adjacent auditory-nerve fibers. There is a report of low correlation between adjacent auditory-nerve fibers in cat (Johnson and Kiang, 1976), in which stimulus levels ranged from 25–95 dB SPL. They stated: ''Whatever correlations exist for the tone-stimulated activity can be attributed to the fact that there is a common stimulus.'' Any mechanical variability would act as

such a common stimulus, so that interpretation of the possibility of partially correlated firing patterns remains unclear, and there may also be species differences.

The present results do not allow the apportioning of the elevated on-noise level fluctuations into possible level and phase fluctuations in the input. Any lack of synchrony in the SFOAE would result in an increased noise signal, because the noise was calculated as described in the Appendix as a variance with respect to a time-averaged response.

E. Notched SFOAE I/O functions

Level-dependent notches in SFOAE I/O functions were observed in some responses, and the notches were found to be repeatable across test sessions separated by 1 month or more. These notches did not occur at the same frequencies or stimulus levels across subjects. Because the effects were not consistent across subjects, the notches did not affect the median SFOAE results described, and had minimal influence on the 25th and 75th percentiles. In some ears, there was rollover at the highest stimulus levels about which it was impossible to conclude whether this was the low side of a notch or an attenuation in SFOAE SPL with increasing stimulus level (one minor example is shown in the SF2000 response in Fig. 2).

Notches have been observed in individual responses in animal studies of DPOAEs, for example, in gerbil (Norton and Rubel, 1990), and in humans (Moulin *et al.*, 1992). However, there are apparently no published reports of notched SFOAE I/O functions in humans. Several possible mechanisms might account for such notches.

Activation of acoustic reflex might influence measurements of SFOAE I/O functions. However, the acoustic reflex typically is not activated at the low levels at which notches were observed, e.g., 55 dB SPL in Fig. 7, and if it were activated, it is difficult to understand how a sharp notch would result. If the acoustic reflex is activated, then the change in admittance magnitude should modify the pressure level at the probe microphone, and the actual SPL associated with the stimulus should differ from the expected SPL. The differences in the actual and expected L_1 and L_2 for the SF1000 condition were analyzed for nine of the subjects from the right-ear data set. This frequency was selected because acoustic reflex shifts are large at this frequency (Feeney *et al.*, 2003). The differences increased slightly with increasing level in a similar manner for all subjects, but the same trend also was observed in the coupler. This indicated that the source of the difference was not biological but rather accounted for in terms of level-dependent effects in the source impedance of the measurement system. Thus, the acoustic reflex does not account for notched SFOAE I/O functions.

The presence of a synchronized spontaneous otoacoustic emission (SSOAE) at a nearby frequency might influence SFOAE measurements, which might include entrainment effects. Definite statements are not possible because SSOAE measurements were not performed. This mechanism appears unlikely in such cases as shown in Fig. 7, in which notched I/O functions were observed at multiple frequencies in the same ear and it would appear that such effects would require multiple SSOAEs that all happened to produce a cancellation at roughly the same L_2 level. In addition, the cancellation, or entrainment, of the SFOAE would likely require a SSOAE at a level at least as large as the SFOAE. These SFOAEs had levels near 15 dB SPL, which is fairly large for SSOAEs in humans. Thus, an explanation of the notches based on SSOAEs appears unlikely.

Notches might be attributed to shifts in the fine structure (Talmadge, 2003). That is, as level increases, the minima might shift to nearby frequencies. When recording at a fixed frequency, as level increases, a minimum might shift to the recording frequency. This possibility cannot be ruled out in the current study because no SFOAE fine-structure responses were measured.

At a phenomenological level, the presence of a notched I/O function can be explained in terms of a sum of two components, each represented by magnitude and phase components. Representing the nonlinear source term by $G_{\mathrm{NL}}(Q)$, a SFOAE model incorporating coherent reflection and nonlinear distortion in additive form is obtained by generalizing Eqs. (6) and (7)

$$P_{\mathrm{oae}} = M\left\{\frac{RQ}{1-R^{-}R} + G_{\mathrm{NL}}(Q)\right\}. \tag{9}$$

Both terms share the middle-ear filtering associated with M; the potential role of multiple internal reflections is neglected (for simplicity) in the nonlinear source term. If the coherent reflection emission mechanism is presumed to act at higher stimulus levels at which the cochlear transmission is compressive, a nonlinear apical reflectance would result. Thus, both terms on the right-hand side of the above equation are nonlinear. In addition, any level-dependent action of the MOC system may further influence the generation of SFOAEs. This simple additive model predicts that a notch is produced when the components have opposite phase and comparable magnitude. Such a two-component model provides a reasonable description of the data, including the observation of 180-deg phase shifts (Fig. 7). The presence of I/O functions without notches is explained by the predominance of one component over the other, or the absence of phase cancellation in the parameter range in which the components are of comparable level.

One component is attributed to the coherent reflection model used to describe low-level SFOAEs (Zweig and Shera, 1995). The other component may be attributed to a nonlinear distortion model of SFOAE generation that becomes important at higher stimulus levels (e.g., Brass and Kemp, 1993). A plausible basis for such a model derives from the observation that saturation of the OHC receptor current accounts for two-tone suppression (Geisler *et al.*, 1990). Such a saturating OHC model can account for notched DPOAE I/O functions, but the model predicts strictly monotonic I/O functions at the fundamental frequency of the OHC receptor potential (Lukashkin and Russell, 1998). Thus, the saturating OHC model predicts strictly monotonic SFOAE I/O functions. If, however, the input to

the OHC has a level-dependent notch at the fundamental frequency, e.g., as produced by a zero in the mechanical motion at the OHC location, the OHC receptor potential would reflect that nonmonotonicity. Some measurements of OHC receptor potentials have such notches. Such a resonant mechanical motion producing a zero might arise from a resonant tectorial membrane motion (Allen, 1980; Zwislocki, 1980), or from a more general multiple-resonance model of the cochlear partition (e.g., Lin and Guinan, 2000).

Another example of a notched SFOAE I/O function, described in a two-tone suppression paradigm, was reported for a cat using a fixed-level probe condition with a probe frequency of 1800 Hz and a suppressor frequency of 2000 Hz (Guinan, 1990). Qualitatively similar notched I/O functions have been obtained in human ears with fixed-level suppressors at frequencies slightly different than the probe frequency (Schairer and Keefe, 2002).

The action of the MOC efferent system might also be involved in the observed SFOAE I/O functions. The efferent system does suppress DPOAE responses in the moderate stimulus level range. While efferent suppression effects tend to be small, on the order of 2–3 dB for contralateral efferent suppression of DPOAEs in humans using broadband noise at levels below the acoustic reflex level (Moulin *et al.*, 1992), large-amplitude SFOAEs in cats are generated by electrical stimulation of the MOC efferent system in cats (Guinan, 1990). Guinan points out that a continuous-tone protocol to measure SFOAEs may produce significant activation of the efferent system so that the resulting SFOAEs may include both suppressive and efferent contributions. This may provide an alternative pair of components contributing to the creation of notched SFOAE I/O functions.

Furthermore, an activation of the MOC efferent system, which has time constants longer than the duration of the elementary stimulus buffer (23.2 ms), is a potential source of the signal-dependent source of variability in SFOAEs at high stimulus levels. Unlike OHC activity, the efferent system cannot track the signal on a cycle-by-cycle basis so that there are possibilities for longer-term fluctuations. Because efferents produce suppression, the presence of any efferent fluctuations might influence the level of the traveling wave and thus the level of the coherently reflected SFOAE.

The present data suggest that the explanation of SFOAE generation at moderate and high levels in the human ear may involve a two-component model. The shift in phase that occurs in the same region as the level-dependent notch supports this type of model.

VI. CONCLUSIONS

In general, the measured SFOAE and DPOAE I/O functions recorded were consistent with functions reported in the literature. SFOAEs were recorded across a wider stimulus range, and in a larger number of normal-hearing ears than in previous reports. SFOAEs were elicited with equal-level (0 to 85 dB SPL) and equal-frequency (half octaves between 500 and 8000 Hz) primaries. SFOAEs at 500 Hz were contaminated by high biological noise levels, and at 8000 Hz by excessive system distortion at moderate to high stimulus levels. For the remaining conditions, responses recorded in a test cavity revealed an increased system distortion at higher stimulus levels for SFOAEs; responses in ears with cochlear implants suggested sources of biological and/or system distortion that cannot be accounted for in terms of the OHC mechanism producing SFOAEs in normal-hearing ears. SFOAE I/O functions have larger absolute levels, but smaller SNRs, in comparison to DPOAE I/O functions, particularly at higher stimulus levels, and are less compressive than DPOAE I/O functions. The measured I/O functions were fitted using a two-slope model that estimated the breakpoint of cochlear nonlinearity, the compressive slope, and the round-trip attenuation of the low-level SFOAE relative to the stimulus level. An increase in biological noise with increases in stimulus level was present only at the frequency of the SFOAE response, but not in DPOAE responses, and no increase was observed at adjacent frequencies. This finding suggests the presence of variability in cochlear mechanics at higher stimulus levels. Hypotheses regarding the potential effects of such a noise source at the level of cochlear encoding may be testable for mechanical measurements of basilar-membrane motion, auditory-nerve fiber recordings, and behavioral studies of intensity discrimination. This increased biological noise contaminated fixed-level primary conditions in some measurements of SFOAEs, with increased noise present at all points across the I/O function. Finally, level-dependent notches were observed in some individual SFOAE I/O functions. This finding suggests the interference of two sources of SFOAE generation, for which the second source may involve nonlinear distortion on the BM or activation of the MOC efferent system.

ACKNOWLEDGMENTS

Fei Zhao participated in preliminary stages of measurements, and in the interpretation of measurement-system distortion and calibration. This research was supported by the NIH (R01 CD003784 and T32 DC00013). We thank Carrick Talmadge and one other anonymous reviewer for providing helpful suggestions and raising important questions.

APPENDIX: NOISE ESTIMATION ASSOCIATED WITH OAE RESPONSES

This Appendix describes the method by which the noise of the OAE and the noise of each stimulus response were calculated. The generic variable P is used in this Appendix to denote the complex DFT of any of the 2E stimulus responses (P_1, P_2, P_{12}) or the associated OAE (P_d). The spectral pressure in the kth frequency bin associated with the mth repetition of the response is $P^m[k]$, such that m varies from 1 to $M=32$ and k varies from 1 to $N/2-1$ (both the zero-frequency bin $k=0$ and the Nyquist rate bin $N/2$ are outside the measurement bandwidth of interest).

The coherent sum (or average complex spectrum) is defined by

$$\langle P[k]\rangle = \frac{1}{M}\sum_{m=1}^{M} P^m[k]. \tag{A1}$$

The coherent signal energy of the signal is defined by

$$|\langle P[k]\rangle|^2=\frac{1}{M^2}\left|\sum_{m=1}^{M} P^m[k]\right|^2. \tag{A2}$$

The (real) variance of the signal is defined as

$$\mathrm{Var}(P[k])=\frac{\Sigma_{m=1}^{M}|P^m[k]|^2-M|\langle P[k]\rangle|^2}{M-1}. \tag{A3}$$

With multiplication by $M/(M-1)$, the above variance equals the incoherent (noise) energy from the nonlinear coherence technique described in Keefe (1998). The noise energy $|P_N[k]|^2$ in the kth spectral bin is defined as the standard error of the mean

$$|P_N[k]|^2=\frac{\mathrm{Var}(P[k])}{M}. \tag{A4}$$

The signal energy $|P[k]|^2$ in the kth spectral bin is

$$|P[k]|^2=|\langle P[k]\rangle|^2. \tag{A5}$$

Defining the reference pressure as $P_{\mathrm{ref}}=0.0002$ dyne/sq cm, the spectral sound-pressure levels of the noise and signal defined from bins $k=1$ up to $N/2-1$ are

$$\begin{aligned} SPL_N[k]&=10\log\left(\frac{2}{N^2}\frac{|P_N[k]|^2}{P_{\mathrm{ref}}^2}\right),\\ SPL_P[k]&=10\log\left(\frac{2}{N^2}\frac{|P_N[k]|^2}{P_{\mathrm{ref}}^2}\right). \end{aligned} \tag{A6}$$

The total SPL is defined in terms of summing energy across frequency:

$$\mathrm{SPL}_{P,\mathrm{tot}}=10\log\left(\frac{2}{N^2P_{\mathrm{ref}}^2}\sum_{k=1}^{N/2-1}|P[k]|^2\right). \tag{A7}$$

This total SPL is used within the program to calibrate the levels of each of the stimuli p_1, p_2, and p_{12}, and a similar relation is obtained for the total noise SPL.

The relevant property of calculating OAE noise using Eqs. (A4) and (A6) is that it is an ensemble average over $M=32$ independent buffers so that random errors in estimating noise diminish as $1/\sqrt{M}$. This is contrasted with algorithms to estimate noise at the OAE frequency based on the use of two independent ensembles ($M=2$), or algorithms that estimate the noise at the OAE frequency by averaging over frequency bins adjacent to the signal frequency. The former method has a small M so that the random error in estimating the noise is large. The latter method assumes that the noise at the OAE frequency is identical to the noise in adjacent bins, an assumption that is tested in the Results section and found to be true for DPOAEs but sometimes false for SFOAEs. The reduction of the random error by $M=32$ is observable in the smooth variation of the noise SPL in an individual SFOAE response such as Fig. 5. This figure also shows the elevation in the noise at the frequency of the SFOAE. The ability to detect such a change in noise level in the SFOAE bin depends on controlling the random errors in estimating noise at all frequencies.

As a practical manner in acquiring these data, each of the 32 repetitions was actually a time average of four valid buffers. The resulting number of buffers in the time average was $N_{\mathrm{Tot}}=128$ responses while the number of averages in calculating the variance, and hence the noise, was 32 responses. The ensemble-averaged noise level is determined by N_{Tot}, while the variability in this noise estimate is reduced for fixed N_{Tot} by increasing M.

Allen, J. B. (**1980**). "Cochlear micromechanics—A physical model of transduction," J. Acoust. Soc. Am. **68**, 1660–1670.

Baer, T., Moore, B. C., and Marriage, J. (**2001**). "Detection and intensity discrimination of brief tones as a function of duration by hearing-impaired listeners," Hear. Res. **159**, 74–84.

Brass, D., and Kemp, D. T. (**1991**). "Time-domain observation of otoacoustic emissions during constant tone stimulation," J. Acoust. Soc. Am. **90**, 2415–2427.

Brass, D., and Kemp, D. T. (**1993**). "Suppression of stimulus frequency otoacoustic emissions," J. Acoust. Soc. Am. **93**, 920–939.

Buus, S., Obeling, L., and Florentine, M. (**2001**). "Can basilar-membrane compression characteristics be determined from distortion-product otoacoustic-emission input-output functions in humans?" in *Proceedings of the 12th International Symposium on Hearing: Physiological and Psychophysical Bases of Auditory Function*, edited by D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker, Maastricht, Netherlands), pp. 373–381.

Dorn, P. A., Konrad-Martin, D., Neely, S. T., Keefe, D. H., Cyr, E., and Gorga, M. P. (**2001**). "Distortion product otoacoustic emission input/output functions in normal-hearing and hearing-impaired human ears," J. Acoust. Soc. Am. **110**, 3119–3131.

Dorn, P. A., Piskorski, P., Keefe, D. H., Neely, S. T., and Gorga, M. P. (**1998**). "On the existence of an age/threshold/frequency interaction in distortion product otoacoustic emissions," J. Acoust. Soc. Am. **104**, 964–971.

Dreisbach, L. E., Siegel, J. H., and Chen, W. (**1998**). "Stimulus-frequency otoacoustic emissions measured at low- and high-frequencies in untrained human subjects," Abstracts of the Twenty-First Annual Midwinter Research Meeting of the Association for Research in Otolaryngology, 88.

Feeney, M. P., Keefe, D. H., and Marryott, L. P. (**2003**). "Contralateral acoustic reflex thresholds for tonal activators using wideband energy reflectance and admittance," J. Speech Lang. Hear. Res. **46**, 128–136.

Geisler, C. D., Yates, G. K., Patuzzi, R. B., and Johnstone, B. M. (**1990**). "Saturation of outer hair cell receptor currents causes two-tone suppression," Hear. Res. **44**, 241–256.

Gorga, M. P., Neely, S. T., Bergman, B. M., Beauchaine, K. L., Kaminski, J. R., and Liu, Z. (**1994**). "Towards understanding the limits of distortion product otoacoustic emission measurements," J. Acoust. Soc. Am. **96**, 1494–1500.

Guinan, J. J. (**1990**). "Changes in stimulus frequency otoacoustic emissions produced by two-tone suppression and efferent stimulation in cat," in *The Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele (Springer, Madison, WI), pp. 170–177.

Johnson, D. H., and Kiang, N. Y. S. (**1976**). "Analysis of discharges recorded simultaneously from pairs of auditory nerve fibers," Biophys. J. **16**, 719–734.

Kalluri, R., and Shera, C. A. (**2001**). "Distortion-product source unmixing: A test of the two-mechanism model for DPOAE generation," J. Acoust. Soc. Am. **109**, 622–637.

Keefe, D. H. (**1998**). "Double-evoked otoacoustic emissions. I. Measurement theory and nonlinear coherence," J. Acoust. Soc. Am. **103**, 3489–3498.

Keefe, D. H. (**2002**). "Spectral shapes of forward and reverse transfer functions between ear canal and cochlea estimated using DPOAE input/output functions," J. Acoust. Soc. Am. **111**, 249–260.

Keefe, D. H., and Ling, R. (**1998**). "Double-evoked otoacoustic emissions. II. Intermittent noise rejection, calibration, and ear-canal measurements," J. Acoust. Soc. Am. **103**, 3499–3508.

Keefe, D. H., Piskorski, P., and Gorga, M. P. (**1999**). "Double-evoked otoacoustic emissions: Time-frequency representations and high-frequency transient-response measurements," Abstracts of the Twenty-Second Annual Midwinter Research Meeting of the Association for Research in Otolaryngology, 99.

Keefe, D. H., Schairer, K. S., and Jesteadt, W. (**2003**). "Is there an OAE

correlate to behavioral overshoot?'' Paper presented at the 26th Midwinter Research Meeting of the Association for Research in Otolaryngology, Daytona Beach, FL, February, 2003.
Keefe, D. H., Schairer, K. S., Zhao, F., and Simmons, J. (**2001**). ''Measurement and interpretation of stimulus frequency otoacoustic emissions: Input/output functions and transient analyses,'' J. Acoust. Soc. Am. **109**, 2373–2374.
Kemp, D. T. (**1978**). ''Stimulated acoustic emissions from within the human auditory system,'' J. Acoust. Soc. Am. **64**, 1386–1391.
Kemp, D. T. (**1979**). ''The evoked cochlear mechanical response and the auditory microstructure—Evidence for a new element in cochlear mechanics,'' Scand. Audiol. Suppl. **9**, 35–47.
Kemp, D. T. (**1986**). ''Otoacoustic emissions, traveling waves and cochlear mechanisms,'' Hear. Res. **22**, 95–104.
Kemp, D. T., Brass, D., and Souter, M. (**1990**). ''Observations on simultaneous SFOAE and DPOAE generation and suppression,'' in *The Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele (Springer, New York), pp. 202–209.
Kemp, D. T., and Brown, A. M. (**1983**). ''A comparison of mechanical nonlinearities in the cochleae of man and gerbil from ear canal measurements,'' in *Hearing Physiological Basis and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer, Germany), pp. 82–88.
Kemp, D. T., and Chum, R. A. (**1980**). ''Observations on the generator mechanism of stimulus frequency acoustic emissions—Two tone suppression,'' in *Physiological, Psychological, and Behavioral Studies in Hearing*, edited by G. van der Brink and F. A. Bilsen (Delft University Press, Delft, Netherlands), pp. 34–42.
Knight, R. D., and Kemp, D. T. (**2001**). ''Wave and place fixed DPOAE maps of the human ear,'' J. Acoust. Soc. Am. **109**, 1513–1525.
Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., Cyr, E., and Gorga, M. P. (**2002**). ''Sources of DPOAEs revealed by suppression experiments, inverse fast Fourier transforms, and SFOAEs in impaired ears,'' J. Acoust. Soc. Am. **111**, 1800–1809.
Kummer, P., Janssen, T., and Arnold, W. (**1998**). ''The level and growth behavior of the 2 $f1-f2$ distortion product otoacoustic emission and its relationship to auditory sensitivity in normal hearing and cochlear hearing loss,'' J. Acoust. Soc. Am. **103**, 3431–3444.
Lin, T., and Guinan, J. J. (**2000**). ''Auditory-nerve-fiber responses to high-level clicks: Interference patterns indicate that excitation is due to the combination of multiple drives,'' J. Acoust. Soc. Am. **107**, 2615–2630.
Lonsbury-Martin, B. L., Cutler, W. M., and Martin, G. K. (**1991**). ''Evidence for the influence of aging on distortion-product otoacoustic emissions in humans,'' J. Acoust. Soc. Am. **89**, 1749–1759.
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. Suppl. **147**, 3–14.
Lukashkin, A. N., and Russell, I. J. (**1998**). ''A descriptive model of the receptor potential nonlinearities generated by the hair cell mechanoelectrical transducer,'' J. Acoust. Soc. Am. **103**, 973–980.
Moore, B. C., and Glasberg, B. (**1997**). ''A model of loudness perception applied to cochlear hearing loss,'' Aud. Neurosci. **3**, 289–311.
Moore, B. C., and Oxenham, A. J. (**1998**). ''Psychoacoustic consequences of compression in the peripheral auditory system,'' Psychol. Rev. **105**, 108–124.
Moulin, A., Collet, L., and Morgon, A. (**1992**). ''Influence of spontaneous otoacoustic emissions (SOAE) on acoustic distortion product input/output functions: Does the medial efferent system act differently in the vicinity of an SOAE?'' Acta. Otolaryngol. (Stockh) **112**, 210–214.
Norton, S. J., and Rubel, E. W. (**1990**). ''Active and passive ADP components in mammalian and avian ears,'' in *Lecture Notes in Biomathematics: The Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele (Springer, New York), pp. 219–226.
Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.
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., and Oxenham, A. J. (**1998**). ''Basilar-membrane nonlinearity and the growth of forward masking,'' J. Acoust. Soc. Am. **103**, 1598–1608.
Rhode, W. S. (**1971**). ''Observations of the vibration of the basilar membrane in squirrel monkeys using the Mossbauer technique,'' J. Acoust. Soc. Am. **49**, 1218–1231.
Rhode, W. S., and Recio, A. (**2000**). ''Study of mechanical motions in the basal region of the chinchilla cochlea,'' J. Acoust. Soc. Am. **107**, 3317–3332.
Robles, L., and Ruggero, M. A. (**2001**). ''Mechanics of the mammalian cochlea,'' Physiol. Rev. **81**, 1305–1352.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). ''Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input–output functions, tuning curves, and response phases,'' J. Acoust. Soc. Am. **80**, 1364–1374.
Ruggero, M. A. (**1992**). ''Responses to sound of the basilar membrane of the mammalian cochlea,'' Curr. Opin. Neurobiol. **2**, 449–456.
Ruggero, M. A., and Rich, N. C. (**1991**). ''Furosemide alters organ of corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane,'' J. Neurosci. **11**, 1057–1067.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
Schairer, K. S., and Keefe, D. H. (**2002**). ''SFOAE input/output functions elicited by slightly off-frequency suppressors,'' J. Acoust. Soc. Am. **111**, 2356.
Schlauch, R. S., DiGiovanni, J. J., and Ries, D. T. (**1998**). ''Basilar membrane nonlinearity and loudness,'' J. Acoust. Soc. Am. **103**, 2010–2020.
Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (**1982**). ''Measurement of basilar membrane motion in the guinea pig using the Mossbauer technique,'' J. Acoust. Soc. Am. **72**, 131–141.
Shera, C. A., and Guinan, J. J. (**1999**). ''Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs,'' J. Acoust. Soc. Am. **105**, 782–798.
Shera, C. A., Guinan, J. J., and Oxenham, A. J. (**2002**). ''Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements,'' Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–3323.
Shera, C. A., and Zweig, G. (**1993**). ''Noninvasive measurement of the cochlear traveling-wave ratio,'' J. Acoust. Soc. Am. **93**, 3333–3352.
Souter, M. (**1995**). ''Stimulus frequency otoacoustic emissions from guinea pig and human subjects,'' Hear. Res. **90**, 1–11.
Strube, H. W. (**1989**). ''Evoked otoacoustic emissions as cochlear Bragg reflections,'' Hear. Res. **38**, 35–45.
Talmadge, C. L. (**2003**). Personal communication.
Talmadge, C. L., Tubis, A., Long, G. R., and Tong, C. (**2000**). ''Modeling the combined effects of basilar membrane nonlinearity and roughness on stimulus frequency otoacoustic emission fine structure,'' J. Acoust. Soc. Am. **108**, 2911–2932.
Teich, M. C., and Khanna, S. M. (**1985**). ''Pulse-number distribution for the neural spike train in the cat's auditory nerve,'' J. Acoust. Soc. Am. **77**, 1110–1128.
Whitehead, M. L., Stagner, B. B., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995**). ''Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears. II. Asymmetry in L1,L2 space,'' J. Acoust. Soc. Am. **97**, 2359–2377.
Wilson, J. P. (**1980**). ''Evidence for a cochlear origin for acoustic re-emissions, threshold fine-structure and tonal tinnitus,'' Hear. Res. **2**, 233–252.
Withnell, R. H., and Yates, G. K. (**1998**). ''Onset of basilar membrane nonlinearity reflected in cubic distortion tone input–output functions,'' Hear. Res. **123**, 87–96.
Yates, G. K., Winter, I. M., and Robertson, D. (**1990**). ''Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range,'' Hear. Res. **45**, 203–219.
Zweig, G., and Shera, C. A. (**1995**). ''The origin of periodicity in the spectrum of evoked otoacoustic emissions,'' J. Acoust. Soc. Am. **98**, 2018–2047.
Zwicker, E., and Schloth, E. (**1984**). ''Interrelation of different otoacoustic emissions,'' J. Acoust. Soc. Am. **75**, 1148–1154.
Zwislocki, J. J. (**1980**). ''Five decades of research on cochlear mechanics,'' J. Acoust. Soc. Am. **67**, 1679–1685.

Discriminating harmonicity

Gerald Kidd, Jr.,[a] Christine R. Mason, Andrew Brughera, and Chung-Yiu Peter Chiu[b]
Hearing Research Center, Sargent College, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215

(Received 2 August 2002; revised 13 February 2003; accepted 12 May 2003)

Simultaneous tones that are harmonically related tend to be grouped perceptually to form a unitary auditory image. A partial that is mistuned stands out from the other tones, and harmonic complexes with different fundamental frequencies can readily be perceived as separate auditory objects. These phenomena are evidence for the strong role of harmonicity in perceptual grouping and segregation of sounds. This study measured the discriminability of harmonicity directly. In a two interval, two alternative forced-choice (2I2AFC) paradigm, the listener chose which of two sounds, signal or foil, was composed of tones that more closely matched an exact harmonic relationship. In one experiment, the signal was varied from perfectly harmonic to highly inharmonic by adding frequency perturbation to each component. The foil always had 100% perturbation. Group mean performance decreased from greater than 90% correct for 0% signal perturbation to near chance for 80% signal perturbation. In the second experiment, adding a masker presented simultaneously with the signals and foils disrupted harmonicity. Both monaural and dichotic conditions were tested. Signal level was varied relative to masker level to obtain psychometric functions from which slopes and midpoints were estimated. Dichotic presentation of these audible stimuli improved performance by 3–10 dB, due primarily to a release from ''informational masking'' by the perceptual segregation of the signal from the masker. [DOI: 10.1121/1.1587734]

PACS numbers: 43.66.Ba, 43.66.Dc, 43.66.Fe [NFV]

I. INTRODUCTION

In multisource acoustic environments, it is often a listener's task to detect, locate, identify, and receive information from a specific sound source in the presence of competing sound sources. The interference in the reception and processing of one particular sound—the signal or target—due to competing sounds, or maskers, is thought to consist of a peripheral component, sometimes called ''energetic masking,'' and a central component referred to as ''informational masking'' (e.g., Watson and Kelly, 1981; Watson, 1987; Durlach *et al.*, 2003a). Previous work has suggested that the informational component of masking can be reduced by exploiting signal presentation schemes that perceptually segregate the signal from the masker(s) (e.g., Kidd *et al.*, 1994, 1998; Neff, 1995; Freyman *et al.*, 1999; Brungart, 2001; Durlach *et al.*, 2003b). Several aspects of an acoustic signal, or the way the signal is presented, have been identified as leading to perceptual segregation. These include *harmonicity*, which is the degree to which the frequency components in a complex spectrum fall at integer multiples of the fundamental frequency (e.g., Yost, 1991; Darwin and Carlyon, 1995). Frequencies that fall at harmonic intervals tend to be perceived as having originated from the same sound source, especially if they also begin and end at the same time, and thus may form an auditory object distinct from other objects.

The evidence that the harmonic–inharmonic distinction is important in the perceptual organization of sounds is relatively strong. A harmonic relation among the frequency components of a complex sound is generally considered compelling evidence that the components arise from the same source. In contrast, mistuning a partial of a harmonic complex can cause the partial to perceptually segregate from the complex and stand out as a separate auditory object. Moore *et al.* (1984) found that subjects could reliably discriminate mistunings smaller than 1% for some partials of harmonic complexes. However, their results did not necessarily reflect the minimum amount of mistuning necessary for the partial to perceptually segregate from the remaining tones because the mistuning could exert an influence on the overall pitch or quality of the complex (e.g., Moore *et al.*, 1985; Darwin, 1992). Subsequent studies specifically intended to examine this issue found that the amount of mistuning necessary to segregate a partial was slightly larger than the just-noticeable difference (jnd) reported by Moore *et al.* (1984) ranging roughly from about 1%–2.5% (e.g., Moore *et al.*, 1986; Hartmann *et al.*, 1990) depending on fundamental frequency and harmonic number. Although these values are larger than the best frequency difference limen for tones played in isolation (e.g., Weir *et al.*, 1977), the ability to segregate a component of a complex sound based on a mistuning of 1 or 2 percent must be considered an impressive demonstration of the resolving power of a perceptual system given a complex input. It is of interest to note that recent work by Roberts and Brunstrom (1998) has indicated that a harmonic relation among the partials of a complex sound is not necessary for the ability to discriminate slight deviations from a regular spectral pattern.

Related work on the role of harmonicity in the grouping and segregation of elements of speech sounds has indicated that ''mistuning'' the frequency components of a single for-

[a]Electronic mail: gkidd@bu.edu
[b]Current address: Department of Psychology, University of Cincinnati, Cincinnati, OH.

mant of a speech sound from the fundamental frequency of the remainder of the sound can lead to the perception of the mistuned formant having arisen from a separate source (e.g., Darwin, 1981). For example, Gardner *et al.* (1989) demonstrated that small mistunings of one formant of a four-formant syllable led to the perceptual segregation of that formant, although it did not alter phonetic identity. However, greater mistuning changed the perceived identity of the speech sound. For the smaller mistunings they speculated that, despite the apparent segregation of the formant, other cues such as common onset time and ''plausibility,'' preserved phoneme identity.

Other research linking harmonicity to auditory object perception has employed tasks in which multiple independent sounds are played to the listener simultaneously. If two sounds are harmonic, but have different fundamental frequencies, they may be perceived as two distinct objects rather than one. The evidence that the harmonic sounds are perceived as separate entities comes in part from speech identification experiments in which the task is to label which phonemes are heard in a double-vowel context (e.g., Scheffers, 1983; Assman and Summerfield, 1990; Culling and Darwin, 1993) or to report the semantic content of a target sentence (e.g., Brokx and Nooteboom, 1982; Bird and Darwin, 1998) in the presence of a competing sentence. In both types of experiments, increasing the difference in the fundamental frequency between sounds generally leads to improved identification performance. Other cues, such as modulation in frequency or amplitude, can interact with harmonic structure to affect the prominence or masked threshold of the target vowel (e.g., McAdams, 1989; Summerfield, 1992; Culling and Summerfield, 1995). Thus, the ability to perceptually segregate simultaneous complex sounds, and to receive the information they convey, can depend on harmonicity.

Carlyon (1994) has demonstrated that, under certain circumstances, the perception of a mistuned partial in an otherwise harmonic complex can be masked if it is pulled into an auditory ''stream'' separate from the remaining tones in the complex. In his task the mistuning of the partial was a 180-deg phase difference in frequency modulation between the target and the remaining harmonic components. When the masking sound (e.g., an unmodulated tone equal to the nominal frequency of the target harmonic) began before and ended after the test stimulus, it caused the mistuned component to perceptually group with the masker and segregate from the harmonic complex, and the just-noticeable difference in mistuning significantly increased. His findings indicate that the ability to discern an inharmonic relation among partials depends to some degree on the grouping of the target component with the remaining tones in the complex. It is also of interest to note that the masking effect he found does not appear to be ''energetic masking'' in the sense that the masker obscured the peripheral representation of the target, but rather it apparently caused a perceptual effect that interfered with the ability of the listener to extract information about the target.

If we accept that there is a perceptual dimension corresponding to ''harmonicity'' which can be used as a basis for separating auditory objects in a multisource environment, then it would be useful to vary the strength of this perception in a way that is a monotonic function of a simple acoustic variable. The procedures used previously to examine the effects of mistuning a single partial or formant of a harmonic complex, or to determine the difference in recognition or quality of speech sounds having different fundamental frequencies, were not intended to generate such a function. It would be of interest to measure the ability of listeners to discriminate varying degrees of harmonicity over a wide range of values so that more extensive and general measures of harmonicity would be possible (i.e., measure just-noticeable differences as a function of the reference harmonicity, determine greater-than/less-than relationships, construct a psychophysical scale of subjective magnitude, etc.).

In order to pursue such studies, however, it would seem that a new measurement approach is required that does not depend on the mistuning of single partials or formants of a complex or on a difference in the fundamental frequency of two separate speech sounds. Further, as suggested by the study by Carlyon (1994), it should be possible to interfere with the perception of harmonicity separately from the detection of the presence of the signal itself. That is, the stimulus is heard but the property of harmonicity is masked. This type of interference would seem to be similar in some respects to conditions producing informational masking where the masker interferes with the extraction of some aspect of an above-threshold sound.

There were two purposes to the present study: the first purpose was to design and evaluate a means for varying the ''degree of harmonicity'' along a continuum that would produce a corresponding perceptual experience in the listener and which could provide the basis for the study of the discrimination of harmonicity. Towards that end, experiment 1 extended the work on inharmonicity perception to the case where all of the partials are mistuned, rather than a single partial. The ability to discriminate between pairs of sounds that differed only in their degree of harmonicity was measured for signals comprised of frequencies that were progressively ''perturbed'' from their expected harmonic intervals. The perturbation that was employed was similar, although not identical, to that which has been used in studies of the role of harmonicity in concurrent vowel identification (e.g., de Cheveigne *et al.*, 1995, 1997). The hope was that it would be possible to vary the discriminability of inharmonicity over a wider range than has been possible using previous methods. The purpose of experiment 2 was to examine how well listeners can extract the property of harmonicity from complex sounds masked by other complex sounds. As noted above, current theory suggests that masking is composed of a peripherally based energetic component and a centrally based informational component. Although there is no doubt that energetic masking could interfere with discriminating among sounds based on the harmonic–inharmonic distinction, much less is known about how informational masking affects such tasks. In energetic masking, it is assumed that performance is limited because of insufficient data available to a central processor to solve the task—the necessary information simply is obscured in the peripheral neural representation. However, in contrast, it is assumed in informational

masking conditions that there *is* sufficient information available in the peripheral representation of the stimulus to solve the task and that it is the central processing that limits performance. Understanding the reasons for this nonoptimal processing provides the motivation for the study of informational masking. It is of particular interest to study informational masking in harmonicity discrimination tasks. This is because we assume that computation of harmonicity depends on information about the frequency components of the stimulus. It would seem plausible that such a computation could be interfered with by irrelevant frequency components, especially if they occur in a highly uncertain context where the listener would be unable to construct a rejection filter matched to the masker (e.g., de Cheveigne *et al.*, 1995; Durlach *et al.*, 2003a). In experiment 2, we used maskers that differed in the ratio of energetic-to-informational masking by varying the number of components of randomly drawn multitone maskers and superimposing them on the signal and foil. This type of multitone masker is similar to that which has been used extensively by Neff and colleagues (e.g., Neff and Green, 1987; Neff *et al.*, 1993; Neff, 1995; see also Oh and Lutfi, 1998; Wright and Saberi, 1999; Richards *et al.*, 2002; Durlach *et al.*, 2003b) to study informational masking in the detection of a pure-tone signal of known frequency. In addition, a dichotic manipulation was introduced that had the effect of promoting perceptual segregation of signal/foil and masker without disturbing the stimulus in the test ear, thereby potentially reducing the informational component of masking.

II. METHODS

Four young adults (ages 18–32 years) including author P.C. served as listeners. All subjects had normal hearing from 250–8000 Hz as determined by standard audiometric assessment. The listeners were tested individually in a double-walled sound-attenuating booth and listened to the stimuli through matched and calibrated TDH-50 earphones. Responses were registered on a computer keyboard, and an LCD monitor provided visual display of trial-by-trial information including the demarcation of presentation and response intervals and response feedback. All stimuli were computer generated at a 20-kHz rate and were subsequently low-pass filtered at 7.5 kHz. The sounds had a total duration of 520 ms including 20-ms $\cos^2$ gating ramps. The task of the listener was to judge which of two sounds—a signal and a foil—was ''more harmonic.'' The subjects were told that the sounds would vary in pitch and that they should use the response feedback to learn to solve the task. A two-interval, two-alternative forced-choice procedure was employed with the signal played in one interval and the foil in the other interval of each trial. Trials were arranged in blocks of 50. The subjects practiced the experimental tasks for several hundred trials prior to collection of the data reported here.

The signals were comprised of 11 equal-level tones. The frequency relationship between the signal tones was either an exact harmonic series (i.e., every tone was an integer multiple of the fundamental frequency, F_0) or was a perturbed harmonic series in which the tones were perturbed in frequency so that they were unlikely to equal integer multiples of the fundamental. The perturbation was implemented by drawing a random sample from a uniform distribution of frequencies and adding that value to the expected harmonic frequency. A different sample was drawn for the fundamental and each harmonic. The range of the uniform distribution of perturbation was a controlled variable extending from the unperturbed case to a case where the range equaled ± 0.4 times F_0 (defined as 80% perturbation). The foils, which were presented in every nonsignal interval, were the same as signals with 100% perturbation (i.e., ± 0.5 times F_0). The fundamental frequencies for both signals and foils were randomly chosen on every interval from a set of 15 frequencies spaced 20 Hz apart ranging from 200–480 Hz. Figure 1 illustrates the signals and foils.

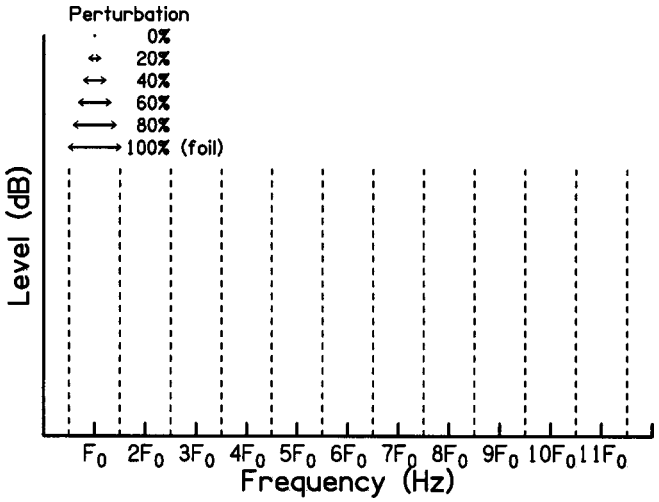


FIG. 1. A schematic illustration of the frequency bins used to generate signals and foils. The abscissa shows the fundamental frequency, F_0, and the frequencies that are integer multiples—harmonics—of F_0. Perfectly harmonic stimuli fall exactly at the integer multiples of F_0, while perturbed signals have F_0's and harmonics selected from the ranges shown above the first bin. Foils have 100% perturbation.

The schematic shows a nominal F_0 and integer multiples of the F_0. The dashed lines indicate the boundaries within which tones comprising the foils may fall, which span a range of ± 0.5 F_0. The arrows above the frequency bin associated with F_0 indicate the range of the foil and the lesser amounts of perturbation, ranging from 20%–80% of F_0, for the perturbed signal. Thus, there is a continuum of values ranging from perfectly harmonic, 0% perturbation, to maximally inharmonic,[1] 100% perturbation. The frequencies of the tones were always spaced more than 2 Hz apart.

There were two experiments. In the first experiment, the variable under study was the degree of perturbation of the signal. The level of the stimuli was held constant at 60 dB SPL. There were five amounts of perturbation: 0%, 20%, 40%, 60%, and 80%. The amounts of perturbation were mixed and randomized throughout each block of trials. The top and middle panels of Fig. 2 schematically illustrate the pairing of a signal and foil on representative trials.

In the top left panel, the signal has 0% perturbation. Note that the nominal F_0 is different for the signal and foil. In the middle left panel, a similar illustration is shown but for a signal having 40% perturbation. In the first experiment,

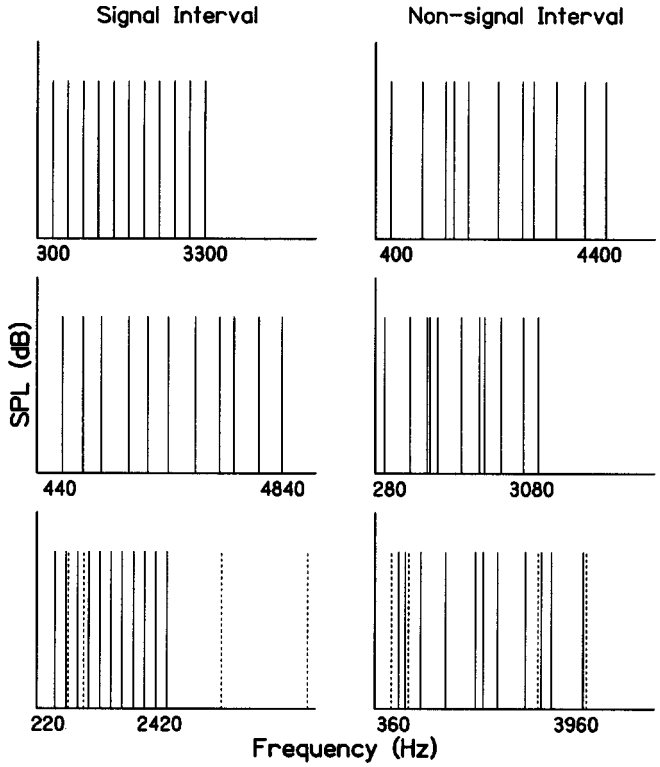


FIG. 2. Examples of signal and foil stimuli that might be presented in the two intervals of a trial in three conditions. Signals are illustrated in the left column, while foils are illustrated in the right column. First, the signal is not perturbed and is paired with a foil (upper two panels); second, the signal is a drawn from the 40% perturbation distribution, again paired with a typical foil (middle two panels); and third, random draws of the four-tone masker are shown as dashed lines overlaid on the no-perturbation signal and typical foil (lower two panels).

listening was monaural.

In the second experiment a signal was presented in one interval of every trial and a foil was presented in the other interval. The signal was always 0% perturbation and the foil was 100% perturbation, with both having randomly drawn nominal F_0's as in experiment 1, but here masking sounds were added. The maskers were comprised of 4, 8, or 12 equal-level tones randomly drawn from 100 to 5520 Hz having a uniform probability density function on a logarithmic frequency axis. A different masker sample was drawn for each stimulus presentation (i.e., different masker samples across intervals of a trial as well as across trials).

The lower panels of Fig. 2 illustrate a signal and a foil with randomly chosen four-tone maskers added. The maskers were constant in level at 60 dB SPL per component for the four-tone masker and 50 dB SPL per component for the eight- and 12-tone maskers. The purpose of the higher level per component for the four-tone masker was to minimize possible ''floor effects'' due to too little masking. Several signal levels were tested in the presence of the fixed-level maskers generating performance-level functions. During each run for a given masker, the signal levels were sampled randomly on each trial. There were two modes of signal presentation: monotic and dichotic. The dichotic condition was signal (or foil) monaural, masker diotic (S_mM_0).

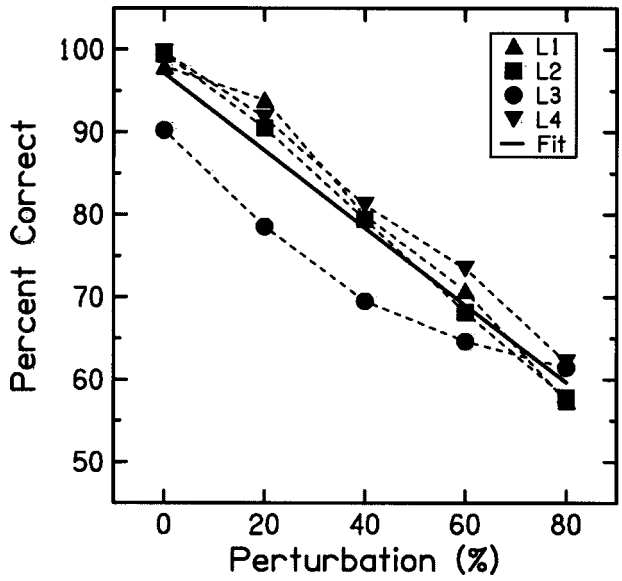


FIG. 3. The results from experiment 1 showing performance for individual listeners (symbols with dashed lines). The abscissa is percent signal perturbation and the ordinate is percent-correct performance. The solid line is a linear least-squares fit to the data (fit). The symbol key identifies individual subjects.

III. RESULTS

A. Experiment 1

Figure 3 shows the individual percent-correct scores plotted as a function of the amount of perturbation. The straight line is a linear least-squares fit to the data.

Three of the listeners were similar in performance for all values of perturbation, while one listener was poorer than the others for all amounts of perturbation up to 80%. However, the trend in the data was very similar for all four listeners: discrimination performance declined monotonically with increasing perturbation over the range of values tested. The line fit the group results extremely well, accounting for 99% of the variance in the data. On average, discrimination performance was about 97% correct in quiet, and decreased about 0.49% for every percent increase in perturbation. Thus, the manipulation of adding random perturbation to the components of a harmonic complex resulted in an orderly and systematic discrimination function that was well described by a straight line having a slope of about -0.49. A small (about 3%, on average) but statistically significant (two-tailed t-test, $t=-9.7$, df$=3$, $p=0.002$) improvement was observed between performance on the first half of the experiment and the last half of the experiment, indicating that some additional learning occurred during data collection. However, there was not any evidence for the more substantial and longer term learning found for tasks such as for the jnd in fundamental frequency (e.g., Grimault *et al.*, 2002). The smaller learning effect here may be due to the fact that it is the statistical variation of the stimuli that limits performance in most conditions.

In Fig. 3 the data from all values of F_0 were pooled. We next sorted the data according to F_0 for the ranges from 200–280 Hz, 300–380 Hz, and 400–480 Hz. The sorted data

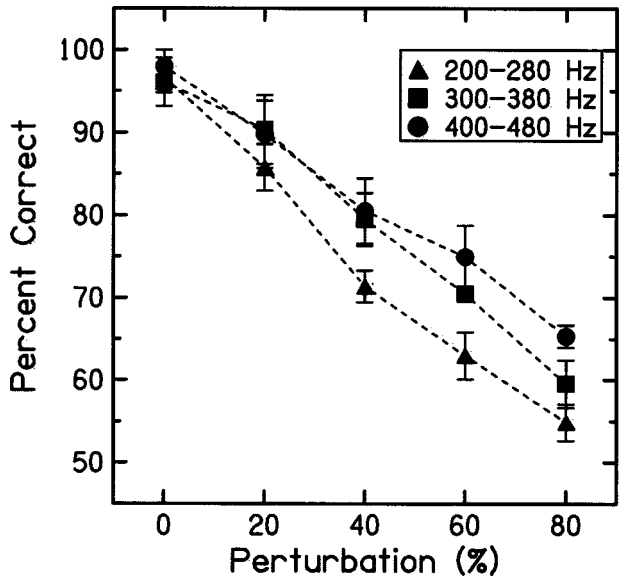


FIG. 4. Group mean percent-correct values and standard errors for the results from experiment 1. The different points and the line indicate partitioning of the data into three F_0 ranges: 200–280 (triangles); 300–380 (squares); and 400–480 (circles).

averaged across subjects are shown in Fig. 4.

For 0% perturbation, all mean scores were above 95% correct. For all nonzero values of perturbation, discrimination performance was poorest for the lowest F_0 range. The difference in discriminability for the various F_0 ranges increased as the amount of perturbation increased. These results are likely due to the denser spacing of the harmonics of the lower F_0 signals. The density of the spacing of the partials may affect the discriminability of harmonicity in several ways. First, the lower the F_0 the more likely that the low-frequency partials will interact, as the filters tend to be relatively broader at the very low frequencies. For example, rough estimates of equivalent rectangular bandwidth indicate that the ''auditory filter'' ranges from about 23% of the center frequency to about 16% of the center frequency over the range of F_0's tested here (e.g., Glasberg and Moore, 1990). Thus, interference among low-frequency partials would be greater for the lower F_0 range. Also, the strength of the pitch of the sounds likely was greater for the sounds with higher F_0's because they would be comprised of more resolved harmonics. It is also possible that roughness is a cue to discrimination. When the sounds are increasingly perturbed, the likelihood increases that the partials will fall close enough in frequency to produce audible beating of the envelope. For the lower F_0's, the closer spacing of the harmonics means that perturbation would produce frequency differences in the range likely to create the sensation of roughness due to the beating of the envelope. Thus, perturbing the signal for lower F_0's would cause the signals to sound more like the foils along the roughness dimension than for higher F_0's, conceivably reducing one cue to discrimination.

B. Experiment 2

Figure 5 displays the results from experiment 2. The percent correct obtained for each listener is plotted as a function of signal-to-masker ratio (S/M) specified as the level per component of signal and masker, for the three different numbers of maskers for both monotic and dichotic presentation conditions. A logistic function[2] was used to fit the data. Generally, the functions increase with increasing S/M in an orderly way. Table I contains the *m* (midpoint) and *k* (slope) values from the fitted functions and estimates of goodness of fit of the lines. Because of the findings from experiment 1 these data are also subdivided according to F_0 range.

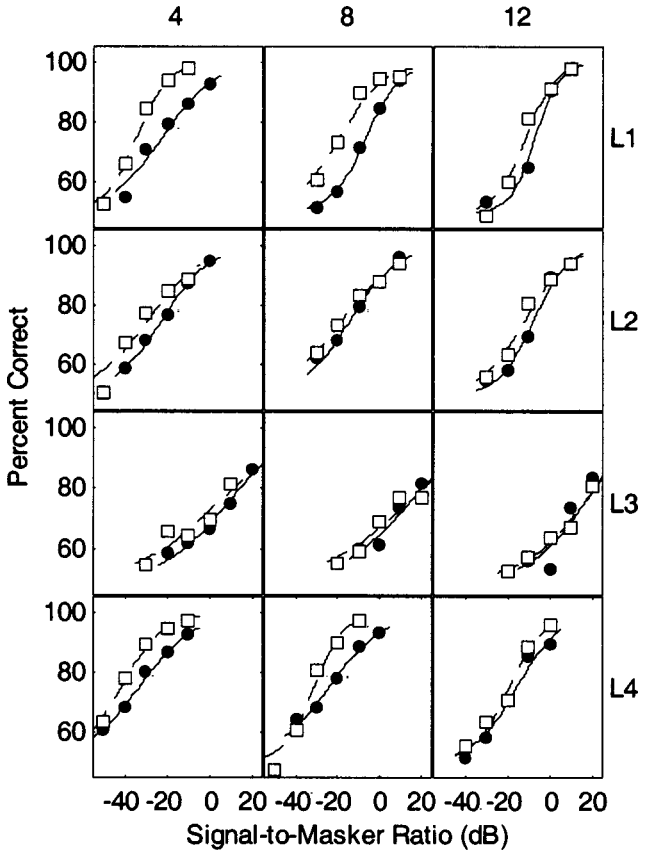


FIG. 5. Results from experiment 2. The abscissa is the signal-to-masker ratio in dB and the ordinate is percent-correct performance. The four subjects' data are contained in the four rows of panels. The three columns, reading left to right, are for conditions in which there were four, eight, or 12 masker tones. The open squares are for dichotic presentation and the filled circles are for monotic presentation. The lines are logistic fits to the data.

Note that the level of the individual masker tones was 10 dB higher for the four-tone masker than for the other two maskers, so the data will be discussed in terms of S/M rather than signal SPL. For monotic presentation, the group mean midpoints of the fitted functions computed across the entire range of F_0's were −17.6, −7.9, and −3.7 dB for the four-, eight-, and 12-tone maskers, respectively. The corresponding values for the dichotic conditions were −25.5, −13.9, and −6.9 dB. For both monotic and dichotic conditions, the same ordering of performance with respect to F_0 was found here as in experiment 1: as the F_0 increased, performance improved. This was somewhat surprising given that the lower F_0 signals would overlap the masker components less than the higher F_0 signals despite the logarithmic spacing of the distribution of masker tones. However, the tendency for the

TABLE I. Individual and group (means and standard errors) midpoint, m, and slope, k, parameters from the logistic function fits to the data. The values are reported for the two interaural presentation conditions, monotic (S_mM_m) and dichotic (S_mM_0) for the three fundamental frequency ranges. PV stands for the proportion of the variance accounted for by the fits.

		S_mM_m									S_mM_0								
		4			8			12			4			8			12		
$F0$	L	m	k	PV	m	k	PV	m	k	PV	m	k	PV	m	k	PV	m	k	PV
200	1	−15.1	0.06	0.88	2.0	0.12	0.96	−1.6	0.20	0.97	−30.9	0.10	0.99	−14.9	0.07	0.82	−8.6	0.16	0.90
	2	−17.0	0.07	0.94	−8.3	0.13	0.95	−5.5	0.13	0.97	−18.1	0.08	0.89	−10.9	0.08	0.95	−6.4	0.10	0.94
	3	10.6	0.07	0.98	13.6	0.07	0.96	19.6	0.09	0.81	13.2	0.05	0.51	17.7	0.06	0.74	18.3	0.10	0.90
	4	−29.1	0.08	0.92	−20.2	0.07	0.85	−11.7	0.10	0.95	−34.8	0.09	1.00	−27.5	0.09	0.90	−15.0	0.14	0.97
	Avg	−12.6	0.07	0.93	−3.2	0.10	0.93	0.2	0.13	0.93	−17.7	0.08	0.85	−8.9	0.07	0.85	−2.9	0.13	0.93
	(se)	(8.4)	(0.00)	(0.02)	(7.2)	(0.02)	(0.03)	(6.8)	(0.03)	(0.04)	(10.9)	(0.01)	(0.12)	(9.5)	(0.01)	(0.05)	(7.3)	(0.01)	(0.02)
300	1	−23.8	0.10	0.95	−11.1	0.14	0.97	−9.8	0.21	0.92	−34.7	0.15	0.97	−19.9	0.10	0.97	−12.5	0.11	0.98
	2	−22.4	0.10	0.98	−15.9	0.09	0.98	−6.8	0.14	0.91	−31.7	0.07	0.94	−22.0	0.08	0.95	−13.5	0.12	0.94
	3	9.9	0.06	0.62	16.8	0.05	0.90	15.2	0.08	0.91	0.7	0.03	0.51	13.3	0.06	0.79	11.2	0.03	0.70
	4	−33.5	0.10	0.96	−25.8	0.09	0.99	−20.4	0.13	0.97	−43.6	0.12	1.00	−32.1	0.19	1.00	−21.9	0.13	1.00
	Avg	−17.4	0.09	0.88	−9.0	0.09	0.96	−5.4	0.14	0.93	−27.3	0.09	0.85	−15.2	0.11	0.93	−9.2	0.10	0.90
	(se)	(9.5)	(0.01)	(0.09)	(9.1)	(0.02)	(0.02)	(7.5)	(0.03)	(0.01)	(9.7)	(0.03)	(0.12)	(9.9)	(0.03)	(0.05)	(7.1)	(0.02)	(0.07)
400	1	−26.2	0.11	0.87	−10.1	0.17	1.00	−10.2	0.17	0.93	−37.6	0.16	1.00	−21.5	0.13	1.00	−14.7	0.16	0.97
	2	−26.7	0.10	0.97	−20.8	0.06	0.95	−10.6	0.10	0.96	−36.4	0.09	0.86	−22.0	0.06	0.94	−16.0	0.09	0.94
	3	4.9	0.08	0.98	11.6	0.08	0.96	12.7	0.08	0.83	−1.0	0.10	0.95	7.2	0.05	0.98	13.6	0.11	0.85
	4	−39.4	0.05	0.96	−26.1	0.08	0.98	−14.9	0.08	0.99	−46.0	0.09	0.92	−32.7	0.17	0.93	−21.7	0.09	0.82
	Avg	−21.9	0.09	0.94	−11.3	0.10	0.97	−5.7	0.11	0.93	−30.3	0.11	0.93	−17.3	0.10	0.96	−9.7	0.11	0.89
	(se)	(9.4)	(0.01)	(0.02)	(8.3)	(0.02)	(0.01)	(6.2)	(0.02)	(0.04)	(10.0)	(0.02)	(0.03)	(8.6)	(0.03)	(0.02)	(7.9)	(0.02)	(0.04)
ALL	1	−22.5	0.08	0.96	−6.6	0.12	1.00	−7.2	0.18	0.97	−34.7	0.13	0.99	−19.4	0.09	0.96	−11.6	0.14	0.98
	2	−22.5	0.09	1.00	−14.3	0.09	0.99	−7.7	0.12	0.98	−29.1	0.08	0.92	−17.8	0.07	0.99	−12.0	0.10	0.98
	3	8.1	0.07	0.97	13.7	0.06	0.97	15.9	0.08	0.88	3.6	0.06	0.87	12.9	0.05	0.91	15.3	0.07	0.97
	4	−33.4	0.08	0.99	−24.3	0.08	0.99	−15.7	0.10	0.98	−41.8	0.10	0.99	−31.2	0.14	0.97	−19.3	0.12	0.98
	Avg	−17.6	0.08	0.98	−7.9	0.09	0.98	−3.7	0.12	0.95	−25.5	0.09	0.94	−13.9	0.09	0.96	−6.9	0.11	0.98
	(se)	(8.9)	(0.00)	(0.01)	(8.0)	(0.01)	(0.01)	(6.8)	(0.02)	(0.02)	(10.0)	(0.02)	(0.03)	(9.4)	(0.02)	(0.02)	(7.6)	(0.01)	(0.00)

lower F_0 signals to be less discriminable was maintained in the presence of the various maskers, possibly because the masker tones had a similar effect to perturbation in the first experiment. When present, they would be more likely to fall nearer a partial of the denser spectra, possibly causing audible beats. For the monotic condition, the group mean m values were −5.2, −10.6, and −13.0 dB, for the F_0 ranges of 200–280, 300–380, and 400–480 Hz, respectively. Corresponding values for dichotic presentation were −9.8, −17.2, and −19.1 dB. Note that L3, as in experiment 1, was much poorer than the other three listeners were, although her data generally followed the same trends as the other subjects.

An analysis of variance revealed that all three main factors: number of masker tones, monotic vs dichotic presentation, and F_0 range were statistically significant ($p<0.001$, $p<0.05$, and $p<0.0001$, respectively) predictors of the variation in the m values. None of the interactions was significant. An analysis of first-half vs second-half performance, collapsed over the levels tested, indicated a small (3.8%) but statistically nonsignificant (two-tailed t-test, $t=-2.75$, df=3, $p=0.07$) improvement in performance during data collection.

The individual and group mean values of m are illustrated in Fig. 6. The abscissa is the number of masker components and the ordinate is the m value expressed as S/M. The different symbols are for different subjects (same code as in Fig. 2), while the lines connect group mean values. For both monotic and dichotic conditions, the group mean values increase with increasing number of masker tones.

Figure 7 displays the slopes of the performance-level functions. As was the case for the midpoints of the functions,

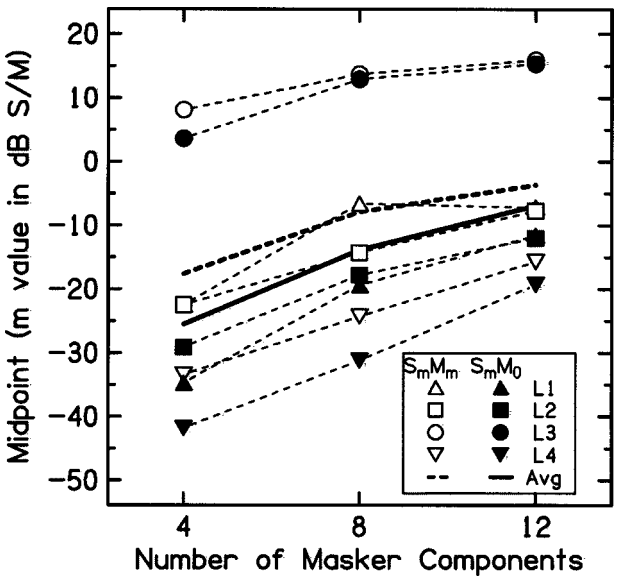


FIG. 6. Individual and group mean values of m (midpoint of the logistic function) for experiment 2. The abscissa is the number of masker tones and the ordinate is signal-to-masker ratio in dB. Symbols correspond to listeners as in Fig. 3. The heavy solid line represents group mean dichotic presentation results and the heavy dashed line represents group mean monotic presentation results.

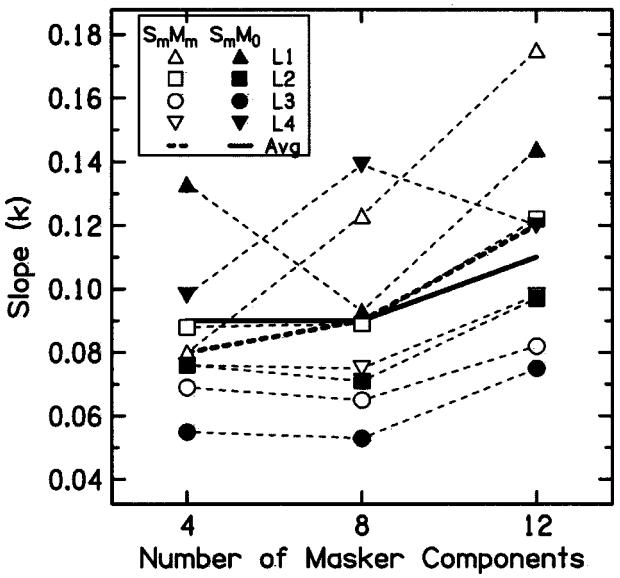


FIG. 7. Same as Fig. 6 except that the values plotted along the ordinate are the slopes (k's) from the fits to the data in experiment 2. The heavy solid line connects the group mean slope values for the S_mM_0 condition while the heavy dashed line connects the group mean slope values for the S_mM_m condition.

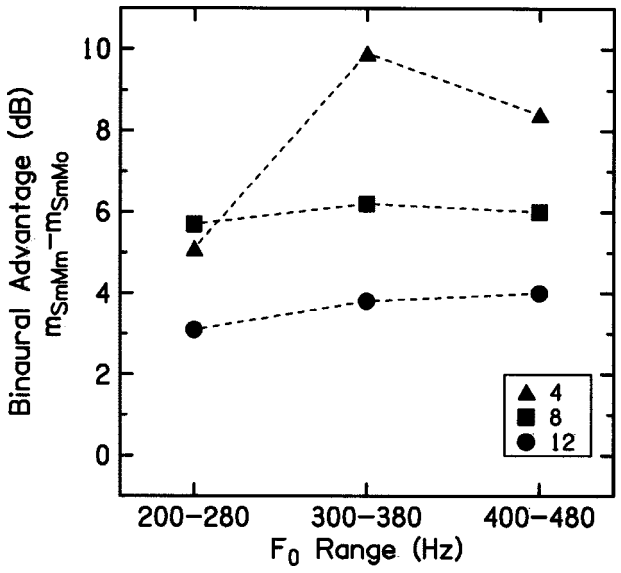


FIG. 8. The advantage of dichotic presentation (difference in m values for dichotic vs monotic presentation) plotted as a function of the F_0 range. The data are group means. The different symbols represent different numbers of tones in the masker as indicated in the symbol key.

the values of k also tended to increase as the number of tones in the masker increased, especially for monotic presentation.

Here, the individual data are plotted as points together with lines indicating group means. L3, who had the highest m values, had the lowest k values in all conditions, although the difference in slope was much less pronounced than in the midpoints. For the monotic condition, the slope of the psychometric function increased, on average, from about 0.08 for the four-tone masker to about 0.12 for the 12-tone masker. For the dichotic condition, the dependence of slope on the number of masker tones was less than for the monotic condition. An analysis of variance indicated that only masker type was a significant main variable ($p<0.05$).

Figure 8 shows the amount of improvement due to dichotic presentation relative to monotic presentation. The abscissa is the F_0 range and the ordinate is the advantage due to dichotic presentation as computed from the difference between group mean m values. For the 12-tone masker, the dichotic advantage varied between 3 and 4 dB across the different F_0 ranges. For the eight-tone masker, the dichotic advantage was nearly constant at about 6 dB. For the four-tone masker, the dichotic advantage was about 5 dB for the lowest F_0 range, nearly 10 dB for the middle F_0 range, and just over 8 dB for the highest F_0 range. It is not clear why the dichotic advantage for the lowest F_0 range of the four-tone masker should be the poorest of the three F_0 ranges. The individual differences in the amount of dichotic advantage were greatest here too, ranging from more than 15 dB for L1 to about 1 dB for L2. Large individual differences are common in informational masking experiments (e.g., Neff and Dethleffs, 1995; Durlach *et al.*, 2003b).

Following experiments 1 and 2, masked detection thresholds for the perfectly harmonic signals were measured for L1 and L3 (L2 and L4 were no longer available) in each of the three maskers. The thresholds were measured using a 2I2AFC adaptive tracking procedure that estimates the 70.7%-correct point on the psychometric function (Levitt, 1971). The average thresholds monaurally (specified as level per component re: masker level per component) were -35.4, -18.6, and -11.1 dB for the signals in the four-, eight-, and 12-tone maskers, respectively. The average m values for those subjects in the corresponding conditions were thus 28.3, 22.2, and 15.5 dB higher than the masked thresholds. As in the other conditions, L3's performance was poorer than the other subjects, and her masked detection thresholds averaged about 16.5 dB higher than L1. However, because the S/M ratio required to achieve comparable discrimination performance (m values as seen in Fig. 6) was also much higher for L3 than L1, her m values re: thresholds were only about 8 dB higher on average. It is of interest to note that there typically are large intersubject differences found in the susceptibility to informational masking (cf. Neff and Dethleffs, 1995; Durlach *et al.*, 2003b) and that L3 may be considered a ''high threshold'' listener. The fact that she was poorer in experiment 1 is also interesting but we cannot draw any general conclusions based on the trend from a single subject. The results above suggest that the signals and foils were presented well above masked detection threshold for most of the levels tested in the second experiment.

IV. DISCUSSION

The way in which the degree of harmonicity was varied in the present study produced orderly discrimination results. Perceptually, the stimuli appeared to vary in the strength of the impression of harmonicity: as the frequencies of the harmonics were increasingly perturbed, the stimuli sounded less harmonic. The variation in the perception of harmonicity with these stimulus manipulations probably has at least two dimensions, pitch and roughness. Generally, the signal would tend to have a stronger pitch and less roughness than the foil. As the harmonics comes closer to the exact integer multiples

of the fundamental, the pitch strength increases. As the components are increasingly perturbed, instances of tones falling close enough in frequency to cause audible beating increases causing the signals to be more similar to the foils. A third possibility that must be considered as an explanation for these results is that the listeners were discriminating between signal and foil by using differences in spectral shape, i.e., profile analysis (e.g., Green, 1988). Although the acoustic stimuli had equal-level components in both signal and foil, the difference in the regularity of the spacing of the tones would produce a difference in the regularity of the internal representation of the sound spectra. Harmonic sounds would be expected to produce a generally flatter excitation pattern with regularly spaced peaks in the lower frequency region with a smoother shape in the higher frequencies where the harmonics fall closer together (cf. Glasberg and Moore, 1990). In contrast, the irregular frequency spacing of the foils would be expected to produce larger peaks and valleys in the excitation pattern. Thus, in addition to differences in pitch and roughness, the listeners could have been using information about the shape of the spectral profile in basing their decisions.

A frequency dependence was observed in the current results such that poorer performance overall was consistently observed for the lowest fundamental frequency range of 200–280 Hz. This was apparent for all of the perturbed F_0 conditions and all of the masked conditions. In both cases, F_0 perturbation and added masker tones, the explanation for the poorer performance at low frequencies may be related to the greater likelihood of the signal containing audible beats that sounded subjectively like the foil. This would be especially true for the higher harmonics of the low F_0 signals (e.g., Moore *et al.*, 1984). Other types of interactions, such as partial masking due to overlapping patterns of excitation and decreased pitch strength due to fewer resolved partials, may also affect performance more for the lower fundamental frequencies than for the higher fundamental frequencies. The notion that the partials of the lower F_0 signals will interact more at lower frequencies than for higher F_0 signals is consistent with masking data indicating that the tuning of the auditory system is relatively broader at low frequencies than at higher frequencies (e.g., Kidd and Feth, 1981) and with computational models of auditory filter shape (e.g., Glasberg and Moore, 1990).

In masked conditions, the masking that occurred was likely a combination of both energetic and informational masking. Unlike most previous studies of informational masking, except for some recent studies using multiple talkers (Freyman *et al.*, 1999, 2001; Brungart, 2001; Brungart *et al.*, 2001; Arbogast *et al.*, 2002), here the signals and maskers overlapped considerably in the frequency domain. For a given frequency range, we would expect that the proportion of the total amount of masking attributable to energetic masking would increase as the number of masker tones increased and the spectral overlap of signal and masker increased. Consistent with earlier work (e.g., Kidd *et al.*, 1998; Wright and Saberi, 1999; Freyman *et al.*, 1999; Brungart, 2001; Kidd *et al.*, 2002), changes in the proportion of energetic/informational masking affects the slopes of the psychometric functions and thus the range of levels over which masking occurs.

In the present study the slopes of the functions were steeper and the midpoints of the functions lay nearer to masked detection thresholds as the number of masker components, and presumably the concomitant amount of energetic masking, increased. Further, the fact that the task is suprathreshold discrimination, averaging about 15–30 dB above detection threshold at the midpoints of the functions, suggests that the binaural advantage is largely due to perceptual processes rather than to traditional binaural analysis mechanisms, such as that described by the equalization-cancellation model (Durlach, 1972; also, Colburn and Durlach, 1978). The interpretation of the current results is that the primary advantage of dichotic presentation was to strengthen the listener's perception of the signal/foil as separate objects from the maskers improving the ability of the listener to focus attention on, and extract information from, the target sound. When the sounds were presented monotically, the signal/foil and masker tended to fuse such that only a single auditory image was formed. When masker and signal formed a single image, the listener had to determine that a subset of the frequencies were harmonically related, extracting those frequencies from other unrelated tones. Our data suggest that adding even a few random-frequency components can greatly disrupt the computation of harmonicity. However, when the signal and masker were two images, the focus of attention could be directed toward the signal, and determining the harmonic relationship among the tones forming that object was much easier. Consider first the results from the 12-tone masker. Because of the relatively high degree of spectral overlap between the masker and the signal/foil, we would expect that energetic masking would dominate the masking that occurred. A small dichotic presentation advantage was observed. The small advantage may have been due to the fact that the signal/foil were presented well above detection threshold where binaural analysis is thought to decline (e.g., Henning, 1973; Yost, 1997). Because binaural analysis improves the S/M in the channels containing signal and noise, it effectively reduces the energetic masking that is present. Thus, only a small dichotic advantage would be expected. At the other extreme, the four-component masker would produce much less spectral overlap, on average, with the signal/foil so that most of the masking that occurred was likely informational masking. In contrast to energetic masking, informational masking is thought to be very effective in interfering with suprathreshold tasks (e.g., Kidd *et al.*, 1998; Freyman *et al.*, 1999, 2001; Brungart, 2001; Kidd *et al.*, 2002; Arbogast *et al.*, 2002). Thus, for a condition in which informational masking dominates, perceptual segregation of the signal from the masker resulting from dichotic presentation would be expected to be very effective in improving performance. In this case, the reduction in informational masking attributable to perceptual segregation facilitated the ability of the listener to extract the property of harmonicity from a suprathreshold stimulus.

Beerends and Houtsma (1989) examined pitch identification for pairs of simultaneously presented two-tone complexes that were drawn from harmonic series having differ-

ent fundamental frequencies. They compared identification performance when the four tones (two partials from each of the two harmonic series) were presented diotically with conditions where the four tones were distributed across the ears in three dichotic combinations. Somewhat surprisingly, given the current results, diotic performance was nearly as good as for the dichotic conditions as long as some allowance was made for cases where the partials from different F_0's fell too near in frequency. Extrapolating to the current results, it might have been predicted that little advantage would be found for our S_mM_0 condition relative to S_mM_m in discriminating the more harmonic of two sounds. However, in our conditions, the uncertainty introduced by the frequency randomization, for the four-component masker in particular, would likely be much greater than was the case for the tone pairs that were subsets of the two harmonic series. We would argue that our stimuli created considerable informational masking in the monaural case that then was reduced when the masker and signal/foil images were separated interaurally by dichotic presentation. Although Beerends and Houtsma (1989) do attribute some of the performance differences to ''central interference'' (p. 818), the uncertainty inherent in their experiment was probably much less than here. Further, the precise way that dichotic information is presented can have a significant effect on whether the sound in the ear opposite a target is beneficial or harmful. For example, Darwin (1992) has shown that playing a mistuned partial to the ear contralateral to the remaining tones in a harmonic complex has much less of an influence on the pitch of the complex than if it is played in the same ear. In more complex listening environments, though, such as the case where a target talker is played with a masking talker in one ear and an unrelated talker is added to the opposite ear, contralateral stimulation can exert an equal or even greater interfering effect than the same stimulus played ipsilaterally (Brungart and Simpson, 2002). And, recent work has indicated that the ''contralateral masking effect'' found by Brungart and Simpson (2002) for speech recognition appears not to be restricted to speech maskers or speech signals, but instead is a more general perceptual phenomenon (e.g., Kidd *et al.*, 2003).

The conclusion above regarding the mechanism responsible for the dichotic advantage found here is consistent with that from a recent study reported by Arbogast *et al.* (2002). In their work, conducted in a mildly reverberant sound field, much greater improvements in S/M from spatial separation of sources in a speech recognition task were observed for predominantly informational maskers than for predominantly energetic maskers. In addition, for the condition producing the greatest amount of informational masking, spatially separating the target speech sound from the masker also significantly increased the slope of the psychometric functions, resulting in greater advantages at high, as compared to low, signal-to-masker ratios. The steeper slope for the spatially separated case, combined with the reduction in the amount of masking, was thought to indicate a shift toward a stronger influence of energetic masking relative to the nonseparated case. In that study, as here, there was no evidence for the sort of plateau or nonmonotonicity reported in some other speech masking studies (e.g., Egan *et al.*, 1954; Dirks and Bower, 1969; Freyman *et al.*, 1999; Brungart, 2001).

Oh and Lutfi (2000) have also examined the relationship between harmonicity and informational masking, although in a different context than that of the present study. They measured detection threshold for a tone of known frequency when embedded in a randomized multitone masker that was comprised of tones that were from a harmonic series (''harmonic masker'') or were inharmonic. In the harmonic masker, the signal could be either harmonically or inharmonically related to the other tones. Their hypothesis was that, in particular, less informational masking would occur for an inharmonic signal in the harmonic masker than in the inharmonic masker because the signal would perceptually segregate more easily from a harmonic complex than from an inharmonic complex. Their results supported that hypothesis. In the monotic masked conditions of the present study, it probably is more accurate to characterize the perception of the listener as that of hearing a single auditory object that had a stronger sense of harmonicity when the signal was added to the masker than when the foil was added to the masker. However, in the dichotic conditions producing large performance advantages, the signal clearly was heard separately from the masker, and it is likely that it was the combination of both the harmonicity of the signal and the dichotic presentation that led to the sense of perceptual segregation.

The present study shares some similarities with the work of de Cheveigne *et al.* (1995, 1997) on the role of harmonicity in the identification of concurrent vowels. As in this study, adding random frequency perturbation to the individual partials of multitone complexes produced the harmonic–inharmonic distinction. Also, as in experiment 2 of the present study, two sounds were played simultaneously with the listener attempting to extract some feature of one or both sounds. In their work the feature to be extracted was the spectral pattern of formants signaling vowel identity with harmonicity used as a means for segregation, whereas here it was the harmonicity of the sounds *per se* that was the feature to be extracted. There are other important differences, as well. First, the perturbation used by de Cheveigne *et al.* was proportional to frequency rather than applied equally to each partial as was the case here. The effect of the different ways of perturbing the signals would largely be that the low-frequency tones in our stimuli would have been perturbed much more than those in de Cheviegne *et al.*'s inharmonic vowels, likely causing greater disruption in pitch and more roughness in our stimuli. Second, the vowel sounds were shaped by spectral envelopes extracted from actual speech rather than being comprised of equal-level tones as in this study. Although, as discussed above, we expect that there may be sufficient information in the difference in the internal representation of spectral shape to discriminate between signal and foil, the differences in the spectral patterns of actual vowel sounds would be much greater. Finally, the task of concurrent vowel identification would seem to be fundamentally different than the 2AFC discrimination procedure used in this study. Correctly labeling speech tokens according to phoneme identity is presumably a much more sophisticated

task than simply choosing which of two sounds had a stronger pitch, less roughness, and/or a more regular spectral shape. It is interesting to note, though, that de Cheveigne *et al*. found that inharmonic vowels were at least as identifiable as harmonic vowels, perhaps suggesting that our foils may have conveyed as much information in solving the discrimination task as did the signals.

V. CONCLUSIONS

Human listeners can accurately and reliably discriminate between sounds on the basis of the degree of harmonicity. Perceptually, pitch, roughness, and spectral shape may contribute cues toward this discrimination, at least for these experimental conditions. Informational maskers may mask the perception of the degree of harmonicity even for sounds presented well above detection threshold. As in other studies of informational masking, stimulus manipulations which promote the perceptual segregation of the signal from the masker—in this case dichotic presentation—significantly improve discrimination performance.

ACKNOWLEDGMENTS

This work was supported by Grants DC00100 and DC04545 from NIH/NIDCD and by the ONR-managed MURI Grant Z883402. The authors also wish to thank Erin Ellis for her assistance on this project, and are grateful to Peter Cariani, Christopher Darwin, and an anonymous reviewer for comments on an earlier version of this manuscript.

[1]"Maximally inharmonic" is accurate only within the context of the current stimulus generation procedure that is constrained by design considerations. If there were no need to control for factors such as stimulus duration and the number, range, and density of the distributions of frequencies comprising the signals and foils, the sounds arguably could be made "less harmonic."

[2]The lines fit to the data took the form $p(c)=\alpha+[(1.0-\alpha)/(1.0+e^{-k(m-x)})]$, where $\alpha=0.5$ is chance performance, m is the level corresponding to the midpoint of the function (about 75% correct), and k is the slope. The best-fitting values of k and m were obtained using the FMINS routine in MATLAB (product of Mathworks, Inc., Cambridge, MA).

Arbogast, T. L., Mason, C. R., and Kidd, Jr., G. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Assman, P. F., and Summerfield, Q. (**1990**). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," J. Acoust. Soc. Am. **88**, 680–697.

Beerends, J. G., and Houtsma, A. J. M. (**1989**). "Pitch identification of simultaneous diotic and dichotic two-tone complexes," J. Acoust. Soc. Am. **85**, 813–819.

Bird, J., and Darwin, C. J. (**1998**). "Effects of a difference in fundamental frequency in separating two sentences," *in Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 263–269.

Brokx, J. P. L., and Noteboom, S. G. (**1982**). "Intonation and the perceptual segregation of simultaneous voices," J. Phonetics **10**, 23–36.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

Brungart, D. S., and Simpson, B. D. (**2002**). "Within-ear and across-ear interference in a cocktail party listening task," J. Acoust. Soc. Am. **112**, 2985–2995.

Carlyon, R. P. (**1994**). "Detecting mistuning in the presence of synchronous and asynchronous interfering sounds," J. Acoust. Soc. Am. **95**, 2622–2630.

Colburn, H. S., and Durlach, N. I. (**1978**) "Models of binaural interaction," in *Handbook of Perception, Vol. IV Hearing* (Academic, New York), pp. 467–518.

Culling, J. F., and Darwin, C. J. (**1993**). "Perceptual separation of simultaneous vowels: Within and across formant grouping by F_0," J. Acoust. Soc. Am. **93**, 3454–3467.

Culling, J., and Summerfield, Q. (**1995**). "The role of frequency modulation in the perceptual segregation of concurrent vowels," J. Acoust. Soc. Am. **98**, 837–846.

Darwin, C. J. (**1981**). "Perceptual grouping of speech components differing in fundamental frequency and onset-time," Q. J. Exp. Psychol. A **33**, 185–207.

Darwin, C. J. (**1992**). "Listening to two things at once," in *The Auditory Processing of Speech*, edited by M. E. H. Schouten (Mouton and Gruyer, Berlin), pp. 133–147.

Darwin, C. J., and Carlyon, R. P. (**1995**). "Auditory Grouping," in *Hearing: From Handbook of Auditory Perception*, 2nd ed. edited by B. C. J. Moore (Academic, London), pp. 387–424.

de Cheveigne, A., McAdams, S., Laroche, J., and Rosenberg, M. (**1995**). "Identification of concurrent and inharmonic vowels: A test of the theory of harmonic cancellation and enhancement," J. Acoust. Soc. Am. **97**, 3736–3748.

de Cheveigne, A., McAdams, S., and Marin, C. M. H. (**1997**). "Concurrent vowel identification. II. Effects of phase, harmonicity, and task," J. Acoust. Soc. Am. **101**, 2848–2856.

Dirks, D. D., and Bower, D. R. (**1969**). "Masking effects of speech competing messages," J. Speech Hear. Res. **12**, 229–245.

Durlach, N. I. (**1972**). "Binaural detection: Equalization and cancellation theory," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. 2.

Durlach, N. I., Mason, C. R., Kidd, Jr., G., Arbogast, T. L., Shinn-Cunningham, B. C., and Colburn, H. S. (**2003a**). "Note on informational masking," J. Acoust. Soc. Am. (in press).

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B., Arbogast, T. L., Colburn, H. S., and Kidd, Jr., G. (**2003b**). "Informational masking: Reducing the effects of uncertainty by increasing target–masker dissimilarity" J. Acoust. Soc. Am. **113**, 2984–2987.

Egan, J. P., Carterette, E. C., and Thwing, E. J. (**1954**). "Some factors affecting multi-channel listening," J. Acoust. Soc. Am. **26**, 774–782.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.

Gardner, R. B., Gaskill, S. A., and Darwin, C. J. (**1989**). "Perceptual grouping of formants with static and dynamic differences in fundamental frequency," J. Acoust. Soc. Am. **85**, 1329–1337.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Green, D. M. (**1988**). *Profile Analysis Auditory Intensity Discrimination* (Oxford University Press, Oxford).

Grimault, N., Micheyl, C., Carlyon, R. P., and Collet, L. (**2002**). "Evidence for two pitch encoding mechanisms using a selective auditory training paradigm," Percept. Psychophys. **64**, 189–197.

Hartmann, W. M., McAdams, S., and Smith, B. K. (**1990**). "Matching the pitch of a mistuned harmonic in an otherwise periodic complex tone," J. Acoust. Soc. Am. **88**, 1712–1724.

Henning, G. B. (**1973**). "Effect of interaural phase on frequency and amplitude discrimination," J. Acoust. Soc. Am. **54**, 1160–1178.

Kidd, Jr., G., and Feth, L. L. (**1981**). "Patterns of residual masking," Hear. Res. **5**, 49–67.

Kidd, Jr., G., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (**1994**). "Reducing informational masking by sound segregation," J. Acoust. Soc. Am. **95**, 3475–3480.

Kidd, Jr., G., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (**1998**). "Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns," J. Acoust. Soc. Am. **104**, 422–431.

Kidd, Jr., G., Mason, C. R., and Arbogast, T. L. (**2002**). ''Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns,'' J. Acoust. Soc. Am. **111**, 1367–1376.
Kidd, Jr., G., Mason, C. R., Arbogast, T. L., Brungart, D. S., and Simpson, B. D. (**2003**). ''Informational masking caused by contralateral stimulation,'' J. Acoust. Soc. Am. **113**, 1594–1603.
Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
McAdams, S. E. (**1989**). ''Segregation of concurrent sounds: Effects of frequency modulation coherence and a fixed resonance structure,'' J. Acoust. Soc. Am. **86**, 2148–2159.
Moore, B. C. J., Goldberg, B. R., and Shailer, M. J. (**1984**). ''Frequency and intensity difference limens for harmonics within complex tones,'' J. Acoust. Soc. Am. **75**, 550–561.
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.
Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1986**). ''Thresholds for hearing mistuned partials as separate tones in harmonic complexes,'' J. Acoust. Soc. Am. **80**, 479–483.
Neff, D. L. (**1995**). ''Signal properties that reduce masking by simultaneous, random-frequency maskers,'' J. Acoust. Soc. Am. **98**, 1909–1920.
Neff, D. L., and Green, D. M. (**1987**). ''Masking produced by spectral uncertainty with multicomponent maskers,'' Percept. Psychophys. **41**, 409–415.
Neff, D. L., and Dethleffs, T. M. (**1995**). ''Individual differences in simultaneous masking with random-frequency, multicomponent maskers,'' J. Acoust. Soc. Am. **98**, 125–134.
Neff, D. L., Dethleffs, T. M., and Jesteadt, W. (**1993**). ''Informational masking for multicomponent maskers with spectral gaps,'' J. Acoust. Soc. Am. **94**, 3112–3126.
Oh, E., and Lutfi, R. A. (**1998**). ''Nonmonotonicity of informational masking,'' J. Acoust. Soc. Am. **104**, 3489–3499.
Oh, E. L., and Lutfi, R. A. (**2000**). ''Effect of masker harmonicity on informational masking,'' J. Acoust. Soc. Am. **108**, 706–709.
Richards, V. M., Tang, Z., and Kidd, Jr., G. (**2002**). ''Informational masking with small set sizes,'' J. Acoust. Soc. Am. **111**, 1359–1366.
Roberts, B., and Brunstrum, J. M. (**1998**). ''Perceptual segregation and pitch shifts of mistuned components in harmonic complexes and in regular inharmonic complexes,'' J. Acoust. Soc. Am. **104**, 2326–2338.
Scheffers, M. T. (**1983**). ''Sifting vowels: Auditory pitch analysis and sound segregation,'' doctoral dissertation, University of Groningen, The Netherlands.
Summerfield, Q. (**1992**). ''Role of harmonicity and coherent frequency modulation in auditory grouping,'' in *The Auditory Processing of Speech*, edited by M. E. H. Schouten (Mouton and Gruyer, Berlin), pp. 157–166.
Watson, C. S., and Kelly, W. J. (**1981**). ''The role of stimulus uncertainty in the discriminability of auditory patterns,'' in *Auditory and Visual Pattern Recognition*, edited by D. J. Getty and J. H. Howard, Jr. (Erlbaum, Hillsdale, NJ), pp. 37–59.
Watson, C. S. (**1987**). ''Uncertainty, informational masking and the capacity of immediate auditory memory,'' in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ), pp. 267–277.
Weir, C. C., Jesteadt, W., and Green, D. M. (**1977**). ''Frequency discrimination as a function of frequency and sensation level,'' J. Acoust. Soc. Am. **61**, 178–184.
Wright, B. A., and Saberi, K. (**1999**). ''Strategies used to detect auditory signals in small sets of random maskers,'' J. Acoust. Soc. Am. **105**, 1765–1775.
Yost, W. A. (**1991**). ''Auditory image perception and analysis,'' Hear. Res. **56**, 8–18.
Yost, W. A. (**1997**). ''The cocktail party problem: Forty years later,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. A. Gilkey and T. R. Anderson (Erlbaum, Hillsdale, NJ), pp. 329–348.

Louder sounds can produce less forward masking: Effects of component phase in complex tones[a)]

Hedwig Gockel[b)]
CNBH, Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, England

Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

Roy D. Patterson
CNBH, Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, England

Ray Meddis
CNBH at Essex, Department of Psychology, University of Essex, Colchester CO4 3SQ, England

(Received 11 October 2002; revised 23 May 2003; accepted 30 May 2003)

The influence of the degree of envelope modulation and periodicity on the loudness and effectiveness of sounds as forward maskers was investigated. In the first experiment, listeners matched the loudness of complex tones and noise. The tones had a fundamental frequency (F0) of 62.5 or 250 Hz and were filtered into a frequency range from the 10th harmonic to 5000 Hz. The Gaussian noise was filtered in the same way. The components of the complex tones were added either in cosine phase (CPH), giving a large crest factor, or in random phase (RPH), giving a smaller crest factor. For each F0, subjects matched the loudness between all possible stimulus pairs. Six different levels of the fixed stimulus were used, ranging from about 30 dB SPL to about 80 dB SPL in 10-dB steps. Results showed that, at a given overall level, the CPH and the RPH tones were louder than the noise, and that the CPH tone was louder than the RPH tone. The difference in loudness was larger at medium than at low levels and was only slightly reduced by the addition of a noise intended to mask combination tones. The differences in loudness were slightly smaller for the higher than for the lower F0. In the second experiment, the stimuli with the lower F0s were used as forward maskers of a 20-ms sinusoid, presented at various frequencies within the spectral range of the maskers. Results showed that the CPH tone was the least effective forward masker, even though it was the loudest. The differences in effectiveness as forward maskers depended on masker level and signal frequency; in order to produce equal masking, the level of the CPH tone had to be up to 35 dB above that of the RPH tone and the noise. The implications of these results for models of loudness are discussed and a model is presented based on neural activity patterns in the auditory nerve; this predicts the general pattern of loudness matches. It is suggested that the effects observed in the experiments may have been influenced by two factors: cochlear compression and suppression.
 [DOI: 10.1121/1.1593065]

PACS numbers: 43.66.Cb, 43.66.Dc, 43.66.Nm [MRL]

I. INTRODUCTION

Most models of loudness are based on the power spectrum of the sound of interest (Fletcher and Munson, 1933; Zwicker, 1958; Zwicker and Scharf, 1965; Stevens, 1972; Moore *et al.*, 1997). For example, the loudness models presented by Zwicker and Scharf (1965) and by Moore *et al.* (1997) first derive an excitation pattern from the power spectrum of the sound. Then, using a modified power law transformation, a specific loudness pattern is calculated from the excitation pattern. The overall loudness of the sound is assumed to be proportional to the area under the specific loudness pattern. For steady sounds, these models give good predictions. However, there is evidence that for fluctuating sounds they may be inadequate. While the data are not entirely consistent across studies, recent studies suggest that, for narrowband sounds, amplitude modulation at medium rates (from 10 Hz up to about 100 Hz) results in a slight decrease of loudness relative to that of an unmodulated sound (Bauch, 1956; Fastl, 1975; Hellman, 1985; Zhang and Zeng, 1997; Moore *et al.*, 1998, 1999; Grimm *et al.*, 2002). The difference in loudness corresponds approximately to a change in level of about 1 dB. For broadband sounds the effect goes in the opposite direction; introducing amplitude modulation at medium rates increases the loudness of the sound slightly (Zhang and Zeng, 1997; Moore *et al.*, 1999;

[a)]Portions of these results were presented at the symposium on ''Psychoacoustics, physiology and models of the central auditory system'' [Gockel *et al.*, Acta Acust. Acust. **88**, 369–377 (2002b)].

[b)]Current address: MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, England. Electronic mail: hedwig.gockel@mrc-cbu.cam.ac.uk

Grimm *et al.*, 2002). The model proposed by Glasberg and Moore (2002), which is based on the calculation of excitation patterns from the short-term spectrum of the stimulus, can predict the effect for narrowband sounds, but not for broadband sounds.

Another psychophysical measure assumed to reflect the excitation evoked in the auditory system is forward masking. The threshold of a sinusoidal signal in forward masking is assumed to be monotonically related to the excitation evoked by the masker at the signal frequency (Houtgast, 1974; Moore and Glasberg, 1983; Moore, 1997; Plack and Oxenham, 1998). As Carlyon and Datta (1997) pointed out, the excitation produced by a stimulus with a temporally fluctuating envelope is likely to be affected by fast-acting compression in the cochlea (Robles *et al.*, 1986; Ruggero, 1992; Ruggero *et al.*, 1997; Recio *et al.*, 1998). Carlyon and Datta (1997) compared the amount of forward masking produced by two harmonic tone complexes, one with components added in Schroeder-positive phase (S+), which is thought to evoke a ''peaky'' waveform on the basilar membrane (BM), and the other with components added in Schroeder-negative phase (S−), which produces a less peaky waveform. The S+ complex produced less forward masking than the S− complex of a sinusoidal signal with frequency corresponding to one of the central harmonics. This was explained by fast-acting compression which would result in lower average excitation for the peaky waveform evoked by the S+ complex than for the less peaky waveform evoked by the S− complex.

Carlyon and Datta (1997) also compared the loudness of the S+ and S− complexes, or rather of a mid-range subset of their components; the subset was turned on after the other components. The results showed that the subset was louder for the S− than for the S+ complex. Thus, both measures, loudness and ''effectiveness as a forward masker,'' were consistent with the idea that the S− complex evokes more excitation than the S+ complex.

In this paper, we present two experiments designed to further our understanding of the relationship between the loudness of *broadband* sounds and their effectiveness as forward maskers. Using broadband stimuli with the same power spectrum, but different crests factors (ratio of peak amplitude to root-mean-square, rms, amplitude), we addressed the following questions: (1) Which is louder? (2) Which produces more forward masking?

In contrast to the study of Carlyon and Datta (1997), it was the loudness of the whole sound that was compared instead of a subset of the components. We would not, however, expect this to change the direction of the effects of compression on loudness and forward masking. We have previously presented results (Gockel *et al.*, 2002b) indicating that, for bandpass filtered complex tones with a low F0 (62.5 Hz), adding the components in cosine phase (large crest factor) led to greater loudness values than adding the components in random phase (small crest factor); this appears inconsistent with the idea that a large crest factor leads to lower excitation and that this in turn leads to lower loudness. It also appears inconsistent with the results of Carlyon and Datta (1997). We also found (Gockel *et al.*, 2002b) that the random phase complex tone had slightly greater loudness than a random noise that was bandpass filtered to produce a similar excitation pattern. However, the difference was not statistically significant.

Here, we present results for more subjects to increase the statistical power. We also present results for both a low F0 (62.5 Hz) and a higher F0 (250 Hz). In a previous experiment (Gockel *et al.*, 2002a) we showed that, for the 62.5-Hz F0, a cosine-phase complex tone was a much less effective simultaneous masker of a noise (filtered into the same frequency region) than the noise was as a masker of the complex tone. This effect was much reduced when the F0 was 250 Hz. We argued that the greater effect for the lower F0 was partly caused by the greater peak factor of the waveforms evoked on the BM by the cosine-phase tone at the lower F0 (as, for a given center frequency, the number of components within the passband of the auditory filter is four times greater for the low F0 than for the high F0). Based on the arguments presented above, one might expect the peak factor to influence loudness. That expectation was assessed here.

II. EXPERIMENT 1

A. Method

1. Rationale and stimuli

The first experiment investigated the influence of the degree of envelope modulation and periodicity on loudness. Listeners matched the loudness of complex tones and noise. The stimuli were the same as those used as simultaneous maskers by Gockel *et al.* (2002a). The complex tones had a fundamental frequency (F0) of 62.5 or 250 Hz and were bandpass filtered between the 10th harmonic and 5000 Hz (3-dB down points, 100 dB/oct slope). The Gaussian noise was filtered in the same way as the complex tone with which it was to be compared in loudness. The components of the complex tones were added either in cosine phase (CPH), giving a large crest factor, or in random phase (RPH), giving a small crest factor. For each F0, subjects were presented with every possible stimulus pair, resulting in three stimulus combinations: CPH vs. RPH, CPH vs. noise, and RPH vs. noise. The long-term excitation patterns of the stimuli, calculated according to the procedure of Glasberg and Moore (1990), were essentially identical within each F0 condition (see Fig. 1 of Gockel *et al.*, 2002a). Thus, for a given F0, loudness models based on the long-term excitation pattern would not predict any differences in loudness between these stimuli. In contrast, models specifically taking into account the effects of fast-acting compression would be expected to predict that the CPH tone will be less loud than the RPH tone or the noise. This is discussed in more detail below.

In the loudness-matching procedure, one stimulus was fixed in level and the other was varied in level. The overall rms level of the fixed stimulus varied from 30 to 80 dB SPL in 10-dB steps for the 62.5-Hz F0, and from 27.6 dB SPL to 77.6 dB SPL in 10-dB steps for the 250-Hz F0. The 2.4-dB level adjustment between the two F0s was included in order to keep the level per ERB constant (Glasberg and Moore,

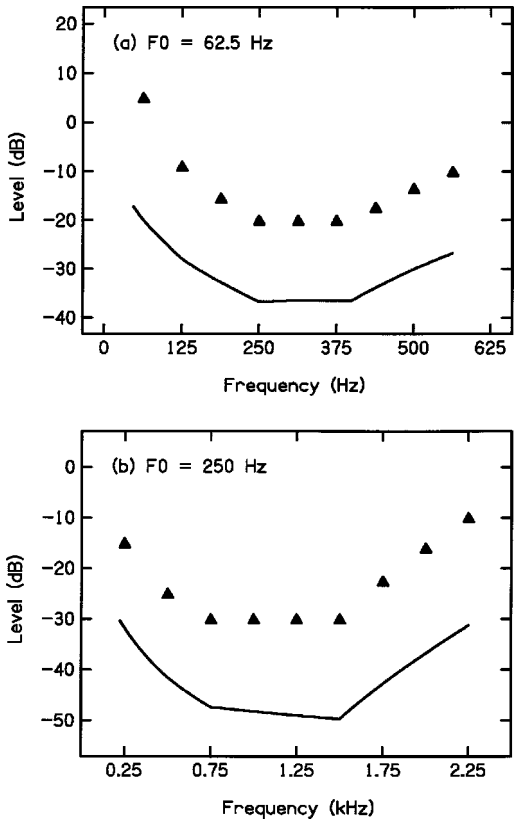


FIG. 1. Derivation of the spectral shape of the masking noise used in experiment 1 for the 62.5-Hz F0 (top) and the 250-Hz F0 (bottom). The filled upward-pointing triangles show the assumed levels of the tones needed to cancel quadratic and cubic distortion products, based on the data of Pressnitzer and Patterson (2001). The solid line shows the derived spectral density of the masking noise which was used to mask the combination tones.

1990); with this adjustment, the level per component within the passband was 6 dB lower for F0 = 62.5 Hz than for F0 = 250 Hz.

There is evidence that the strength of combination tones can depend on the phase of the primary components (Buunen *et al.*, 1974; Pressnitzer and Patterson, 2001). This could lead to differences in loudness between our stimuli. To test this possibility, in half of the conditions a noise was used to mask combination tones. Figure 1 shows the spectral shape of this noise for the 62.5-Hz F0 (top) and the 250-Hz F0 (bottom). The strongest combination tones were assumed to be quadratic distortion tones occurring at F0 and low harmonics of F0, and cubic distortion tones which are most intense for frequencies just below the lower edge of the passband of the stimulus (Goldstein, 1967; Greenwood, 1971; Zwicker and Fastl, 1999). The shape and level of the masking noise relative to the level of the components in the target tones were based partly on the measurements of Pressnitzer and Patterson (2001) of the level of a sinusoidal tone required to cancel combination tones produced by harmonic CPH tones. The filled, upward-pointing triangles in Fig. 1 show the estimated cancellation tone level relative to the level per component in our CPH tones as a function of frequency. These relative levels are applicable for moderate levels of the primary components. The continuous curve in Fig. 1 shows the derived spectrum level of the masking noise which would just mask the cancellation tones and, thus, the distortion tones. The latter derivation included the finding that, at very low frequencies, a higher signal-to-noise ratio at the output of the auditory filter is required for threshold. When present, the masking noise was continuous and at a constant level during one run; the fixed level of the masking noise was equal to that calculated to just mask the distortion products. Across runs, the noise level was kept constant relative to the level of the fixed stimulus. The noise should have substantially reduced the contribution of combination tones to loudness.

The duration of each signal was 700 ms, including 40-ms raised-cosine onset and offset ramps. The stimuli were generated digitally in advance using a sampling rate of 25 kHz. The tones were generated by adding sinusoids with frequencies ranging from F0 up to 10 kHz, while the noise was generated in the time domain by sampling from a Gaussian distribution. Bandpass filtering was then performed with a 900-tap, digital finite-impulse-response (FIR) filter with a linear phase response. For each F0, ten different realizations were produced for the RPH tone and for the Gaussian noise; one of the ten was chosen at random for each presentation. The fixed-level and variable-level signals were played through one channel of a 16-bit digital-to-analog converter (Tucker-Davis Technologies, TDT, DD1), lowpass filtered at 10 kHz (TDT FT6-2), and attenuated separately using a programmable attenuator (TDT PA4). The background noise was recorded onto CDR and played back with the test stimuli in the appropriate conditions. Stimuli were fed to a headphone buffer (TDT HB6) and presented monaurally via headphones with a diffuse-field response (AKG K 240 DF). Subjects were seated individually in an IAC double-walled, sound-attenuating booth.

2. Procedure

The two sounds (X and Y) which had to be matched in loudness were presented monaurally in regular alternation and were separated by a 200-ms silent interval. Synchronously with the two sounds, two yellow LEDs on a response box came on in regular alternation. Subjects indicated which sound was louder by pressing the button underneath the LED which accompanied that sound. Within a given run, either sound X or sound Y was fixed in level; the level of the other sound was varied to match the loudness. The starting level of the variable sound was chosen randomly from a range extending from 20 dB below to 20 dB above the level of the fixed sound, except for the lowest and the highest levels of the fixed sound. For these, the range extended from 10 dB below to 10 dB above the fixed level, to avoid inaudible or uncomfortably loud sounds. When the variable level sound was judged to be louder, it was decreased in level. When the fixed level sound was judged to be louder, the variable level sound was increased in level. The attenuator setting was changed during the silent interval between stimuli. When no button was pressed, sound presentation continued without any change. The initial step size was 5 dB. After two turnpoints it was reduced to 3 dB, and after two more turnpoints to 1 dB. Subjects were encouraged to bracket the point of equal loudness several times, i.e., to go from ''sound X is louder'' to ''sound Y is louder,'' before making their final

adjustment. To indicate that they were satisfied with their match, they pressed a button underneath a green LED, which stopped the run. The variable level at this point was taken as the matching level. This ''satisfied'' button press was only accepted after the final step size was reached. If it was pressed earlier, sound presentation continued without any change. In order to balance bias effects, eight matches were done varying sound X in level, and eight matches varying sound Y. In addition, for four of these eight matches the variable-level sound was presented with the first LED, and for the other four matches it was presented together with the second LED. No feedback was provided.

For a given F0, the same subjects ran in all conditions. Half of the subjects started with the masking noise present, and half without the noise. The conditions were presented in a counterbalanced order. Within a given condition, which was determined by the specific combination of the two sounds, the identity of the fixed stimulus, and the order of the two sounds, the level of the fixed stimulus was chosen randomly from the range of fixed levels to be tested. One match was obtained for each condition and level in turn before additional measurements were obtained in any other condition. To familiarize subjects with the procedure and equipment, they participated in one loudness match for each condition before the experiment proper commenced.

3. Subjects

For the 62.5-Hz F0, six subjects participated in all conditions. For the 250-Hz F0, five subjects participated in all conditions, three of whom also took part in the 62.5-Hz F0 conditions. Their ages ranged from 20 to 50 years, and their quiet thresholds were better than 15 dB HL at audiometric frequencies between 500 and 5000 Hz.

B. Results

Figures 2 and 3 show the mean results for the 62.5-Hz F0 and the 250-Hz F0, respectively. The left and right columns give the results without and with masking noise, respectively. The first, second, and third rows show the results from the three stimulus pairs, i.e., noise vs. CPH tone, RPH vs. CPH tone, and noise vs. RPH tone. The difference between the overall rms levels of the two stimuli at the point of equal loudness is plotted as a function of the level of the fixed stimulus; the direction of the difference is indicated in each panel. For example, the upper row shows the level of the noise minus the level of the CPH complex. Results are plotted separately for the case when X was varied in level and when Y was varied in level. For example, in the upper row, the open circles indicate the difference in level for matches when the CPH tone was fixed and the noise was varied, whereas the filled upward-pointing triangles show the difference when the CPH tone was varied. Each symbol shows the difference averaged across subjects and across the two orders of presentation, and is thus based on 48 matches for the 62.5-Hz F0, and on 40 matches for the 250-Hz F0. The error bars show the standard errors of the means across subjects. The solid lines show the mean for the two cases of the variable sound. The dashed lines in the right-hand column illustrate predictions which will be discussed later.

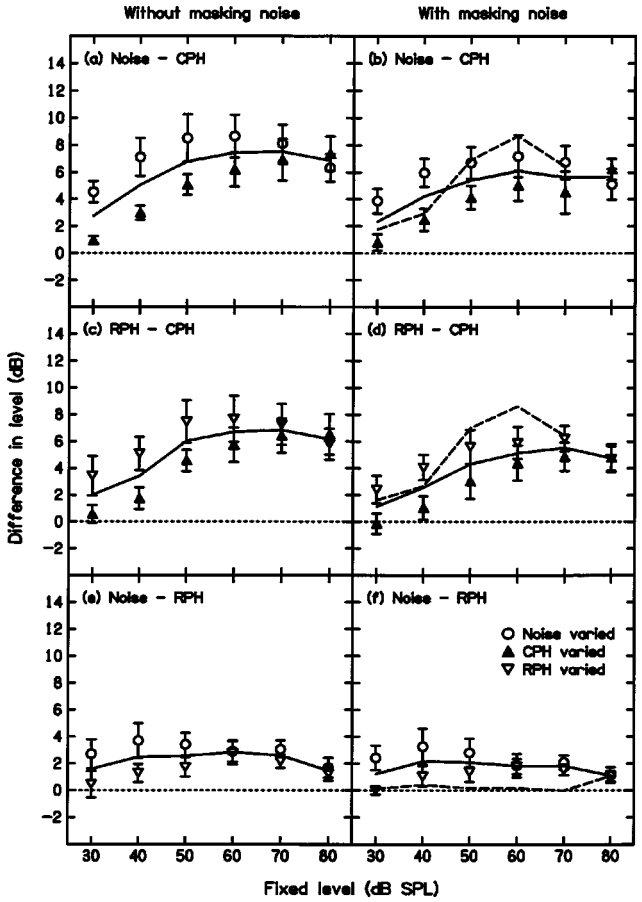


FIG. 2. Loudness results (experiment 1) for F0 = 62.5 Hz, averaged across six subjects. The left and right columns show results without and with the noise used to mask combination tones. Panels (a) and (b) (first row) show results of loudness comparisons between the Gaussian noise and the CPH tone. The difference in level (noise minus CPH tone) at the point of equal loudness is plotted. The open circles show results when the noise was varied in level; the filled upward-pointing triangles show results when the CPH tone was varied. The solid line shows the average of the two cases. Error bars show the standard error of the mean across subjects. The dashed line shows the predictions of the loudness model described in Sec. II D. Panels (c) and (d) show the results of loudness comparisons between the RPH and CPH tones. Panels (e) and (f) show the results for loudness comparisons between the Gaussian noise and the RPH tone.

1. 62.5-Hz F0

Consider the results for the 62.5-Hz F0 first (Fig. 2). For each fixed level, at the point of equal loudness, the noise level was greater than the CPH tone level [Fig. 2(a)], the RPH tone had a higher level than the CPH tone [Fig. 2(c)], and the noise had a higher level than the RPH tone [Fig. 2(e)]. These effects were strongest at medium to high levels, and there was a trend for them to be slightly reduced in the presence of the masking noise [Figs. 2(b), (d) and (f)]. A clear bias effect can be seen. At nearly all levels, listeners tended to set the level of the variable sound to a level higher than that ''required for equal loudness'' (as indicated by the solid lines). A similar bias effect has been observed in previous studies on loudness matching (Stevens, 1956; Zwicker *et al.*, 1957). The bias decreased with increasing level of the

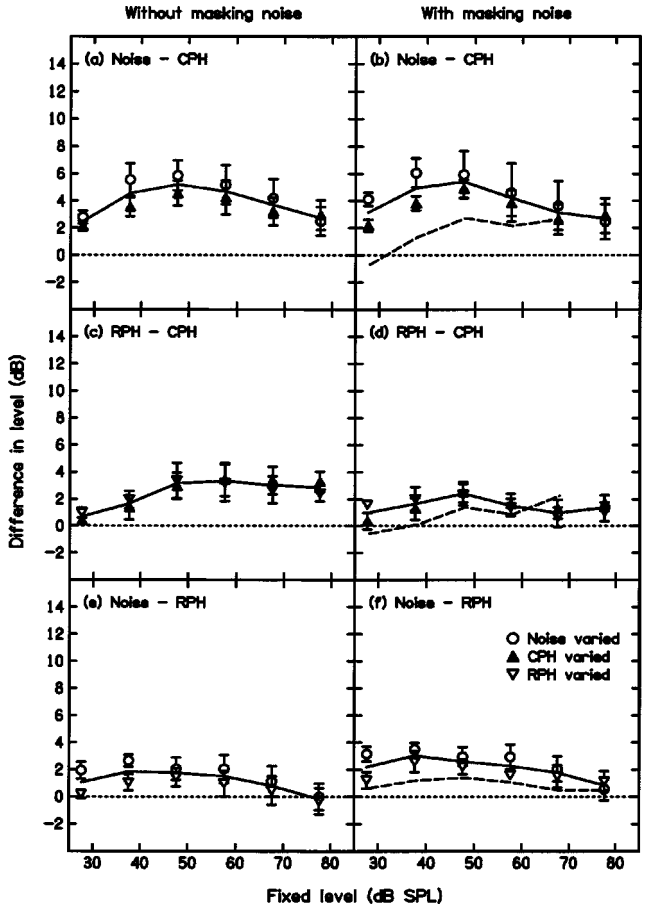


FIG. 3. As Fig. 2, but for F0 = 250 Hz.

fixed stimulus, perhaps because subjects avoided adjusting the variable sound to a high listening level. This ''regression effect'' has also been observed previously (Stevens, 1956; Stevens and Greenbaum, 1966; Stevens and Guirao, 1967).

Analyses of variance (ANOVAs) for repeated measures were conducted on the data for the three stimulus pairings separately. The mean value of the difference in level at the point of equal loudness for each individual subject in each condition was used as the input. Throughout this paper, the Huynh–Feldt correction was used when the condition of sphericity was not satisfied (see Howell, 1997). The results are summarized in Table I. The main effect of level was significant for the noise vs CPH comparison and for the RPH vs CPH comparison, but not for the RPH vs noise comparison. The main effect of presence or absence of the masking noise was significant only for the RPH vs CPH comparison. The main effect associated with which sound was varied was significant for the noise vs CPH and noise vs RPH comparisons, indicating that the bias was significant. The interaction between the identity of the varied sound and the fixed level was significant for all three comparisons, reflecting the reduction of the bias with increasing level of the fixed stimulus.

For the noise vs CPH comparison, the grand mean was significantly above zero ($p<0.0012$), indicating that, at the point of equal loudness, the level of the noise was significantly greater than that of the CPH tone. For all fixed levels tested, the difference in matching level was significantly above zero ($p<0.01$ for all levels). For the RPH vs CPH comparison, the grand mean was also significantly above zero ($p<0.003$), indicating that, at the point of equal loudness, the level of the RPH tone was significantly greater than that of the CPH tone. The difference between matching levels was significantly above zero ($p<0.05$) for all but the lowest fixed level tested. For the noise vs RPH comparison, the grand mean was also significantly above zero ($p<0.05$). As the ANOVA showed no effect of level for this comparison, no *post hoc* tests were performed.

2. 250-Hz F0

Next, consider the results for the 250-Hz F0 (Fig. 3). The general pattern of the results was similar to that found for F0 = 62.5 Hz, but the differences in level at the point of equal loudness were somewhat smaller. The results of the ANOVAs are summarized in Table I. The main effect of level was significant only for the RPH vs CPH comparison. The main effect of presence or absence of the masking noise was not significant for any comparison. The main effect of identity of the varied sound was also not significant for any comparison. The interaction between the identity of the varied sound and the fixed level was significant only for the RPH vs CPH comparison. The grand mean was significantly above zero ($p<0.05$) for all three comparisons, indicating differences in loudness at equal rms levels. Thus, as for the

TABLE I. Results of three-way ANOVAs (identity of varied sound×masker×fixed level) on level differences at the point of equal loudness measured in experiment 1. Only significant effects are listed.

	F0 = 62.5 Hz	F0 = 250 Hz
Noise vs CPH		
Varied sound	$F(1,5)=8.1$, $p<0.05$	···
Fixed level	$F(5,25)=9.6$, $p<0.001$	···
Varied sound×fixed level	$F(5,25)=11.1$, $p<0.001$	···
RPH vs CPH		
Masker	$F(1,5)=17.1$, $p<0.01$	···
Fixed level	$F(5,25)=7.9$, $p<0.01$	$F(5,20)=3.7$, $p<0.05$
Varied sound×fixed level	$F(5,25)=4.2$, $p<0.05$	$F(5,20)=3.8$, $p<0.05$
Noise vs RPH		
Varied sound	$F(1,5)=6.6$, $p<0.05$	···
Varied sound×fixed level	$F(5,25)=3.9$, $p<0.05$	···

62.5-Hz F0, the results indicate that at the point of equal loudness, the level of the noise was above that of the CPH and RPH tones and the level of the RPH tone was significantly above that of the CPH tone.

C. Discussion

At equal rms level, the CPH tone sounded louder than the noise or the RPH tone, and the RPH tone sounded louder than the noise. This was true for both F0s, although, for the 250-Hz F0, the effects were somewhat smaller than for the 62.5-Hz F0. It is not surprising that the effects of component phase are larger for the lower F0, because the number of components interacting at the output of a given auditory filter is four times greater. Thus, the envelope of the CPH tone with the lower F0 would exhibit greater modulation after auditory filtering than that of the CPH tone with the higher F0. The effects for the 62.5-Hz F0 were similar to those that we reported previously (Gockel *et al.*, 2002b), although in the former study the difference in loudness between RPH tones and noise was not significant (although it was present), whereas in the present study the difference was significant. The noise designed to mask combination tones only reduced the size of the effects slightly at the lower F0, and had no significant effect at the higher F0. Therefore, while there might be a small contribution of audible distortion products to the observed loudness differences, this cannot be the main factor involved. The effects of component phase on loudness were clear and statistically significant, but the direction of the effects was opposite to what would be predicted from BM compression. The CPH tones, the sounds with the greater crest factor (both physically and in the waveform evoked on the BM), were judged as louder than the RPH tone and the noise, both of which have lower crest factors.

The present results contrast with those of Carlyon and Datta (1997). They found the loudness of a subset of components covering the mid range of a complex tone to be greater for an S− complex than for an S+ complex. Their complex tone had an F0 of 100 Hz, with components ranging from 200 to 2000 Hz, and the central subset of components whose loudness had to be judged ranged from 900 to 1300 Hz. Thus, the 400-Hz bandwidth of the central subset of components was relatively small compared to that of our stimuli. Their results are consistent with the finding mentioned in the Introduction, that for narrowband sounds, amplitude modulation at medium rates (10–100 Hz) reduces loudness slightly, relative to that of the unmodulated sound (Zhang and Zeng, 1997; Moore *et al.*, 1998, 1999; Grimm *et al.*, 2002).

Our own results are also in the same direction as the findings mentioned in the Introduction; for broadband sounds, amplitude modulation at medium rates slightly increases loudness relative to that of the unmodulated sounds (Zhang and Zeng, 1997; Moore *et al.*, 1999; Grimm *et al.*, 2002). However, the present effect is larger than the 1–2-dB difference usually observed, especially for the lower F0. Since most earlier studies used sinusoidal amplitude modulation, one possible reason for the larger effect with the current stimuli might be the relatively large crest factor of our CPH stimuli; it was 12.4 for the 62.5-Hz F0 and 5.0 for the 250-Hz F0 (of course, the crest factor on the BM would be different from this). For broadband sounds, it appears that a large crest factor can lead to an increase in loudness. However, the loudness does not correspond to the peak value of the (physical) waveform. For the 62.5 Hz-F0, the peak value of the CPH wave is about four times greater than the peak value of the RPH wave. This corresponds to about a 12-dB difference in level, which is much larger than the observed level differences at the point of equal loudness.

Another factor which might have influenced our results is the fact that our CPH stimuli had a strong and distinct pitch corresponding to the F0 (Schouten, 1940, 1970). In contrast, for a high-frequency sinusoidal carrier modulated at a low rate, the pitch corresponding to the modulation rate is weak or absent (Ritsma, 1962; de Boer, 1976). Possibly, the existence of a strong residue pitch might increase the loudness of a sound relative to one with the same power spectrum but a weak or absent pitch. As the pitch of our CPH tones was stronger than that of the RPH tones (Lundeen and Small, 1984; Patterson, 1987; Roberts *et al.*, 2002), especially for the lower F0 (Warren and Bashford, 1981), this might have contributed to the differences in loudness at equal rms level. A problem with this account is that the subjective difference in pitch strength between the RPH tone and noise was larger than the difference in pitch strength between the CPH and RPH tones, but the mean difference in level at the point of equal loudness was larger for the latter pair of sounds. However, loudness might be a nonlinear function of pitch strength, growing rapidly with increasing pitch strength.

A third factor which might have contributed to our results is related to the annoyance of sounds. The CPH tones would have had a greater roughness than the RPH tones or the noise (Terhardt, 1974; Hellman, 1985; Zwicker and Fastl, 1999) and this might have led to greater annoyance of the CPH stimuli, and an effect on loudness. Subjects were instructed to judge loudness and ignore all other factors but they were not specifically instructed to ignore annoyance. However, it is unlikely that annoyance is the only factor influencing the effects of phase on loudness. First, several experienced listeners (including the authors), who knew about the difference between annoyance and loudness, judged the CPH stimuli to be louder than the RPH or noise stimuli at equal rms level. Second, the perceived roughness for a modulation rate of 250 Hz is substantially lower than for a modulation rate of 62.5 Hz (Zwicker and Fastl, 1999). The CPH and RPH stimuli with F0 = 250 Hz did not appear to differ from each other or from the noise in roughness. Thus, annoyance due to roughness is unlikely to have influenced the loudness matches for the 250-Hz F0, but effects of phase on loudness were observed.

D. Model predictions

Models of loudness perception for steady sounds generally ignore the phase spectrum of the sound (Zwicker and Scharf, 1965; Zwicker *et al.*, 1984; Moore *et al.*, 1997), and thus are unable to account for the present results. A more recent model for fluctuating sounds described by Glasberg and Moore (2002) is based on the short-term spectrum of the waveform and takes into account the effects of peripheral

compression. This model predicts a *reduction* in loudness of the CPH complex relative to that of the RPH complex or the noise by about 5 to 6 phons for the 62.5-Hz F0 at medium levels. For the 250-Hz F0 a difference of about 1 phon in the same direction is predicted. For both F0s, no difference in loudness is predicted between the RPH tone and the noise. Thus, the model fails to account for our results. It is not clear whether other models for the loudness of fluctuating sounds (Zwicker, 1977; Fastl, 1993; Chalupper and Fastl, 2002; Grimm *et al.*, 2002) would do any better.

In what follows, we describe a model which correctly predicts the direction of the observed loudness differences between CPH tones and RPH tones or noise. However, the magnitudes of the effects are not accurately predicted for all conditions. The modeling was performed with the AMS/DSAM software package[1] and MATLAB.

The waveforms of the sounds used in our experiment were used as input to the model. The sampling rate was 50 kHz. The input level was varied over a wide range in 1-dB steps for each waveform. The first stage of the model is an FIR filter which simulates the combined outer and middle ear response as specified in Glasberg and Moore (2002), except that here, as our headphones had a diffuse-field response, we used a diffuse-field correction instead of the correction for frontal free-field sound incidence. The next stage was a dual-resonance nonlinear filterbank (Lopez-Poveda and Meddis, 2001) that simulates the nonlinear response of the BM at different points. We used 30 filters with characteristic frequency (CF) varying between 40 Hz and 10 kHz. The CFs were spaced according to Greenwood's (1990) frequency-place map for humans. The filter bank takes stapes motion as input, which is a linear function of the sound pressure. The output is BM velocity. The remaining stages are calculated in parallel for the 30 filter outputs. The BM velocity is lowpass filtered to simulate the displacement of the inner hair cell stereocilia, according to Eq. (1) in Sumner *et al.* (2002). The stereocilia response is converted into a receptor potential according to Eqs. (2) and (3) in Sumner *et al.* (2002). Then, the receptor potential is converted into auditory nerve (AN) spike probability (with parameters according to Table II of Sumner *et al.*). This latter stage has three parallel parts, which generate the spike probabilities for high-, medium-, and low-spontaneous rate fibers, respectively. Each of the three spike probabilities is converted into a stream of spikes, each of which is the sum of the spike activity for a population of fibers; 60, 25, and 15 fibers were used for the high-, medium-, and low-spontaneous rate fibers, respectively. The three streams of spikes were then added in each frequency channel and the simulated AN activity was summed across all 30 channels. The result is an overall neural activity pattern (NAP) for each sound as a function of time. All stages to this point were simulated with AMS/DSAM. The following calculations were done in MATLAB.

For each waveform, various measures based on the NAP were calculated. These measures were chosen as they had been suggested as correlates of loudness by previous researchers. Call a given measure (e.g., the overall mean of the NAP) *M*. For a given pair of sounds (e.g., CPH vs RPH with F0=62.5 Hz), we determined the difference in level required to give the same value of *M* for the two sounds, for each level of the fixed sound used in the experiment. If *M* is closely related to loudness, then the differences in level so determined should match those obtained in the experiment at the point of equal loudness. We tried to find a measure for which this was the case, as closely as possible, over a wide range of levels, for both F0s and for all pairs of sounds that were compared. The spike generation process in the model is stochastic and so we averaged across simulations to produce stable response values. For the CPH stimulus, ten NAPs were calculated for each input level. For each of the ten NAPs, the measures were calculated individually and then averaged. The measures which will be discussed are all based on this averaging process. For the RPH and the noise stimuli, three NAPs were calculated for each input level for each of the ten different waveforms used in the experiment, and the measures were calculated for each of them individually. The measures which will be discussed are the averages across the 30 values.

The first measures considered were the maximum value in the NAP, the average value of the NAP, the mean value of all peaks in the NAP, and the mean value of all peaks above the 80th percentile. Measures comparable to these have been used previously to model the loudness of steady and fluctuating sounds (Fastl, 1977; Zwicker, 1977; Howes, 1979; Fastl, 1993; Zwicker and Fastl, 1999). For inputs with equal rms values, the mean of the NAP for the CPH tone was below that for the RPH tone or the noise. This was true for all input levels, though the amount of the difference varied as a function of level. Thus, this measure would predict an effect on loudness in the opposite direction to that found. This indicates that, contrary to what is often assumed, loudness is not directly related to the overall activity in the auditory nerve. The two measures involving peaks in the NAP also predicted an effect in the wrong direction. The measure based on the single largest value in the NAP, which occurred close to the onset of the response, did predict an effect in the right direction. However, this is unlikely to be the basis for loudness judgments, since loudness increases with duration up to 200 ms (Scharf, 1978; Florentine *et al.*, 1996), and the size of the predicted effect was much too large.

The above measures were derived directly from the raw NAP and they do not reflect the fact that the NAP of the CPH stimulus is clearly periodic; it looks like a saw-tooth function with a period corresponding to 1/F0. Moreover, perceived loudness is presumably based on some form of running average of the NAP. To accommodate these factors, we examined the effect of asymmetric temporal smoothing of the NAP, using an operation resembling the calculation of gain in an automatic gain control system with different attack and release times. The operation is similar to the one used by Glasberg and Moore (2002) to determine short-term loudness from what they called ''instantaneous loudness'' (which was assumed not to be directly accessible for conscious processing). The temporal smoothing of the NAP was done in the following way. We define SM_n as the smoothed NAP at the nth sample point, and NAP_n as the raw NAP at the nth sample point.

If $NAP_n > SM_{n-1}$ (corresponding to an attack), then

$$SM_n = a\mathrm{NAP}_n + (1-a)\mathrm{SM}_{n-1}, \quad (1)$$

where a is a constant related to the attack time, T_a (in ms):

$$a = 1 - e^{1/(Fs \cdot T_a)} \quad (2)$$

with Fs corresponding to the number of samples per ms (50 in this case).

If $\mathrm{NAP}_n < \mathrm{SM}_{n-1}$ (corresponding to a release), then

$$SM_n = r\mathrm{NAP}_n + (1-r)\mathrm{SM}_{n-1}, \quad (3)$$

where r is a constant related to the release time, T_r (in ms):

$$r = 1 - e^{1/(Fs \cdot T_r)}. \quad (4)$$

Based on the smoothed NAPs, the same measures were calculated as before. The mean across the smoothed NAPs gave predictions in the right direction. The attack and release times that produced the best fits to our data were 3.1 and 49.49 ms, respectively. The release time happens to be identical to the one chosen by Glasberg and Moore (2002), while the attack time is about one-seventh of their value. The predictions obtained with these values are plotted as dashed lines in the right-hand columns of Figs. 2 and 3. For loudness matches involving the CPH tone at medium levels, the predicted effect is larger than the obtained effect for the low F0, but it is smaller than the obtained effect for the higher F0. No predictions are plotted for a fixed input level of 80 dB SPL for the CPH tone. The reason is that, for input levels above about 80 dB SPL, the smoothed (and the raw) NAP statistics of the RPH tone and the noise hardly increased with increasing level; to get the same mean value as for the CPH tone, the level of the RPH tone had to be about 40 dB above that of the CPH tone. This may reflect neural saturation in the model, perhaps resulting from inadequate numbers of fibers or insufficient dynamic range in the fibers with low spontaneous rates. For loudness matches between RPH tones and noise, the model failed to explain the effect observed with the lower F0; for the higher F0, there was a small effect in the right direction. The predicted effect for the higher F0 does not seem to be a consequence of the temporal smoothing, since there was a similar effect with the ''raw'' (unsmoothed) NAPs.

In summary, a model based on the ''raw'' (unsmoothed) NAPs produced results opposite in direction to the observed effects. This indicates that, contrary to what is often assumed, loudness is not directly related to the overall activity in the auditory nerve. A similar conclusion has been reached by Relkin and Doucet (1997). However, a model based on NAPs smoothed with a fast attack and slow release produced results in qualitative agreement with the data.

III. EXPERIMENT 2

The second experiment was designed to test whether the effectiveness of our stimuli as forward maskers would reflect their loudness, i.e., whether the CPH tone would be a more effective forward masker than the RPH tone and the noise. In simultaneous masking, using the same complex tones as in the present study, Gockel *et al.* (2002a) found that the CPH tone was a less efficient masker of a noise signal than the RPH tone for F0 = 62.5 Hz, while the opposite was true for F0 = 250 Hz. This was explained in terms of the temporal structure of the stimuli, which allows ''listening in the dips'' for the lower F0. Such listening in the dips is usually assumed to be irrelevant in forward masking, since the amount of forward masking reflects the average excitation or neural activity evoked by the masker over a relatively long time period (Zwicker, 1977; Carlyon and Datta, 1997). Therefore, it is not obvious that the effects of masker component phase will be the same for simultaneous masking and for forward masking. The question addressed here was whether the longer term activity evoked by the masker, as measured in forward masking, is directly related to loudness.

As mentioned in the Introduction, Carlyon and Datta (1997) measured the forward masking produced by S+ and S− complexes and found that, at least for high masker levels, the S+ complex (large peak factor on the BM) produced less forward masking than the S− complex (low peak factor on the BM). They interpreted this result in terms of the effects of fast-acting compression on the effective excitation evoked by the complexes. Their results led us to expect that the CPH tone would be a less effective forward masker than the RPH tone or noise, despite the greater loudness of the CPH tone.

A. Method

1. Stimuli

The signal was a 20-ms sinusoid with a frequency of 702, 1114, 1768, 2806, or 4454 Hz. The middle frequency, 1768 Hz, corresponded to the geometric mean of the frequency range covered by the forward masker, which extended from 625 to 5000 Hz; the other signal frequencies were offset by two semitones, and ten semitones, respectively, from the lower and upper end of the masker's frequency range. The signal had 10-ms, raised-cosine onset and offset ramps and no steady-state portion; it followed the masker without a silent gap.

Three stimuli were used as forward maskers: the CPH and RPH tones with an F0 of 62.5 Hz, and the Gaussian noise. They were filtered in the same way as in experiment 1. Their duration was 208 ms (corresponding to 13 periods), including 10-ms raised-cosine onset and offset ramps. To test whether there was an effect of time elapsed between the last peak in the masker waveform and the onset of the signal, the CPH masker began at two different points in its 16-ms period; either 0 or 12 ms after the first peak (conditions CPH 0 and CPH 12, respectively). To measure level dependence and to allow differences in masker effectiveness to be expressed as differences in effective masker level (Houtgast, 1974; Moore and Glasberg, 1983), the masker was presented at 30, 50, 68, 77, and 85 dB SPL. If compression plays a major role, then its effects should be most obvious at medium to high levels, and less clear or absent at low levels. The method of stimulus generation and the equipment were the same as for experiment 1.

2. Procedure

A two-interval two-alternative forced-choice task was used. The forward masker was presented in both intervals, which were marked by lights and separated by 300 ms of

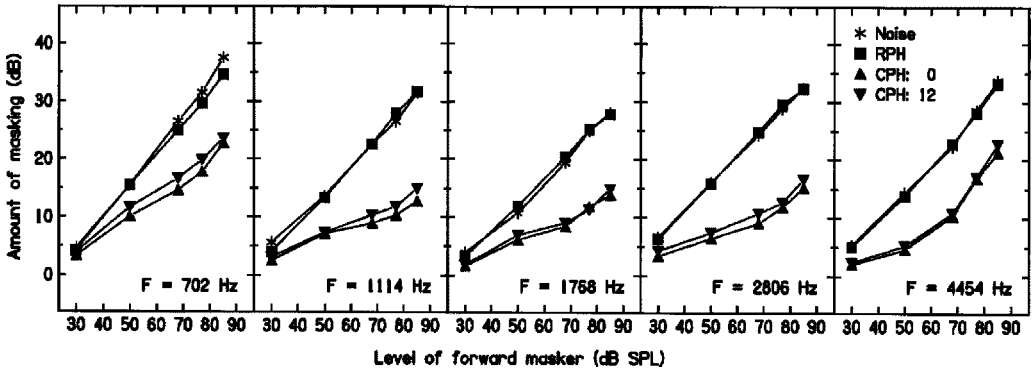


FIG. 4. Forward masking results (experiment 2), averaged across six subjects, as a function of masker level. The asterisks, squares, upward-pointing triangles, and downward-pointing triangles indicate results for the noise, the RPH tone, the CPH tone with 0-ms offset in the period, and the CPH tone with 12-ms offset in the period, respectively. The five panels show the results for the five different signal frequencies, as indicated in each panel.

silence. The subject was required to indicate which interval contained the signal. Feedback was provided following each response. The level of the signal was adjusted using a three-down one-up adaptive rule (Levitt, 1971) tracking 79% correct responses. At the beginning of each threshold measurement, the signal level was 10 dB above that of the masker, except for the lowest masker level, where it was 20 dB above the level of the masker. Initially, the signal level was increased or decreased in 8-dB steps. Following two reversals, the step size was reduced to 4 dB, and following two more reversals, it was reduced to 2 dB. Eight reversals were obtained with this final step size. The threshold was defined as the mean of the signal levels at the last eight reversals.

At least three threshold estimates were obtained for each condition from each subject. If the thresholds for a given subject and condition varied by more than 6 dB, two additional threshold estimates were obtained. The thresholds reported correspond to the mean of these three to five estimates, for each condition and subject. The total duration of a single session was about 2 h, including rest times. For a given signal frequency, one threshold was collected for each condition before the signal frequency was changed. Within a block with fixed signal frequency, one threshold was collected for each of the four maskers before the masker level was changed. Within the four runs at a given level, the identity of the masker was chosen randomly. The masker level was changed in either ascending or descending order between groups of four runs, in order to avoid having a run with a high-level masker immediately before a run with a low-level masker. The direction of level change was reversed when the next thresholds were collected for this signal frequency. The order of signal frequencies was counterbalanced across subjects. One threshold was obtained for each condition in turn, before additional measurements were obtained in any other condition. Subjects participated in at least one threshold measurement in each condition before data collection proper commenced. This took approximately 8 h. Some of the subjects showed strong learning effects, in which case their second threshold measurement in each condition was regarded as an additional practice run and the experiment proper started after that.

Absolute thresholds for each signal were measured after the main experiment was completed, using the same three-down one-up procedure as before. Four threshold estimates were obtained for each signal and each subject.

3. Subjects

Six subjects participated in all conditions, two of whom had substantial experience in other psychoacoustic experiments. One of them was the first author, and one of them was a musician. Their ages ranged from 18 to 41 years, and they all had normal hearing at audiometric frequencies between 500 and 4000 Hz.

B. Results

Figure 4 shows the masking produced by each masker as a function of masker level. Each panel shows results averaged across the six subjects for one signal frequency. The RPH tone (squares) and the noise (asterisks) produced similar amounts of masking, increasing at a steady rate from about 5 to between 27 and 36 dB with increasing masker level; the size of the increase depended on the signal frequency. The two CPH maskers (upward and inverted triangles) produced far less masking than the RPH tone or noise. The amount of masking increased from about 3 to between 14 and 23 dB with increasing masker level. The rate of increase depended on signal frequency and masker level. In summary, the CPH tones produced far less forward masking than the RPH tone or the noise, despite the fact that they were clearly louder.

To assess the statistical significance of these effects, a three-way ANOVA (masker identity×signal frequency ×masker level) for repeated measures was conducted on the amount of masking. The main effect of masker identity was highly significant [$F(3,15)=330.6$, $p<0.001$], as were the main effects of signal frequency [$F(4,20)=7.2$, $p<0.001$] and of level [$F(4,20)=280.4$, $p<0.001$]. There was also a significant interaction between masker identity and signal frequency [$F(12,60)=3.4$, $p<0.05$], indicating that the relative amount of masking produced by the different maskers depended on the signal frequency. The interaction between level and signal frequency was highly significant [$F(16,80)=11.6$, $p<0.001$], indicating that the increase in amount of masking with level depended on signal frequency. Furthermore, the interaction between masker identity and masker level was highly significant [$F(12,60)=81.6$, p

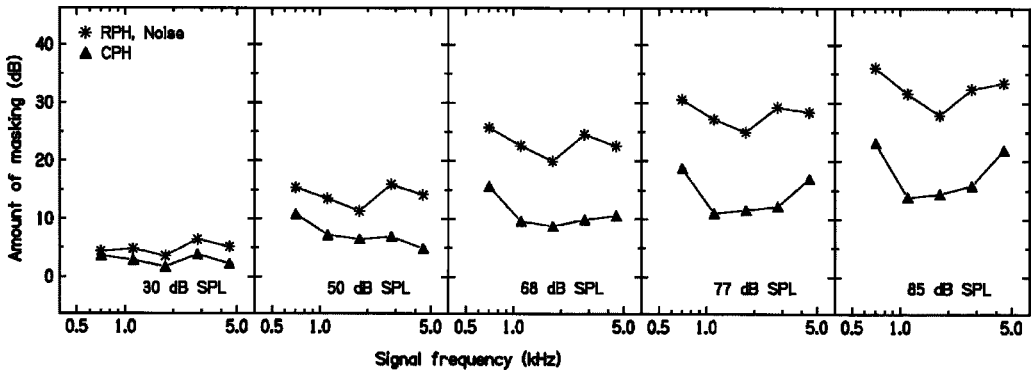


FIG. 5. Mean amount of forward masking, plotted as a function of the frequency of the signal. The asterisks show results averaged for the RPH-tone and noise maskers; the upward-pointing triangles show results averaged for the two CPH-tone maskers. The five panels show the results for the five different masker levels, as indicated in each panel.

<0.001], showing that the increase in amount of masking with increasing level differed across maskers. Finally, the three-way interaction was significant [$F(48,240)=2.9$, $p=0.001$].

Post hoc paired comparisons based on Fisher's least significant difference procedure were made across the different masker conditions. As they showed no significant difference between the RPH tone and noise masker, the data were averaged across these two masker conditions for subsequent analyses. Averaged across level and signal frequency, the amount of masking produced by the CPH 0 masker (10.1 dB) was slightly smaller than that produced by the CPH 12 masker (11.1 dB). Thus, the time between the last envelope peak in the masker's waveform and the beginning of the signal, being shorter for the CPH 12 condition, did have a small but significant effect ($p<0.001$). The fact that the difference was very small is consistent with the idea that forward masking depends upon the average effect of the masker over a relatively long time interval. Since thresholds for the CPH 0 and CPH 12 maskers were almost the same, the data for these two conditions were averaged for subsequent analyses.

To enable comparison of the masking effects across the different signal frequencies, Fig. 5 shows the amount of masking, averaged across RPH tone and noise maskers (asterisks), and averaged across the two CPH tone maskers (upward-pointing triangles) as a function of signal frequency; each panel shows the data for one masker level. Even for the lowest masker level of 30 dB SPL (left-most panel), the amount of masking produced by the RPH tone and noise masker was higher than for the CPH tone maskers, albeit only by about 2 dB. This difference was, however, significant; a two-way ANOVA (masker identity $\times$signal frequency) for the masker level of 30 dB SPL showed a highly significant main effect of masker identity [$F(1,5)=112.9$, $p<0.001$]. The interaction between masker identity and signal frequency was also significant [$F(4,20)=6.1$, $p<0.01$]. With increasing masker level (from left to right panels) the difference in masking between the RPH/noise masker and the CPH maskers increased markedly. At the highest masker level of 85 dB SPL, the RPH and noise masker produced about 32 dB of masking (averaged across signal frequency), while the CPH maskers produced about 17.8 dB of masking. The effect of masker identity was largest at 77 and 85 dB SPL for the signal frequencies of 1114 and 2806 Hz (the frequencies adjacent to the center frequency), amounting to 16–18 dB.

For the CPH masker, the masking pattern became more bowl shaped with increasing masker level, i.e., at high levels the amount of masking was greater for the outermost signal frequencies than for the three mid-frequency signals. For the RPH tone and the noise masker, it was a bent V-shaped pattern, with a clear minimum at the center frequency and a "bump" at 2806 Hz. The increased masking at 2806 Hz may be related to the peak in the response of the headphones around 3000 Hz. This peak is reflected in the long-term excitation pattern of the masker, as shown in Fig. 1(a) of Gockel *et al.* (2002a). Perhaps this peak was enhanced by a release from lateral suppression leading to increased forward masking (Houtgast, 1974; Moore and Glasberg, 1983). If this is the case, it remains unclear why no "bump" occurred for the CPH stimuli. See below for further discussion of the effects of suppression.

To express the differences in amount of masking as differences in effective level of the maskers (Houtgast, 1974; Moore and Glasberg, 1983), we estimated the increase in level of the CPH masker necessary for it to produce the same amount of masking as the RPH tone and noise masker, for the four higher levels of the CPH masker. The estimates were based on interpolation, using the short line segments connecting the data for adjacent masker levels, as in Fig. 4. Figure 6 shows these increases plotted as a function of the level of the CPH masker, with signal frequency as parameter. The CPH masker had to have a level up to 35 dB above the level of the RPH or noise masker to produce equal masking. For all levels except the lowest, the necessary increase in CPH masker level was greatest for the 2806-Hz signal frequency. The second largest difference occurred for the 1114-Hz signal frequency. This is a consequence of the V-shaped masking pattern for the RPH and noise maskers, and the bowl-shaped masking pattern of the CPH maskers. So, interestingly, the effectiveness of the CPH maskers was reduced most relative to that of the RPH or noise maskers at the two signal frequencies adjacent to the center frequency, rather than at the center frequency.

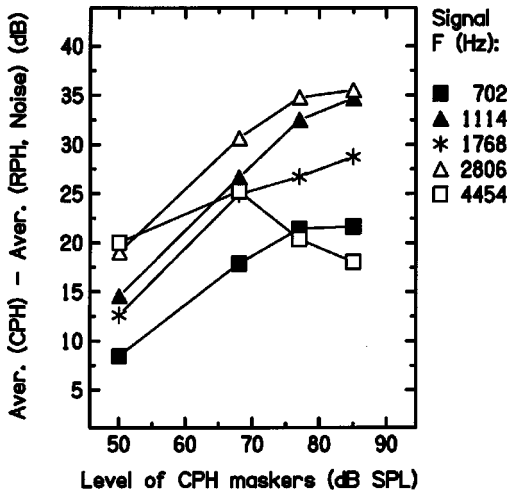


FIG. 6. Differences in effective level of the forward maskers. The difference in level of the CPH maskers and the RPH/noise maskers at the point where they produced an equal amount of forward masking is plotted as a function of the level of the CPH maskers. The five curves correspond to the five different signal frequencies, as indicated in the legend.

C. Discussion

The results of the forward-masking experiment are consistent with those of Carlyon and Datta (1997) indicating that stimuli evoking waveforms with a large crest factor on the BM (S+ or CPH tones) are less effective forward maskers than stimuli evoking waveforms with small crest factors (S− or RPH tones and noise). Carlyon and Datta explained this effect in terms of fast-acting compression on the BM, which would result in lower average excitation for waveforms with large peak factors. However, it is remarkable that the differences in effective level of the CPH and RPH maskers could be as much as 35 dB. It is difficult to conceive how such a large effect could be explained in terms of BM compression alone.

To see why this is so, consider a hypothetical example. In what follows, for simplicity, we consider only the envelopes of the hypothetical signals, but the arguments would be similar if the waveforms were considered. Assume that a signal has an envelope that is very peaky: the envelope consists of low-amplitude portions (amplitude=0.01 units) lasting 0.99 time units, and brief high-amplitude portions (amplitude=1 unit) lasting 0.01 time units. The rms value of the envelope is 0.10049 units and the crest factor of the envelope is 9.95. This signal is compared with a signal of constant envelope amplitude and the same rms value. Assume that each of these two signals is passed through an instantaneous compressor with a compression ratio of 3 over the whole range of input amplitudes and which applies a gain of unity (0 dB) when the input envelope amplitude is one unit. Following compression, the low-amplitude portions of the first signal have an amplitude of 0.215, the peak amplitude stays the same (unity), and the rms value of the envelope becomes 0.236. For the second signal, the (constant) envelope amplitude is increased to 0.465. To equate the rms output values of the two signals (allowing for the compression ratio of 3), the input level of the peaky stimulus would have to be increased by 17.7 dB. This is about one-half of the largest difference in effective masker level inferred from the forward masking data. The required change in input level needed to equate the output levels of the two stimuli would be somewhat larger if the compression ratio were higher. For example, a compression ratio of 5 would lead to a required change in input level of the peaky signal of 18.9 dB. We conclude that it is unlikely that the differences in effective level of our maskers inferred from the forward masking data can be accounted for solely by the effects of fast-acting compression on the BM.

It is possible that suppression has a strong influence on the masking produced by broadband CPH maskers, as used in the present experiments. When the masker evokes a highly peaky waveform on the BM, and the peaks are approximately synchronous at the outputs of different auditory filters, the peaks at the outputs of filters tuned away from the signal frequency may suppress peaks at the output of the on-frequency filter, and this may reduce the forward masking produced by the peaks. Also, if the masker dips are synchronous across channels, then suppression will be weak during the dips, as the strength of suppression depends partly on overall level (Javel, 1981; Sellick *et al.*, 1982; Recio and Rhode, 2000). If suppression is at its strongest when the masker itself is at its highest short-term level, and is weakest when the masker is at its lowest short-term level, this could produce a strong reduction of the effective level of a peaky masker, and thus contribute to differences in effectiveness of the CPH tones and RPH tones/noise as forward maskers.

Suppression within a broadband sound occurs both from low frequencies towards higher ones and vice versa, although the former effect occurs mainly for medium to high overall levels (Houtgast, 1974; Javel, 1981; Ruggero *et al.*, 1992). Suppression will be strongest in a given frequency region when there is suppression from both lower and higher regions, i.e., when the frequency is well within the spectral range of the sound. Towards the spectral edge of the sound, there is effectively a release from suppression, which results in enhancement of the edge (Houtgast, 1972, 1974). This can account for our finding that the amount of masking was more for the signal frequencies towards the edge of the spectral range of the maskers than for the signal frequencies well within that spectral range.

IV. GENERAL DISCUSSION

At equal rms level, the CPH maskers clearly produced less forward masking than the RPH or noise maskers, even though the former were louder than the latter. Thus, the loudness of the stimuli and their effectiveness as forward maskers cannot be based on identical processes. It is generally assumed that both the loudness of a sound and the effectiveness of that sound as a forward masker are related to the excitation evoked by the sound in the peripheral auditory system. Our results suggest that one or both of these assumptions is incorrect.

The model based on neural activity patterns (NAPs), introduced in Sec. II D, was able to predict some aspects of the loudness-matching data. To achieve this, we assumed that loudness is based on the output of a smoothing process applied to the NAP. The smoothing was implemented using a

sliding temporal integrator analogous to an automatic gain control system, with a fast attack time and slower decay time. Effectively, this system responds strongly to the peaks in the NAP evoked by the CPH tone, thus increasing the overall response to that tone relative to the response evoked by the RPH tone or noise.

It is usually assumed that forward masking is related to the excitation or neural activity evoked by the masker, averaged or smoothed over a time interval of a few tens of milliseconds (Zwicker, 1977; Carlyon and Datta, 1997; Zwicker and Fastl, 1999). In fact, forward masking is often modeled using the concept of a sliding temporal integrator or window (Moore *et al.*, 1988; Plack and Moore, 1990; Moore and Oxenham, 1998; Plack and Oxenham, 1998). The temporal integrator in these models is asymmetric in time, but is usually assumed to be linear. Such models can correctly predict that CPH stimuli produce less forward masking than RPH or noise stimuli, provided that they include a compressive nonlinearity prior to the temporal integrator. However, a linear temporal integrator would not account for our loudness data. To model these data, we had to assume a nonlinear smoothing process, which behaves differently depending on whether the momentary input is increasing or decreasing. This mechanism emphasizes the peaks in the input (in our model the peaks in the NAP), and this leads to a greater output for the CPH than for the RPH or noise stimuli.

V. SUMMARY AND CONCLUSIONS

In experiment 1, we obtained loudness matches between complex tones with components added in cosine phase (CPH) and random phase (RPH), and between those complex tones and noise. Two fundamental frequencies (F0s) were used: 62.5 and 250 Hz. The stimuli were bandpass filtered so that only unresolved harmonics were present for the tones. For a given F0, excitation patterns calculated from the power spectrum were essentially identical for all stimuli. The results showed the following.

(1) For a given overall level, the CPH tone was louder than the RPH tone and the noise, and the RPH tone was louder than the noise. At the point of equal loudness, the level of the noise was greater than that of the CPH tone by up to 7 dB.
(2) The effects were greatest for medium to high levels.
(3) The effects were only slightly reduced by the addition of a background noise intended to mask combination tones.
(4) The effects were greater for F0 = 62.5 Hz than for F0 = 250 Hz.

The general pattern of the results could be accounted for using a model based on simulated neural activity patterns (NAPs), using a temporal smoothing mechanism resembling the operation of an automatic gain control system with a fast attack time and slow decay time. However, the model failed to predict the observed difference in loudness of the RPH-tone and noise for F0 = 62.5 Hz.

Experiment 2 compared the effectiveness of the same sounds (F0 = 62.5 Hz only) as forward maskers, using several signal frequencies within the spectral range of the maskers. The results showed the following.

(1) The noise and RPH tone produced similar amounts of forward masking and both produced more masking than the CPH tone.
(2) The results were only slightly affected by the time during the period of the CPH masker when that masker was turned off.
(3) The rate of growth of forward masking was markedly greater for the noise and RPH tone than for the CPH tone.
(4) For high masker levels, the amount of masking was greater for signal frequencies close to the spectral edges of the masker than for signal frequencies well within the spectral range of the masker.
(5) For high masker levels, the CPH masker needed to have a level as much as 35 dB above that of the RPH or noise maskers to produce equal amounts of masking.

Taken together, the results of the two experiments indicate that, for a given overall level, peaky sounds (CPH tones) are louder and produce less forward masking than less peaky sounds (RPH tones or noise). Thus, it cannot be the case that loudness and forward masking are determined by the same peripheral processes, such as the level of excitation or neural activity. We suggest that the representation of these sounds in the auditory system may be affected by at least two processes: fast-acting cochlear compression and suppression.

ACKNOWLEDGMENTS

This work was supported by the MRC (Grant No. G9900362). We thank Brian Glasberg for calculating the predictions of the Glasberg and Moore loudness model for our stimuli. We also thank Marjorie Leek, Rhona Hellman, and an anonymous reviewer for helpful comments on an earlier version of this paper.

[1]The AMS/DSAM software is available over the Internet at http://www.essex.ac.uk/psychology/hearinglab/dsam/index.htm

Bauch, H. (**1956**). ''Die Bedeutung der Frequenzgruppe für die Lautheit von Klängen,'' Acustica **6**, 40–45.
Buunen, T. J. F., Festen, J. M., Bilsen, F. A., and van den Brink, G. (**1974**). ''Phase effects in a three-component signal,'' J. Acoust. Soc. Am. **55**, 297–303.
Carlyon, R. P., and Datta, A. J. (**1997**). ''Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system,'' J. Acoust. Soc. Am. **101**, 3636–3647.
Chalupper, J., and Fastl, H. (**2002**). ''Dynamic loudness model (DLM) for normal and hearing-impaired listeners,'' Acust. Acta Acust. **88**, 378–386.
de Boer, E. (**1976**). ''On the 'residue' and auditory pitch perception,'' in *Handbook of Sensory Physiology, Vol. 3*, edited by W. D. Keidel and W. D. Neff (Springer, Berlin).
Fastl, H. (**1975**). ''Loudness and masking patterns of narrow noise bands,'' Acustica **33**, 266–271.
Fastl, H. (**1977**). ''Loudness of running speech,'' J. Audiol. Tech. **16**, 2–13.
Fastl, H. (**1993**). ''Loudness evaluation by subjects and by a loudness meter,'' in *Sensory Research-Multimodal Perspectives*, edited by R. T. Verrillo (Erlbaum, Hillsdale, NJ).
Fletcher, H., and Munson, W. A. (**1933**). ''Loudness, its definition, measurement and calculation,'' J. Acoust. Soc. Am. **5**, 82–108.
Florentine, M., Buus, S., and Poulsen, T. (**1996**). ''Temporal integration of loudness as a function of level,'' J. Acoust. Soc. Am. **99**, 1633–1644.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.
Glasberg, B. R., and Moore, B. C. J. (**2002**). ''A model of loudness applicable to time-varying sounds,'' J. Audio Eng. Soc. **50**, 331–342.
Gockel, H., Moore, B. C. J., and Patterson, R. D. (**2002a**). ''Asymmetry of masking between complex tones and noise: The role of temporal structure and peripheral compression,'' J. Acoust. Soc. Am. **111**, 2759–2770.
Gockel, H., Moore, B. C. J., and Patterson, R. D. (**2002b**). ''Influence of component phase on the loudness of complex tones,'' Acust. Acta Acust. **88**, 369–377.
Goldstein, J. L. (**1967**). ''Auditory nonlinearity,'' J. Acoust. Soc. Am. **41**, 676–689.
Greenwood, D. D. (**1971**). ''Aural combination tones and auditory masking,'' J. Acoust. Soc. Am. **50**, 502–543.
Greenwood, D. D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.
Grimm, G., Hohmann, V., and Verhey, J. L. (**2002**). ''Loudness of fluctuating sounds,'' Acust. Acta Acust. **88**, 359–368.
Hellman, R. P. (**1985**). ''Perceived magnitude of two-tone-noise complexes: Loudness, annoyance, and noisiness,'' J. Acoust. Soc. Am. **77**, 1497–1504.
Houtgast, T. (**1972**). ''Psychophysical evidence for lateral inhibition in hearing,'' J. Acoust. Soc. Am. **51**, 1885–1894.
Houtgast, T. (**1974**). ''Lateral suppression in hearing,'' Ph.D. thesis, Free University of Amsterdam.
Howell, D. C. (**1997**). *Statistical Methods for Psychology* (Duxbury, Belmont, CA).
Howes, W. L. (**1979**). *Overall Loudness of Steady Sounds According to Theory and Experiment* (National Aeronautics and Space Administration, Washington, DC).
Javel, E. (**1981**). ''Suppression of auditory nerve responses: I. Temporal analysis, intensity effects and suppression contours,'' J. Acoust. Soc. Am. **69**, 1735–1745.
Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Lopez-Poveda, E. A., and Meddis, R. (**2001**). ''A human nonlinear cochlear filterbank,'' J. Acoust. Soc. Am. **110**, 3107–3118.
Lundeen, C., and Small, A. M. (**1984**). ''The influence of temporal cues on the strength of periodicity pitches,'' J. Acoust. Soc. Am. **75**, 1578–1587.
Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing*, 4th ed. (Academic, San Diego).
Moore, B. C. J., and Glasberg, B. R. (**1983**). ''Masking patterns of synthetic vowels in simultaneous and forward masking,'' J. Acoust. Soc. Am. **73**, 906–917.
Moore, B. C. J., and Oxenham, A. J. (**1998**). ''Psychoacoustic consequences of compression in the peripheral auditory system,'' Psychol. Rev. **105**, 108–124.
Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). ''A model for the prediction of thresholds, loudness and partial loudness,'' J. Audio Eng. Soc. **45**, 224–240.
Moore, B. C. J., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (**1988**). ''The shape of the ear's temporal window,'' J. Acoust. Soc. Am. **83**, 1102–1116.
Moore, B. C. J., Launer, S., Vickers, D., and Baer, T. (**1998**). ''Loudness of modulated sounds as a function of modulation rate, modulation depth, modulation waveform and overall level,'' in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London).
Moore, B. C. J., Vickers, D. A., Baer, T., and Launer, S. (**1999**). ''Factors affecting the loudness of modulated sounds,'' J. Acoust. Soc. Am. **105**, 2757–2772.
Patterson, R. D. (**1987**). ''A pulse ribbon model of monaural phase perception,'' J. Acoust. Soc. Am. **82**, 1560–1586.
Plack, C. J., and Moore, B. C. J. (**1990**). ''Temporal window shape as a function of frequency and level,'' J. Acoust. Soc. Am. **87**, 2178–2187.
Plack, C. J., and Oxenham, A. J. (**1998**). ''Basilar-membrane nonlinearity and the growth of forward masking,'' J. Acoust. Soc. Am. **103**, 1598–1608.
Pressnitzer, D., and Patterson, R. D. (**2001**). ''Distortion products and the pitch of harmonic complex tones,'' in *Physiological and Psychophysical Bases of Auditory Function*, edited by D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker, Maastricht).
Recio, A., and Rhode, W. S. (**2000**). ''Basilar membrane responses to broadband stimuli,'' J. Acoust. Soc. Am. **108**, 2281–2298.
Recio, A., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (**1998**). ''Basilar-membrane responses to clicks at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **103**, 1972–1989.
Relkin, E. M., and Doucet, J. R. (**1997**). ''Is loudness simply proportional to the auditory nerve spike count?'' J. Acoust. Soc. Am. **191**, 2735–2740.
Ritsma, R. J. (**1962**). ''Existence region of the tonal residue. I,'' J. Acoust. Soc. Am. **34**, 1224–1229.
Roberts, B., Glasberg, B. R., and Moore, B. C. J. (**2002**). ''Primitive stream segregation of tone sequences without differences in fundamental frequency or passband,'' J. Acoust. Soc. Am. **112**, 2074–2085.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). ''Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves, and response phases,'' J. Acoust. Soc. Am. **80**, 1364–1374.
Ruggero, M. A. (**1992**). ''Responses to sound of the basilar membrane of the mammalian cochlea,'' Curr. Opin. Neurobiol. **2**, 449–456.
Ruggero, M. A., Robles, L., Rich, N. C., and Recio, A. (**1992**). ''Basilar membrane responses to two-tone and broadband stimuli,'' Philos. Trans. R. Soc. Lond. B **336**, 307–315.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
Scharf, B. (**1978**). ''Loudness,'' in *Handbook of Perception, Volume IV. Hearing*, edited by E. C. Carterette and M. P. Friedman (Academic, New York).
Schouten, J. F. (**1940**). ''The residue and the mechanism of hearing,'' Proc. Kon. Akad. Wetenschap **43**, 991–999.
Schouten, J. F. (**1970**). ''The residue revisited,'' in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden, The Netherlands).
Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (**1982**). ''Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique,'' J. Acoust. Soc. Am. **72**, 131–141.
Stevens, S. S. (**1956**). ''Calculation of the loudness of complex noise,'' J. Acoust. Soc. Am. **28**, 807–832.
Stevens, S. S. (**1972**). ''Perceived level of noise by Mark VII and decibels (E),'' J. Acoust. Soc. Am. **51**, 575–601.
Stevens, S. S., and Greenbaum, H. B. (**1966**). ''Regression effect in psychophysical judgment,'' Percept. Psychophys. **1**, 439–446.
Stevens, S. S., and Guirao, M. (**1967**). ''Loudness functions under inhibition,'' Percept. Psychophys. **2**, 459–465.
Sumner, C. J., Lopez-Poveda, E. A., O'Mard, L. P., and Meddis, R. (**2002**). ''A revised model of the inner-hair cell and auditory-nerve complex,'' J. Acoust. Soc. Am. **111**, 2178–2188.
Terhardt, E. (**1974**). ''On the perception of periodic sound fluctuations (roughness),'' Acustica **30**, 201–213.
Warren, R. M., and Bashford, Jr., J. A. (**1981**). ''Perception of acoustic iterance: pitch and infrapitch,'' Percept. Psychophys. **29**, 395–402.
Zhang, C., and Zeng, F.-G. (**1997**). ''Loudness of dynamic stimuli in acoustic and electric hearing,'' J. Acoust. Soc. Am. **102**, 2925–2934.
Zwicker, E. (**1958**). ''Über psychologische und methodische Grundlagen der Lautheit,'' Acustica **8**, 237–258.
Zwicker, E. (**1977**). ''Procedure for calculating loudness of temporally variable sounds,'' J. Acoust. Soc. Am. **62**, 675–682.
Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics—Facts and Models* (Springer-Verlag, Berlin).
Zwicker, E., and Scharf, B. (**1965**). ''A model of loudness summation,'' Psychol. Rev. **72**, 3–26.
Zwicker, E., Fastl, H., and Dallmayr, C. (**1984**). ''BASIC-Program for calculating the loudness of sounds from their 1/3-oct. band spectra according to ISO 532B,'' Acustica **55**, 63–67.
Zwicker, E., Flottorp, G., and Stevens, S. S. (**1957**). ''Critical bandwidth in loudness summation,'' J. Acoust. Soc. Am. **29**, 548–557.

Suprathreshold effects of adaptation produced by amplitude modulation[a)]

Magdalena Wojtczak[b)] and Neal F. Viemeister
Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, Minnesota 55455

(Received 13 February 2003; revised 18 May 2003; accepted 23 May 2003)

This work extends the study of adaptation to amplitude modulation (AM) to the perception of highly detectable modulation. A fixed-level matching procedure was used to find perceptually equivalent modulation depths for 16-Hz modulation imposed on a 1-kHz standard and a 4-kHz comparison. The modulation depths in the two stimuli were compared before and after a 10-min exposure to a 1-kHz tone (adaptor) 100% modulated in amplitude at different rates. For modulation depths of 63% ($20 \log m = -4$) and smaller, the perceived modulation depth was reduced after exposure to the adaptor that was modulated at the same rate as the standard. The size of this reduction expressed as a difference between the post- and pre-exposure AM depths was similar to the increase in AM-detection threshold observed after adaptation. Postexposure suprathreshold modulation depth was not appreciably reduced when the modulation depth of the standard was large (approached 100%). A much smaller or no reduction in the perceived modulation depth was also observed when the modulation rates of the adaptor and the standard tone were different. The tuning of the observed effect of the adaptor appears to be much sharper than the tuning shown by modulation-masking results. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1593067]

PACS numbers: 43.66.Ed, 43.66.Mk [MRL]

I. INTRODUCTION

Dynamic changes in amplitude and frequency are important features used in perception of speech, music, and other sounds. For this reason, numerous studies have investigated mechanisms that might be involved in the processing of these dynamic changes. In particular, it has been suggested that there exist channels that specialize in the processing of specific features of sound. Evidence for existence of the ''feature channels'' is from experiments showing that prolonged exposure to a specific feature of sound, such as amplitude or frequency modulation, decreases sensitivity to that feature but does not influence sensitivity to other features or affects them to a much lesser extent. The observed decrease in sensitivity is presumed to reflect a decrease in response due to selective adaptation of neural channels tuned to a specific feature (Kay and Matthews, 1972; Regan and Tansley, 1979; Gardner and Wilson, 1979; Tansley and Suffield, 1983).

Later studies have cast doubt on such interpretation of the results obtained from experiments that were presumed to show the effects of adaptation. Wakefield and Viemeister (1984) measured psychometric functions for detection of FM upsweeps before and after exposure to adapting upsweeps. They found that threshold elevation observed after the exposure (expressed in terms of percent of the pre-exposure threshold frequency sweep) corresponds to a relatively small change in performance, and that this change could be explained in terms of nonsensory factors such as the use of an ''inappropriate'' reference for the detection of the signal. Moody *et al.* (1984) observed strong training effects in the detection of FM upsweeps. They found that elevated thresholds observed after long exposure decreased more rapidly with the amount of training than pre-exposure thresholds. After a sufficient number of experimental blocks no difference between pre- and postexposure thresholds was observed. Thus, the effect that was interpreted as selective adaptation and was taken as a strong support for the existence of a sensory channel sensitive to frequency sweeps, disappeared with training. In the study of Regan and Tansley (1979), which used tones modulated sinusoidally in amplitude and frequency, each subject completed as many as 90 sessions but no strong training effects that would eliminate the difference between pre- and postexposure thresholds were mentioned.

Support for the notion that channels specializing in the processing of AM may exist in the auditory system was provided by studies that recorded neural responses to stimuli with fluctuating envelopes. Langner and Schreiner (1988) found a map of best modulation frequencies in the inferior colliculus of the cat that resembled the tonotopic map of responses to auditory frequencies at different stages of auditory processing. Consistent with the tuning in the modulation-frequency domain, it was found that modulation-masking patterns measured psychophysically show a clear peak around the modulation frequency of the masker (Bacon and Grantham, 1989; Houtgast, 1989). These masking results inspired development of models that use a bank of modulation filters to process envelope fluctuations (Dau *et al.*, 1996; Ewert and Dau, 2000). Existence of modulation channels that selectively process different modulation rates implies that if these channels adapt as a result of prolonged stimulation, the observed adaptation should be modulation-rate specific, i.e., an adapting sound modulated at a given rate would affect sensitivity to similar rates but not to rates remote from

[a)]A portion of this work was presented at the 22nd ARO meeting [M. Wojtczak and N. F. Viemeister (1999), ''Adaptation produced by amplitude modulation,'' p. 59].

[b)]Electronic mail: wojtc001@umn.edu

the adapting rate. We are not aware of any study that systematically examined the effect of the adaptor as a function of the difference between the adapting and test modulation rate.

Richards *et al.* (1997) used two different modulation rates for the adapting tone, and found that when the adaptor and the test tone were modulated at the same rate, threshold for detecting AM was elevated by about 7.5 dB (20 log *m*). In contrast, only about 1-dB elevation was observed when the adapting modulation rate (56 Hz) differed substantially from the test rate (16 Hz). Thus, the effect of the modulated adaptor appears to be modulation-rate specific. Richards *et al.* also measured pre- and postadaptation AM detection for different frequency separations between the adapting carrier and the test carrier, both modulated at the same rate. For a carrier frequency of 1500 Hz of the adapting tone, no effect of adaptation was observed for test tones with carrier frequencies two octaves below and one octave above the frequency of the adaptor. This suggests that AM detection threshold is elevated by an adapting sound when the carrier frequency and the modulation rate of the adaptor are similar to those of the test tone.

All experiments aimed at investigating the effects of adaptation to modulation have measured differences between pre- and postexposure thresholds for the detection of modulation. Detection threshold is a measure of sensitivity, and as such it provides information about the limits of the auditory processing. More directly related to the daily tasks that the auditory system performs is information about how prolonged exposure to modulated stimuli affects perception of well-detectable (suprathreshold) modulations. The present study used a subjective matching procedure to investigate this issue. In experiment 1, the adapting stimulus and the test stimulus had the same carrier frequencies and modulation rates and thus, presumably, they were processed by the same ''feature channel.'' The comparison had a carrier frequency two octaves higher than the adapting tone and a modulation rate identical to the modulation rate of the adaptor. Because of the large separation between the carrier frequencies, the frequency channel processing the comparison was unaffected by the adapting stimulus, as was shown by Richards *et al.* and confirmed in our experiment 1. Thus, it was reasoned that if the adaptor has an effect on the perceived modulation depth in the test tone, changes in the subjectively equivalent modulation depth of the comparison would be observed. In experiment 2, different modulation rates of the adapting stimulus were used to examine whether the effect of the adaptor is modulation-rate specific.

II. EXPERIMENT 1: THE EFFECT OF THE SAME-RATE ADAPTING AM

A. Stimuli and procedure

A fixed-level modulation-matching procedure was used to study the effect of prolonged exposure to AM. Listeners were asked to compare modulation depths between a 1-kHz AM standard and a 4-kHz AM comparison, both presented for 500 ms at 60-dB SPL and both modulated at a rate of 16 Hz. The adapting stimulus was a 60-dB 1-kHz tone, 100% modulated in amplitude by a 16-Hz sinusoid. The adaptor

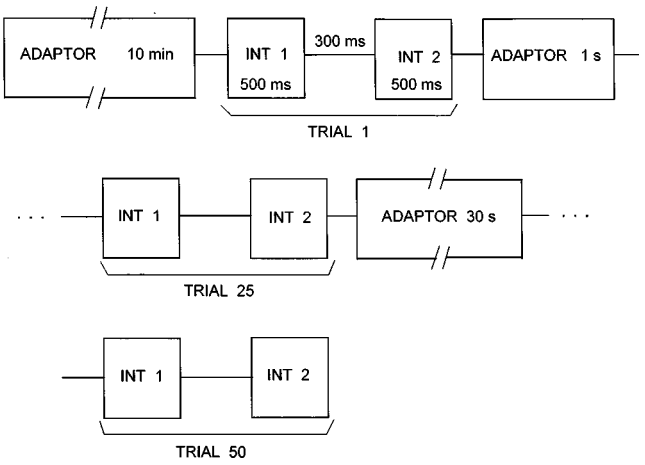


FIG. 1. Schematic illustration of the experimental procedure used with the adapting stimulus.

was played for 10 min before the testing began. A warning signal was displayed on a computer screen 30 s before the first trial was presented.

Three modulation depths of the 1-kHz standard and eight modulation depths of the 4-kHz comparison were used within a block. The standard modulation depths were −4, −6, and −8 dB (20 log *m*) in one set of blocks, and 0, −1, and −2 dB in another. For the three smaller standard modulation depths, the comparison modulation depths were selected so that they bracketed the point of subjective equality (PSE) estimated for each subject during pilot testing. Blocks with the three larger modulation depths used comparison modulation depths from −14 through 0 dB in steps of 2 dB because this range covered the PSE estimates observed in pilot runs for all subjects. On each trial, one standard and one comparison modulation depth were drawn randomly for presentation. The order in which the 1-kHz standard and the 4-kHz comparison were presented within a trial was also random. The listener made a judgment as to whether the first or the second interval contained stronger fluctuations.[1] Because the task was subjective in nature, no feedback indicating the correct response was provided. The schematic illustration of the experimental procedure in the adapting condition is presented in Fig. 1. After each trial, the adapting sound was reintroduced for 1 s, and after the first 25 trials, testing was interrupted and the adapting stimulus was played for 30 s to reinforce possible adaptation to modulation. Each block consisted of 50 trials. Each subject completed 25 blocks that were used to estimate the PSE.

The same fixed-level matching procedure was used to find the pre-exposure PSEs for amplitude modulation imposed on the 1- and 4-kHz carriers. Prolonged exposure to the unmodulated adaptor was not used because other studies have demonstrated that silence is an equivalent reference for the adaptation condition, since the unmodulated tone has no effect on AM detection (Regan and Tansley, 1979; Tansley and Suffield, 1983), and using such a tone would unnecessarily extend the duration of the experiment.

To compare the effect of the adapting stimulus on the perceived suprathreshold AM with its effect on modulation detection, pre- and postexposure AM detection thresholds

were measured using an adaptive 2-down, 1-up, three-interval forced-choice (3IFC) procedure. In this task, a visual feedback indicating the correct response was provided. The adaptor and the standard tone were the same as in the matching task, except that the modulation depth of the standard was varied adaptively to obtain threshold. The step sizes used in the adaptive procedure were 4 dB for the first four reversals and 2 dB for the remaining eight reversals. AM-detection threshold from a single run was computed based on the last eight reversals. The final threshold estimate was obtained by averaging thresholds from three separate runs. Pre- and postexposure AM-detection thresholds were also measured with the 4-kHz test tone. This was done to assess whether the 1-kHz AM adaptor affects the processing of AM in the frequency channel tuned to 4 kHz.

The stimuli were generated digitally with a sampling rate of 44.1 kHz on a NeXT computer equipped with a 16-bit D/A converter. They were attenuated to the desired level and directed to a Sony MDR-V6 earphone for monaural presentation.

B. Subjects

Four listeners participated in the study. Their hearing thresholds were normal as determined by comparing the thresholds measured at octave frequencies between 250 and 8000 Hz to the lab norms. Not all listeners were run in every condition. Three listeners were compensated for their services. One listener (S1) was the first author.

C. Results

Pre- and postexposure thresholds for detecting AM confirmed the finding by Richards *et al.* (1997) that the adapting modulated tone increases thresholds by about 8 dB for the standard with a carrier frequency identical to that of the adaptor (1 kHz), but does not affect sensitivity to AM measured for the standard with a carrier frequency two octaves above (4 kHz) that of the adaptor.

For suprathreshold AM, PSEs were estimated from functions which showed the proportion of responses, for which the comparison was judged more modulated than the standard, plotted for each of the eight comparison depths used in a given set of blocks. Because three standard modulation depths were used within each block, three functions were obtained, one for each modulation depth of the standard.

Figure 2 shows an example of pre- (left panel) and postexposure results (right panel) for one listener. Data for each standard modulation depth were separately fitted with a sigmoidal function and the PSE was estimated based on this fit

$$P(m_c > m_{st}) = \frac{1}{1+\left(\frac{m_c}{m_{c(0.5)}}\right)^{\alpha}}, \tag{1}$$

where P is the predicted proportion of responses where the modulation depth of the comparison (m_c) sounded like a larger modulation depth than that of the standard (m_{st}). The parameters $m_{c(0.5)}$ and α were varied until the sum of squared deviations between the predicted P values and the data was minimum.[2] The value of $m_{c(0.5)}$ that produced the best fit was taken as the PSE between the standard and the comparison modulation depths. This value represented the predicted modulation depth of the comparison that would be evaluated by the listener as greater than the modulation depth of the standard on 50% of trials.

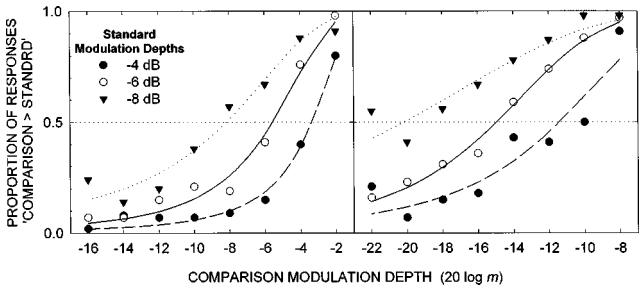


FIG. 2. The ordinate is the proportion of times the listener chose the comparison modulation depth, plotted on the abscissa, as larger than the standard modulation depth. Different symbols show data for different standard modulation depths. The left panel shows pre-exposure matching results; the right panel shows results obtained after exposure to a 10-min 1-kHz tone that was 100% modulated at a rate of 16 Hz.

The pre-exposure PSEs fall very close to the standard modulation depths, −4, −6, and −8 dB (left panel in Fig. 2). This result implies that the same physical modulation depth imposed on a 1-kHz carrier and a 4-kHz carrier produces a similar percept of the depth of fluctuations. Other listeners showed very similar results that indicated no dependence of the perceived modulation depth on the carrier frequency, at least for the carrier frequencies of 1 and 4 kHz tested within this study. Paired t-tests performed separately for each standard modulation depth on the data from all the listeners show no significant difference ($p<0.01$) between the perceived modulation depths in the 1- and 4-kHz carriers prior to adaptation.

The three functions in Fig. 2 are shifted relative to one another with the leftmost function representing results for the smallest standard modulation depth and the rightmost function representing results for the largest standard modulation depth. This suggests that the listener actually used the standard to make judgments about the fluctuation depth in the comparison.

The right panel of Fig. 2 shows data obtained after the listener has been exposed to an adapting AM tone. All three functions are shifted toward smaller modulation depths, suggesting that the adapting sound had an effect of reducing the perceived fluctuations. For each of the three standard modulation depths, the PSE was smaller than the standard modulation depth. Paired t-tests performed separately for each standard modulation depth using data from all subjects revealed a significant effect due to the exposure to the adapting stimulus (obtained t-values corresponded to $p>0.01$).

Similar analysis was performed on the results obtained from the set of blocks that used the standard modulation depths of 0, −1, and −2 dB. In this case, all three listeners who performed the task showed a very small shift in the perceived modulation toward smaller depths, indicating very little or no reduction of the perceived depth of fluctuations

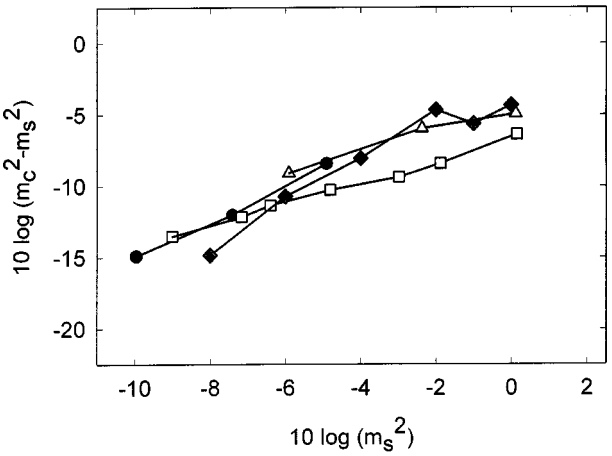


FIG. 3. Modulation-depth discrimination thresholds estimated from the curves fitted to pre-exposure AM matching data (diamonds). The estimated thresholds for the modulation depth of 0, −1, and −2 dB correspond to detecting a decrement, whereas the thresholds for −4, −6, and −8 dB, correspond to detecting an increment in modulation depth. For comparison, data from objective modulation-depth discrimination tasks are plotted: open squares (von Fleischer, 1980), open triangles (Ozimek and Sek, 1988), and filled circles (Wakefield and Viemeister, 1990). (From Fig. 5 in Wakefield and Viemeister.)

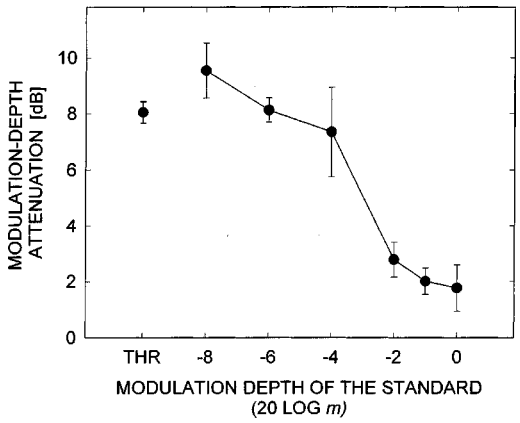


FIG. 4. Attenuation of the perceived modulation depth computed from the difference between the pre- and postexposure PSE, averaged across listeners. The leftmost unconnected point shows the difference between pre- and postadaptation AM detection thresholds. The vertical bars show ± one standard deviation from the mean.

after the 10-min exposure to the 100% modulated tone. Paired t-tests showed that for these standard modulation depths the effect of the adaptor was not significant ($p < 0.01$).

Since before adaptation, no significant difference was found between the standard modulation depth in the 1-kHz carrier and the perceptually equivalent modulation depth in the 4-kHz carrier, the pre-exposure functions obtained from the modulation-matching procedure could be used to estimate modulation-depth discrimination thresholds. These thresholds were computed by solving Eq. (1) for the modulation depth of the comparison (m_c) and assuming $P(m_c > m_{st}) = 0.707$ for the standard modulation depths of −4, −6, and −8 dB, and $P(m_c > m_{st}) = 0.293$ for the standard modulation depths of 0, −1, and −2 dB. Thus, for the three lower modulation depths, depth discrimination was estimated for an increment in the modulation depth, while for the large modulation depths, depth discrimination thresholds were estimated for a decrement in modulation depth. Figure 3 shows modulation-depth discrimination thresholds estimated from the preadaptation matching functions (diamonds) presented along with published modulation-depth discrimination thresholds for increments and decrements in modulation depth (von Fleischer, 1980; Ozimek and Sek, 1988; Wakefield and Viemeister, 1990), for a range of the standard modulation depths relevant to the present data. The thresholds for modulation-depth discrimination estimated from the subjective AM matching procedure fall very close to the thresholds measured directly in the objective modulation-depth discrimination tasks. This supports the validity of the subjective procedure and also provides further support for the inference that prior to adaptation, equal physical modulation depths in the 1-kHz and 4-kHz carriers are perceptually equivalent.

Comparison of the pre- and postexposure PSEs allows for estimation of the attenuation in the perceived (output) modulation depth. This can be evaluated indirectly by computing the reduction in terms of the physical (input) modulation depth. For each subject and every standard modulation depth, a difference between the PSE obtained before the exposure to the adapting stimulus and the PSE obtained after the 10-min exposure was computed to represent the reduction in perceived modulation depth. Attenuations averaged across listeners, plotted as a function of the modulation depth of the standard are shown in Fig. 4.

For standard modulation depths of −4 dB and smaller, an average reduction between 7 to 9 dB was observed. There was a tendency for the attenuation to decrease with increasing standard modulation depth. A shift in AM detection threshold, shown by the unconnected point on the left, also falls within that range, i.e., after a 10-min exposure to the adapting modulated tone, the modulation depth at AM-detection threshold was 8 dB higher than the modulation depth measured pre-exposure. As the modulation depth of the standard approaches the modulation depth of the adapting stimulus (0 dB), the reduction in the perceived modulation depth becomes statistically insignificant. For standard modulation depths of 0, −1, and −2 dB the computed attenuation is only around 2 dB, with a slight tendency to decrease with increasing modulation depth.

For the three lower modulation depths of the standard, the postexposure PSEs were computed for each individual listener based on separate successive sets of five blocks (the results are not shown here). This was done to make sure that no major shifts in the obtained PSEs occurred with increasing amount of training. No systematic shifts of the PSE toward the modulation depth equal to that of the standard were observed for any of the subjects and any of the three standard modulation depths. In fact, the PSEs exhibited a high degree of stability across all sets of blocks.

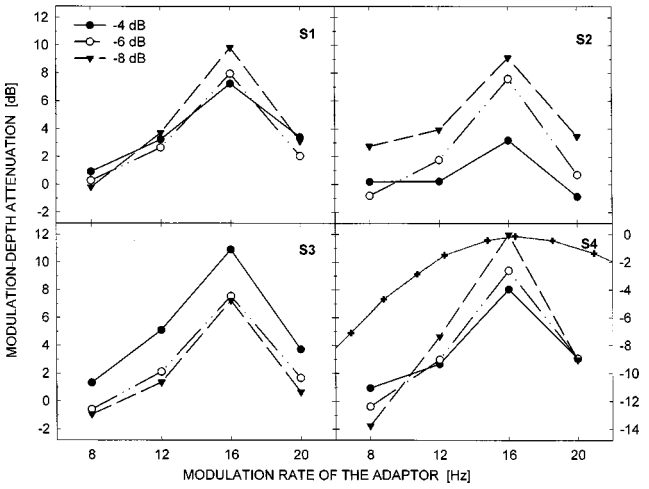


FIG. 5. Attenuation of the perceived modulation depth, estimated from the difference between the pre- and postexposure PSE, plotted as a function of the modulation rate of the adapting stimulus. The modulation rate of the standard was 16 Hz. Each panel shows results for a different subject. Different symbols and types of connecting lines correspond to different modulation depths of the standard. The cross-hair symbols connected by the gray solid line (bottom right panel) represent the transfer characteristic of a 16-Hz modulation filter with a Q-value of 2. The rightmost *y* axis shows the attenuation by the modulation filter.

III. EXPERIMENT 2: TUNING IN SUPRATHRESHOLD EFFECTS OF ADAPTATION

A. Method

In this experiment, three different modulation rates (8, 12, and 20 Hz) of the adaptor whose carrier frequency was 1 kHz were used to examine rate specificity of the effects observed in experiment 1. The 1-kHz standard and the 4-kHz comparison were modulated at 16 Hz. Only the three lower modulation depths of the standard (-4, -6, and -8 dB), for which the effect of the adaptor was significant, were used to study tuning. The method, the equipment, and the subjects used were the same as in experiment 1.

B. Results

Figure 5 shows differences between the pre- and postexposure PSEs plotted as a function of the modulation rate of the adapting tone, for each individual listener. These differences, representing the magnitude of adaptation, are the largest when the modulation rate of the adapting tone is the same as that of the standard. When the adapting modulation rate differs from the rate of modulation in the standard, prolonged exposure to the adaptor produces a smaller reduction of the perceived standard modulation depth.

As shown by all these individual data, even a difference in modulation rate between the adaptor and the standard as small as 4 Hz causes a substantial reduction in the effect of the adaptor. The ''tuning'' of the effect of the adapting tone appears symmetrical around the modulation rate of the standard. For the adapting rate of 8 Hz, no consistent effect of the adapting stimulus is observed. As will be discussed later, the lower right panel also shows the transfer characteristic of a modulation filter centered on 16 Hz, with a Q-value of 2 (cross-hair symbols).

IV. DISCUSSION

This study examined whether adaptation, a phenomenon that has been observed for threshold modulation depths, is general, i.e., whether it is also present in the processing of well-detectable fluctuations. Large fluctuations are important in everyday listening and thus, studying mechanisms that affect their perception is crucial for characterizing the envelope processing in the auditory system. The experiments performed within this study used a 10-min adapting modulated stimulus. While in everyday listening we may not be exposed to a constant-rate modulations continuously over such a long period of time, previous studies have shown that adaptation asymptotes after about 10 min of exposure. Thus, the magnitude of adaptation reported in this study should be the maximum adaptation produced by AM tonal stimuli.

The results indicate that the magnitude of adaptation produced by a 100% AM adaptor depends on the modulation depth of the standard. This indicates that the effect of the adaptor is not equivalent to a simple, fixed attenuation of the envelope fluctuations of the standard: As the modulation depth of the standard approaches that of the adaptor, the amount of attenuation of the modulation becomes negligible (Fig. 4). One explanation is that the attenuation is, in fact, fixed, but that because of threshold/floor effects the perceived modulation of the standard is not appreciably affected at large modulation depths of the standard. More specifically, if the effect of the adaptor is primarily to reduce the effective depth of the envelope minima, then for large modulation depths these minima may be below threshold, either from forward masking by the envelope maxima or the absolute detection threshold for the minima. Thus, an increase in the ''internal'' minima produced by adaptation may have a negligible effect on perceived modulation depth at large modulation depths because the elevated minima are obscured by threshold/floor effects. This conjecture relies upon a differential effect of adaptation on envelope minima and maxima. How such a differential effect might occur is unclear.

Another explanation is in terms of ''intensity''-dependent adaptation. If the attenuation of the perceived modulation depth is due to adaptation, it is possible that the amount of adaptation would decrease as the intensity of the adapted stimulus approaches the intensity of the adaptor. (Intensity is meant here as the magnitude of the parameter that is subjected to adaptation.)

The crucial issue is whether or not the reduction of the perceived modulation depth after prolonged exposure to a modulated tone with a similar carrier frequency and modulation rate is indeed a result of fatigue (adaptation) of neural channels that specialize in the processing of this modulation. As mentioned in the Introduction, Wakefield and Viemeister (1984) suggested that increased thresholds for detection of frequency sweeps might result from the use of an improper reference after exposure to a highly detectable adapting frequency sweep. Their conclusion was based on the observation that the increase in threshold corresponded to a change in performance from 75% to only 65%. In the present study, psychometric functions for AM detection were not measured. However, it has been shown that for AM detection, $d' = k\,m^2$ (Moore and Sek, 1992; Edwards and Viemeister,

1994). The 8-dB increase in threshold after exposure thus corresponds to a decrease in d' of 0.8 log units, a factor of 6.3. In terms of a change in performance, this is a large effect [d' would decrease from 1.28 to 0.2; $P(C)$ would decrease from 70.7% to 39%]. It seems unlikely that such a large change results from the use of an improper reference. It is possible that neural fatigue or adaptation is responsible for the elevation of AM-detection threshold. The same mechanism is likely to cause a reduction in the perceived suprathreshold modulation depth.

Within the present study, postexposure AM detection thresholds were not measured for modulation rates of the adaptor that differed from the modulation rate of the test stimulus. However, the results in Fig. 5 show clearly that the reduction in the perceived suprathreshold modulation is much less for adaptor rates differing by only 4 Hz from the test modulation rate. This tuning in the effect of the adaptor resembles the tuning observed in modulation masking (Bacon and Grantham, 1989; Takahashi and Bacon, 1992). A direct comparison of sharpness of the tuning between the present data and the modulation-masking data is difficult. Bacon and Grantham used a signal modulation rate of 16 Hz, but only relatively remote masker modulation rates (4 and 64 Hz) were used in their study to measure the amount of masking. For these two masker rates, negligible or no masking was observed. Takahashi and Bacon used a range of masker modulation rates but they used only a signal modulation rate of 8 Hz. Their data for young adults show that maskers with modulation rates one octave below (4 Hz) and slightly more than a half-octave above (12 Hz) the signal rate still produce about 12 to 15 dB of masking. The data in Fig. 5 show no effect of adaptation for the adaptor rate an octave below the modulation rate of the standard, and only a very small effect (less than 4 dB) for the adaptor modulation rate only 0.3 octave above the rate of the standard. Under the assumption that the internal amplitude of the standard modulation after adaptation is proportional to the amplitude of the adapting modulator at the output of the ''adaptation filter'' in the modulation frequency domain, the derived tuning of the hypothetical adaptation filter is sharper than tuning of the modulation filter derived based on modulation-masking data. To illustrate this, the transfer characteristic of a modulation filter centered on 16 Hz, with a Q-value of 2 is plotted with the adaptation data in Fig. 5. Typically, modulation filters symmetrical on a log frequency scale with a Q-value of 2 or 1 are used to account for modulation masking (e.g., Ewert and Dau, 2000).

The differences in the observed tuning between modulation-masking and modulation adaptation data may reflect two different levels of modulation processing. For example, a sharper tuning in adaptation would be observed if adaptation followed the modulation-filtering process and if the magnitude of adaptation decreased with decreasing level at the output of the modulation filter. Different tuning may also result from the different nature of the detection (objective) and matching (subjective) tasks. It should be noted, however, that the magnitude of adaptation estimated by comparing pre- and postexposure AM detection thresholds was very similar to the magnitude of adaptation estimated from the subjective matching task for three lower modulation depths of the standard. This result suggests that the subjective nature of the task probably was not responsible for the observed differences in tuning.

V. FINAL REMARKS

In summary, the size of AM-detection threshold elevation suggests that the decreased sensitivity likely does not reflect a nonsensory effect such as a change in the detection criterion or the use of the adaptor rather than a comparison as a reference for detection. The reduction in the perceived suprathreshold AM, and the robustness of the results (lack of considerable training effects) appears to support the notion that there may be a primary neural basis for the observed effects. Tuning of the effect is generally consistent with the idea that there exist ''hard-wired'' channels tuned to specific modulation rates, and that those channels are associated with specific auditory-frequency channels. The difference between the degree of tuning observed for the effect of adaptation versus that for modulation masking is probably due to different levels at which the two processes occur.

The similarity between the amount of adaptation seen in the detection and the matching data at the lower modulation depths, and the size and robustness of the effect argue for its sensory basis. Although we believe that the adaptation effects shown in these experiments are primarily sensory, a cognitive component, based on perceived adaptor-standard similarity, cannot be dismissed. Since the matching task is subjective and requires judgments across different dimensions (carrier frequency), fairly complex cognitive strategies are likely to be involved. Further investigation using objective psychophysical techniques should help to evaluate the sensory basis for adaptation to suprathreshold AM.

ACKNOWLEDGMENTS

We thank Dr. Mark Stellmack and Dr. Martin Rickert for stimulating discussions about this research. We also thank two anonymous reviewers for their helpful and constructive comments. This work was supported by Grant No. DC 00683 from NIDCD.

[1]After several unsuccessful attempts at using adaptive matching procedures, a fixed-level procedure with multiple standards and comparisons was adopted. It was observed when using matching procedures, that because of a difference in the perceptual quality between the standard and the comparison, some listeners did not use the standard as a reference but instead, provided consistent responses by apparently matching each comparison to some memorized ''internal standard.'' In the fixed-level procedure, on a given trial the modulation depths of the standard and the comparison were chosen randomly, thus encouraging the listener to compare the two stimuli. A monotonic relation between the pre-exposure standard modulation depths and the perceptually equivalent comparison modulation depths confirmed that the standard modulation depths were used as references in evaluating the perceived modulation depths in the comparison.

[2]Sometimes pre-exposure comparisons during runs that used standard modulation depths of 0, −1, and −2 dB did not yield even one comparison modulation depth producing larger than 0.5 proportion of responses where the comparison was judged to be more modulated. In those cases, the sigmoidal function defined by Eq. (1) was fit to the data and the PSE was estimated by extrapolation of the function toward larger modulation depths with the constraint that the modulation depth in dB could not exceed zero. In these cases, the portion of the matching function representing values

above 0.5 was very steep. In the postexposure condition, there was always at least one modulation depth of the comparison that was evaluated as producing stronger perceived fluctuations than the standard on more than 50% of trials.

Bacon, S. P., and Grantham, D. W. (**1989**). ''Modulation masking: Effects of modulation frequency, depth, and phase,'' J. Acoust. Soc. Am. **85**, 2575–2580.

Dau, T., Püschel, D., and Kohtrausch, A. (**1996**). ''A quantitative model of the effective signal processing in the auditory system. I. Model structure,'' J. Acoust. Soc. Am. **99**, 3615–3622.

Edwards, B. W., and Viemeister, N. F. (**1994**). ''Modulation detection and discrimination with three-component signals,'' J. Acoust. Soc. Am. **95**, 2202–2212.

Ewert, S. D., and Dau, T. (**2000**). ''Characterizing frequency selectivity for envelope fluctuations,'' J. Acoust. Soc. Am. **108**, 1181–1196.

Gardner, R. B., and Wilson, J. P. (**1979**). ''Evidence for direction-specific channels in the processing of frequency modulation,'' J. Acoust. Soc. Am. **66**, 704–709.

Houtgast, T. (**1989**). ''Frequency selectivity in amplitude-modulation detection,'' J. Acoust. Soc. Am. **85**, 1676–1680.

Kay, R. H., and Matthews, D. R. (**1972**). ''On the existence in human auditory pathways of channels selectively tuned to the modulation present in frequency modulated tones,'' J. Physiol. (London) **225**, 657–667.

Langner, G., and Schreiner, C. (**1988**). ''Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanism,'' J. Neurophysiol. **60**, 1799–1822.

Moody, D. B., Cole, D., Davidson, L. M., and Stebbins, W. C. (**1984**). ''Evidence for a reappraisal of the psychophysical selective adaptation paradigm,'' J. Acoust. Soc. Am. **76**, 1076–1079.

Moore, B. C. J., and Sek, A. (**1992**). ''Detection of combined frequency and amplitude modulation,'' J. Acoust. Soc. Am. **92**, 3119–3131.

Ozimek, E., and Sek, A. (**1988**). ''AM difference limens for noise bands,'' Acustica **66**, 153–160.

Regan, D., and Tansley, B. W. (**1979**). ''Selective adaptation to frequency-modulated tones: Evidence for an information-processing channel selectively sensitive to frequency changes,'' J. Acoust. Soc. Am. **65**, 1249–1257.

Richards, V. M., Buss, E., and Tian, L. (**1997**). ''Effects of modulator phase for comodulation masking release and modulation detection interference,'' J. Acoust. Soc. Am. **102**, 468–476.

Takahashi, G. A., and Bacon, S. P. (**1992**). ''Modulation detection, modulation masking, and speach understanding in noise in the elderly,'' J. Speech Hear. Res. **35**, 1410–1421.

Tansley, B. W., and Suffield, J. B. (**1983**). ''Time course of adaptation and recovery of channels selectively sensitive to frequency and amplitude modulation,'' J. Acoust. Soc. Am. **74**, 765–775.

von Fleischer, H. (**1980**). ''Subjektive grösse von unterschieden im amplituden-modulationsgrad von sinustönen,'' Acustica **46**, 31–37.

Wakefield, G. H., and Viemeister, N. F. (**1984**). ''Selective adaptation to linear frequency-modulated sweeps: Evidence for direction-specific FM channels?,'' J. Acoust. Soc. Am. **75**, 1588–1592.

Wakefield, G. H., and Viemeister, N. F. (**1990**). ''Discrimination of modulation depth of sinusoidal amplitude modulation (SAM) noise,'' J. Acoust. Soc. Am. **88**, 1367–1373.

On the detection of dispersion in the head-related transfer function

Zachary A. Constan[a)]
Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824

William M. Hartmann[b)]
Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215and Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824 [c)]

(Received 22 November 2002; revised 21 May 2003; accepted 21 May 2003)

Because of dispersion in head-related transfer functions (HRTFs), the interaural time difference (ITD) varies with frequency. This physical effect ought to have consequences for the size or shape of the auditory image of broadband noise because different frequency regions of the noise have different ITDs. However, virtual reality experiments suggest that human listeners are insensitive to head-related dispersion. The experiments of this article test that suggestion by experiments that isolate dispersion from amplitude effects in the HRTF and attempt to optimize the opportunity for detecting it. Nevertheless, the experiments find that the only effect of dispersion is to shift the lateralization of the auditory image. This negative result is explained in terms of the cross-correlation function for head-dispersed noise. Although the broad-band cross-correlation function differs considerably from 1.0, the cross-correlation functions within bands characteristic of auditory filters do not. A detailed study of the lateralization shifts show that the experimental shifts can be successfully calculated as an average of stimulus ITDs as weighted by Raatgever's frequency-weighting function (Thesis, Delft, The Netherlands, 1980). © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1592159]

PACS numbers: 43.66.Pn, 43.66.Ba, 43.66.Qp [LRB]

I. INTRODUCTION

As a sound wave approaches a listener's ear, it is diffracted by the listener's head. Diffraction causes the sound wave to be filtered, as characterized by the head-related transfer function (HRTF). The HRTF is dispersive, i.e., the phase shift is not a linear function of the frequency. Instead, the phase shift increases more slowly than linearly with increasing frequency. Because the ratio of the phase shift to frequency is the phase delay, the phase delay is not constant but decreases with increasing frequency. It is as though high-frequency sound waves traveled faster around the head than low-frequency waves.

The effects of diffraction by the head, including dispersion, can be well approximated by a diffraction formula for the sound pressure on the surface of a sphere. The formula for an incident plane wave from a distant source (Kuhn, 1977, 1979, 1982, 1987; Kuhn and Guernsey, 1983) appears as Eq. (A1) in the Appendix. This series formula gives the complex sound pressure as a function of azimuth, θ_i, frequency, f, and head radius, a. The frequency and radius enter only as the dimensionless variable $c/(2\pi af)$, where c is the speed of sound. For an incident wave from a nearby source, a corresponding series formula was given by Brungart (1999), Brungart and Rabinowitz (1999), and Brungart *et al.* (1999). In that case the distance from the source to the head is another variable.

In connection with binaural hearing, where two ears are involved, it is evident that diffraction will have an effect on *interaural* properties. In particular, dispersion affects the interaural time difference (ITD), defined here as the interaural phase delay. The ITD is an important cue used by listeners to determine the azimuth of a sound source (Strutt, 1907), and dispersion causes the ITD to depend, not only on the azimuth, but also on the frequency.

Figure 1 shows the ITD as a function of frequency for seven different azimuths of incidence (measured from the forward direction) as calculated from the spherical-head formula. The high-frequency limit of the ITD, shown by filled squares on the right of Fig. 1, is two-thirds of the low-frequency limit, shown by open circles on the left. A comparison between the curves and the high- and low-frequency limits shows that most of the change in ITD occurs in a rather narrow frequency region. For frequencies around 1 kHz, where much of the dispersion occurs, the wavelengths of the sound waves are large compared to head anatomical details and very large compared to details of the pinna. Therefore, as observed by Kuhn (1982) or Brungart and Rabinowitz (1999), the ITDs calculated from the spherical-head formula turn out to be in reasonable agreement with ITDs measured on real heads or on artificial heads such as KEMAR.

The question posed in this article is whether human listeners are sensitive to the dispersion created by the head. One line of reasoning suggests that dispersion should, in-

a)Present address: Department of Physics, Albion College, Albion, MI 49224.
b)Author to whom correspondence should be addressed. Electronic mail: hartmann@pa.msu.edu
c)Present address.

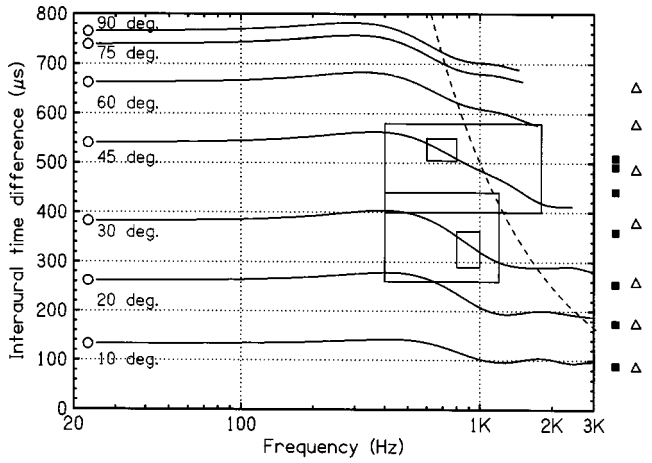


FIG. 1. Head-related dispersion according to Kuhn's spherical-head model with the ears at antipodes. This graph plots the ITD as a function of log frequency for sound incident on a spherical head from various azimuths, measured from the forward direction (nose). The dashed curve indicates ITDs that equal one-half period for a particular frequency. Points to the left of this line are lateralized as expected according to the sign of the ITD. Points on this line have ambiguous lateralization. The open circles on the left show the low-frequency limit, ITD$=(3a/c)\sin(\theta_i)$. The filled squares on the right show the high-frequency limit, ITD$=(2a/c)\sin(\theta_i)$. The open triangles on the right show the predictions of the Woodworth formula, ITD $=(a/c)[\sin(\theta_i)+\theta_i]$, included for reference only. Rectangles show the frequency ranges used in experiment 1: Broadband: 20–3000 Hz. midband 30°: 400–1200 Hz. Midband 45°: 400–1800 Hz. Narrowband 30°: 800–1000 Hz. Narrowband 45°: 600–800 Hz.

deed, be perceptible in human binaural hearing. Figure 1 shows that the ITD depends strongly on frequency for an incident azimuth of 30°. The ITD changes by about 80 μs (from 399 to 320 μs) as the frequency increases from 500 to 1000 Hz. By comparison, it is known that listeners can detect a change in the ITD as small as 10 μs when the baseline ITD is 400 μs (Domnitz, 1973; Domnitz and Colburn, 1977). The ITD shown in Fig. 1 is the *phase delay*, which describes the delay in the fine structure of a narrowband signal. Listeners are sensitive to such ITDs in waveform fine structure at frequencies near 800 Hz, where the dispersion occurs for 30°, and even more sensitive at lower frequencies where dispersion occurs for larger azimuths. Therefore, the variation in ITD caused by head dispersion is clearly in the range of phase delays that are perceptible.

By numerical differentiation of Eq. (A1) in the Appendix we learned that head dispersion causes a frequency variation in *group delay* about 50% greater than the variation in phase delay. The group delay describes the delay of peaks and valleys in the envelope of the signal. Therefore, the interaural comparison of envelope features would be at least as inconsistent across frequency as the comparison of waveform features.

The frequency dependence of the ITD leads to binaural incoherence, characterized by a normalized interaural cross-correlation function with a peak that is less than unity. In turn, binaural incoherence leads to a broadened auditory image, an effect known as apparent source width (ASW) (Blauert and Lindemann, 1986; Ando, 1998). Listeners are extremely sensitive to the ASW introduced by small amounts of binaural incoherence (Gabriel and Colburn, 1981; Constan, 2002). Therefore, one might expect that head-related dispersion would become evident to listeners through a change in the ASW.

A logical gap in the above reasoning is that it concerns two different kinds of incoherence. The ASW experiments showing keen sensitivity to binaural incoherence have studied incoherence that occurs within a single frequency region, perhaps producing moment-to-moment variations in the lateral position of an image. By contrast, the incoherence introduced by dispersion is across different frequencies, perhaps producing different lateral positions for different frequency regions. One can view the present article as an attempt to discover whether what is learned about perceived incoherence from ASW experiments can be applied across bands to dispersion.

Although the auditory sensitivities described above suggest that listeners might be able to detect head-related dispersion, the evidence from several virtual reality experiments indicates that listeners cannot do so. Experiments by Kistler and Wightman (1992), Hartmann and Wittenberg (1996), and Kulkarni *et al.* (1999) asked listeners to discriminate between accurate HRTFs and HRTFs in which the ITDs were made frequency independent. The experiments found that it does not seem to matter perceptually if the details of ITD frequency dependence are suppressed in favor of frequency-independent ITDs. Equivalently, these experiments show that it does not matter if the true interaural phase shifts, determined experimentally from HRTFs, are replaced by phase shifts that increase linearly with frequency with a suitable slope.

The virtual reality experiments suggest that listeners cannot detect the frequency dependence of the ITD typical of human heads. However, additional work is needed to test this generalization because the virtual reality experiments have not been conclusive. First, head diffraction and corresponding virtual reality experiments do not present dispersion in isolation. Instead, head dispersion is always accompanied by an interaural level difference (ILD), and diffraction causes the ILD itself to depend on frequency. Second, the virtual reality experiments have not been optimized for revealing the effects of dispersion.

The experiments of the present article test the ability to detect dispersion when it is the only frequency-dependent variable and under conditions expected to be optimum. The experiments use headphones to produce isolated dispersion, i.e., frequency-dependent ITDs according to the spherical-head calculation without the corresponding ILDs. The experiments use noise bands with frequency ranges and simulated azimuths of incidence intended to give the listener the best chance of detecting dispersion.

II. EXPERIMENT 1: DISCRIMINATION AT TWO INCIDENT AZIMUTHS

In experiment 1, listeners compared noise bands with spherical-head-related dispersion, as shown in Fig. 1, against noise bands with ITDs that were constant (zero dispersion). Listeners had the task of recognizing the dispersive (head-delayed) noise. It was expected that our listeners would use differences in apparent source width to recognize dispersion.

A. Method

Experiment 1 featured two incident azimuths for the head-delayed noise, 30° and 45°. Figure 1 shows that both have an interesting ITD characteristic. Over the frequency range of interest (20–3000 Hz), the 30° plot exhibits the steepest change in ITD as a function of frequency while the 45° plot has the most extreme ITD variation (from 563 to 411 μs). The large variations for these two azimuths would be expected to give listeners a good chance of detection.

The experiment tested listeners at three different bandwidths, identified as broadband, midband, and narrowband. The purpose of the three was to provide the listener with different perspectives on the head-delayed interval. The broadband (BB) (20–3000 Hz) stimulus included the dispersive frequency region along with higher and lower frequencies, where the ITD was nearly constant. The midband (MB) stimulus was defined by the region where the ITD varied from its maximum value to a much lower high-frequency shelf value, presenting listeners with only the band of greatest change. Finally, the narrowband (NB) stimulus was limited to the steepest part of the ITD change, the region of greatest slope. Because the experimental signals were produced digitally, it was easy to produce precise noise bands with frequency ranges shown by rectangles in Fig. 1 with numerical values given in the caption.

Because our goal was to determine whether listeners can detect dispersion present in head-delayed noise and not present in constant-ITD noise, it was important to control for the fact that listeners could distinguish between the two noises by the lateralization of their images if the constant ITD was not chosen correctly. The correct value of constant ITD would be that value which leads to the same lateral position as the head-delayed noise. However, we did not know what the correct value was. Therefore, we used ten different equally spaced constant ITDs in random order during the course of an experiment run. The constant ITD ranges were as follows: 30° BB/MB: 260–440 μs; 30° NB: 290–362 μs; 45° BB/MB: 400–580 μs; and 45° NB: 505–550 μs. These regions are shown by rectangles in Fig. 1. For each range the largest and smallest of the ten values were outside the range of ITDs spanned by the head-delayed noise bands.

Both constant-ITD noises and head-delayed noises were generated digitally by a Tucker-Davis Technologies Array Processor, AP2. The processor constructed noises in the frequency domain, setting the upper and lower band limits and filling between them with equal-amplitude random-phase components. For a head-delayed interval, the processor then introduced the precalculated interaural phases. For a constant-ITD interval, it imposed one of the constant ITD values. Both head-delayed and constant-ITD noises were recomputed for each experimental trial by using a different basis noise waveform and adding frequency-dependent or frequency-independent delay as required. The frequency dependence of the ITD was verified with a digital delay line, signal subtractor, and a spectrum analyzer, independent of the signal generating equipment.

B. Listeners and procedure

There were four listeners, each identified by a single letter: T was a male age 24 with normal hearing and no previous experience in listening experiments. W was a male age 60 with typical middle-age high-frequency hearing loss but normal hearing in the range of the present experiments and with extensive previous listening experience. X was a male age 26 with normal hearing and previous experience. Z was a male age 27 with normal hearing and previous experience. Listeners Z and W were the authors.

During runs, listeners sat in a double-walled sound room listening with Sennheiser HD-480 II headphones. The experimental runs consisted of 100 two-interval forced-choice (2IFC) trials (ten for each constant ITD value). On each trial, the program presented the listener with two 500-ms intervals in random order. One was a constant-ITD stimulus, and the other was head delayed. Both intervals had rise/fall times of approximately 30 ms, and had simultaneous onsets/offsets in both ears.

After presenting both intervals, the program prompted for a response, and the listener pressed one of two buttons to indicate which interval was head delayed. It was expected that listeners would learn to recognize the head-delayed noise by feedback given by pilot lamps on the listener's response box after every trial. Following several training runs, listeners participated in four runs at each combination of incident azimuth (30° and 45°) and bandwidth (BB, MB, NB).

C. Results—45°

Experiment 1 found the percentage of trials in which the listener correctly identified the head-delayed interval as a function of the ITD of the alternative, namely the constant-ITD interval. The experimental results for 45° are shown in Fig. 2, one panel for each listener. Each panel includes three plots, one for each bandwidth. The data are averaged over the four runs for each bandwidth and show the percentage of trials answered correctly. The error bars on each point have an overall length of two standard errors of the mean, based on the four runs. The horizontal bars labeled ''BB/MB range'' and ''NB range'' delineate the set of ITDs included in head-delayed stimuli of those bandwidths. The dashed vertical line labeled ''LFL'' corresponds to the low-frequency ITD limit of the head-delayed noise, as represented by the open circles in Fig. 1. The dashed horizontal line indicates the 50% correct level (guessing), which is the performance expected if listeners cannot distinguish between constant-ITD and head-delayed intervals.

The percentages of correct responses in Fig. 2 show that listeners were often confused. The data for listeners W and Z tended to be a U-shaped function of the constant ITD, with best performance at the extremes of the range where the constant ITD most differed from ITDs characteristic of the head-delayed noise. These listeners achieved performance rates well in excess of 75% for extreme ITDs but only chance performance (50%) near the center of the range. The data for listeners T and X show that performance hovered around chance, exhibiting only a partial U shape.

It is of interest to compare the ITD of minimum perfor-

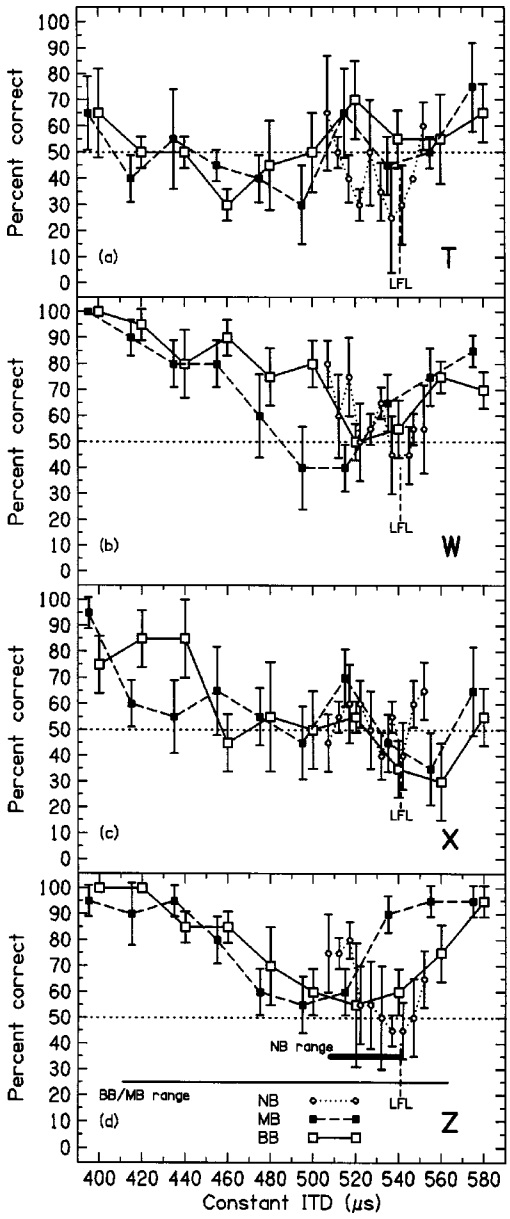


FIG. 2. Results of experiment 1 for 45° head dispersion for listeners T, W, X, and Z. This graph plots the percentage of trials in which the listener correctly identified the head-delayed stimulus versus the constant-ITD interval. The three functions represent results from broadband (BB), midband (MB), and narrowband (NB) conditions. The dashed horizontal line at 50% correct indicates chance performance (guessing). The thin horizontal line marks the range of ITDs included in the BB and MB head-delayed stimuli, while the thick horizontal line indicates those ITDs included in the NB head-delayed stimulus. The ten experimental values of constant ITD span these ranges for each bandwidth. The dashed vertical line labeled "LFL" marks the low-frequency limit ITD for a sound incident at 45°. Error bars on data points have an overall length of two standard errors.

mance for listeners W and Z with the low-frequency limit, indicated by the vertical dashed line (LFL), especially for the broadband (BB) case because the BB noise included the low-frequency region. The data show that the minimum occurs to the left of the vertical line, i.e., the most confusable constant ITD was less than the low-frequency limiting value of head-delay. A similar conclusion can be drawn for listener T, though with less confidence. The data for listener X do not show any clear frequency dependence.

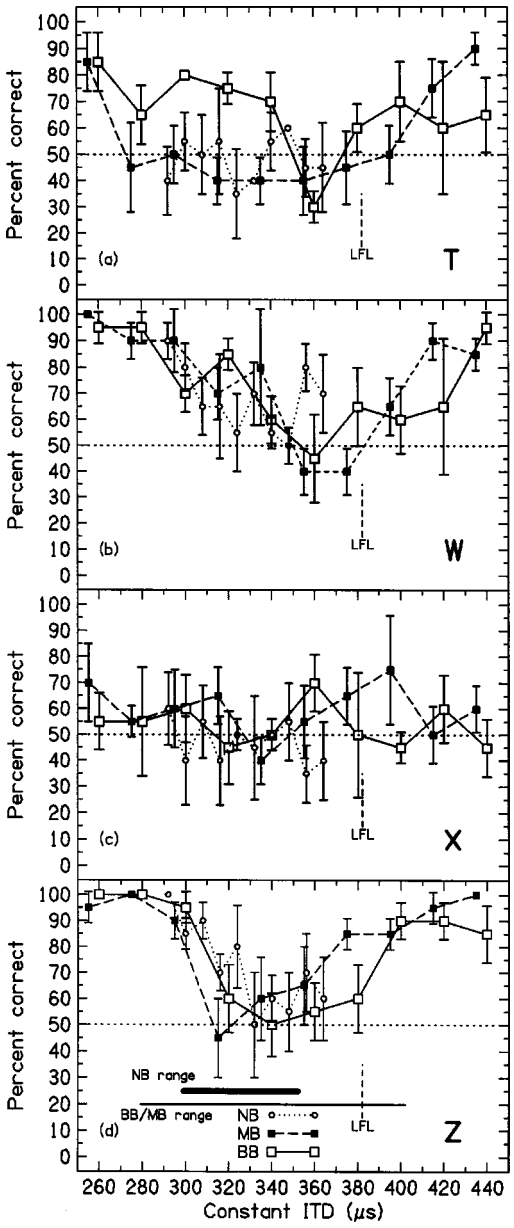


FIG. 3. Same as Fig. 2 but for a 30° head dispersion.

D. Results—30°

The experimental results for an azimuth of 30° are shown in Fig. 3, entirely parallel to Fig. 2 for 45°. The data for 30° resembled the data for 45° in that performance was at chance near the middle of the range of constant ITD. For listeners W and Z, the U-shape was narrower and better defined for 30° than for 45°. The data for listener T were also better described as U-shaped for 30° than for 45°.

Similar to the 45° data, the minima in the U-shaped functions did not coincide with the low-frequency limit. Instead, the BB data for listeners T, W, and Z had a minimum that was again to the left of the low-frequency limit, i.e., the most confusable constant ITD was less than the low-frequency limit of the head-delayed ITD. Because the U-shaped functions were somewhat narrower for 30° compared to 45°, the shift in the minimum away from the low-frequency limit was more convincing for 30° than for 45°.

E. Discussion

The results shown in Figs. 2 and 3 suggest a simple interpretation of experiment 1. The minima of the U-shapes near chance performance for listeners W and Z, and the generally chance performance for listeners T and X, suggest that listeners cannot distinguish between head-delayed noise and noise with a constant ITD. The high performance for listeners W and Z for ITDs at the extremes of the range suggest that these listeners made decisions based on the different lateralizations of the two kinds of noise. Specifically, noises with constant ITDs that differed greatly from those ITDs characteristic of the head-delayed noise were systematically heard to the right or the left of the head-delayed noise.

Informally, all the listeners reported that they heard lateralization differences in the noises and that their judgments were affected by that cue. Apparently listeners W and Z, who were more familiar with the experiment, made more consistent use of the lateralization cue.

III. EXPERIMENT 2: ROVED AZIMUTH

Experiment 1 suggested that listeners are unable to distinguish noise with head-related dispersion from noise with constant ITD, consistent with the results of virtual reality experiments. However, experiment 1 is not entirely satisfying because it included a useful lateralization cue, and it was evident that some listeners were influenced by that cue. The worry about experiment 1 is that in paying attention to the lateralization of the noise images, listeners may have missed more subtle differences associated with apparent source width. Therefore, we are not prepared to say that listeners cannot detect dispersion solely on the basis of experiment 1.

A better experiment would ask listeners to discriminate between head-delayed noise and noise with constant ITD in a context in which lateralization cues are of no use. Therefore, we designed experiment 2 to eliminate the usefulness of lateral position, leaving listeners with only interaural correlation as a means for discrimination. Our technique comprised several features. First, we roved the azimuth of the head-delayed noise perceptibly but slightly to prevent listeners from using an arbitrarily-small shift in perceived location as a cue. Second, we chose values of the constant ITD to be maximally confusable with the set of azimuthal locations. Third, we provided feedback after the response. The combination of the small rove in azimuth and the feedback was intended to actively discourage the listener from attempting to use laterality as a cue in the task.

A. Method

Experiment 2 used six different stimuli—three head-delayed bands associated with various azimuths and three corresponding constant-ITD bands. According to the design, each trial selected the two intervals randomly, one from each group so there would be a comparison between head-delayed and constant-ITD stimuli. In this way, the lateral positions of the two were randomized, and lateralization provided no consistent cues for the listener. It was expected that listeners would quickly learn from the feedback that small lateralization cues were useless in performing the task. Aside from the roving image positions and the adjusted constant ITDs, experiment 2 was identical to experiment 1, with the same task—identify the head-delayed interval.

TABLE I. Ideal performance in experiment 2A, equivalence of lateral position. In this two-interval task, complete confusion between head-delays at azimuths of 42°, 45°, and 49° and three constant ITDs is shown by 50% in the diagonal cells. This indicates that the noise with constant ITD appears to the left of the head-delayed noise on exactly half the trials. In addition the adjustments attempt to get the off-diagonal corners to be complementary and fairly close to zero and 100%.

	ITD1	ITD2	ITD3
42°	50%		ϵ%
45°		50%	
49°	(100−ϵ)%		50%

The experiment included three azimuths, near 45°, for head-delayed stimuli based on previous experience in experiment 1. From Fig. 1 it is clear that 45° data exhibit a large overall variation in ITD, producing considerable decorrelation over a relatively small frequency range and a good chance for our listeners to distinguish head-delayed noises from constant-ITD noises.

To set up experiment 2 we first needed to choose incidence azimuths above and below 45°. We chose 49° and 42° because for both of these azimuths the ITDs differed from the ITD for 45° by about 30 μs, a small but noticeable difference. Next, we needed to choose a set of corresponding constant ITDs. According to results from experiment 1, when listeners attempted to identify the 45° head-delayed stimulus, the ITD of greatest confusion was 520 μs. That value established a starting point for a preliminary experiment, experiment 2A.

B. Experiment 2A—Equivalence of lateral position

The purpose of experiment 2A was to choose three constant ITDs that would lead to lateral positions equivalent to head-delayed lateral positions for azimuths of 42°, 45°, and 49°. In experiment 2A the listeners were asked to choose which of two noises was heard further to the left. One of the noises was head-delayed, the other had constant ITD.

The goal of finding three constant ITDs (ITD1, ITD2, and ITD3) that were lateralization-equivalent to head delays could be expressed in the form of an ideal response matrix with azimuths in the rows and constant ITDs in the columns as shown in Table I. Each cell of the matrix gives the percentage of responses indicating that the constant-ITD noise is heard to the left of the head-delayed noise. Most important, the diagonal elements of the matrix are 50% indicating complete confusion about whether the constant-ITD stimulus is to the left or right of the head-delayed stimulus. Within that 50% constraint for the diagonal elements, further centering leverage can be obtained by having the left-most constant-ITD reliably appear to the left of the right-most head-delay, and to have the right-most constant ITD appear to the right of the left-most head delay by an equal amount.

With the ideal of Table I in mind, we performed runs of 99 trials, 11 repetitions of all possible combinations of constant ITD and head delay in random order. The listeners were

TABLE II. Results of experiments 2A and 2B for three bandwidths and four listeners (T, W, X, and Z). The results of experiment 2A are shown as the best matching constant ITDs (units of μs) for three azimuths (42°, 45°, and 49°). The results of experiment 2B are the diagonal elements of the confusion matrices, percentages showing the listener's ability to distinguish between a head-delayed stimulus (azimuth) and the constant ITD. A score of 50% is expected for random guessing. Average ITDs over the four listeners and the most confusing ITD from experiment 1 (45°) are given for comparison.

Listener	42° vs ITD1		45° vs ITD2		49° vs ITD3	
	Broadband (20–3000 Hz)					
	2A	2B	2A	2B	2A	2B
T	475 μs⇒49%		510 μs⇒52%		545 μs⇒55%	
W	490 μs⇒59%		520 μs⇒47%		550 μs⇒53%	
X	490 μs⇒53%		520 μs⇒52%		550 μs⇒52%	
Z	480 μs⇒51%		510 μs⇒54%		540 μs⇒56%	
AV	484 μs⇒53%		515 μs⇒51%		546 μs⇒54%	
Experiment 1			520 μs			
	Midband (400–1800 Hz)					
	2A	2B	2A	2B	2A	2B
T	460 μs⇒52%		505 μs⇒48%		550 μs⇒56%	
W	490 μs⇒43%		525 μs⇒54%		545 μs⇒67%	
X	485 μs⇒60%		515 μs⇒51%		545 μs⇒45%	
Z	480 μs⇒50%		510 μs⇒49%		540 μs⇒50%	
AV	479 μs⇒51%		514 μs⇒50%		545 μs⇒54%	
Experiment 1			510 μs			
	Narrowband (600–800 Hz)					
	2A	2B	2A	2B	2A	2B
T	490 μs⇒57%		535 μs⇒48%		580 μs⇒49%	
W	500 μs⇒50%		530 μs⇒58%		560 μs⇒43%	
X	490 μs⇒47%		520 μs⇒51%		550 μs⇒52%	
Z	500 μs⇒45%		530 μs⇒48%		560 μs⇒55%	
AV	495 μs⇒50%		529 μs⇒51%		562 μs⇒50%	
Experiment 1			540 μs			

the same as in experiment 1. We made adjustments to the various ITD values until the listener produced two runs with results corresponding approximately to the ideal. Different listeners required slightly different constant ITDs to achieve lateralization equivalence. The final values of ITD for each bandwidth and listener from experiment 2A are given in Table II (dimensions of μs). The last line of each section of the table shows the most confusing constant ITDs from experiment 1 for comparison with the 45° results of experiment 2A. The results are very similar.

The percentages of constant-ITD-on-the-left responses for experiment 2A were collected into three-by-three confusion matrices, in all, a set of 12 arrays—four listeners times three bandwidth conditions, broadband (BB, 20–3000 Hz), midband (MB, 400–1800 Hz), and narrowband (NB, 600–800 Hz). For listeners W and Z, the arrays resembled Table I rather well. The mean of the 18 diagonal elements for these two listeners was 47% ($\pm$16), and the corner elements were equal or close to 100% and 0%. For listeners T and X, the arrays approximated the form of Table I, but the table entries all tended more toward a mean value of 50%, as might be expected from the results of experiment 1. Therefore, it is probable that the sets of constant ITDs in Table II effectively eliminated lateralization as a useable cue in discrimination experiments for both pairs of listeners. Although changes in noise image location might have been perceptible for listeners W and Z, those changes were unlikely to be of any value in performing the task because of the symmetry of the design.

C. Experiment 2B—Discrimination

After the appropriate noise signals were established in experiment 2A, the four listeners were presented with those noises, in pairs, in experiment 2B. The task in experiment 2B was the same as in experiment 1—in a randomly ordered trial, discriminate the head-delayed interval from the constant-ITD interval by any means possible. With the utility of a lateralization cue likely eliminated, it was expected that listeners could only depend on the stimulus dispersion.

Experiment 2B comprised four runs for the three bandwidth conditions (BB,MB,NB)—a total of 12 runs for each listener. Each run consisted of 99 trials, 11 occurrences of the 9 possible pairs of incident azimuth and constant ITD, presented in random order.

The results for experiment 2B were displayed in 12 arrays, as for experiment 2A, a total of 108 values. Of these, only one combination of head-delay and constant ITD led to more than 75% correct (77%) and only one combination led to more than 75% wrong (76%). The 12 arrays are summarized in the 12 lines of Table II by giving only the diagonal elements of the arrays. If listeners cannot distinguish between constant ITDs and head-delays, these elements should be 50%. The table entries are, in fact, close to 50%. None of the entries approach threshold values of 25% or 75%. The conclusion of experiment 2B was that listeners cannot distinguish between head-delayed noise and constant-ITD noise when these noises are similarly lateralized.

IV. EXPERIMENT 3—BROADBAND, WRONG HEAD

Experiments 1 and 2 showed that listeners are insensitive to dispersion resulting from head diffraction. We suspect that this insensitivity arises because the head dispersion does not introduce a detectable form of incoherence. There is another possibility, however. It is possible that listeners are insensitive to head-related dispersion because they are so familiar with it. In real life, this dispersion, and the resulting binaural incoherence, is an unavoidable consequence of having a head. Experiment 3—broadband, wrong head was performed to test the second possibility.

A. Method

To test the possibility that listeners did not detect the head-related dispersion because it was so natural, we presented listeners with dispersion that had the same magnitude and yet was unnatural. The wrong-head experiment used an inverted dispersion curve. The stimuli were created by starting with the values for 45° incidence, as shown in Fig. 1, and simply reversing the sign of the difference from the low-frequency limit. The procedure is illustrated in Fig. 4, which shows the 45° part of Fig. 1 together with the wrong-head function. The same transformation was applied to the 42° and 49° stimuli as well to produce three perceptually-separate wrong-head images. Only the broadband frequency range (20–3000 Hz) was used.

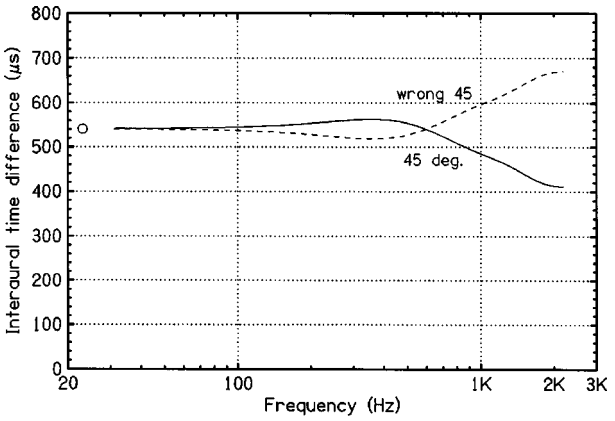


FIG. 4. The wrong-head stimulus for 45° is a reflection of the stimulus for 45° about the low-frequency limit for the ITD. Thus, the magnitude of the dispersion is the same but the signs are opposite.

Aside from changes in the stimuli, experiment 3 ran exactly the same as experiment 2, with parts A and B. Again, runs consisted of 11 presentations of each of the 9 pairings. The four listeners each did four runs in 3B.

B. Results of Experiment 3A—Equivalence of lateral position

As for experiment 2A, constant ITDs were adjusted to approach the ideal in Table I. After a series of runs established the three values of constant ITDs that were most confused with the three wrong-head delays, the percentage of judgments with the constant ITD on the left were put into three-by-three confusion matrices. Ideally the diagonal elements of these matrices should be 50%. The averages of diagonal elements actually obtained for the four listeners were T 55 (±17), W 50 (±9), X 55 (±4), and Z 48 (±5). The final constant-ITD values appear in Table III.

It is of interest to compare the optimized constant ITDs in Table III (wrong head) with the optimized constant ITDs in the broadband part of Table II (head-related) because the low-frequency limits were the same for these two cases. The constant ITDs that optimally matched the wrong-head delays were always greater, by about 8% on the average. This result indicates that the perceived equivalent ITD is not established by the low-frequency limit. Instead, the equivalent ITD seems to reflect some form of average of ITDs present in the band.

TABLE III. Results of experiments 3A and 3B (wrong head) for four listeners (T, W, X, and Z). The results of experiment 3A are shown as the best matching constant ITDs (units of μs) for three azimuths (42°, 45°, and 49°). The results of experiment 3B are the diagonal elements of confusion matrices, percentages showing the listener's ability to distinguish between a wrong-head-delayed stimulus (azimuth) and the constant ITD. A score of 50% is expected for random guessing. Average ITDs and percentages of correct identifications over the four listeners are given to summarize.

	Wrong head—Broadband (20–3000 Hz)					
	42° vs ITD1		45° vs ITD2		49° vs ITD3	
Listener	3A	3B	3A	3B	3A	3B
T	500 μs⇒53%		545 μs⇒46%		590 μs⇒58%	
W	530 μs⇒21%		560 μs⇒54%		590 μs⇒53%	
X	530 μs⇒55%		565 μs⇒48%		600 μs⇒56%	
Z	530 μs⇒53%		560 μs⇒57%		590 μs⇒63%	
AV	522 μs⇒46%		558 μs⇒51%		592 μs⇒58%	

C. Results of Experiment 3B—Discrimination

After choosing appropriate constant ITDs in experiment 3A, we asked listeners to discriminate between constant ITDs and head-delays in experiment 3B, parallel to experiment 2B. The results for experiment 3B were displayed in three-by-three confusion matrices as for experiment 3A. It was expected that the diagonal elements of these matrices would be about 50% because lateralizations should be most equivalent for those elements. The values of diagonal elements actually obtained appear in Table III. Except for one anomalously low value (21% for listener W) the values were close to 50% as expected. In fact, the off-diagonal elements were also close to 50%. Of 24 such elements for the four listeners, the largest was 67% and the smallest was 37%. The mean (±sd) was 52% (±9%). There was only one combination of head-delay and constant ITD that led to more than 75% correct or incorrect. The conclusion of experiment 3B was the same as for experiment 2B, namely that listeners cannot distinguish between head-delayed noise and constant-ITD noise when these noises are similarly lateralized. Therefore, the conjecture that head-related dispersion is not detected only because it is familiar is disproved.

V. LATERAL POSITION

The lateral position experiments, 2A and 3A, were necessary preliminary experiments for the discrimination experiments that followed. However, the lateral position experiments proved to have value in their own right. We used these data to find the value of the constant-ITD that best matched the simulated 42°, 45°, and 49° head-delayed functions.

Because the percentage of constant-ITD-left responses in the lateralization confusion matrices (not shown here) did not agree exactly with the ideal in Table I, we used the data in the matrices to make straight line fits. We performed best fits independently for the four experimental conditions: broadband, midband, narrowband, and wrong-head. The best-matching ITD was obtained by drawing a straight line through the two percentages closest to 50% and interpolating or extrapolating to the 50% point. The three azimuths, four experimental conditions, and four listeners led to a total of 48 such calculations.[1] The 48 best-matching ITDs are shown by open symbols in Fig. 5.

The best-matching ITDs from Fig. 5 can be compared with expectations about lateralization based on the stimulus. The dashed lines in the figure give the values of the ITDs at the highest frequency and the lowest frequency in the stimulus as well as the maximum value of ITD in the stimulus. (It is the minimum value for the wrong head.) Figure 5 shows that the best-matching ITDs do not agree with any of these limiting values except, perhaps, for the narrowbands, where everything is close together. It is particularly interesting that the best-matching ITDs do not agree with the low-frequency values in the stimulus. A similar conclusion was reached in experiments 1 and 3.

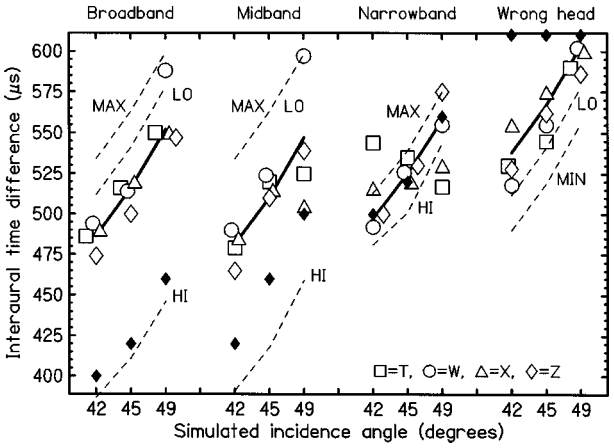


FIG. 5. Open symbols show the best-matching ITDs, i.e., values of constant ITD matching the lateralization of noise synthesized with incident azimuths of 42°, 45°, and 49° for four conditions: broadband, midband, narrowband, and broadband wrong head. The four open symbols are for the four different listeners. (Some points have been slightly displaced horizontally for clarity.) Dashed lines are labeled ''HI,'' ''LO,'' and ''MAX.'' These show the ITD at the highest frequency in the stimulus (HI), the ITD at the lowest frequency in the stimulus (LO), and the maximum value of the ITD in the stimulus (MAX). (For midband noise the MAX and LO values are the same.) (The maximum is turned into a minimum by the wrong-head manipulation.) The highest-frequency values for the wrong-head stimuli are above 600 μs. The solid line shows the weighted average of stimulus ITDs, as weighted by Raatgever's asymmetrical Gaussian function. The solid diamonds show the peaks of cross-correlation functions measured on the stimuli. Solid diamonds on the top axis are off the chart at ITD values (from left to right) 620, 660, and 700 μs.

An alternative estimate of the expected best-matching ITD based on the stimulus is the value of the lag at the peak of the cross-correlation function. The filled diamonds in Fig. 5 show this lag of the peak measured on the experimental stimuli. It is evident that the peaks of the cross-correlation functions agree with the best-matching ITDs only for the narrowbands. For the broadband and midband cases, the peaks of the cross-correlation occur at ITD values that are much lower than the best-matching constant ITD values. As might be expected, for the wrong-head case the cross-correlation peaks occur at ITD values that are much higher.

The problem with the peak in the cross-correlation function is that the cross-correlation calculation gives too much weight to the high frequencies, as noted by Kulkarni *et al.* (1999). An improved estimate would weight lower frequencies preferentially prior to cross-correlation (e.g., Bernstein and Trahiotis, 1996). Alternatively, a prediction for the best-matching ITD can be obtained from a frequency-weighted average of the stimulus ITDs. The solid line in Fig. 5 was calculated from a weighted average of the values of ITD present in the bands. The weighting function emphasized the frequencies around 600 Hz and rolled off with different slopes above and below 600. The function was an asymmetrical Gaussian:

$$u(f) = \exp[-(f/600-1)^2], \quad f \geqslant 600,$$

$$u(f) = \exp[-(f/300-2)^2], \quad f < 600.$$

This function was developed by Raatgever (1980) to describe broadband noise experiments in which one-third octave bands of noise were given delays different from the rest of the noise spectrum and then increased in intensity until they dominated the perception of lateralization. The frequency bands that required the least increase were given the greatest weight in function $u(f)$. It is evident that Raatgever's weighting procedure provides an excellent fit to the lateralization results in Fig. 5. This weighting function has also been incorporated into models that have successfully accounted for the lateralization of bands of noise narrower than our broadband noise. (Stern *et al.*, 1988; Shackleton *et al.*, 1992).

VI. COHERENCE

Data from the spherical-head experiments showed that listeners could not distinguish between noise bands of constant-ITD and head-delayed noises with equivalent lateral positions. That result is somewhat surprising because the head-delayed stimuli led to a large change in ITD across frequency.

To further pursue the binaural physical properties of the stimulus we measured the coherence (defined as the height of the peak of the normalized cross-correlation function) of 1 s samples of our stimuli. The coherence values for head-delayed noises at 45° were as follows: for broadband noise, 0.963; for midband noise, 0.972; and for narrowband noise, 0.999. Values for 42° and 49° differed by less than 0.01 from the values at 45°. By contrast, the measured coherence values of our constant-ITD stimuli were all 1.000, as expected.

A. Comparative coherence

The measured values of reduced coherence can be compared with the just-detectable incoherence found in experiments designed to study the detection of incoherence *per se*. (Note: Quantitatively, one can think of incoherence as 1.0 minus the coherence.) These experiments (called ''added-noise'' experiments) begin with a common noise in both ears (reference coherence of 1.0) and then reduce the coherence by adding an independent noise to one or both ears. The just-detectable level of the independent noise immediately leads to the threshold for incoherence detection.

In such experiments, Pollack and Trittipoe (1959a,b) found that listeners could detect a decorrelation in broadband noise that reduces the coherence to 0.96. Gabriel and Colburn (1981) found detectable decorrelation when coherence was reduced to 0.975. Constan (2002) found detectable decorrelation when coherence was reduced to 0.96 for nine listeners out of nine. For two of the listeners who participated in the present study decorrelation could be detected at a coherence value greater than 0.98. These threshold coherence values are comparable to the coherence of our head-delayed noises—0.963 for broadband and 0.972 for midband. Therefore, one might reasonably expect the head-delayed incoherence to have been detected in experiments 1–3.

A possible explanation for our negative detection result lies in the fact that experiments 1–3 required coherence discrimination when the baseline ITDs were near 500 μs. Pollack and Trittipoe (1959a,b) reported some loss in resolution for images lateralized by an ITD, and the review by Durlach and Colburn (1978) noted that coherent noise begins to lose its compact character for delays greater than 1000 μs. Such

an effect is expected on the basis of models of the binaural delay line in which the density of EE cells becomes progressively sparser as their characteristic delay increases. The consequent emphasis on cells with small delays has been called ''centrality'' (Stern and Trahiotis, 1995).

An experimental study of the ability to detect decorrelation as a function of ITD was made by Constan (2002). The results showed that the value of coherence required for detection did decrease slightly as the ITD increased from 0 to 500 μs, as expected, but for two listeners out of three, the reduced coherence remained above 0.97. Therefore, an appeal to the baseline ITD as an explanation for the results of experiment 1–3 was not persuasive.

B. Incoherence within bands and across bands

There is an important difference between incoherence caused by dispersion and incoherence caused by adding independent noise. Dispersion creates incoherence because different frequency regions lead to different ITDs, perhaps perceived as different directions by the listener. However, within any narrowband, the change in coherence tends to be very small. There are mathematical reasons why this is so. Over a narrowband, it can be assumed that the ITD varies linearly with frequency. If the band is centered at frequency f_0 and the extent of the variation in ITD over the band is $\Delta\tau$, then it can be shown that the peak of the normalized cross-correlation function differs from unity by an amount proportional to $(f_0\Delta\tau)^2$. Thus, incoherence varies only as the square of the dispersion, tending to make it small. Dramatic evidence of this effect occurred for the narrowbands used in our experiments. These bands were chosen to have the greatest possible concentrated variation in ITD, and yet when their cross-correlation functions were measured the coherence turned out to be 0.999 or 0.998.

The small incoherence found in the narrowbands was surprising, but it might have been predicted. In the limit of an infinitesimal bandwidth the signal becomes a pure tone, and the coherence for a pure tone is always unity. The moral of this story is that it is difficult to get much incoherence from a smooth ITD variation over a small bandwidth.

In fact, the narrowbands used in our experiments were not particularly narrow by the standards of auditory filtering. The band limits were 600 and 800 Hz, a bandwidth of 200 Hz centered on 700, for a relative bandwidth of 29%, or 0.42 oct, or 2.0 Cams.[2] Therefore, even with a large amount of dispersion there is very little incoherence within each critical band—even less than we measured in our narrowband stimuli. By contrast, when noise is made binaurally incoherent by adding a little independent noise, the incoherence is statistically identical in each critical band and equal to the overall incoherence.

VII. SUMMARY AND CONCLUSION

After presenting listeners with a range of theoretical head-shift stimuli (covering several different azimuths, three bandwidths, and one inversion), this study concludes that listeners cannot distinguish a signal with head-related dispersion from one that is perfectly coherent. Thus, it seems that listeners are insensitive to the decorrelation introduced by the presence of a spherical head. It is possible for listeners to detect dispersion based on lateral position, as suggested by Kulkarni *et al.* (1999), but when lateral position is eliminated as a useable cue listeners cannot discriminate between noises with constant ITD and noises with dispersion. Unfortunately, it is statistically difficult to demonstrate a negative result. Confidence in the conclusion reported here derives partly from the consistency of results from three rather similar experiments.

Initially, this conclusion was surprising because the experimental head-delayed stimuli had been chosen to try to maximize the role of dispersion in several ways, and the values of incoherence measured on the stimuli were comparable to known thresholds for incoherence detection. These thresholds were based on added-noise experiments in which incoherence was created by starting with diotic noise as a reference (zero incoherence) and adding independent noise. However, dispersion creates incoherence that is quantitatively evident only when computed across a frequency range that is wide compared to auditory filter widths. By contrast, the added-noise experiments create incoherence in each critical band. We conclude that listeners are sensitive to incoherence within critical bands, but are sufficiently insensitive to incoherence across bands that they cannot detect head-related dispersion.

Additional experiments studied the lateralization of dispersive noises. These experiments bear directly on the shift in the location of an auditory image caused by head dispersion in free-field listening. The shifts were found to be large enough to recommend that future work on localization or lateralization of broadband sounds cannot simply associate an ITD with direction without considering the range of the power spectrum. The low-frequency limit is not accurate enough. Further, the effective ITD cannot be estimated successfully from the peak of the cross-correlation function. Instead, the effective ITD for image location turns out to be a weighted average of ITDs actually present in the stimulus. An asymmetrical Gaussian weighting function, proposed by Raatgever (1980) for a different purpose, gives a good account of our lateralization data.

ACKNOWLEDGMENTS

Dr. Brad Rakerd wrote the software to measure cross-correlation functions, and Tim McCaskey helped write the programs to compute spherical-head-related transfer functions. The authors are also grateful to Dr. H. S. Colburn for helpful discussions of the results. This work was mainly supported by the NIDCD Grant No. DC00181. The article was written while the second author was visiting the Hearing Research Center at Boston University and partially supported by NIDCD Grant No. DC00100.

APPENDIX: ROUND-HEAD MATHEMATICS

The dispersion calculations of this article follow Kuhn (1977) in assuming a plane wave incident on a spherical head with ears separated by 180°.

Kuhn's formulation of the pressure at the surface of a sphere is as follows:

$$\left(\frac{p_i+p_s}{p_o}\right)_{r=a}=\left(\frac{1}{ka}\right)^2\sum_{n=0}^{n_{\max}}\frac{i^{n+1}(2n+1)P_n(\cos\theta)}{j_n'(ka)-iy_n'(ka)}. \tag{A1}$$

The symbols p_i, p_s, and p_o refer to incident, scattered, and free-field pressures respectively. The argument ka in this equation consists of the wave number k, defined as 2π divided by the wavelength, and the head radius a, which is nominally 0.0875 m. Function P is a Legendre polynomial. Functions j' and y' are derivatives of spherical Bessel functions and spherical Neuman functions respectively (Abramowitz and Stegun, 1964).

Angle θ is the azimuth of the source, measured with respect to the interaural axis. For antipodal ears, one can set θ to be 90° or less for the near ear, and set it to the complement $(180-\theta)$ for the far ear. Because Eq. (A1) only involves $\cos(\theta)$, evaluations for the near ear and far ear only require the calculation of Legendre polynomials for $c=\cos(\theta)$ and $-c=\cos(180-\theta)$, respectively. The usual definition of incident angle in binaural research is with respect to the forward direction, called θ_i. Angle θ in Eq. (A1) is computed from the equation: $\theta=90-\theta_i$.

Equation (A1) is a transfer function H for a given incident angle θ_i and angular frequency ω, where ω is k times the speed of sound. With subscript e indicating either left L or right R ears, the transfer function can be written in terms of magnitude and phase,

$$H_e=|H_e(\omega)|\exp[i\phi_e(\omega)]. \tag{A2}$$

The log magnitude and phase can be extracted by taking the natural log,

$$\ln H_e(\omega)=n_e(\omega)+i\phi_e(\omega), \tag{A3}$$

where n_e is the transfer function gain in nepers. Interaural properties can be completely described by the interaural transfer function, which is the ratio H_R/H_L, or

$$\ln[H_R(\omega)-H_L(\omega)]=\Delta n(\omega)+i\Delta\phi(\omega), \tag{A4}$$

where

$$\Delta n(\omega)=n_R(\omega)-n_L(\omega), \tag{A5}$$

and

$$\Delta\phi(\omega)=\phi_R(\omega)-\phi_L(\omega). \tag{A6}$$

The interaural time difference (ITD$=\Delta t$) then is the phase delay for a sine of frequency ω, $\Delta t=\Delta\phi/\omega$.

1. Our stimuli

Stimuli used in this paper always took $\Delta n=0$ because the focus was on the ITD and its frequency dependence. Our experiments contrasted two forms of ITD, head-delayed—as computed from the phase shifts in Eq. (6)—and constant, $\Delta t=\Delta t_0$. The constant ITD condition was called "linear phase" by Kulkarni *et al.* (1999) because it leads to an interaural phase that is a linear function of frequency, $\Delta\phi=\Delta t_0\omega$. This condition was also used by McKinley and Ericson (1997).

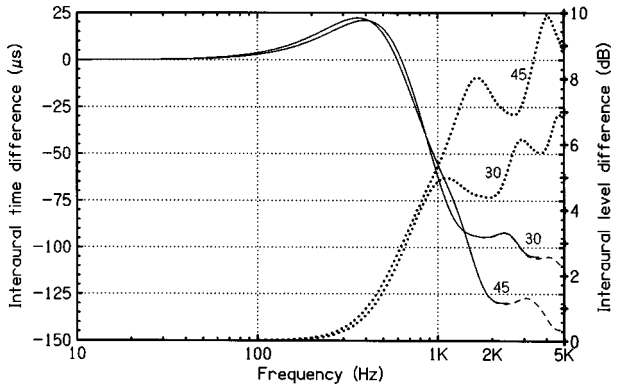


FIG. 6. Interaural time difference (ITD or Δt) for 30° and 45°. The ITD shown by the solid line is from the imaginary part of the log of the interaural transfer function [Eq. (A6)]. The ITD shown by the dashed line is calculated from the Hilbert transform of the real part of the log of the interaural transfer function [Eq. (A7)]. The low-frequency limit is subtracted off in both calculations. The agreement between the two calculations is perfect, evidence that the interaural transfer function is minimum phase. The dotted lines show the interaural level difference (ILD or Δn), the real part of the log of the interaural transfer function. These ILDs were used in the Hilbert transform calculation.

2. Minimum phase

The concept of minimum phase is not particularly helpful in the context of our work. It is discussed here in order to make contact with previous work (Kulkarni *et al.*, 1999; Kistler and Wightman, 1992). Minimum phase values of phase shifts for left and right ears, ψ_L and ψ_R, are Hilbert transforms of n_L and n_R (e.g., Hartmann, 1997, pp. 580 ff). Because the Hilbert transform is a linear operation, the minimum phase function for the interaural phase shift is

$$\Delta\psi=\psi_R-\psi_L=\mathcal{H}(\Delta n), \tag{A7}$$

where $\mathcal{H}$ is the Hilbert transform operator.

Previous authors have studied interaural phase shifts of the form

$$\Delta\phi=\Delta\psi+\omega t_0, \tag{A8}$$

i.e., minimum phase plus linear phase (frequency-independent delay), where $\Delta\psi$ is calculated from a measured log magnitude HRTF. This form has been justified because it is believed that the HRTFs are well described as minimum phase plus linear phase (e.g., Mehrgardt and Mellert, 1977). In our case, Eq. (A8) is exact because the spherical head transfer functions, as described by Eq. (A1) are minimum phase. To illustrate that idea, Fig. 6 shows the ITD for 30° and 45° calculated in two ways. First (solid line), the ITD was calculated from Eq. (A6), essentially the difference of left and right phase shifts calculated from Eq. (A1). Second (dashed line), the ITD was calculated from the Hilbert transform of the log magnitude of the interaural transfer function, i.e., from Eq. (A7). (In both cases, the low-frequency limit of the ITD was subtracted off.) The agreement between the two different kinds of calculation is so good that the dashed line cannot be seen except at the high frequencies where the solid line stops, somewhere above 2 kHz.

In connection with minimum phase calculations, a semantic difficulty has arisen when Eq. (A8) is used to describe the interaural time difference by dividing by ω,

$$\Delta t = \Delta\phi/\omega = \Delta\psi/\omega + t_0 . \quad \text{(A9)}$$

This ITD, Δt, has been described in the literature as frequency independent. In fact, it is only the t_0 part that is frequency independent. The minimum phase part $\Delta\psi/\omega$ has an important frequency dependence that cannot be ignored.

[1]Of these 48 fits, 23 were interpolated, 11 were extrapolated, and 7 were exactly on 50%. There were seven comparisons that were not monotonic and defied the fitting procedure. For those seven values we took the values of ITD1, ITD2, or ITD3 from the stimuli themselves because the course of the equivalence-of-position experiments leads naturally to these associations.

[2]Auditory filter units of *Cams* come from integrating the reciprocal of the critical bandwidth, as measured on the banks of the River Cam in England. (Glasberg and Moore, 1990; Moore, 1995). The integral is not hard to do (Hartmann, 1997, p. 251). The term ''Cam'' is preferred by the authors as a name for units on this particular critical band scale. It is preferred over the more widely used ''ERB'' (stands for equivalent rectangular bandwidths) because all critical band units since the time of Barkhausen (1926) and Fletcher (1938) have been equivalent rectangular bandwidths.

Abramowitz, M., and Stegun, I. (**1964**). *Handbook of Mathematical Functions*, National Bureau of Standards Applied Math Series No. 55 (USGPO, Washington, DC).

Ando, Y. (**1998**). *Architectural Acoustics* (Springer, New York).

Barkhausen, H. (**1926**). ''Ein neuer Schallmesser für die Praxis,'' Z. Tech. Phys. (Leipzig) **7**, 599–601.

Bernstein, L. R., and Trahiotis, C. (**1996**). ''The normalized correlation: Accounting for binaural detection across center frequency,'' J. Acoust. Soc. Am. **100**, 3774–3784.

Blauert, J., and Lindemann, W. (**1986**). ''Auditory spaciousness: Some further psychoacoustic analyses,'' J. Acoust. Soc. Am. **80**, 533–542.

Brungart, D. S. (**1999**). ''Auditory localization of nearby sources. III. Stimulus effects,'' J. Acoust. Soc. Am. **106**, 3589–3602.

Brungart, D. S., and Rabinowitz, W. M. (**1999**). ''Auditory localization of nearby sources—HRTFs,'' J. Acoust. Soc. Am. **106**, 1465–1479.

Brungart, D. S., Durlach, N. I., and Rabinowitz, W. M. (**1999**). ''Auditory localization of nearby sources II: Localization of a broadband source,'' J. Acoust. Soc. Am. **106**, 1956–1968.

Constan, Z. A. (**2002**). ''All Things Coherence,'' doctoral thesis, Michigan State University, unpublished.

Domnitz, R. (**1973**). ''The interaural time jnd as a simultaneous function of interaural time and interaural amplitude,'' J. Acoust. Soc. Am. **53**, 1549–1552.

Domnitz, R., and Colburn, H. S. (**1977**). ''Lateral position and interaural discrimination,'' J. Acoust. Soc. Am. **61**, 1586–1598.

Durlach, N. I., and Colburn, H. S. (**1978**). ''Binaural Phenomena,'' *Handbook of Perception*, Vol. 4, edited by E. Carterette (Academic, San Francisco), pp. 365–466.

Fletcher, H. (**1938**). ''Loudness, masking and their relation to the hearing process and the problem of noise measurement,'' J. Acoust. Soc. Am. **9**, 275–293.

Gabriel, K. J., and Colburn, H. S. (**1981**). ''Interaural correlation discrimination: I. Bandwidth and level dependence,'' J. Acoust. Soc. Am. **69**, 1394–1401.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched noise data,'' Hear. Res. **47**, 103–138.

Hartmann, W. M. (**1997**). *Signals, Sound and Sensation* (AIP, Springer-Verlag, New York).

Hartmann, W. M., and Wittenberg, A. (**1996**). ''On the externalization of sound images,'' J. Acoust. Soc. Am. **99**, 3678–3688.

Kistler, D. J., and Wightman, F. L. (**1992**). ''A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction,'' J. Acoust. Soc. Am. **91**, 1637–1647.

Kuhn, G. F. (**1977**). ''Model for the interaural time differences in the azimuthal plane,'' J. Acoust. Soc. Am. **62**, 157–167.

Kuhn, G. F. (**1979**). ''The pressure transformation from a diffuse sound field to the external ear and to the body and head surface,'' J. Acoust. Soc. Am. **65**, 991–1000.

Kuhn, G. F. (**1982**). ''Towards a model for sound localization,'' in *Localization of Sound: Theory and Applications*, edited by R. W. Gatehouse (Amphora, Groton, CT), pp. 51–64.

Kuhn, G. F. (**1987**). ''Physical acoustics and measurements pertaining to directional hearing,'' in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer, New York), pp. 3–25.

Kuhn, G. F., and Guernsey, R. M. (**1983**). ''Sound pressure distribution about the human head and torso,'' J. Acoust. Soc. Am. **73**, 95–105.

Kulkarni, A., Isabelle, S. K., and Colburn, H. S. (**1999**). ''Sensitivity of human subjects to head-related transfer-function phase spectra,'' J. Acoust. Soc. Am. **105**, 2821–2840.

McKinley, R. L., and Ericson, M. A. (**1997**). ''Flight demonstration of a 3-D audio display,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. A. Anderson (Erlbaum, Mahwah, NJ), pp. 683–699.

Mehrgardt, S., and Mellert, V. (**1977**). ''Transformation characteristics of the external human ear,'' J. Acoust. Soc. Am. **61**, 1567–1576.

Moore, B. C. J. (**1995**). ''Frequency analysis and masking,'' in *Hearing, Handbook of Perception and Cognition*, 2nd ed., edited by B. C. J. Moore (Academic, San Diego).

Pollack, I., and Trittipoe, W. J. (**1959a**). ''Binaural listening and interaural noise cross correlation,'' J. Acoust. Soc. Am. **31**, 1250–1252.

Pollack, I., and Trittipoe, W. J. (**1959b**). ''Interaural noise correlations: Examination of variables,'' J. Acoust. Soc. Am. **31**, 1616–1618.

Raatgever, J. (**1980**). ''On the binaural processing of stimuli with different interaural phase relations,'' thesis, Dutch Efficiency Bureau, Pijnacker, Delft, The Netherlands.

Shackleton, T. M., Meddis, R., and Hewitt, M. J. (**1992**). ''Across frequency integration in a model of lateralization,'' J. Acoust. Soc. Am. **91**, 2276–2279.

Stern, R. M., Zeiberg, A. S., and Trahiotis, C. (**1988**). ''Lateralization of complex binaural stimuli: A weighted image model,'' J. Acoust. Soc. Am. **84**, 156–165.

Stern, R. M., and Trahiotis, C. (**1995**). ''Models of naural Interaction,'' in *Handbook of Perception and Cognition—Hearing*, edited by B. Moore (Academic, San Diego), pp. 347–386.

Strutt, J. W. (**1907**). ''On our perception of sound direction,'' Philos. Mag. **13**, 214–232.

Auditory spatial resolution in horizontal, vertical, and diagonal planes[a)]

D. Wesley Grantham,[b)] Benjamin W. Y. Hornsby, and Eric A. Erpenbeck
Vanderbilt Bill Wilkerson Center for Otolaryngology and Communication Sciences, Department of Hearing and Speech Sciences, VUMC, Nashville, Tennessee 37232-8700

(Received 28 February 2003; revised 13 May 2003; accepted 16 May 2003)

Minimum audible angle (MAA) and minimum audible movement angle (MAMA) thresholds were measured for stimuli in horizontal, vertical, and diagonal (60°) planes. A pseudovirtual technique was employed in which signals were recorded through KEMAR's ears and played back to subjects through insert earphones. Thresholds were obtained for wideband, high-pass, and low-pass noises. Only 6 of 20 subjects obtained wideband vertical-plane MAAs less than 10°, and only these 6 subjects were retained for the complete study. For all three filter conditions thresholds were lowest in the horizontal plane, slightly (but significantly) higher in the diagonal plane, and highest for the vertical plane. These results were similar in magnitude and pattern to those reported by Perrott and Saberi [J. Acoust. Soc. Am. **87**, 1728–1731 (1990)] and Saberi and Perrott [J. Acoust. Soc. Am. **88**, 2639–2644 (1990)], except that these investigators generally found that thresholds for diagonal planes were as good as those for the horizontal plane. The present results are consistent with the hypothesis that diagonal-plane performance is based on independent contributions from a horizontal-plane system (sensitive to interaural differences) and a vertical-plane system (sensitive to pinna-based spectral changes). Measurements of the stimuli recorded through KEMAR indicated that sources presented from diagonal planes can produce larger interaural level differences (ILDs) in certain frequency regions than would be expected based on the horizontal projection of the trajectory. Such frequency-specific ILD cues may underlie the very good performance reported in previous studies for diagonal spatial resolution. Subjects in the present study could apparently not take advantage of these cues in the diagonal-plane condition, possibly because they did not externalize the images to their appropriate positions in space or possibly because of the absence of a patterned visual field. [DOI: 10.1121/1.1590970]

PACS numbers: 43.66.Qp, 43.66.Pn [LRB]

I. INTRODUCTION

Since the classic study by Mills in 1958, auditory spatial resolution has been extensively investigated for sources in the horizontal plane. In the case of stationary targets, the task typically involves the presentation of two discrete and temporally nonoverlapping sounds from different azimuthal positions, and the subject is asked to indicate whether the second sound is to the left or right of the first. The threshold angle for discriminating the positions is called the minimum audible angle (MAA). Under optimum circumstances (i.e., with broadband or low-frequency sources presented from directly in front of a subject) the MAA is about 1° (Mills, 1958; Grantham, 1986; Hafter *et al.*, 1988; Litovsky and Macmillan, 1994). Performance with narrow-band stimuli is worse in the midfrequency range from 1.5 to 3.0 kHz than at higher or lower frequencies (Mills, 1958), and performance generally deteriorates as the reference stimulus moves off midline (Mills, 1958; Hafter *et al.*, 1988) or as the duration of the silent interval between the two presentations is decreased (Perrott and Pacheco, 1989).

Horizontal-plane spatial resolution has also been investigated with *moving* targets. In this case a single target is presented that is moving either to the left or the right, and the subject must indicate on each trial the direction of travel. Here, the threshold angle for discriminating direction is referred to as the minimum audible movement angle (MAMA). The MAMA depends on the frequency content and velocity of the stimulus; under the best circumstances (velocity $<20°$/s; low-frequency or wideband source) the MAMA is 2–5° (Harris and Sergeant, 1971; Grantham, 1986; Perrott and Tucker, 1988; Perrott and Marlborough, 1989; Grantham, 1997). In general, it has been found that the magnitude of the optimum MAMA can approach, but is never smaller than, the value of the optimum MAA (see Grantham, 1995, for a review).

The MAA and MAMA have also been measured for sources in the vertical plane. For sources in the median sagittal plane (i.e., sources directly in front of the subject), the MAA under optimal conditions is about 4° (Perrott and Saberi, 1990; Blauert, 1997), and the MAMA under optimal conditions is about 6° (Saberi and Perrott, 1990).

It is generally accepted that spatial resolution in the horizontal plane is mediated by discrimination of interaural temporal differences and interaural level differences, while spatial resolution in the vertical plane is mediated by spectral change cues produced by the directional filtering of the pinnae. Thus, we can consider that two different systems are

[a)]Portions of this work were presented at the 137th meeting of the Acoustical Society of America, Berlin, March, 1999 [J. Acoust. Soc. Am. **105**, 1344(A) (1999)].

[b)]Electronic mail: d.wesley.grantham@vanderbilt.edu

involved in horizontal- and vertical-plane auditory spatial resolution. The question naturally arises, then, how humans spatially resolve sounds that are aligned in *diagonal* planes, which would involve both the horizontal-plane and the vertical-plane systems.

Using a wideband source, Perrott and Saberi (1990) measured MAAs as a function of the angle of the plane of presentation of their loudspeaker array, which varied in 10° steps from 0° (horizontal plane) to 90° (vertical plane). In agreement with previously reported results, these authors found the MAA in the horizontal plane was about 1° and the MAA in the vertical plane was about 4°. The thresholds for the intermediate (diagonal) planes were somewhat surprising. As the angle of the plane increased from 0° to 90°, the MAA, rather than gradually increasing from 1° to 4°, remained fairly constant at 1° until the plane reached 70°–80°. In other words, spatial resolution was relatively *independent of* plane of presentation until the array was almost vertical.

In a subsequent paper (Saberi and Perrott, 1990), these authors measured spatial resolution in diagonal planes for *moving* stimuli. In this case they found that MAMAs, at least for the slower velocities employed, were as good at a 45° plane as at a 0° plane. Furthermore, MAMAs measured at *near*-vertical planes (80° and 87°) were substantially lower than those at 90°. The authors concluded that ''the ability to detect motion is essentially independent of the path traveled by the source, with one noted exception, sources moving within a few degrees of the vertical plane.''

In a third paper in this series Saberi *et al.* (1991) measured MAAs for sources presented from *behind* the subject at planes of 0°, 60°, and 90°. Here, the pattern of results was slightly different from that obtained in the first two papers for signals presented from in front of subjects. While MAAs at 60° were about the same as those at 0° for sources in front (Perrott and Saberi, 1990), they were almost twice as large as the 0° MAAs for sources in the rear (Saberi *et al.*, 1991).

A. Two hypotheses to account for diagonal-plane spatial resolution

Saberi and Perrott considered two hypotheses to explain diagonal-plane spatial resolution. According to the first hypothesis, subjects combine information from the horizontal and vertical dimensions independently, such that

$$d'_{\text{diag}} = \sqrt{d'^2_{\text{HP}} + d'^2_{\text{VP}}}, \tag{1}$$

where d'_{diag}, d'_{HP}, and d'_{VP} represent performance in the diagonal, horizontal, and vertical planes, respectively. According to our analysis, this ''independent-contributions'' hypothesis consistently predicted larger thresholds than observed for diagonal orientations greater than about 45°, at least for the cases in which stimuli were presented from in front of the subject (Perrott and Saberi, 1990; Saberi and Perrott, 1990). On the other hand, for the case in which stimuli were behind the subject (Saberi *et al.*, 1991), this hypothesis appeared to account well for the diagonally presented signals.

As an alternative to the independent-contributions hypothesis, Saberi and Perrott considered an ''interaction model,'' which stated that interaural difference cues (based on azimuthal differences) might *interact* with spectral cues (based on elevation differences). Because this hypothesis was proposed based on data indicating resolution was relatively independent of plane of presentation (Saberi and Perrott, 1990), we refer to it here as the ''constant-resolution'' hypothesis. According to this hypothesis, there is a physical interaction of the cues underlying horizontal-plane and vertical-plane spatial resolution. For example, as a result of facial features, torso effects, and position of the pinnae on the head, a change in diagonal position of a sound source may result in greater changes in interaural differences at certain frequencies than would occur just considering the horizontal component of the trajectory alone. This hypothesis appeared to account better for the data in these authors' first two papers (in which signals were presented in the frontal field) than did the independent-contributions hypothesis.

The present experiment was undertaken to further explore human auditory spatial resolution in diagonal planes and to attempt to understand what mechanisms underlie this sensitivity. Specifically, we wanted to attempt to replicate the surprising finding reported by Perrott and Saberi (1990) and Saberi and Perrott (1990) that auditory spatial resolution could be as good in diagonal planes as in the horizontal plane. Accordingly, MAAs and MAMAs were measured for sources in the horizontal, vertical, and diagonal (60° from the horizontal) plane in the subjects' frontal field. The obtained diagonal thresholds were compared to predictions of the independent-contributions hypothesis and the constant-resolution hypothesis. In an attempt to better understand the potential cues that underlie spatial resolution in diagonal planes, data were obtained, and these comparisons were made, not only for wideband sources, but also for low-pass and high-pass stimuli that would allow some differential control over the availability of horizontal-plane and vertical-plane cues. Finally, physical measurements were made of the stimuli arriving to KEMAR's ears to determine the extent to which interaural difference cues may interact with elevation-based spectral cues produced by diagonally presented signals.

II. METHODS

A. Stimuli

A pseudovirtual technique was employed for stimulus presentation. Stimuli were created by recording sounds through the KEMAR manikin's two ears while it was positioned in a 6×6×6-m anechoic chamber [see Fig. 1(a)]. All sounds were presented through a single 10-cm JBL 8110H loudspeaker suspended from the end of a movable boom whose pivot point was directly over KEMAR's head. The loudspeaker was at KEMAR's ear level and could rotate a full 360° around him in the horizontal plane at a distance of about 1.75 m.

The resulting stimuli arriving at KEMAR were transduced by two ER-11 microphones at the positions of KEMAR's ear canals, low-pass filtered at 15 kHz, digitized at a 40-kHz sampling rate through a two-channel 16-bit A/D converter, and stored on computer disk. The stored two-channel files were subsequently played back to subjects through a 16-bit D/A converter as they sat in a darkened sound-

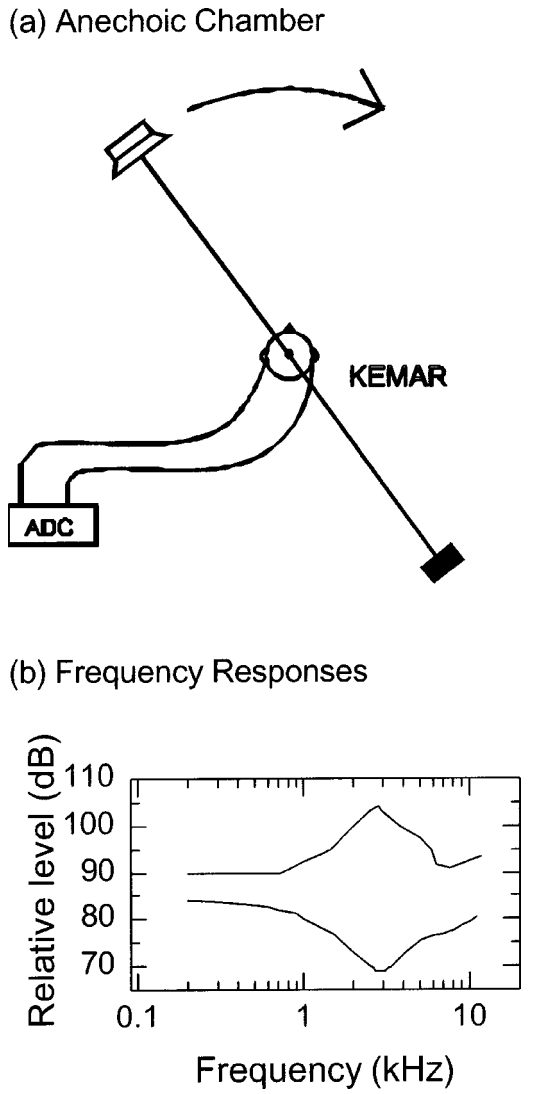


FIG. 1. (a) Schematic diagram showing a binaural recording being made through the KEMAR manikin while a moving loudspeaker is activated in the anechoic chamber. (b) Frequency response of the ER-11 microphones employed in KEMAR's ears (with diffuse-field equalization filter applied) (lower curve) and of the ER-4S insert earphones (in a Zwislocki coupler) employed to play the recorded files back to subjects (upper curve).

insulated test booth. The frequency response of the ER-4S insert earphones in a Zwislocki coupler, shown as the upper curve in Fig. 1(b), roughly mirrored that of the ER-11 mics that were employed (with the diffuse-field equalization filter activated) for recording [lower curve in Fig. 1(b)]. Thus, the stimuli arriving at subjects' ear canals incorporated the dynamic cues (interaural temporal differences, interaural level differences, and monaural spectral features) provided by KEMAR's head, torso, and pinnae.

Two-channel recordings were thus made as the active loudspeaker traversed a 180° arc from −90° to +90° [for the case of horizontal-plane stimulation, the endpoints corresponded to positions opposite KEMAR's left ear and right ear, respectively, and the arc traversed KEMAR's front hemifield [see Fig. 2(a)]. For the moving targets stimulus velocity was ±20°/s; thus, the duration of a traversal was 9.0 s. For each direction of travel four independent recordings were made of each traversal. In addition to the moving targets,

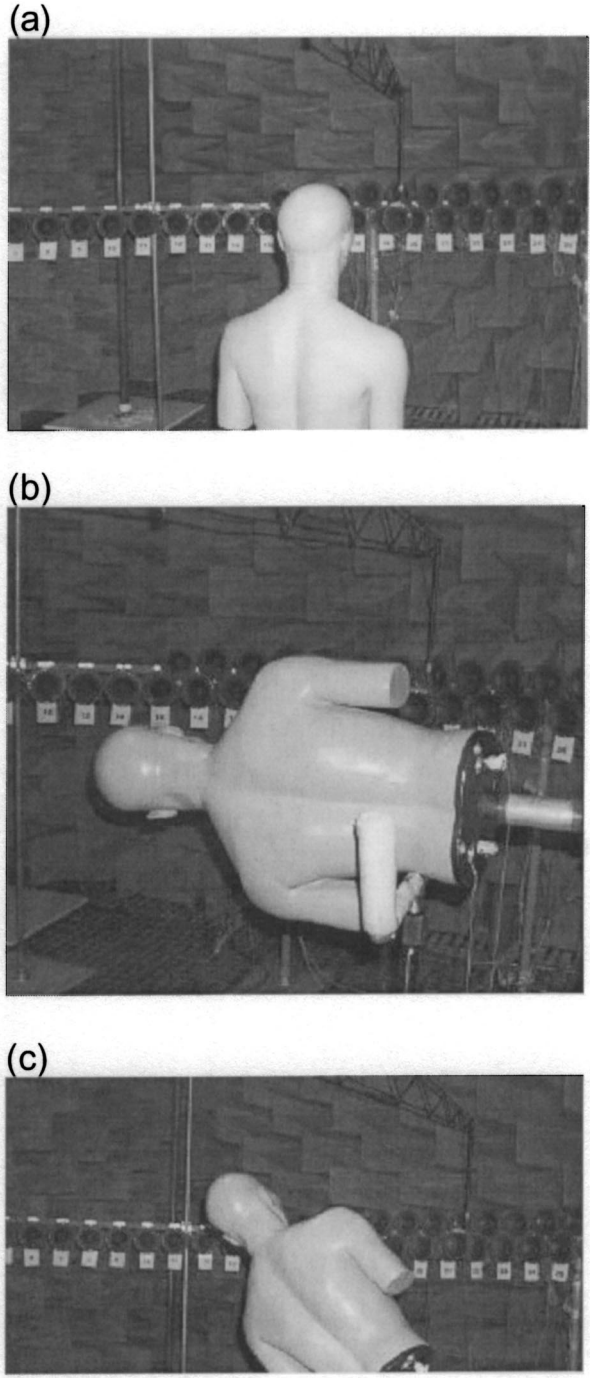


FIG. 2. KEMAR positioned for recording in the anechoic chamber. (a) In the upright position; the loudspeaker trajectory is in the horizontal plane. (b) In the side position; the trajectory is in KEMAR's median sagittal plane. (c) In the diagonal position; the trajectory is at a 60° angle relative to KEMAR. The center of KEMAR's head remained in the same position in all cases.

recordings were made for stationary targets at 4° azimuthal separations between approximately ±90°, and, because of better human spatial resolution in the frontal region, at 2° azimuthal separations between ±26°. The duration of each of the stationary target recordings was 2.0 s.

Three different stimuli were recorded in this manner:

(i) Wideband: Gaussian noise bandpass filtered from 70 to 15 000 Hz;
(ii) High-pass: Gaussian noise bandpass filtered from 6000 to 15 000 Hz;

(iii) Low-pass: Gaussian noise bandpass filtered from 70 to 2000 Hz.

In all cases, the filters employed (Kemo VBF/25.13) had nominal low- and high-pass attenuation rates of 90 dB/octave. Recording levels were set such that when the stationary stimulus from 0° azimuth was played continuously through the ER-4S earphones, the levels were approximately 78, 70, and 77 dB SPL for the wideband, high-pass, and low-pass stimuli, respectively (measured in a Zwislocki coupler). These values were chosen to accommodate within a comfortable listening range the variation in stimulus level that occurred in each ear as the target signal traversed the 180° arc.

For each of the stimuli and velocities described above, recordings were made with KEMAR in three different orientations; in each case KEMAR's nose was pointed toward the midpoint of the 180° arc, as shown in Fig. 2. In the horizontal orientation [Fig. 2(a)], KEMAR was upright; in the vertical orientation [Fig. 2(b)], he was supported on his side; in the diagonal orientation [Fig. 2(c)], he was set to lean at a 60° angle from the vertical. The orientation of the loudspeaker arc with respect to KEMAR in these three cases was thus horizontal, vertical, and 60° diagonal (upper-left to lower-right), respectively. For all three orientations, the center of KEMAR's head was maintained at a constant position, directly under the pivot point of the moving boom and at the same height from the floor as the plane described by the moving loudspeaker.

B. Procedure

As described above, all stimuli, moving and stationary, were prerecorded and stored on computer disk. The stored files were output via a two-channel 16-bit D/A converter at a rate of 40 kHz, low-pass filtered at 15 kHz (Kemo VBF/25.01, 135-dB/octave skirts), and presented to subjects through ER-4S insert earphones as they sat, individually, in a darkened sound booth.

1. Moving targets (measuring the MAMA)

For the case of moving targets, an adaptive, single-interval, two-alternative forced-choice procedure was employed to determine the minimum audible movement angle (MAMA) for each of the conditions under study. A single velocity (20°/s) was employed. Stimulus type (wideband, high-pass, low-pass) was held constant within a threshold run. On each trial a single stimulus was presented via the earphones that ''moved'' at either a positive or a negative velocity, and the subject had to push one of two buttons to indicate which direction s/he perceived the movement to be. No feedback was given. Note that the perceived directions depended on which plane of presentation was being employed; for horizontal, the possible responses were ''left''/''right;'' for vertical, they were ''up''/''down;'' for diagonal, they were ''up-left''/''down-right.'' The plane of the trajectory was fixed within a threshold run, and subjects were always told in advance which plane was being tested.

Task difficulty was manipulated by varying the extent of the arc presented on any given trial. To obtain a sweep of a given angular extent, the appropriate segment was excised from one of the four (randomly determined from trial to trial) 180° arcs stored on the disk; the midpoint of the desired arc was jittered to be within ±4° (also randomly determined from trial to trial) of the midpoint of the entire 180° arc. After excision of the desired arc, a 10-ms rise-decay time was digitally applied to the two-channel stimulus.

On the first trial in a run, a relatively large arc was presented, so that the subject could easily identify the direction. Following three correct responses, the extent of the presented arc was reduced, and following each incorrect response, the arc was increased, thus converging on the angle (the MAMA) whose direction could be identified 79% of the time (Levitt, 1971). Note that because velocity was constant (±20°/s), changing the extent of the arc from trial to trial resulted in a concomitant change in stimulus duration.

After the first two reversals in the direction of tracking of the angular change, the angular step size was reduced; tracking continued for six additional reversals, at which point the run terminated, and the threshold (MAMA) was computed as the mean of the final six reversal points. The magnitudes of the starting (first-trial) arc, the initial angular step size, and the final angular step size depended on the condition being run, and were selected for each subject and each condition based on trial and error in practice runs. For example, for the horizontal conditions (for which MAMAs were generally small), the starting arc, the initial step size, and the final step size were typically 12°, 4°, and 2°, respectively. For some of the more difficult conditions (such as vertical), the starting arc, initial step size, and final step size were sometimes as large as 64°, 16°, and 8°, respectively.

2. Stationary targets (measuring the MAA)

Measurement of the minimum audible angle (MAA) was analogous to the measurement of the MAMA described above. The primary difference was that, instead of there being a single presentation of a moving target on each trial, two stationary stimuli from different positions were presented on each trial, and the subject had to press a button to indicate whether the second stimulus appeared to be left or right (horizontal plane), up or down (vertical plane), or up-left or down-right (diagonal plane) relative to the first stimulus. Duration of each stimulus was 300 ms, and the interstimulus interval was also 300 ms.

Task difficulty during an adaptive run was manipulated by varying the angular separation between the two stimuli, whose positions could be independently selected from the set of stationary recordings (obtained at 2°–4° separations) described in Sec. II A. As in the MAMA procedure, for a given desired separation, the positions of the two stimuli presented on each trial were chosen such that the midpoint was jittered (randomly from trial to trial) within ±4° of the midpoint of the 180° range over which signals were stored. Each 300-ms stationary target was excised from the appropriate 2.0-s stationary recording (with the starting point randomly determined from trial to trial), and a 10-ms rise–decay time was digitally applied prior to presentation. Unlike the case with the moving targets, of course, changes in the angular extent

of the separation of the two signals were not accompanied by changes in stimulus or trial duration.

As with the moving targets, stimulus type (wideband, high-pass, low-pass) and the plane of loudspeaker presentation (horizontal, vertical, diagonal) were held constant within any given threshold run. Rules for adaptive tracking and the procedure for choosing starting angular separation and initial and final angular step sizes were the same as those described for the moving targets. A threshold for each run was computed as the average of the final six reversals, except for the special case of the horizontal-plane MAA (described below).

3. Computation of threshold MAMAs and MAAs

At least three ''good'' threshold runs were obtained for each condition (''good'' is defined below), with the stipulation that not all three could be obtained on the same day. If the standard deviation across the first three runs was greater than one-third of the mean across the three runs, a fourth run was obtained; if the standard deviation across the four was still greater than one-third of the mean, a fifth run was obtained, and, if necessary, a sixth. Testing ceased after the sixth run regardless of the size of the standard deviation.

For all conditions except the horizontal-plane MAA (see below), final computed thresholds were thus taken as the mean of three to six ''good'' threshold runs. A good run was defined as one for which the standard deviation over the reversal points was less than one-half of the mean over the reversal points (unless the threshold was less than 8°, in which case it was always considered good). The rationale for rejecting thresholds (over 8°) for which the standard deviation was greater than one-half the computed threshold was that a high standard deviation (i.e., large tracking excursions) usually reflected cases in which the subject intermittently ''lost'' and ''found'' the cue during a run. Such runs were likely to terminate normally, yet were assumed not to be representative of optimum or stable performance.

A different threshold estimation procedure was undertaken for the horizontal-plane MAA condition. In this condition subjects' thresholds were generally smaller than the smallest nonzero azimuthal separation employed (2°), and, although 0° was also used in the tracking, averaging over reversal points in this case may not have yielded accurate estimates of threshold. Accordingly, a combined psychometric function was generated for each subject from the trials across all (3–6) runs. Generally, this function was based on 3–4 levels of azimuthal separation and a total of about 150 trials. A three-parameter sigmoid function was fit to the data (r^2 was generally greater than 0.96), and the azimuthal separation corresponding to 79% correct was taken from the fitted function as threshold.

4. Order of testing

Subjects were initially given at least two sessions of practice prior to data collection. The practice sessions began with exposure to the horizontal condition, for which all subjects were able to obtain stable and reliable thresholds from the beginning (for all stimuli: wideband, high-pass, and low-pass, and for both moving and stationary conditions). Once subjects had become familiar with the task in the horizontal condition, sample runs from each of the other two planes were presented. In each case, during practice as well as during actual data collection, subjects were informed prior to each run of which plane was to be presented.

Generally, for a given loudspeaker plane a set of up to 12 threshold runs was programed to be presented in sequence. These runs consisted of one or two runs of each of the three stimulus types (wideband, high-pass, and low-pass) and each of the two tasks (MAA and MAMA), where the order of presentation of the 12 runs was shuffled by the computer. The sequence of 12 runs took about 1 h; subjects took rest breaks, as needed, during the set of runs. If time permitted, a set of 8–12 additional runs was programed for the second part of the session, generally for a different loudspeaker plane from that employed in the first part.

With the exception that the first day of data collection began with the horizontal plane, the order of presentation of the different planes was pseudorandom across subjects. Typically, two threshold runs were completed for all conditions in a given plane (taking most of the 1.5- to 2-h session), and then the different planes were revisited (in different random orders) as many times as it took in succeeding sessions to fulfill the criterion for stopping data collection.

C. Subject selection

Initially, seven female subjects were employed (ages 18–36). All had normal hearing bilaterally, as indicated by screening at 20 dB HL at octave frequencies from 250–8000 Hz (ANSI, 1989). They were tested in sessions that generally lasted 1 to 1.5 h. Frequent breaks were provided during a session as needed to avoid fatigue.

During the practice phase and early data collection it became clear that these seven subjects could be divided into two groups based on their performance in the vertical-plane MAA task with the broadband noise stimulus. Three of the seven had MAA thresholds of 10° or less in this condition, and all three stated that they had a clear and immediate perception of elevation change with these signals (although the sounds were not always externalized—see below). The other four subjects stated that they did not always perceive elevation changes, or, if they did notice elevation differences between the two signals on a given trial, they were not always certain which signal had the higher elevation. Vertical-plane MAA thresholds for these four subjects were 20° or higher.

Based on this dichotomy among our first group of subjects, it was decided to restrict further testing to subjects who could obtain, in a single test session without prior practice, a vertical-plane MAA with the wideband noise stimulus of 10° or less—i.e., to subjects who could perform in line with previously reported results (Perrott and Saberi, 1990). The rationale for this selection was based on the assumption that only subjects who met this criterion were actually able to effectively use the elevation-based spectral cues provided by KEMAR's pinnae, head, and torso. According to this assumption, subjects who could not meet the criterion, perhaps as a result of the shape of their pinnae and heads, were not receiving familiar elevation cues and thus could not be as-

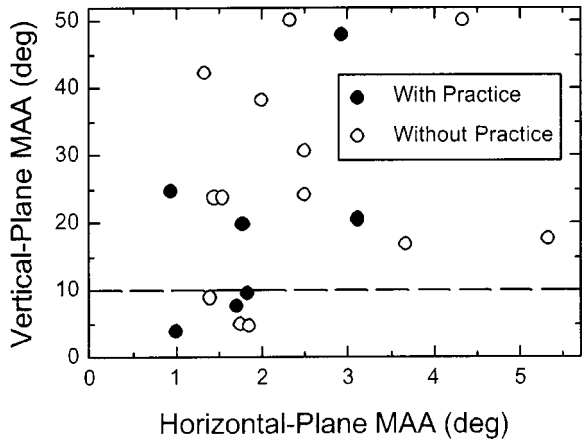


FIG. 3. Scatterplot showing vertical-plane MAA vs horizontal-plane MAA for 20 subjects. Thresholds were determined with the broadband noise stimuli. Filled symbols (''with practice'') show the seven subjects originally tested; open symbols (''without practice'') show the 13 subjects who were subsequently screened for participation in the main experiment.

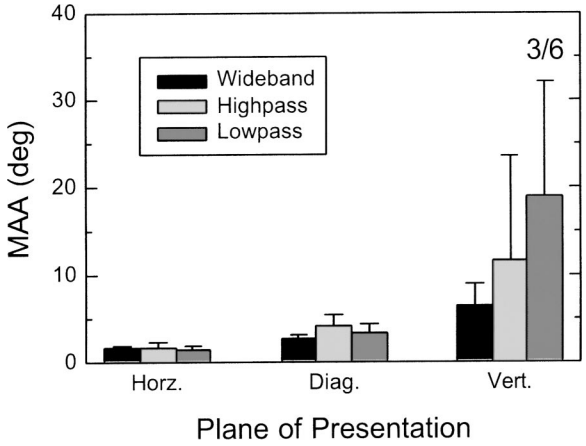


FIG. 4. Mean MAA thresholds across the six subjects for the three planes of presentation: horizontal (0°), diagonal (60°), and vertical (90°). The different shadings represent the different filter conditions. Error bars indicate one standard deviation. The fraction shown over the right-most bar indicates that only three of the six subjects could obtain a threshold with the low-pass signal in the vertical plane.

sumed to be performing normally in the vertical-plane MAA task.

Accordingly, we screened 13 additional normal-hearing subjects to find those who could meet our criterion. The screening consisted of 1 or 2 h, during which 1–6 adaptive MAA thresholds with the wideband noise stimulus were obtained for both horizontal-plane and vertical-plane conditions. As will be seen in the next section, three of these 13 subjects met the criterion. Thus, counting the three initial subjects who had met the criterion, the final selected group consisted of six subjects (four females and two males), ranging in age from 23 to 58.

The screening was considered part of the practice phase for the three subjects in this latter group who met the criterion. Subsequent testing in all the conditions of interest required 10–12 sessions, which were completed within 3–4 weeks.

III. RESULTS AND DISCUSSION

A. Subject screening

As described in the previous section, a total of 20 subjects was run, of whom six met our criterion of having, with the wideband stimulus, a vertical-plane MAA of 10° or less. Data from all 20 subjects are shown in Fig. 3 in the form of a scatterplot showing each individual's vertical-plane MAA vs his or her horizontal-plane MAA.

The filled symbols (''with practice'') represent the original seven subjects, whose data were collected after they had received practice runs, and before we decided to screen for additional subjects. The open symbols (''without practice'') represent the 13 subjects who went through the screening; except for the three subjects who met the criterion (i.e., whose data points lie below the dashed line), these data points represent MAAs obtained in a single test session. The vertical-plane MAAs for two of these subjects who could not do the task at all are indicated at 50°.

It is noteworthy than only six out of 20 subjects met the vertical-plane MAA criterion. This suggests that, in general, recording through KEMAR may not preserve the important elevation-dependent spectral information that underlies human spatial resolution in the vertical plane. To the extent that the use of individualized head-related transfer functions (HRTFs) is important for the preservation of elevation information, it is possible that only individuals who happen to have pinna and head features similar to those of KEMAR (as perhaps do the six represented by the data points below the dashed line) would benefit from recordings made through KEMAR.

On the other hand, it is also noteworthy that the horizontal-plane MAAs for the 20 subjects are quite consistent: a majority shows a threshold less than 2°, and all but three subjects show a threshold of 3° or less. This finding is consistent with the notion that the interaural cues underlying horizontal-plane spatial resolution depend little on the idiosyncrasies of different individuals' HRTFs.

B. Spatial perception of the sound images: Informal reports

None of the six subjects in the final group reported the sound images from these signals to be in front of them at the appropriate distance (e.g., about 1.75 m from the head). Some of the subjects reported that the images were in front of them, but that they seemed only inches from the head; others reported the images were either inside the head, or a few inches behind the head. These observations are in agreement with those of Koehnke and Besing (1996), who reported that their subjects achieved good externalization of KEMAR-recorded stimuli in reverberant, but not in anechoic environments. In any case, despite the failure of our subjects to externalize the signals appropriately, all reported that, for angular distances greater than threshold, they perceived clear motion of the images in the appropriate directions.

C. MAA and MAMA thresholds

Mean thresholds across the six subjects who met the screening criterion are shown in Fig. 4 (MAA) and Fig. 5

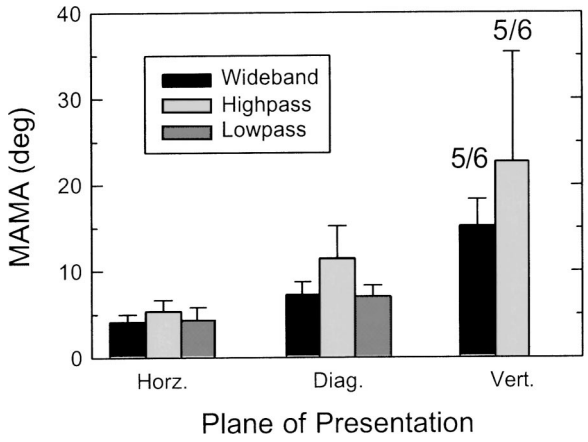


FIG. 5. Mean MAMA thresholds across the six subjects for the three planes of presentation (see the caption of Fig. 4). The fractions shown over the two vertical-plane bars indicate the number of subjects out of six who were able to do the task. Note that vertical-plane thresholds for the low-pass stimulus are not shown, because none of the six subjects could obtain a threshold in this condition.

(MAMA), with error bars representing standard deviations. In each figure the speaker plane (horizontal, diagonal, vertical) is indicated along the abscissa, and the different bars represent the three different filter conditions (wideband, high-pass, low-pass). Fractions appearing above some of the bars indicate the number of subjects, out of six, for whom thresholds could be obtained for the indicated condition (where no fractions appear, all six subjects could do the task). Mean thresholds across subjects in all conditions (with standard deviations) are also listed in the first three columns of Table I.

A three-way repeated-measures analysis of variance was conducted on the data, where the factors were velocity (MAA vs MAMA), filter condition (wideband, high-pass, low-pass), and plane (horizontal vs diagonal). The vertical-plane thresholds were excluded from the analysis because there were several missing data points in this condition [see footnote (a) in Table I]. For those cases in which vertical-plane thresholds *were* obtained, they were generally much higher than the corresponding thresholds in the horizontal and diagonal planes, which is evident from the mean values shown in the figures and in Table I.

The ANOVA revealed that all three main effects were significant (discussed below), as were two of the interactions: velocity×plane [$F(1,5)=29.219$; $p<0.01$] and velocity×filter condition [$F(2,10)=11.894$; $p<0.02$]. All p values to be reported take into account the Greenhouse–Geisser correction.

1. Effect of velocity

The MAMAs (Fig. 5) were consistently and significantly larger than the MAAs (Fig. 4) [$F(1,5)=93.485$; $p<0.001$]. Although the ANOVA revealed that velocity interacted significantly with both plane and with filter condition, these interactions were not of special interest in the present investigation. It can be seen from the figures that the average MAMA was greater than the average MAA for all filter conditions and all three planes (including the vertical plane).

The fact that the MAMA is larger than the MAA has been previously reported for wideband stimuli in the horizontal plane (Perrott and Musicant, 1977; Perrott and Tucker, 1988; Chandler and Grantham, 1992) and in the vertical and diagonal planes (Perrott and Saberi, 1990; Saberi and Perrott, 1990). The present data show that this spatial resolution advantage with stationary targets over moving targets is present with high-pass and low-pass as well as with wideband stimuli.

2. Effect of plane

Overall, the diagonally presented targets yielded higher thresholds than the horizontally presented targets [$F(1,5)=45.5$; $p<0.01$]. It also seems reasonable to conclude, based on the magnitude of the obtained thresholds and on the fact that in several cases subjects could not perform the task,

TABLE I. Mean thresholds across subjects (with standard deviations) in degrees of arc for the three orientations (columns 1–3). The column labeled "IC hyp" shows the mean predictions across subjects (with standard deviations) for the independent-contributions hypothesis. Note that the constant resolution (CR) prediction is equal to the threshold in the horizontal orientation (first column). The last two columns show the t-ratios for the tests for the difference between diagonal threshold and the constant-resolution hypothesis and the independent-contributions hypothesis, respectively. Degrees of freedom for all tests are 5.

	Horizontal (CR hyp)	Diagonal	Vertical	N[a]	IC Hyp	t ratio Diag vs CR hyp	t ratio Diag vs IC hyp
MAA							
Wideband	1.6(0.3)	2.8(0.4)	6.5(2.4)	6/6	2.9(0.6)	16.08[b]	−0.64
High pass	1.6(0.6)	4.1(1.4)	11.6(12.0)	6/6	2.9(1.1)	3.89[c]	1.69
Low pass	1.5(0.4)	3.4(1.0)	19.0(13.0)	3/6	2.8(0.7)	5.64[b]	1.99
MAMA							
Wideband	4.2(0.9)	7.3(1.5)	15.3(3.0)	5/6	7.7(1.8)	7.43[b]	−0.74
High pass	5.4(1.3)	11.4(3.8)	22.6(12.7)	5/6	9.6(2.2)	4.50[b]	1.38
Low pass	4.5(1.3)	7.2(1.1)	···	0/6	8.9(2.6)	3.88[c]	−1.52

[a]The number of subjects who could perform the vertical-plane task out of the number who ran in each respective condition.
[b]$p<0.01$.
[c]$p<0.05$.

that performance with the vertically presented targets (not included in the ANOVA) is considerably worse than performance in either of the other two planes.

To relate these findings to previous data, we focus first on the condition that has received the most attention in earlier investigations—MAAs obtained with wideband stimuli (black bars in Fig. 4). For horizontal-plane presentation, the mean MAA was 1.6°, which is similar to MAA thresholds that have been reported previously for broadband signals (Perrott and Saberi, 1990; Chandler and Grantham, 1992). This result indicates that the pseudovirtual technique employed for stimulus presentation in the present experiment was successful in preserving the interaural cues underlying horizontal-plane spatial resolution.

Spatial resolution in the diagonal plane was almost as good as in the horizontal plane (mean MAA: 2.8°), while resolution in the vertical plane was markedly worse than in the other two planes (mean MAA: 6.5°). This general pattern of results is in basic agreement with that obtained by Perrott and Saberi (1990) and by Saberi *et al.* (1991).

The main characteristics noted above for wideband MAAs were also observed for MAAs obtained in the low-pass and high-pass conditions, and for the MAMAs in all three filter conditions. Thus, vertical-plane thresholds (when the task was possible) were higher than those for the horizontal and diagonal planes, while thresholds in the latter two planes were similar to each other, although the diagonal plane generally yielded consistently (and significantly) higher thresholds than did the horizontal plane.

3. Effect of filter condition

Overall, there was a significant effect of filter condition [$F(2,10)=13.75$; $p<0.01$]. However, a set of four single-factor ANOVAs revealed that, for the different velocity-plane combinations, this effect was significant only for the diagonal MAMAs (middle set of bars in Fig. 5) [$F(2,10)=11.277$; $p<0.02$]. For the horizontal MAMAs (leftmost bars in Fig. 5) and for the horizontal and diagonal MAAs (leftmost and middle bars in Fig. 4), the effect of filter condition was not statistically significant.

Thus, in the case of horizontal-plane presentation, performance was virtually the same in all three filter conditions for both stationary and moving signals. Previous studies have shown that horizontal-plane MAAs and MAMAs show some frequency dependence for narrow-band stimuli, with performance worse in a midfrequency region (1.5 to 3.0 kHz) than at lower or higher frequencies (Mills, 1958; Perrott and Tucker, 1988; Chandler and Grantham, 1992). In addition, it has been shown that horizontal-plane spatial resolution increases as bandwidth increases for stimuli centered at 3.0 kHz (Chandler and Grantham, 1992). The present data add further to our knowledge about spectral effects in auditory spatial resolution. Specifically, they indicate that restricting the stimulus to either low frequencies, where horizontal-plane spatial resolution is mediated primarily by interaural temporal differences, or to high frequencies, where spatial resolution is mediated primarily by interaural level differences and interaural spectral difference cues, does not degrade spatial resolution ability, at least for stimuli with relatively large bandwidths.

Considering the data in the vertical plane (rightmost bars in Figs. 4 and 5), there may be an advantage for wideband over high-pass stimuli (because of missing data points, a statistical test was not performed for this difference). If the difference is real, it would suggest that spectral information below 6 kHz is useful in making vertical resolution judgments. On the other hand, thresholds in the low-pass conditions were mainly indeterminate: only three of the six subjects could do the MAA task, and none could do the MAMA task. This failure to discriminate vertical-plane changes is not surprising, given that the primary cues underlying localization and discrimination in the median vertical plane are spectral cues above 4000 Hz (Roffler and Butler, 1968; Hebrank and Wright, 1974). The fact that three subjects *were* able to do the MAA task for these low-pass stimuli is somewhat surprising; it is likely they were able to take advantage of the shoulder-bounce cue that can provide some elevation information for low-frequency stimuli (Algazi *et al.*, 2001).

Concerning the significant effect of filter condition alluded to in the opening paragraph of this section (for the diagonal MAMAs—middle bars in Fig. 5), analytical comparisons confirmed what is apparent from the figure: the high-pass thresholds were significantly higher than the wideband and low-pass thresholds, and the latter two were not different from each other. Given that there were practically no differences in the horizontal thresholds across the filter conditions, and given that the low-pass MAMA was unmeasurable in the vertical plane, it was not clear why, in the diagonal plane, the high-pass stimuli produced the highest MAMA thresholds.

D. Predictions for diagonal-plane performance

As described in the Introduction, Saberi and Perrott (1990) considered two hypotheses to account for auditory spatial resolution performance in the diagonal plane. The ''independent contributions'' hypothesis, expressed by Eq. (1), posits that performance in the diagonal plane is based on independent contributions from a horizontal-plane mechanism sensitive to interaural differences and a vertical-plane mechanism sensitive to spectral variations (and assumes no physical interaction between the two types of cues). Alternatively, the ''constant-resolution'' hypothesis states that performance in a spatial resolution task is essentially independent of the plane of the sources, until the plane is nearly vertical (e.g., greater than 70°–80°). This constant sensitivity may be based on the physical interaction of the acoustic cues underlying performance in horizontal-plane and vertical-plane conditions. In other words, the interaural differences for diagonally separated sources with a given horizontal component may be greater (e.g., at particular frequency regions) when there is also an elevation component than when that horizontal separation is presented alone. We were interested in comparing our subjects' performance to the predictions of these two hypotheses.

Mean MAA and MAMA thresholds across the six subjects for the diagonal plane are replotted in Fig. 6, with performance in the three filter conditions represented in the

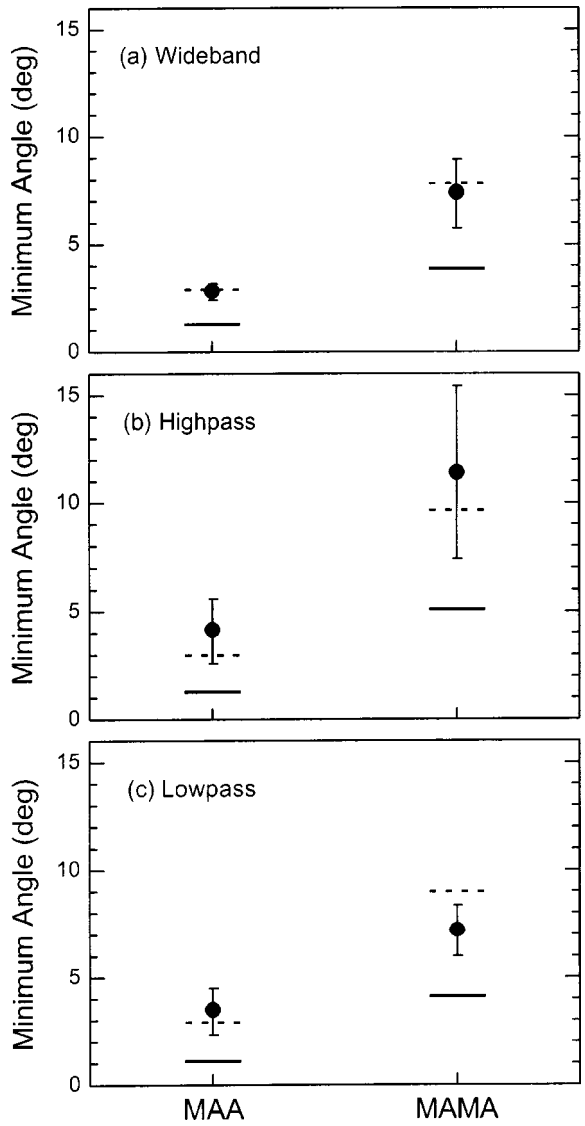


FIG. 6. Mean thresholds across subjects for the diagonal-trajectory condition, plotted with 95% confidence limits. Data are shown for the wideband noise (upper panel), high-pass noise (middle panel), and low-pass noise (lower panel). The lower (solid) horizontal hash mark plotted with each data point denotes the mean prediction of the ''constant-resolution'' hypothesis; the upper (dashed) hash mark denotes the mean prediction of the ''independent-contributions'' hypothesis.

three panels. Error bars indicate 95%-confidence limits (these thresholds are listed in column 2 of Table I). The mean predictions for the constant-resolution and independent-contributions hypotheses are shown for each condition by the lower (solid) and upper (dashed) horizontal lines, respectively (listed in columns 1 and 5 of Table I).

Predictions from the two hypotheses were generated as follows:

(1) For the constant-resolution hypothesis, the prediction for each subject was simply that subject's performance measured in the horizontal-plane condition. That is, according to this hypothesis, resolution in the diagonal plane should be equivalent to performance in the horizontal plane (at least for diagonal planes that do not approach vertical). These predictions are listed in column 1 of Table I, and are displayed as the *solid* horizontal line plotted with each data point in Fig. 6.

(2) For the independent-contributions hypothesis, we assumed reasonable shapes for the underlying psychometric functions for performance in the horizontal- and vertical-plane conditions. For cases in which vertical-plane thresholds were not obtainable, we assumed that $d'_{VP}=0$ for all angular extents. Under these assumptions, predictions were generated from Eq. (1) (details are provided in the Appendix). The predictions are listed in column 5 of Table I (under ''IC Hyp''), and are displayed as the *dashed* horizontal line plotted with each data point in Fig. 6.

To test the significance of the difference between obtained and predicted thresholds, a series of *t*-tests was performed, in which each of the six data points was compared with the two predictions. The results (shown in the last two columns of Table I) revealed that all six conditions were significantly different from the constant-resolution hypothesis (two-tailed test). This result is consistent with the significant main effect of plane (horizontal vs diagonal targets) revealed by the overall ANOVA reported in Sec. III C 2. On the other hand, none of the six cases was significantly different from the independent-contributions prediction according to a two-tailed test, suggesting that for the stimuli and conditions employed in the present experiment, we cannot reject the hypothesis that diagonal-plane spatial resolution is mediated by independent contributions from a horizontal-plane mechanism and a vertical-plane mechanism.

We conclude from these results that spatial resolution in a diagonal plane, when using KEMAR-recorded stimuli, is not consistent with the constant-resolution hypothesis. Specifically, in all cases, for both static and dynamic signals, mean performance was significantly worse when the plane of the trajectory was 60° than when it was horizontal.

IV. GENERAL DISCUSSION

The results of the present experiment differ in one important respect from those reported by Perrott and Saberi (1990). The 60° MAA thresholds from that study were consistent with the constant resolution hypothesis (i.e., they were about the same as the 0° thresholds) and were, on average, about 45% lower than the predictions of the independent contributions hypothesis generated from their data. By contrast, the diagonally measured MAAs and MAMAs from the present study are consistent with the independent-contributions hypothesis and inconsistent with the constant resolution hypothesis. That is, our data suggest that auditory spatial resolution in diagonal planes is mediated by the independent combination of the available interaural cues underlying horizontal-plane performance and the available spectral cues underlying vertical-plane performance. Unlike Perrott and Saberi, we found no evidence that subjects could maintain constant resolution as the diagonal plane increased from the horizontal.

The most obvious methodological difference between the current experiment and those conducted by Perrott and Saberi is the fact that the latter investigators conducted their experiments in real space, while we employed a ''pseudovirtual'' technique and presented stimuli through earphones. It

is known that localization with virtual cues produced using ''nonindividualized'' HRTFs is generally poorer than localization in real environments or than virtual localization ''using one's own ears'' (Wenzel *et al.*, 1993; Middlebrooks, 1999). Although we screened our subjects with the goal of retaining only subjects who received ''true'' vertical-plane cues through KEMAR's ears, it is, of course, not possible to know exactly how the resulting thresholds reflect the sensitivity that would have been measured with the same stimuli in real environments. This is a necessary limitation of employing the pseudovirtual procedure.

A second difference between our study and those conducted by Perrott and Saberi relates to the availability of visual information. While our subjects were tested in the dark and thus had no visual cues, Perrott and Saberi (1990) and Saberi and Perrott (1990) conducted their experiments in a dimly lit room, in which the plane of the speaker array was clearly visible.[1] It has been shown that the presence of a visual target or a patterned visual field can facilitate performance in auditory spatial tasks, even when the available visual cues are uncorrelated with the auditory signals (e.g., Platt and Warren, 1972; Shelton and Searle, 1980; Lakatos, 1995). It is possible that, in Perrott and Saberi's experiments, the visual perception of the diagonally orientated loudspeaker array facilitated subjects' spatial resolution of auditory targets presented from that array, thus yielding smaller MAA thresholds than we observed. We find it intriguing in this regard that the pattern of MAAs measured by Saberi *et al.* (1991), where stimuli were presented from *behind* the subjects (thus not affording visual perception of the loudspeakers) was similar to that obtained in the present experiment: MAAs were about twice as high for the 60° plane as for the 0° plane.

A. Measurements of interaural differences for KEMAR-presented stimuli

We have seen that Perrott and Saberi's subjects demonstrated evidence of constant spatial resolution of sound sources as the plane of presentation changed from 0° to 60° while ours did not. In order to further explore possible reasons for these differences, it was of interest to measure the interaural differences provided by the diagonally presented signals, to learn whether, *in principle*, there might be a clearly identifiable physical basis for superior performance in the diagonal plane over that which would be expected based on simple geometrical considerations. In other words, we wanted to know if the physical features of the human head, pinna, and torso might produce some interaction between spectral shape cues and interaural difference cues produced by sounds presented from diagonal planes, such that the magnitudes of interaural differences might be larger than expected.[2] Accordingly, we measured the interaural differences in the wideband signals that had been recorded through KEMAR's ears from the stationary sources spanning 180°. These were the same recorded sources that had been employed in the determination of the wideband MAA for our subjects (black bars in Fig. 4).

In order to obtain the most useful comparison, we measured the *difference* in interaural differences for the two signals that were presented symmetrically around midline during a typical trial in the MAA experiment. For example, in a horizontal-plane condition, if the signal in the first interval, presented from an azimuth of $-2°$ azimuth, had an interaural temporal difference (ITD) of $+20$ μs (positive values indicate left ear leading) and the signal in the second interval, presented from $+2°$ azimuth, had an ITD of -20 μs, the *difference* in ITD for this 4° separation (which is the candidate cue for discrimination) would be 40 μs. We compare this value to the ITD difference produced by a diagonally presented pair of signals, *whose horizontal extent is equal to that of the horizontally presented signals* (i.e., 4°). Because our diagonal plane was 60°, the horizontal component of any given angle measured along the diagonal arc is exactly half of that angle ($\cos 60° = 0.5$). For the example just given, an extent of 8° along the diagonal plane would yield a 4° projected extent along the horizontal plane. Thus, the appropriate comparison to assess a prediction based on simple geometry is between the ITD difference produced by a horizontal-plane 4° extent and that produced by a diagonal-plane 8° extent.

The interaural differences for any given binaural stimulus were computed in the MATLAB environment by performing FFTs on the left- and right-ear signals. The ITDs were computed by taking the difference between the phase spectra of the left and right waveforms, unwrapping this phase spectrum difference function, and estimating its slope over the frequency range 500–4000 Hz (where the phase spectra were most well behaved). The interaural level differences (ILDs) were computed by taking 20 times the log of the ratio of the left and right magnitude functions from 0 to 15 kHz. In both cases, the FFTs were carried out over successive nonoverlapping windows of the 2-s signals, and the final results were obtained by averaging across these windows. The ILD or ITD *difference* for a given separation $X°$ was taken as the difference between the interaural difference for the signal at $-(X/2)°$ and the interaural difference for the signal at $+(X/2)°$.

The results for the ITD measurements are shown in Fig. 7(a). The circles connected with the dashed line represent the ITD differences for sources presented from the horizontal plane, plotted as a function of the azimuthal separation of the sources. The triangles connected with the solid line show the ITD differences for diagonally presented sources, plotted as a function of the horizontal-plane separation (as projected from the diagonal arc) of the sources. The fact that the two functions are very similar indicates that the ITD cues produced by diagonally separated signals are about what would be expected based on simple geometry.

Overall ILD differences are displayed in Fig. 7(b), where again the circles represent the measurements from horizontal-plane conditions, and the triangles represent the measurements from diagonal-plane conditions (these were based on rms values computed from the digital waveforms). Again, the functions for the two planes are very similar, suggesting that *overall* ILD differences produced by diagonally separated sources are predictable from those produced by horizontally separated sources having the same horizontal extent.

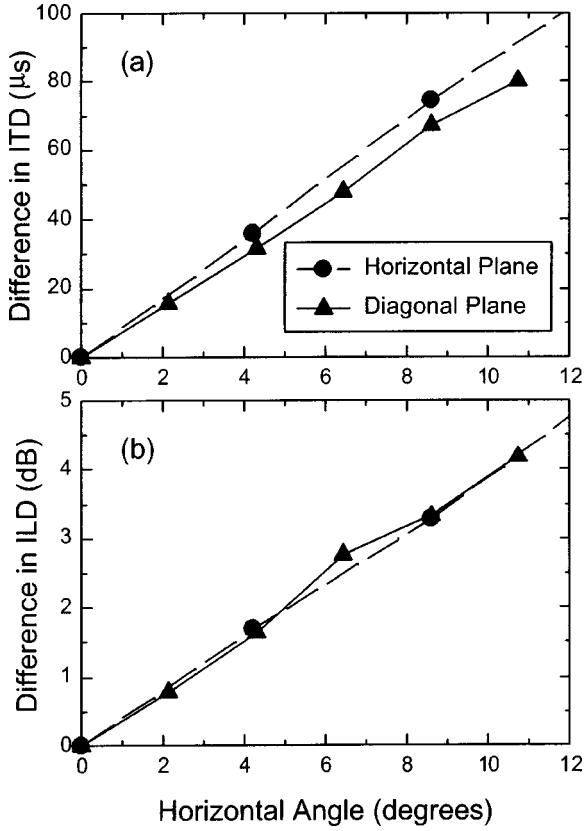


FIG. 7. Results of measurements of the wideband stimuli recorded through KEMAR. (a) The difference between the ITDs of the two signals presented in the two intervals of an MAA task as a function of the horizontal angle between the two signals. (b) The difference between the overall ILDs of the two signals presented in the two intervals of an MAA task as a function of the horizontal angle between the two signals. Circles are for signals presented from the horizontal trajectory; triangles are for signals presented from the diagonal trajectory.

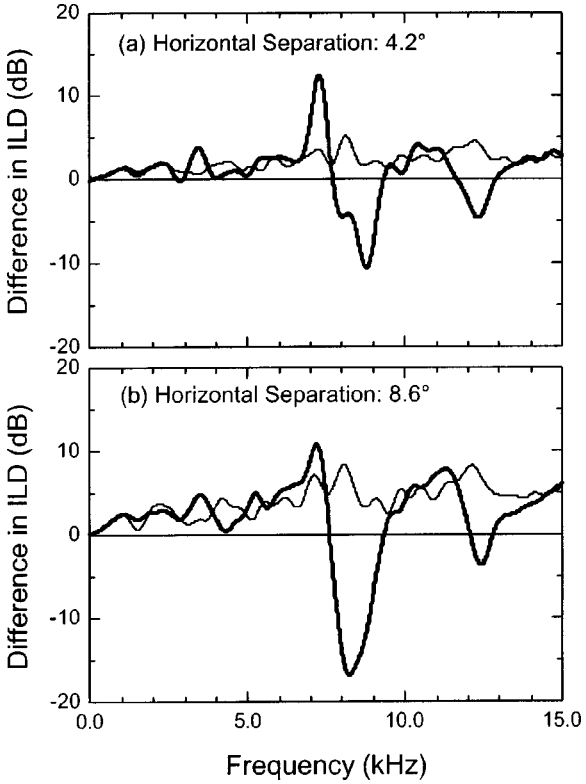


FIG. 8. Results of ILD measurements of the wideband stimuli recorded through KEMAR. Differences between the ILDs of the two signals presented in the two intervals of an MAA task as a function of frequency. Thin line: signals presented from the horizontal trajectory; thick line: signals presented from the diagonal trajectory. (a) Horizontal separation is 4.2°. (b) Horizontal separation is 8.6°.

However, when we look at the ILD differences as a function of frequency, some interesting patterns emerge. Figure 8 displays the ILD differences plotted as a function of frequency for sources with a horizontal-plane separation of 4.2° (upper panel) or 8.6° (lower panel). The thin lines represent horizontal-plane sources, while the thick lines represent diagonal-plane sources. Note that the functions in the upper panel correspond to the points plotted at 4.2° in Fig. 7(b) and those in the lower panel correspond to the points plotted at 8.6°.

While the ILD difference functions for the horizontally presented signals (thin lines) are positive and fluctuate relatively little over the entire frequency range, indicating that the ILD (favoring the left ear) for the leftmost source is consistently larger than that for the right source for all frequencies, this is not the case for the diagonally presented signals (thick lines). For these latter functions there are more dramatic fluctuations as a function of frequency than for the horizontal-plane functions, including one frequency region—near 8 kHz—where the ILD difference is *negative*. This indicates that the ILD in this frequency region from a source ''up-left'' is *smaller* than that for a source ''down-right;'' furthermore, comparing the upper and lower panels in the figure, it appears that the magnitude of this negative difference increases as the spatial separation increases from 4.2° to 8.6°.

In conclusion, these measurements indicate that, for the KEMAR manikin, ITDs and *overall* ILDs measured for diagonally presented sources are not different from the corresponding interaural differences measured for horizontally presented sources having the same separation in horizontal extent. However, when looking within specific frequency regions, ILDs from diagonally presented sources can be larger than, and in some cases can be opposite in sign to, those from horizontally presented sources. Presumably the irregular behavior of the diagonal ILD difference functions is based on the shape of the head, pinnae, and facial features, that, of course, depart markedly from a spherical model. The resulting pattern of ILD differences as seen in Fig. 8 could provide a basis for the surprisingly good performance observed in diagonal planes by Perrott and Saberi (1990) and Saberi and Perrott (1990).

B. Basis for constant resolution

The question of exactly which cues might be responsible for humans' ability to maintain ''constant spatial resolution'' as the plane of sound sources increases from horizontal to near-vertical remains unanswered. Based on the measurements made with KEMAR, it is possible that the 8-kHz frequency region is important for discriminating diagonal sources (at least for a 60° plane); Fig. 8 reveals other fre-

quency regions that might contribute to performance as well (e.g., a peak at 7 kHz). However, despite the availability of these cues, the subjects in our study were apparently not able to take advantage of them to achieve better performance in the diagonal plane than that predicted by the independent contributions hypothesis.

Based on the results of the present and previous studies, it is possible that a critical element in order for subjects to maximize diagonal-plane resolution is that there be a visual field available, and one, in particular, that contains a diagonal feature that matches the trajectory of the sounds to be discriminated. Perhaps only with such visual information can subjects take advantage of the diagonal cues revealed in Fig. 8 in order to achieve constant resolution. In the two studies that reported evidence of constant resolution, the loudspeaker array was visible to the subjects (Perrott and Saberi, 1990; Saberi and Perrott, 1990). In the only other two investigations of diagonal-plane spatial resolution, targets were either presented from behind the subjects (Saberi *et al.*, 1991), or they were presented through earphones in the dark (present study); neither of these latter two studies showed evidence of constant resolution.

V. SUMMARY AND CONCLUSIONS

An experiment measured MAA and MAMA thresholds for stimuli in horizontal, vertical, and diagonal (60°) planes, employing a pseudovirtual technique in which signals were recorded through KEMAR's ears and played back to subjects through insert earphones. Thresholds were obtained for wideband, high-pass, and low-pass noises. The main results may be summarized as follows:

(1) In an initial screening employing a broadband noise stimulus, only six of 20 subjects were able to obtain ''reasonable'' vertical-plane MAAs (i.e., MAAs less than 10°). We believe this reflects a limitation with the ''pseudovirtual'' technique, and that only subjects who have pinnae and heads shaped similarly to those of KEMAR can make full use in a vertical-plane spatial task of the spectral cues provided by KEMAR's ears. All further testing was conducted only on the six subjects who passed the screening.

(2) Even the six subjects who passed the screening reported that the images were not externalized to the appropriate place in space (1.75 m in front of them). Rather, the images were either internalized within their heads, or externalized to within a few inches of the head. This failure to externalize properly is probably the result of the use of KEMAR's HRTFs rather than subjects' own individual HRTFs. Despite the imperfect externalization, however, all six subjects reported clear spatial motion of the images in the appropriate (horizontal, vertical, or diagonal) directions.

(3) For horizontal-plane discrimination, MAAs and MAMAs were independent of filter condition and were in line with previously published values. Thus, restricting the frequency content of signals to frequencies below 2000 Hz or above 6000 Hz had no measurable effect on spatial resolution. This suggests that subjects are equally good at resolving spatial positions based on ITDs (the primary cue underlying performance for low-frequency signals) and on ILDs and spectral cues (the primary cues underlying performance for the high-frequency signals).

(4) As expected, vertical-plane performance depended strongly on spectral content. For low-pass stimuli, spatial discrimination of targets in the vertical plane was extremely difficult or impossible.

(5) MAA and MAMA thresholds for all three filter conditions were lowest for the horizontal plane, slightly (but significantly) higher for the diagonal plane, and highest for the vertical plane. These results were similar in pattern to those reported by Perrott and Saberi (1990) and Saberi and Perrott (1990), except that these investigators generally found no differences between the horizontal and diagonal (up to 70°) thresholds.

(6) The diagonal-plane results of the present study, as well as those reported by Saberi *et al.* (1991), who presented signals from *behind* subjects, are generally consistent with an ''independent-contributions'' hypothesis, according to which spatial resolution in diagonal planes is based on independent contributions from a horizontal-plane mechanism sensitive to interaural difference cues and a vertical-plane mechanism sensitive to head- and pinna-related spectral cues. On the other hand, the diagonal-plane results of Perrott and Saberi (1990) are consistent with a ''constant-resolution'' hypothesis, according to which spatial resolution is independent of the plane of presentation (for planes up to about 70°). A possible reason for the different results is that the studies of Perrott and Saberi were conduced in real space, while ours were conducted with the pseudovirtual method. Additionally, it is possible the presence of a patterned visual field, in particular containing a visual representation of the diagonal trajectory, is necessary in order to obtain optimal thresholds in the diagonal plane, thus enabling a subject to achieve constant resolution.

(7) Measurements performed on the KEMAR-recorded signals indicated that ITDs and *overall* ILDs of diagonally-presented signals are not different from those measured for horizontally presented signals having the same horizontal extent. However, further measurements revealed that ILDs *within specific frequency regions* for diagonally presented signals may be larger (in either a positive or negative direction) than those measured with horizontally presented signals having the same horizontal-plane extent. These larger than expected ILDs may account for the constant resolution observed in the studies of Perrott and Saberi (i.e., for the surprisingly good diagonal-plane performance). Our subjects were unable to take advantage of these frequency-specific ILDs to achieve constant resolution, possibly as a result of the virtual technique employed to present the stimuli, or possibly as a result of the absence of visual information.

ACKNOWLEDGMENTS

The authors express their gratitude to Terrey Penn for her dedicated assistance in creating and measuring the signals. They also thank Dr. Daniel Ashmead for statistical con-

sultation, and Dr. Leslie Bernstein, Dr. Kourosh Saberi, and two anonymous reviewers for many helpful comments on previous versions of this manuscript. This work was supported in part by a grant from NIDCD.

APPENDIX: PREDICTIONS OF DIAGONAL ANGULAR THRESHOLDS UNDER THE "INDEPENDENT-CONTRIBUTIONS" HYPOTHESIS

The ''independent-contributions'' hypothesis, expressed by Eq. (1), states that performance in a diagonal-plane spatial resolution task is based on sensitivities to angular separation in the horizontal plane d'_{HP} and in the vertical plane d'_{VP}. The current study obtained threshold angles for performance in the horizontal and vertical planes. However, in order to generate predictions for diagonal-plane performance, we must know the forms of the psychometric functions underlying performance in these two planes.

In the absence of empirical psychometric functions, we began by assuming a ''reasonable'' shape for the psychometric functions underlying performance in the two-alternative, forced-choice discrimination task employed in both experiments. For this we chose a linear function for d' vs angular separation. This psychometric function, when plotted as percent correct vs angular separation, has the shape of the upper half of the normal probability integral, which has been shown to accurately describe performance in several tasks (including interaural time discrimination) involving nonintensive stimulus changes (Saberi, 1995; Saberi and Green, 1997). The function may be written simply as

$$d' = b \cdot x, \tag{A1}$$

where x represents the angular separation of two sources (or the angular extent of the sweep for MAMAs), and b is the slope.

Because the function passes through the origin, it is completely determined by selecting one additional point, which we take from the data for a particular horizontal-plane condition as $(x_{\mathrm{thr}}, 1.16)$, where x_{thr} represents the estimated threshold in degrees for the condition under study, and the d' score of 1.16 corresponds to $P(C) = 79.4\%$ in the two-interval procedure (the target percentage for threshold).

Example psychometric functions for horizontal-plane (HP) and vertical-plane (VP) performance are shown as the solid and dashed curves, respectively, in Fig. 9. These were generated, as described above, based on a measured HP threshold of 1.3° and a measured VP threshold of 6.0° [note the place of the intersection of the psychometric functions with the horizontal line drawn at $P(C) = 79.4\%$].

We represent the computed psychometric functions as

$$d'_{\mathrm{HP}} = b_{\mathrm{HP}} \cdot x,$$

and (A2)

$$d'_{\mathrm{VP}} = b_{\mathrm{VP}} \cdot x,$$

where x again represents angular separation, and b_{HP} and b_{VP} are the psychometric function slopes for horizontal- and vertical-plane performance, respectively, derived from a subject's thresholds. One may now generate the psychometric function for a diagonal plane as follows: For a given diagonal angular extent x_{D}, we first computed the horizontal and vertical components of the diagonal distance. These are computed as

$$x_{\mathrm{H}} = x_{\mathrm{D}} \cdot \cos(\alpha), \quad x_{\mathrm{V}} = x_{\mathrm{D}} \cdot \sin(\alpha), \tag{A3}$$

where x_{H} and x_{V} represent the component horizontal and vertical extents, and α is the angle of the diagonal plane relative to the horizontal plane (60° in this experiment). Now, d'_{diag} for angular extent x_{D} may be computed by substituting the x values from Eq. (A3) into Eq. (A2), and then substituting the obtained d' values into Eq. (1). This results in one point on the new psychometric function. Repeating the process for different values of x_{D}, one generates the entire psychometric function for the diagonal plane (shown as the dash-dot function in Fig. 9). Predicted diagonal threshold under the independent-contributions hypothesis is then taken as that point on the function that intersects the line $P(C) = 79.4\%$ (2.4° in the example shown).

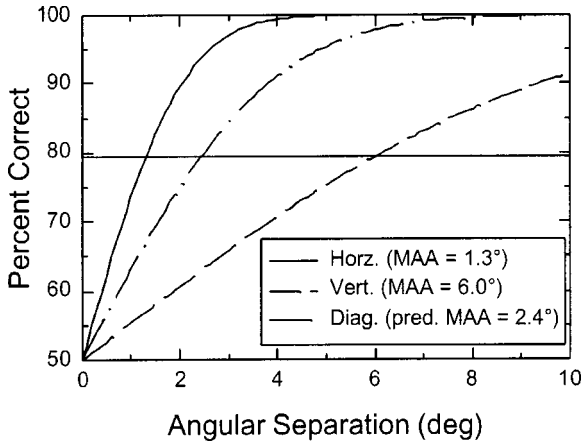


FIG. 9. Hypothetical psychometric functions for a horizontal-plane condition (left-most function) and for a vertical-plane condition (right-most function). Thresholds (the points where the psychometric functions intersect $p(c) = 79.4\%$) are assumed to be 1.3° and 6.0°, respectively. The dash-dot curve shows the predicted psychometric function for the diagonal plane (60°), according to an independent-contributions hypothesis [Eq. (1)]. Predicted threshold is 2.4°.

1. Predictions for diagonal angular thresholds in cases in which vertical-plane thresholds were not obtainable

For those subjects and conditions in which vertical-plane thresholds could not be measured (see Table I), the same basic procedure as described above was followed to generate predicted diagonal angular thresholds, but the slope of the psychometric function for vertical-plane resolution [b_{VP} in Eq. (A2)] was assumed to be 0. In this case, Eq. (1) reduces to

$$d'_{\mathrm{diag}} = d'_{\mathrm{HP}}. \tag{A4}$$

This means that the diagonal-plane prediction is determined entirely by the horizontal component of the diagonal trajectory, with no contribution from the vertical-plane system. This manner of prediction is intuitively reasonable, given that no vertical-plane threshold could be measured in these cases.

Because the horizontal component of a diagonal trajectory (x_H) is related to the diagonal extent (x_D) according to Eq. (A3), the predicted threshold for a diagonal trajectory ($\alpha=60°$) based solely on horizontal-plane discrimination turns out to be 2.0 times the value of the horizontal-plane threshold. By way of comparison, it should be noted that the vertical-plane system generally contributes relatively little to the diagonal-plane prediction even when vertical thresholds are obtained: For the case when the vertical threshold is 4–5 times the horizontal threshold (as in Fig. 9), the predicted diagonal threshold (for $\alpha=60°$) is about 1.8 times the value of the horizontal-plane threshold.

[1]Personal communication from Kourosh Saberi, 2003.

[2]A second type of cue interaction that might occur is that involving spectral changes in one ear produced by sources along a diagonal plane. In particular, the spectral change (in one ear) that accompanies a given position change along the diagonal plane may, as a result of the effects of head, pinnae, and torso for sources that have both horizontal and vertical extents, be greater (in certain frequency ranges) than the changes associated with an elevation displacement of the same magnitude that is confined to the vertical plane. Such an interaction of spectral cues could also lead to better performance in the diagonal plane than that predicted by the independent-contributions hypothesis.

Algazi, V. R., Avendano, C., and Duda, R. O. (**2001**). ''Elevation localization and head-related transfer function analysis at low frequencies,'' J. Acoust. Soc. Am. **109**, 1110–1122.

ANSI (**1989**). ANSI S3.6-1989, ''Specifications for Audiometers'' (American National Standards Institute, New York).

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization, Revised Edition* (MIT, Cambridge, MA).

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.

Grantham, D. W. (**1997**). ''Auditory motion perception: Snapshots revisited,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Mahwah, NJ), pp. 295–313.

Grantham, D. W. (**1986**). ''Detection and discrimination of simulated motion of auditory targets in the horizontal plane,'' J. Acoust. Soc. Am. **79**, 1939–1949.

Grantham, D. W. (**1995**). ''Spatial hearing and related phenomena,'' in *Handbook of Perception and Cognition: Hearing*, edited by B. C. J. Moore (Academic, San Diego), pp. 297–345.

Hafter, E. R., Buell, T. N., Basiji, D. A., and Shriberg, E. E. (**1988**). ''Discrimination of direction for complex sounds presented in the free-field,'' in *Basic Issues in Hearing: Proceedings of the 8th International Symposium on Hearing*, edited by H. Duifhuis, J. W. Horst, and H. P. Wit (Academic, London), pp. 394–401.

Harris, J. D., and Sergeant, R. L. (**1971**). ''Monaural/binaural minimum audible angles for a moving sound source,'' J. Speech Hear. Res. **14**, 618–629.

Hebrank, J., and Wright, D. (**1974**). ''Spectral cues used in the localization of sound sources on the median plane,'' J. Acoust. Soc. Am. **56**, 1829–1834.

Koehnke, J., and Besing, J. M. (**1996**). ''A procedure for testing speech intelligibility in a virtual listening environment,'' Ear Hear. **17**, 211–217.

Lakatos, S. (**1995**). ''The influence of visual cues on the localisation of circular auditory motion,'' Perception **24**, 457–465.

Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Litovsky, R. Y., and Macmillan, N. A. (**1994**). ''Sound localization precision under conditions of the precedence effect: Effects of azimuth and standard stimuli,'' J. Acoust. Soc. Am. **96**, 752–758.

Middlebrooks, J. C. (**1999**). ''Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency,'' J. Acoust. Soc. Am. **106**, 1493–1510.

Mills, A. W. (**1958**). ''On the minimum audible angle,'' J. Acoust. Soc. Am. **30**, 237–246.

Perrott, D. R., and Marlborough, K. (**1989**). ''Minimum audible movement angle: Marking the end points of the path traveled by a moving sound source,'' J. Acoust. Soc. Am. **85**, 1773–1775.

Perrott, D. R., and Musicant, A. D. (**1977**). ''Minimum auditory movement angle: Binaural localization of moving sound sources,'' J. Acoust. Soc. Am. **62**, 1463–1466.

Perrott, D. R., and Pacheco, S. (**1989**). ''Minimum audible angle thresholds for broadband noise as a function of the delay between the onset of the lead and lag signals,'' J. Acoust. Soc. Am. **85**, 2669–2672.

Perrott, D. R., and Saberi, K. (**1990**). ''Minimum audible angle thresholds for sources varying in both elevation and azimuth,'' J. Acoust. Soc. Am. **87**, 1728–1731.

Perrott, D. R., and Tucker, J. (**1988**). ''Minimum audible movement angle as a function of signal frequency and the velocity of the source,'' J. Acoust. Soc. Am. **83**, 1522–1527.

Platt, B. B., and Warren, D. H. (**1972**). ''Auditory localization: The importance of eye movements and a textured visual environment,'' Percept. Psychophys. **12**, 245–248.

Roffler, S. K., and Butler, R. A. (**1968**). ''Factors that influence the localization of sound in the vertical plane,'' J. Acoust. Soc. Am. **43**, 1255–1259.

Saberi, K., Dostal, L., Sadralodabai, T., and Perrott, D. R. (**1991**). ''Minimum audible angles for horizontal, vertical, and oblique orientations: Lateral and dorsal planes,'' Acustica **75**, 57–61.

Saberi, K. (**1995**). ''Some considerations on the use of adaptive methods for estimating interaural-delay thresholds,'' J. Acoust. Soc. Am. **98**, 1803–1806.

Saberi, K., and Green, D. M. (**1997**). ''Evaluation of maximum-likelihood estimators in nonintensive auditory psychophysics,'' Percept. Psychophys. **59**, 867–876.

Saberi, K., and Perrott, D. R. (**1990**). ''Minimum audible movement angles as a function of sound source trajectory,'' J. Acoust. Soc. Am. **88**, 2639–2644.

Shelton, B. R., and Searle, C. L. (**1980**). ''The influence of vision on the absolute identification of sound-source position,'' Percept. Psychophys. **28**, 589–596.

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). ''Localization using nonindividualized head-related transfer functions,'' J. Acoust. Soc. Am. **94**, 111–123.

Effect of the speed of a single-channel dynamic range compressor on intelligibility in a competing speech task

Michael A. Stone[a] and Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

(Received 14 November 2002; revised 24 April 2003; accepted 19 May 2003)

Using a ''noise-vocoder'' cochlear implant simulator [Shannon *et al.*, Science **270**, 303–304 (1995)], the effect of the speed of dynamic range compression on speech intelligibility was assessed, using normal-hearing subjects. The target speech had a level 5 dB above that of the competing speech. Initially, baseline performance was measured with no compression active, using between 4 and 16 processing channels. Then, performance was measured using a fast-acting compressor and a slow-acting compressor, each operating prior to the vocoder simulation. The fast system produced significant gain variation over syllabic timescales. The slow system produced significant gain variation only over the timescale of sentences. With no compression active, about six channels were necessary to achieve 50% correct identification of words in sentences. Sixteen channels produced near-maximum performance. Slow-acting compression produced no significant degradation relative to the baseline. However, fast-acting compression consistently reduced performance relative to that for the baseline, over a wide range of performance levels. It is suggested that fast-acting compression degrades performance for two reasons: (1) because it introduces correlated fluctuations in amplitude in different frequency bands, which tends to produce perceptual fusion of the target and background sounds and (2) because it reduces amplitude modulation depth and intensity contrasts.
 [DOI: 10.1121/1.1592160]

PACS numbers: 43.66.Ts, 43.71.Gv [KWG]

I. INTRODUCTION

Listeners with cochlear hearing impairment require assistive listening devices that reduce the dynamic range of real-world signals (Pearsons *et al.*, 1976; Levitt, 1982; Killion, 1997) so as to fit the residual dynamic range of the impaired cochlea. Dynamic range compressors, sometimes called automatic gain controls (AGCs), have been used to perform this function. There has been much experimenting, and consequent debate, as to the parameter sets that are ''optimal'' for performing the compression (for reviews, see Dillon, 1996; Moore, 1998). Parameters that may be varied include the number of compression channels (whose gain variation may be interlinked), the speed of gain variation, the compression ratio, and the compression threshold. The debate primarily hinges around the compromise between restoring audibility of brief low-level sounds (Villchur, 1973; Yund and Buckles, 1995a, b) and the reduced discrimination that may result from the reduction in signal modulation depth (Lippmann *et al.*, 1981; Plomp, 1988; van Buuren *et al.*, 1999). There appears to be no ideal parameter set for any one configuration of hearing loss.

Early speech researchers considered speech as a series of temporal modulations impressed on narrowband carriers which were either part of a harmonic complex, a noise source, or both (Dudley, 1939). More recently, investigators have used this model to dissociate the temporal envelope from the carrier fine structure and have independently manipulated the two properties, although some cross-effects are unavoidable (Van Tasell *et al.*, 1987; Drullman *et al.*, 1994a, b; Shannon *et al.*, 1995; Loizou *et al.*, 1999). These investigators have shown that, although the fine structure does carry information, most information is carried in envelope modulations with rates between about 2 and 32 Hz.

Drullman (1994a, b) showed that speech intelligibility could be degraded by reducing or removing temporal modulations in the range 2–32 Hz. Dynamic range compressors, by using a control signal derived from the signal envelope, also have the effect of reducing the envelope modulations. The amount of reduction depends on the attack and release time constants, as well as the compression ratio. Braida *et al.* (1982) and Stone and Moore (1992) demonstrated this reduction as a function of envelope modulation frequency for some typical fast-acting compressors (these are often called syllabic compressors as their gain varies significantly over syllabic timescales). However, deterministic reduction of envelope modulation, as produced by compression, has smaller deleterious effects on intelligibility than reduction by addition of noise (Noordhoek and Drullman, 1997).

Commercial cochlear implants have traditionally used a single-channel, fast-acting compressor with a high compression ratio prior to the band-pass filtering necessary to produce a channel signal for each electrode (Clark *et al.*, 1990). A second stage of compression has usually been necessary to map the channel signals into the current range between threshold and discomfort at each electrode. The attack time of the front-end compressor has typically been short, around a few milliseconds, so as to provide protection from sudden intense sounds, which would otherwise sound too loud. The release time has typically been several hundred milliseconds.

[a]Electronic mail: mas19@cam.ac.uk

However, compression thresholds have been surprisingly high (around 70 dB SPL) compared to those found in acoustic hearing aids, so the compression was only activated by signals with moderate to high levels (Seligman and Whitford, 1995). Consequently, large compression ratios have been necessary in the second stage compressors so as to ensure audibility while avoiding discomfort. With the limited dynamic range typical of electrical stimulation (Clark *et al.*, 1990), and the limited number of discriminable levels of electrode current (Loizou *et al.*, 1999), the result may be that only large differences in level within and across frequency regions are detectable. This in turn may adversely affect the ability to understand speech (Loizou *et al.*, 1999). If the front-end compression threshold were to be reduced, then the front-end compressor would become more active, allowing less compression of the electrode signals and possibly giving an improved ability to perceive level contrasts within and across electrodes. In simulations of cochlear implants, Loizou *et al.* (1999, 2000) showed that speech intelligibility improved as the number of discriminable levels was increased, at least up to a certain point. However, greater activity of a front-end *fast* compressor would reduce envelope modulations in the 2–32-Hz range, which might counteract potential improvements resulting from reduced compression of the individual channel signals. A slow-acting front-end compressor would not have this effect, and therefore might lead to better intelligibility.

Only a few studies have examined the effect of fast-acting compression on speech intelligibility using either simulated or real cochlear implants. Van Tasell and Trine (1996) studied the intelligibility of vowel–consonant–vowel clusters and closed-set sentences that were transformed to ''signal-related noise.'' This may be regarded as simulating a single-channel implant. The stimuli were presented in quiet and processed using a single-channel fast compressor (4.5-ms attack and 40-ms release times) with a high compression ratio of 8. The compression led to lower intelligibility than obtained without compression. McDermott *et al.* (2002) tested multi-channel cochlear implantees, using either a single-channel syllabic compressor with a compression ratio of 2 or no processing at the front end. They assessed the perception of speech in quiet for input levels between 45 and 70 dBA. Performance was better for the compression system than for the linear system at all levels tested, but the difference was smaller at the higher levels. The declining benefit of compression with increasing input level suggests that audibility was the primary factor limiting performance rather than the ability to resolve intensity contrasts. The difference between their results and those of van Tasell and Trine (1996) can be attributed to two factors: (1) More information was available to the subjects through the use of multiple channel stimulation, and (2) McDermott *et al.* used a lower compression ratio than van Tasell and Trine. In contrast to the findings for speech in quiet, McDermott *et al.* reported no significant effect of compression for speech presented in a steady background noise.

For a fluctuating signal that is controlled by a single-channel compressor, all components of the signal have their gain controlled by whichever component or mixture of components is most intense at a given time. Since the gain is time varying, this means that amplitude fluctuations become correlated across frequency, even if the fluctuations in the original signal were relatively independent. For a single voice input, correlation in amplitude fluctuations across channels already exists, but the degree of correlation may be increased by the operation of the compressor. A second effect occurs when the input is a mixture of two independent modulated signals, such as would arise from two competing speakers with similar overall levels. The voice which determines the gain signal switches from moment to moment. Hence, amplitude fluctuations in each voice lead to correlated fluctuations in the other voice at the output of the compressor; the two signals become ''comodulated'' and lose part of their independence. Common envelope modulation promotes perceptual fusion of simultaneous sounds (Bregman, 1990). Hence the compressor might be expected to disrupt the perceptual separation of two voices. The major goal of this study was to determine whether fast-acting compression does indeed degrade the ability to identify speech in the presence of a background talker, using a simulation of a cochlear implant.

In a multi-channel cochlear implant, each channel signal can be considered as an envelope that modulates the ''fine structure'' of the signal (Bracewell, 1986). Even when ''high-rate'' pulsatile strategies are employed, implant users appear unable to extract information about the fine structure; essentially, only the envelope information is conveyed (Moore, 2003). A simulation of the transmission of envelope cues has been proposed and explored by Shannon *et al.* (1995). They filtered speech into a small number of broad frequency bands and extracted the envelope in each band. These envelopes were then used to modulate noises with passbands corresponding to the original bandwidths used for filtering. For example, the envelope extracted from speech filtered between 800 and 1500 Hz was used to modulate noise with a passband from 800 to 1500 Hz. The modulated noise bands were then added together.

Shannon *et al.* (1995) and Loizou *et al.* (1999) explored the number of channels necessary to provide comprehension of speech when no background noise was present. They concluded that between three and four channels were necessary to give 50% intelligibility in a speech-in-quiet task. Friesen *et al.* (2001) found that a greater number of channels was required to understand speech in the presence of a continuous background noise for a variety of speech-to-noise ratios. We used such a simulation in our study, but did not use a steady noise background. Such noise lacks the amplitude fluctuations which might comodulate the target signal during the process of compression. We used a single-talker background, which would lead to comodulation of the target and background. The first experiment reported here investigated the intelligibility of sentences in a single-talker background as a function of the number of channels in a simulated cochlear implant. In two further experiments, we explored the effect of two front-end single-channel compression systems, differing in their time constants.

II. EXPERIMENT 1: ESTABLISHING A BASELINE IN PERFORMANCE

A. Speech material

All speech was filtered such that the average long-term spectrum matched that of the HINT sentence material (Nilsson *et al.*, 1994). The power spectral density was approximately constant up to 500 Hz, and changed by about −9 dB/oct above that. All processing was performed at a sampling rate of 16 kHz. For the target sentence material, the ASL lists (MacLeod and Summerfield, 1990) were used. These are composed of 18 lists of 15 short sentences, with three key words in each, spoken by a male talker of British English. The speech used as the background and also for training was recorded from male speakers of British English reading naturally from scripts in a large sound-isolated and sound-treated room with a low reverberation time (RT_{60} <50 ms for 250 Hz and above). Recordings were made direct to digital audio tape using a high-quality, low-noise condenser microphone. Using CoolEdit2000™, stammerings, repetitions, and pauses for breath were removed, but natural-sounding pauses between sentences with durations of 100–140 ms were left. Each speaker used a different script. The recorded length of the interfering signal greatly exceeded the 45-s duration of each target sentence list. The start point of the competing speech was varied randomly within the file for each sentence list.

The level of a given speech (or background talker) file was measured by producing a histogram of the root mean square (rms) values of 10-ms segments of the file. The histograms were generally bimodal with a primary peak representing the speech energy, and a secondary peak at lower levels representing the undesired noise inherent in the recording. A threshold close to the bottom of the valley between the two peaks was selected and the overall rms value of the 10-ms segments for each file that exceeded this threshold was taken as the measure of the speech level for that file. The proportion of segments lying below the threshold was about 20%.

In later experiments, fast or slow compression was applied to the stimuli (see later for details). The histograms of the mixed speech after compression processing were less distinct in shape. A histogram of the rms values of 10-ms segments of the respective speech files was produced. An overall mean was calculated for all of the 10-ms segments. A threshold was selected 15 dB below this mean. Typically, about 20% of segments lay below this threshold. The rms value of the 10-ms segments that exceeded this threshold was used to calculate the overall output level. The value calculated in this way changed by about 0.1 dB per 1-dB change in the threshold value. The output rms value was scaled to be the same for all files used during an experiment. This was done to ensure that all stimuli were roughly equally loud.

B. Signal processing for simulation of a cochlear implant

The speech-plus-background signal to be processed was ''preemphasized'' by applying a gain rising at 3.3 dB/oct between 500 and 4000 Hz, giving a total of 10 dB. Below 500 Hz the gain was 0 dB, and above 4000 Hz it was 10 dB. The signal was then filtered into one of a specified number of channels. The maximum number of channels was chosen such that the analysis/synthesis filters used in the processing were not narrower than the width of a normal human auditory filter at the same center frequency (Glasberg and Moore, 1990). The channel edge frequencies were calculated using the ANSI (1997) standard method for calculating the Speech Intelligibility Index (SII), such that each channel made an equal contribution to SII. We used the ''21 critical band'' procedure, and the band-importance function for ''average speech'' (Pavlovic, 1987). The anti-alias filter used here had a cutoff frequency of 7800 Hz, which leads to a maximum possible SII value of 0.99.

Strictly speaking, the ANSI (1997) method is not applicable to speech processed as described below. However, the intention was to have similar amounts of speech information carried in each analysis channel, so as to maximize the information transmission of the final signal and allow easier interpretation of the effect of varying the number of channels.

Although the low edge of the lowest channel and the high edge of the highest channel for all processing systems had theoretical values defined as described above, in practice the lowest and highest channels were configured as low- and high-pass filters, respectively, and the final stimuli were later bandpass filtered between 100 and 7800 Hz (see below for details). All channel filters were implemented using a finite-impulse-response (FIR) method, each of which introduced a frequency-independent time delay. They were designed such that the low-pass edge of one filter had the complementary response to the high-pass edge of the next higher filter. To ensure smooth recombination, the length of each FIR filter was adjusted so that the response in the transition bands was similar in shape across channels, when plotted on a logarithmic frequency scale. The outputs of all filters were time aligned. At the crossovers between adjacent channels, each filter response was at the −6-dB point, which, given that the two responses were in phase, meant that the response summed to unity. At the cross-over between a channel and its next-but-one neighboring channel, each filter response was below the −23-dB point. Addition of the channel outputs gave virtually perfect recombination: within the frequency range 100 to 7800 Hz, the ripple was less than ±0.2 dB. The channel edge frequencies for the different numbers of channels used are given in Table I.

After filtering, the channel envelopes were extracted by full-wave rectification and low-pass filtering. The low-pass filter had a linear phase response, and a response that was −0.5, −3, and −40 dB at 40, 50, and 100 Hz, respectively. The bandwidth should have been high enough to give near-asymptotic performance (Shannon *et al.*, 1995; Drullman *et al.*, 1994a, b).

The channel envelopes were each used to modulate a broadband white noise, after which the modulated noises were bandlimited to the same widths as those of the corresponding analysis filters. The band-limiting filters each included a shaping filter (i.e., the response was not uniform

TABLE I. Channel edges (Hz) of the analysis/resynthesis filters used in the processing.

4 channels	*100*				759.4				1529.5				2919.5				*7800*
6 channels	*100*		567			974.7			1529.5			2347.5		3662.3			*7800*
8 channels	*100*		484.3		759.4		1093.2		1529.5		2110.7		2919.5		4141.9		*7800*
11 channels	*100*	420.1		599		815.2		1059.4	1358.3	1719.7		2172.7		2749.2		3512.3	4649	*7800*
16 channels	*100*	357.3	484.3	611.5	759.4	920.4	1093.2	1298.9	1529.5	1797.2	2110.7	2478.3	2919.5	3457.7	4141.9	5209	*7800*

across the passband) such that the HINT spectral shape was reapplied to the output of the processing. This was especially important when the number of channels was small. If this shaping had not been performed, the long-term spectral shape of the output signal (after the channels were recombined) would have been a series of rectangular blocks of differing widths depending on the number of analysis filters. Time delays were added to the channel signals to compensate for the time delays introduced by the FIR filters (which introduced delays that varied with center frequency), such that the relative timing of envelope signals across frequency was the same for the input and the output. The filtered noise bands were then added back together and a linear-phase six-pole infinite impulse response high-pass filter with a corner frequency of 100 Hz was applied to remove spurious low-frequency noise.

C. Subjects and method

Ten volunteer subjects (5M, 5F, aged 18–21 years), all university undergraduates, were selected on the basis of their having audiometric thresholds ≤15 dB HL at octave frequencies between 125 and 8000 Hz and at 3000 and 6000 Hz. Subjects attended one session, which lasted just over 1 h after initial audiometric screening. Subjects were paid for their attendance.

A MATLAB™ script was used to control replay of preprocessed audio files through a high-quality audio card (Turtle Beach Montego II) hosted within an IBM-compatible PC. Levels were controlled via a Mackie 1202 mixing desk which also handled microphone signals from both the subject and the experimenter. Each subject was seated in a double-walled booth. The subject was continually monitored by the experimenter, but never heard their own voice through their headphones. The experimenter's microphone was only routed to the subject during pauses in the signal presentation. Signals were presented diotically at 68 dB SPL via Sennheiser HD580 headphones. This ensured near-maximum audibility of all channels, while avoiding a possible rollover in performance that can occur at high presentation levels (Fletcher, 1953; Studebaker *et al.*, 1999). The experimenter operated the equipment from outside of the booth and transcribed the oral feedback.

D. Procedure

Since the processed signals were novel to the subjects, a training period was necessary before testing proper began. Shannon *et al.* (1995) used long training periods, about 8–10 h, before data collection. Dorman *et al.* (1997, 1998) and Loizou *et al.* (1999) used much shorter training periods, varying from several tens of minutes down to several minutes. Where comparisons could be made, Dorman *et al.* (1997) found little difference between their results and those of Shannon *et al.* (1995), apart from with vowel material, for which the shorter training period may have contributed to poorer performance. We opted for a relatively short training period, but counterbalanced the presentation order among subjects so as to control for longer-term learning effects.

Training was performed in two phases before each processing condition was tested. The first training phase was intended to help subjects to get used to and interpret the processed speech in quiet. A 45-s segment of continuous speech was processed and mixed with the original (nonprocessed) speech at four differing ratios of original-to-processed signal. (The original and processed speech were aligned in time to remove the delay introduced by the processing.) For the first presentation, the ratio was 0 dB. At this ratio, the speech was clearly comprehensible, even for the lowest number of channels. After the presentation, the subject could either ask for a repeat of the presentation or continue to a presentation with a more adverse ratio. Subsequent presentations used ratios of −6, −15, and −24 dB. At −24 dB, the unprocessed signal was essentially masked by the processed signal. After each presentation the subject was offered the opportunity for a further presentation. This offer was rarely used. In the second training phase, the subject was presented consecutively with processed stimuli consisting of 12 sentences of the target speaker mixed with the competing speaker at a speech-to-background ratio of +10 dB. The target sentences were recorded at the same time as the original ASL corpus (MacLeod and Summerfield, 1990), but were not released as part of that corpus: in total they comprise four lists of 15 sentences each. A transcription of the content of these sentences was available to the subject. This presentation was always repeated, and an option for a third presentation was available to the subject. The 12 training sentences were not used again either for training or testing.

After this two-stage training phase, the intelligibility test was conducted at a signal-to-background ratio of +5 dB (with the rms level of the target and the background each measured as described earlier). Three consecutive ASL lists were used, making a total of 45 target sentences. For each presentation of a target sentence, the background was ramped up over 0.25 s, the ramp starting 1 s before the target sentence. After the target sentence was completed, the background continued for about a further 0.5 s before being ramped down over 0.25 s. In a subsidiary training phase, to allow subjects to get used to the experimental method and the specific processing condition, the first 12 sentences of each block of 45 sentences were presented twice, with a short pause between each presentation. Subjects were encouraged to respond after each presentation, even if the sentence appeared to be nonsense. During the process of double

presentation of sentences in the first processing condition used with a given subject, and to a lesser extent in the second processing condition, subjects were given feedback when they made incorrect responses. The results from the first 12 sentences of each condition were later discarded, leaving 33 sentences containing 99 words. The remaining 33 sentences were only presented once. For each subject, five conditions differing in channel number (4, 6, 8, 11, or 16) were tested in a counterbalanced design, using a different testing order for the different gender groups.

E. Results

The scores were transformed into rationalized arcsine units (RAU, Studebaker, 1985). Despite the training given, there was evidence that performance increased across successive conditions, i.e., a training effect occurred. To test this, each data point within a subject group was normalized by dividing by the mean score for the processing condition that the point came from. The data were then time-ordered and the mean and standard deviation of the score were calculated for each time slot. Using a *t*-test, we compared the means for the first and last conditions presented. For both gender groups there was a highly significant difference between the two (males, $t=5.38$, $8\,df$, $p<0.005$, one-tailed; females, $t=3.3$; $8\,df$, $p<0.01$, one-tailed). However, there was no significant difference between the training effects for the two groups.

As the two groups showed similar training effects, the time-ordered data in RAU were averaged across all subjects for each time of presentation. The time-order averaged data were then expressed relative to the average score obtained in the last condition. The resulting relative scoring rates, in the first four conditions tested, were 0.70, 0.83, 0.95, and 0.97. We conclude that the training effect occurred primarily during the first two conditions, which lasted in total about 25 min. To correct for the training effect, the scores for each of the first four conditions tested for each subject were divided by the relative score for that condition, e.g., the score for the second condition tested was divided by 0.83. The correction for training removed a large source of variation in the raw (uncorrected) data.

The correction for training is based on the assumption that the training effect was similar across subjects despite the slightly different order of testing conditions experienced by each subject. This is probably not quite correct but may be a reasonable approximation. The corrected scores (in RAUs) were subjected to an analysis of variance (ANOVA), with factors number of channels and gender (now ignoring time order). The ANOVA showed no significant effect of gender, $F(1,8)=0.003$, $p=0.96$, or interaction between number of channels and gender, $F(4,32)=1.97$, $p=0.123$, so all further results will be presented collapsed across gender. There was a significant effect of number of channels, $F(4,32)=461.05$, $p<0.001$. The difference between the scores for 11 and 16 channels was significant at $p<0.025$ (one-tailed *t*-test, $t=2.53$, $9\,df$). All other differences between pairs of channels were significant at $p<0.001$ ($t>6.5$, $9\,df$). The progressive improvement in performance with increasing number of channels is consistent with the results of Qin and Oxenham (2003), although they measured SRTs rather than percent correct for a fixed speech-to-background ratio.

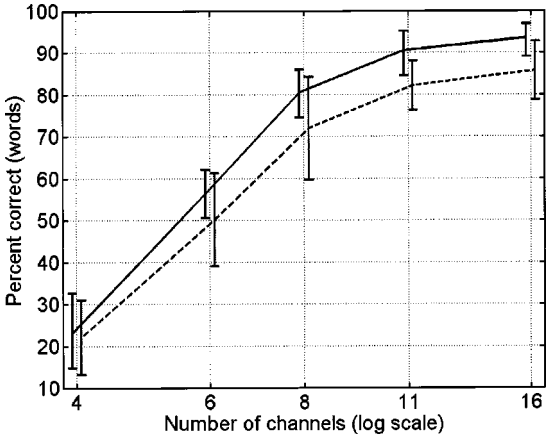


FIG. 1. Results of experiment 1. The dashed line shows results before correction for training effects and the solid line shows results after correction. Lines are horizontally displaced for clarity. Error bars show ±1 standard deviation.

Figure 1 (solid line) shows the mean corrected word intelligibility scores, transformed back from RAUs into percentages, as a function of number of channels. The error bars show standard deviations (SDs) across subjects. The dashed line shows means and SDs for the original (uncorrected) data. The SDs for 6, 8, and 16 channels were much reduced by the compensation for training effects, but the SDs were little changed for 4 and 11 channels.

III. EXPERIMENT 2: INTELLIGIBILITY AS A FUNCTION OF NUMBER OF SIMULATED CHANNELS AND SPEED OF DYNAMIC RANGE COMPRESSION

The aim of this experiment was to assess the effect on speech intelligibility of simulations of two compressors found on commercial cochlear implant systems. The compressors have very different time constants as defined by the 25-dB step technique described in ANSI (1996). Stone *et al.* (1999) pointed out that this technique employs an output measure, so for the same averaging circuit within the compressor, the measured time constants change with the compression ratio used. Since the compression ratios of the systems that we wished to simulate were numerically very different, we used a measure of compression ratio allowing more meaningful comparisons between systems.

A. Characterizing the performance of a dynamic range compressor

We mentioned earlier how compressors reduce the depth of envelope modulation; the amount of the effect depends on the modulation frequency, the compression ratio, and the time constants used in the compressor gain averaging circuit. Both Braida *et al.* (1982) and Stone and Moore (1992) characterized the performance of a compressor by plotting the effective compression ratio as a function of the sinusoidal modulation frequency of the input. Effective compression ratio, cr_e, was defined as the ratio of the peak-to-valley input amplitude in decibels to the peak-to-valley output amplitude, again in decibels. Mathematically,

$$cr_e = \frac{\text{Input modulation depth (dB)}}{\text{Output modulation depth (dB)}} \leqslant cr_{\text{static}}, \qquad (1)$$

where cr_{static} is the asymptotic compression ratio for signals with very slow modulation. The measure cr_e does not take into account the distortion of the envelope that commonly happens with compressors, especially when using high compression ratios, and when attack and release time constants are not the same.

A measure that may be more meaningful than cr_e is the ''fractional reduction of modulation,'' f_r. This measure indicates the relative amount of modulation removed by the compressor:

$$f_r = \frac{\text{Change in modulation depth}}{\text{Original modulation depth}} = \frac{(cr_e - 1)}{cr_e}. \qquad (2)$$

Using Eq. (2), consider the response of some compressors to modulation at very low modulation rates, where $cr_e = cr_{\text{static}}$. With no compression, $f_r = 0$. For a 2:1 compressor, $f_r = 0.5$. For a 10:1 compressor, $f_r = 0.9$, implying that there is little temporal variation left in the envelope. When calculating the gain, *G*, to be applied to a signal by a compressor, f_r appears as the negative of the exponent to which the mean input signal, *I*, is raised (Stone and Moore, 1992):

$$G = C \times I^{(1-cr)/cr}, \qquad (3)$$

where *C* is a constant and cr here refers to cr_{static}. The f_r scale proposed here is not perfect since it does not fully convey the degree of loss of information (Noordhoek and Drullman, 1997). However it does allow meaningful comparisons between compressors, especially those with high compression ratios. Additionally, it provides a method whereby the effect of the concatenation of compressors may be more accurately predicted.

B. Subjects and method

Twelve volunteer subjects (6M, 6F, aged 18–21 years), all university undergraduates, were selected using the same criteria as for experiment 1. Subjects attended one session, which lasted just over 1 h after initial audiometric screening. Subjects were paid for their attendance. Since subjects had exhibited a strong training effect in experiment 1, the subjects for this experiment were assigned to one of two groups according to whether or not they had participated in experiment 1. These two groups will be referred to as ''experienced'' and ''novice.'' Experienced subjects were not tested until at least 6 weeks had elapsed after their participation in experiment 1.

The equipment and method were nearly the same as used in experiment 1 except that, due to a limitation in the amount of sentence material available, only ten sentences (instead of 12) were used in the second stage of training where subjects were given a transcript of what they were hearing. Even for the experienced subjects, the first testing condition used sentence material from the ASL corpus that had not been used in experiment 1. The speech-to-background ratio was +5 dB, the same as in experiment 1.

C. The two different forms of dynamic range compression

The two compression systems used were

(a) a ''fast'' compressor with near syllabic time constants and
(b) a ''slow'' compressor, the ''AGCII'' system implemented in the ''PSP'' processor manufactured by Advanced Bionics Corporation.

In what follows, the words ''fast'' and ''slow'' in quotes will be used to denote the specific systems tested here. These are each described in more detail below. Both the ''fast'' and ''slow'' compressors were simulated in MATLAB™. All speech processing was performed ''off-line.'' The static compression ratio, cr_{static}, was 7. Both compressors operated on the ''preemphasized'' signal described earlier.

1. "Fast" compressor

The ''fast'' compressor was adjusted to have performance between that of the ''AGCI'' system found in the Advanced Bionics ''PSP'' processor and that of the analog compressor found in the older ''S'' series processor from the same manufacturer. These two fast compressors have the same attack and release time constants, as measured using the ANSI (1996) standard method; the attack time was 1–2 ms and the release time was about 360 ms. We measured f_r for both of these systems using the method described in Moore *et al.* (2001). One PC generated a sinusoidally modulated signal that was played from its sound card through the system under test, and a second PC recorded the output via its soundcard. The processed envelope was then extracted from the recorded signal, filtered to reduce noise, and its peak-to-valley ratio measured. Compressors with high ratios (greater than about 3), especially with fast attack times, can produce large distortion of the temporal envelope. To maintain ''fidelity'' of the envelope, the filter cutoff frequency was set to 11 times the modulation frequency. The carrier frequency was 1 kHz. The input level to each compressor was adjusted so that the valley of the modulation was always a few decibels above the compression threshold. We found larger than expected differences between the two fast compressors. For the analog compressor, f_r for high modulation rates decreased with increasing modulation depth of the input signal (implying an effective decrease in the speed of the compressor). The AGCI system did not show such variation with modulation depth. Its performance was similar to that of the analog compressor with a 25-dB peak-to-valley ratio. We chose to simulate a fast system whose variation of f_r with modulation rate was between that for the AGCI system and the analog system when the modulation depth was 25 dB. The final system chosen had attack and release times of 2 and 240 ms, respectively.

2. "Slow" compressor

The ''slow'' compressor was similar to that described as ''DUAL-HI'' in Stone *et al.* (1999) for an acoustic hearing aid application. This compressor employed two control loops acting in concert. The first, a slow loop, with sentence-length

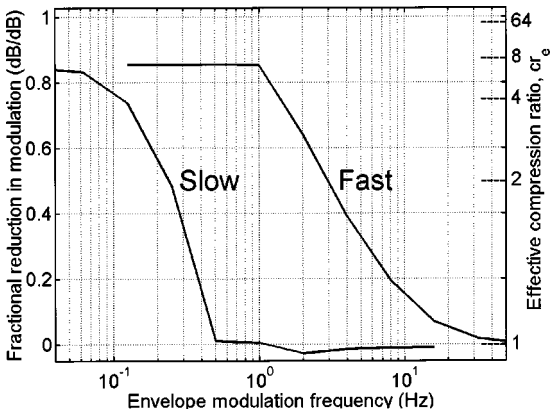


FIG. 2. Effect on modulation of the two compressors under test, as a function of the modulation rate of the input. The left ordinate shows the fractional reduction in modulation depth and the right ordinate shows the effective compression ratio.

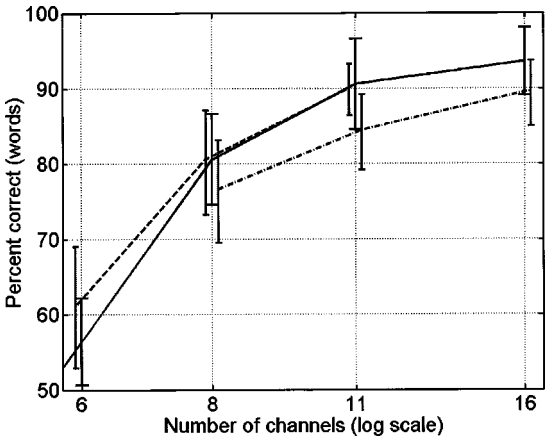


FIG. 3. Results of experiment 2, corrected for training effects. The solid line shows data from experiment 1 (baseline) after correction for training effects. The dashed and dash-dotted lines show results for "slow" and "fast" compression, respectively. Error bars show ±1 standard deviation.

time constants, calculated a running average of the input signal magnitude, and used this to set the gain to be applied to the signal. The second, a fast loop with much shorter time constants, also calculated a running average of the input magnitude. Normally, this loop had no effect. However, if the fast running average exceeded the slow running average by a fixed amount, then the fast loop took over the gain control. The main purpose of the fast loop was to protect the user from sudden increases in sound level which could either be transient, or part of a longer term increase in level. In the latter case, the fast compressor held the output at a safe level until the slow averaging system "caught up."

For the "slow" system, the time constants were about 25% less than published in Stone *et al.* (1999). This choice was made on the basis of field trials of the "PSP" system with implant patients in London and Los Angeles.

3. Choice of compression threshold

We chose compression thresholds that were lower than usually employed in cochlear implant processors, but not unrealistically low. This was done to ensure that the compressors were active during presentation of the target speech. We used input levels of 67 dB SPL (unweighted) for the target speech and 62 dB SPL for the background speech. The compression threshold was 55 dB for the "fast" compressor and 60 dB for the "slow" compressor. With such levels, simultaneous pauses in the competing and the interfering speech were rarely long enough to allow the gain to recover to its maximum value, as defined by the compression threshold.

4. Comparison of f_r for the two compressors

The value of f_r for the two compressors is plotted in Fig. 2 as a function of sinusoidal envelope modulation rate (modulation depth=10 dB). The 10-dB depth is close to the typical modulation depth at the speech-to-background ratio used in these tests. However, the value of f_r varied little with modulation depth. The "slow" compressor produced essentially no reduction in modulation depth for rates above 0.5 Hz. The "fast" compressor produced a reduction greater than 0.5 for rates up to 3 Hz and greater than 0.25 for rates up to 5 Hz.

D. Experimental conditions

Six conditions (two types of compression by three channel numbers) were tested in a counterbalanced design. We wanted to compare results for the two compressors by determining the number of channels required to achieve a given level of performance. Pilot trials indicated that performance would be worse for the "fast" than for the "slow" compressor. For this reason, we used 8, 11, or 16 channels for the "fast" system and 6, 8, or 11 channels for the "slow" system. Based on the results of experiment 1, this choice was expected to give scores in the range 60% to 90%.

E. Results

The scores were transformed into RAUs. Each data point within a subject group was then normalized by dividing by the mean score for the processing condition that the point came from. The data were then time-ordered and the mean and standard deviation of the score were calculated for each time slot. Using a *t*-test, we compared the means for the first and last conditions presented. For both the experienced and novice groups, the difference was highly significant (experienced, $t=7.0$, $10\,df$, $p<0.0005$, one-tailed; novice, $t=3.75$, $10\,df$, $p<0.005$, one-tailed). However, there was no significant difference between the training effects for the two groups. Averaged across the two groups, the scores for successive tests, relative to that for the last condition tested, were 0.85, 0.89, 0.90, 1.00, and 0.96. Prior to further analysis, the scores for both groups were "corrected" using these factors, as described earlier, to compensate for training effects.

The corrected scores, transformed from RAUs back into percentages, are plotted in Fig. 3 for the "fast" system (dashed-dotted line) and the "slow" system (dashed line). The solid line shows the data from experiment 1. The "slow" system did not reduce intelligibility relative to that for the baseline condition (no compression). This is as expected because the "slow" system should have produced little gain variation: its time constants were long compared to the periods of the modulation in the stimulus. Occasionally,

the fast compressor of the dual-loop system would have been activated so as to control signal peaks that were well above the longer-term mean value of the signal. It is encouraging to see that this did not alter intelligibility. For six channels, mean scores were slightly higher for the ''slow'' compressor than in the baseline, but this difference was not statistically significant. The ''fast'' compressor led to lower scores than in the baseline for all channel numbers tested.

To assess the significance of differences between the ''slow'' and ''fast'' compressors, an ANOVA was conducted using data only for channel numbers that were used with both compressors. The factors were speed of compression (''fast'' or ''slow''), number of channels (8 or 11), and group experience. There was a significant effect of experience, $F(1,10)=6.92$, $p=0.025$. The overall corrected means were 87.2 RAU for the experienced group and 80.6 RAU for the novice group. There were significant effects of speed of compression, $F(1,10)=18.13$, $p=0.002$, and number of channels, $F(1,10)=41.06$, $p<0.001$. There was no significant interaction between the speed of compression and number of channels, $F(1,10)=0.3$, $p=0.597$.

F. Interim discussion

The ''fast'' system consistently degraded intelligibility relative to the ''slow'' system (for 8 and 11 channels) and relative to the baseline (for 16 channels). The difference in the latter case was statistically significant ($t=2.34$, $p<0.005$, $20\,df$, one-tailed). Apart from these obvious decreases in percent correct, we can quantify the deleterious effect of the ''fast'' compressor by posing two questions: (a) How many extra channels are needed with the ''fast'' system to give performance equal to that obtained using the ''slow'' system; (b) What is the effective number of channels lost by using the ''fast'' system? To estimate these numbers, we used linear interpolation in the RAU domain: the domain change reduced the large changes in slope between each data point that were found when the data were expressed as percentages, as in Fig. 3. The 8-channel ''slow'' system produced the same performance as a 9.8-channel ''fast'' system, while the 11-channel ''slow'' system produced the same performance as a 16-channel ''fast'' system. The 8-channel ''fast'' system produced the same performance as a 7.5-channel ''slow'' system, while the 11-channel ''fast'' system produced the same performance as an 8.6-channel ''slow'' system. It is clear that the ''fast'' system produced material reductions in performance.

Experiments 1 and 2 both showed significant changes in performance over time, despite the training performed before data collection. In experiment 1, the training effect was small compared to the large effect of the number of channels. In experiment 2, the training effect was large compared to the differences between the compression systems, but it was necessary to use the same experimental procedure, so as to be able to make comparisons between the two experiments. For both experiments, the training reached a plateau after two to three conditions had been tested, i.e., after about 20–25 min of presentation of processed material.

The scores for experiment 2 spanned a range of approximately 28% (or 32 RAU). However, five of the six conditions gave scores that fell within the upper half of this range. There was a need, therefore, for a new experiment that would give a wider range of scores. We also wanted to improve the training before measures were taken. These points were addressed in experiment 3.

IV. EXPERIMENT 3: ADDITIONAL MEASUREMENT OF INTELLIGIBILITY DIFFERENCES BETWEEN THE TWO DYNAMIC RANGE COMPRESSION SYSTEMS

A. Subjects and method

Twelve volunteer subjects (5M, 7F, aged 18–30 years) were selected on the same basis as before. No subject had participated in experiment 1 or 2 and none was familiar with the speech material or processing method. Subjects were paid for their attendance. The subjects were split into two groups, novice and experienced, according to whether they had attended a prior training session. Again, six conditions (two types of compression by three channel numbers) were tested in a counterbalanced design. Six, 8, and 11 channels were used for both compression systems. Otherwise, the method was the same as for experiment 2.

B. Training of the two groups

The novice group underwent identical training as for experiment 2. The experienced group participated in a preliminary experiment using similar processing, the results of which will not be reported here. The training for the preliminary experiment occurred primarily in a block at the start of the session and lasted 15–20 min before the start of data collection. The preliminary experiment itself lasted about 50 min. This training also involved more feedback from the experimenter on correctness of responses than had been used in experiments 1 or 2. When a subject made an incorrect response, the sentence was presented a second time. If the subject still did not produce all of the key words correctly, the subject was informed of the correct answer, and the sentence was repeated before progressing to the next sentence. This use of greater feedback to the experienced group was continued into the experiment reported here which typically took place about 1 week after the preliminary experiment.

C. Results

The scores were transformed into RAUs. As for experiments 1 and 2, we looked for a training effect. There was a significant difference between the first and last conditions for both experienced and novice groups (experienced, $t=3.76$, $10\,df$, $p<0.005$, one-tailed; novice, $t=11.4$, $10\,df$, $p<0.0001$, one-tailed). The scores for each group separately were normalized by dividing by scores for the last condition. The resulting normalized scores for the successive conditions were 0.89, 0.98, 1.01, 0.92, and 0.95 for the experienced group and 0.63, 0.85, 0.92, 0.83, and 0.92 for the novice group. As expected, the training effect was larger for the novice group. The normalized scores for the first condition were significantly different for the two groups; $t=7.1$, $10\,df$,

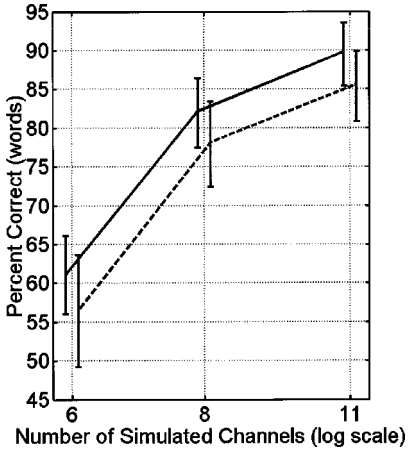


FIG. 4. Results of experiment 3, corrected for training effects. The solid and dashed lines show results for ''slow'' and ''fast'' compression, respectively. Error bars show ±1 standard deviation.

$p<0.0001$, one-tailed. Scores for successive conditions tested were corrected as before, but this time separately for each group.

The corrected scores were subjected to an ANOVA with factors speed of compression (''fast'' or ''slow''), number of channels (6, 8, or 11) and group experience. There was a significant effect of group experience, $F(1,10)=5.52$, $p=0.041$, but the effect was relatively small. The overall time-corrected means were 78.6 RAU for the experienced group and 73.4 RAU for the novice group. There was a significant effect of compression speed, $F(1,11)=28.11$, $p<0.001$. A significant difference was present between the ''fast'' and ''slow'' systems for all numbers of channels: 11 channels, $t=3.08$, $p<0.005$; 8 channels, $t=2.39$, $p<0.025$; and 6 channels, $t=2.22$, $p<0.025$ (all tests with 11 df, one-tailed). There was no significant interaction between compression speed and number of channels; $F(2,22)=0.18$, $p=0.84$.

The corrected scores transformed from RAUs back into percentages are plotted in Fig. 4, for the ''slow'' (solid line) and ''fast'' (dashed line) compression systems. ''Fast'' compression gave lower scores than ''slow'' compression for all channel numbers tested.

Using the same interpolation method as for experiment 2, the ''cost'' in either extra channels required or channels lost can be calculated from Fig. 4. The 6-channel ''slow'' system produced the same performance as a 6.4-channel ''fast'' system, while the 8-channel ''slow'' system produced the same performance as a 9.4-channel ''fast'' system. The 8-channel ''fast'' system produced the same performance as a 7.5-channel ''slow'' system, while the 11-channel ''fast'' system produced the same performance as a 9.2-channel ''slow'' system. These costs are similar to those found in experiment 2.

V. DISCUSSION

The data support our hypothesis that single-channel ''fast'' compression as the front end to a simulated cochlear implant processor would impair intelligibility. We described earlier how the deleterious effect of the ''fast'' compressor could be expressed as the change in number of channels required to achieve performance the same as for the ''slow'' system (or the baseline). The ''cost'' can be described in an alternative way using the data presented in Friesen *et al.* (2001). They measured the intelligibility of HINT sentence material, processed in a similar way to that reported here, as a function of number of channels and signal-to-noise ratio. The noise was unmodulated and was shaped to have the same spectrum as the sentence material. They fitted their data using sigmoidal functions. The parameters defining these functions are given in their Table X. For a given number of channels and a given score for our own data, we determined the signal-to-noise ratio that would lead to the same performance in the study of Friesen *et al.*, using their fitted parameter values. We then calculated the difference in effective speech-to-noise ratio between the ''fast'' and ''slow'' compressors. The results are shown in Table II. For each experiment, the percent correct scores are given for each channel number and compression condition. Adjacent to the columns of percent-correct scores are the speech-to-steady-noise ratios that would lead to the same percent correct. Friesen *et al.* did not simulate an 11-channel system. Parameters for this system were interpolated from those given for their 8- and 12-channel systems. The column labeled ''difference'' shows the differences between the speech-to-noise ratios for the ''fast'' and ''slow'' compressors. Typically, for moderate levels of intelligibility, the cost was a little less than 1 dB. For higher levels of intelligibility, the cost was closer to 2 dB.

As described in the Introduction, the cost of the ''fast'' compression might have occurred for two reasons. First, ''fast'' compression reduces the modulation depth and intensity contrasts in the signal and this might adversely affect

TABLE II. Results from all three experiments showing the percent correct and the equivalent speech-to-steady-noise ratio (SNR) in dB, using Table X of Friesen *et al.* (2001). The difference in equivalent speech-to-noise ratio between ''fast'' and ''slow'' compression is shown in bold.

No. of channels	Experiment 1		Experiment 2					Experiment 3				
			% correct		SNR			% correct		SNR		
	% correct	SNR	Fast	Slow	Fast	Slow	**Difference**	Fast	Slow	Fast	Slow	**Difference**
4	23.1	3.2	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
6	56.4	4.6	⋯	61.1	⋯	5.3	⋯	56.5	61.1	4.6	5.3	**0.7**
8	80.6	6.2	76.7	80.7	5.5	6.3	**0.8**	78.1	82.1	5.8	6.5	**0.7**
11	90.6	9.0	84.5	90.1	6.9	8.8	**1.9**	85.6	89.8	7.2	8.7	**1.5**
16	93.6	8.0	89.7	⋯	6.4	⋯	⋯	⋯	⋯	⋯	⋯	⋯

intelligibility, regardless of the nature of the background sound (Drullman *et al.*, 1994a, b; van Buuren *et al.*, 1999). This might be especially true for the type of signal processing used here, which removed temporal fine structure cues, so that information was conveyed only by the patterns of amplitude modulation in different frequency bands. Second, ''fast'' compression introduces comodulation between the target and background, which might disrupt the perceptual separation of the two, especially when the background itself is modulated, as in our experiments. We will refer to these two factors as ''modulation reduction'' and ''comodulation,'' respectively. To quantify the first factor, we calculated the fractional reduction in modulation, f_r, in the individual channels, using as input continuous speech in a background of an interfering speaker. This was done by sampling the output of each channel using 125-ms windows and comparing the distribution of levels with and without the fast compressor in operation prior to the division into channels. For the six-channel system, f_r was 0.5, 0.39, 0.34, 0.39, 0.18, and 0.05 for channels 1 to 6, respectively (in order of increasing center frequency). For the 11-channel system, f_r was 0.47, 0.42, 0.35, 0.28, 0.30, 0.35, 0.34, 0.34, 0.07, 0.11, and -0.03, for channels 1–11, respectively. For low and medium center frequencies, the fast compression markedly reduced the amount of modulation, although by less than for the broadband signal (equivalent to a single-channel system), for which f_r was 0.71. For the two or three highest center frequencies, the modulation reduction was small.

Our data do not allow us to decide the relative importance of modulation reduction and comodulation. However, some clues can be gained from comparisons with another study. McDermott *et al.* (2002) showed that a fast compressor used with a cochlear implant could improve intelligibility in situations where lack of audibility was the primary problem. Fast compressors are very effective at improving audibility. However, for speech in steady background noise, for which audibility was not a problem, McDermott *et al.* showed that the fast compressor produced a slight, but nonsignificant worsening in performance in comparison to no compression. The fact that they found a nonsignificant worsening in steady noise, while we found a significant worsening in a modulated background (speech), suggests that comodulation of the target and background contributes to the deleterious effect found by us. However, McDermott *et al.* used a lower compression ratio than us (2:1 vs 7:1), and their compression was effective over a narrower range of levels than ours, so the comparison between the two studies needs to be viewed with caution. Lorenzi *et al.* (1999), using a method of signal processing similar to the one employed in this paper, showed that fast-acting expansion of the temporal envelope in each channel of a four-channel system (achieved by raising the envelope magnitude to the power 2) led to a small improvement in the identification of speech in a steady noise background. Since magnification of amplitude modulation can improve intelligibility, one might argue that reduction of amplitude modulation will worsen it, even without significant comodulation.

Another way to determine the relative importance of modulation reduction and comodulation would be to assess the effect of varying the target-to-background ratio. For a background level similar to that of the target, the ''fast'' compressor would produce considerable comodulation. At the signal-to-background ratio of $+5$ dB used here, we would expect a moderate degree of comodulation. For a higher signal-to-background ratio, say $+10$ dB, the ''fast'' compression would introduce less comodulation. If comodulation is the main cause of the deleterious effect of the ''fast'' compressor, the effect should be reduced at high signal-to-background ratios. This prediction remains to be tested.

The ''slow'' compressor used in this study did not degrade performance relative to the baseline condition. The use of such a compressor would allow the user of a cochlear implant or hearing aid to deal with the widely varying overall sound levels that occur in different listening situations (Moore and Glasberg, 1988; Moore *et al.*, 1991; Stone *et al.*, 1999). A potential problem with slow compressors occurs when the signal level decreases abruptly, for example, when leaving a noisy room. In such situations, it takes a significant time for the compressor to adjust to the new signal level so as to restore audibility; the implant or aid may appear to go ''dead'' for a while. However, it is rare in everyday life for environmental sounds to decrease very rapidly in level.

Our results indicate that, for cochlear implants, a slow front-end compressor is likely to lead to better performance than a fast front-end compressor in situations where the signal-to-background ratio is not very high. Ten to 15 years ago, the performance of cochlear implantees was typically very poor when the signal-to-background ratio was low. However, recent improvements in the performance of cochlear implantees mean that they can cope reasonably well in such situations. Hence the front-end compressor should be designed to work well at relatively poor signal-to-background ratios. Most current implant systems make use of fast-acting front-end compressors, which may effectively make signals audible, but which may not be optimal for discrimination.

Finally, we turn to the issue of the training required to achieve stable results when using simulations of cochlear-implant processing. Our subjects usually received only limited experience with the simulation prior to testing. It turned out that this was insufficient to achieve stable results, but to maintain consistency between experiments, we continued to use the same training scheme and later applied ''corrections'' to compensate for training effects. Other researchers have used longer periods of training (Shannon *et al.*, 1995) or none at all (Fu *et al.*, 1998; Fu and Shannon, 1999) with similarly novel stimuli. It appears necessary to use between 30 and 60 min of training before practice effects cease to be important.

VI. CONCLUSIONS

In a simulation of a cochlear implant assessed using normal-hearing listeners, ''slow'' single-channel compression of target speech in a background of speech did not lead to a significant reduction in intelligibility relative to that obtained with no compression. A ''fast'' compressor, with near-syllabic time constants, consistently degraded intelligibility, by 4%–5%, over a wide range of performance levels. The

cost was equivalent to a loss of between 6% and 22% of the number of channels. Expressed in a different way, the cost was equivalent to raising the level of a spectrally matched continuous background noise by 1–2 dB. Our results suggest that slow compression will be more beneficial than fast compression as the front-end of a cochlear-implant processor.

ACKNOWLEDGMENTS

This work was supported by the Medical Research Council (UK). We thank Advanced Bionics Corporation for the loan of hardware and access to technical information on the design of their processors. We thank Dianne van Tasell, Ken Grant, and an anonymous reviewer for helpful comments on an earlier version of this paper.

ANSI (**1996**). ANSI S3.22-1996, ''Specification of Hearing Aid Characteristics'' (American National Standards Institute, New York).

ANSI (**1997**). ANSI S3.5-1997, ''Methods for the Calculation of the Speech Intelligibility Index'' (American National Standards Institute, New York).

Bracewell, R. N. (**1986**). *The Fourier Transform and its Applications* (McGraw–Hill, New York).

Braida, L. D., Durlach, N. I., De Gennaro, S. V., Peterson, P. M., and Bustamante, D. K. (**1982**). ''Review of recent research on multiband amplitude compression for the hearing impaired,'' in *The Vanderbilt Hearing-Aid Report*, edited by G. A. Studebaker and F. H. Bess (Monographs in Contemporary Audiology, Upper Darby, PA).

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (Bradford Books, MIT Press, Cambridge, MA).

Clark, G. M., Tong, Y. C., and Patrick, J. F. (**1990**). *Cochlear Prostheses* (Churchill Livingston, Edinburgh).

Dillon, H. (**1996**). ''Compression? Yes, but for low or high frequencies, for low or high intensities, and with what response times?'' Ear Hear. **17**, 287–307.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). ''Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs,'' J. Acoust. Soc. Am. **102**, 2403–2411.

Dorman, M. F., Loizou, P. C., Fitzke, J., and Tu, Z. (**1998**). ''The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels,'' J. Acoust. Soc. Am. **104**, 3583–3585.

Drullman, R., Festen, J. M., and Plomp, R. (**1994a**). ''Effect of reducing slow temporal modulations on speech reception,'' J. Acoust. Soc. Am. **95**, 2670–2680.

Drullman, R., Festen, J. M., and Plomp, R. (**1994b**). ''Effect of temporal envelope smearing on speech reception,'' J. Acoust. Soc. Am. **95**, 1053–1064.

Dudley, H. (**1939**). ''Remaking speech,'' J. Acoust. Soc. Am. **11**, 169–177.

Fletcher, H. (**1953**). *Speech and Hearing in Communication* (Van Nostrand, New York).

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). ''Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants,'' J. Acoust. Soc. Am. **110**, 1150–1163.

Fu, Q.-J., and Shannon, R. V. (**1999**). ''Effect of acoustic dynamic range on phoneme recognition in quiet and noise by cochlear implant users,'' J. Acoust. Soc. Am. **106**, L65–L70.

Fu, Q.-J., Shannon, R. V., and Wang, X. (**1998**). ''Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing,'' J. Acoust. Soc. Am. **104**, 3586–3596.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.

Killion, M. C. (**1997**). ''Hearing aids: Past, present and future: Moving toward normal conversations in noise,'' Br. J. Audiol. **31**, 141–148.

Levitt, H. (**1982**). ''Speech discrimination ability in the hearing impaired: spectrum considerations,'' in *The Vanderbilt Hearing-Aid Report*, edited by G. A. Studebaker and F. H. Bess (Monographs in Contemporary Audiology, Upper Darby, PA).

Lippmann, R. P., Braida, L. D., and Durlach, N. I. (**1981**). ''Study of multichannel amplitude compression and linear amplification for persons with sensorineural hearing loss,'' J. Acoust. Soc. Am. **69**, 524–534.

Loizou, P. C., Dorman, M., and Tu, Z. (**1999**). ''On the number of channels needed to understand speech,'' J. Acoust. Soc. Am. **106**, 2097–2103.

Loizou, P. C., Dorman, M., Poroy, O., and Spahr, T. (**2000**). ''Speech recognition by normal-hearing and cochlear implant listeners as a function of intensity resolution,'' J. Acoust. Soc. Am. **108**, 2377–2387.

Lorenzi, C., Berthommier, F., Apoux, F., and Bacri, N. (**1999**). ''Effects of envelope expansion on speech recognition,'' Hear. Res. **136**, 131–138.

MacLeod, A., and Summerfield, Q. (**1990**). ''A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: rationale, evaluation, and recommendations for use,'' Br. J. Audiol. **24**, 29–43.

McDermott, H. J., Henshall, K. R., and McKay, C. M. (**2002**). ''Benefits of syllabic input compression for users of cochlear implants,'' J. Am. Acad. Audiol **13**, 14–24.

Moore, B. C. J. (**1998**). *Cochlear Hearing Loss* (Whurr, London).

Moore, B. C. J. (**2003**). ''Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants,'' Otol. Neurotol. **24**, 243–254.

Moore, B. C. J., and Glasberg, B. R. (**1988**). ''A comparison of four methods of implementing automatic gain control (AGC) in hearing aids,'' Br. J. Audiol. **22**, 93–104.

Moore, B. C. J., Glasberg, B. R., and Stone, M. A. (**1991**). ''Optimization of a slow-acting automatic gain control system for use in hearing aids,'' Br. J. Audiol. **25**, 171–182.

Moore, B. C. J., Stone, M. A., and Alcántara, J. I. (**2001**). ''Comparison of the electroacoustic characteristics of five hearing aids,'' Br. J. Audiol. **35**, 307–325.

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). ''Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise,'' J. Acoust. Soc. Am. **95**, 1085–1099.

Noordhoek, I. M., and Drullman, R. (**1997**). ''Effect of reducing temporal intensity modulations on sentence intelligibility,'' J. Acoust. Soc. Am. **101**, 498–502.

Pavlovic, C. (**1987**). ''Derivation of primary parameters and procedures for use in speech intelligibility predictions,'' J. Acoust. Soc. Am. **82**, 413–422.

Pearsons, K. S., Bennett, R. L., and Fidell, S. (**1976**). *Speech Levels in Various Environments. Report No. 3281* (Bolt, Beranek and Newman, Cambridge, MA).

Plomp, R. (**1988**). ''The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function,'' J. Acoust. Soc. Am. **83**, 2322–2327.

Qin, M. K., and Oxenham, A. J. (**2003**). ''Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers,'' J. Acoust. Soc. Am. (in press).

Seligman, P., and Whitford, L. (**1995**). ''Adjustment of appropriate signal levels in the Spectra 22 and mini speech processors,'' Ann. Otol. Rhinol. Laryngol. Suppl. **166**, 172–175.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.

Stone, M. A., and Moore, B. C. J. (**1992**). ''Syllabic compression: Effective compression ratios for signals modulated at different rates,'' Br. J. Audiol. **26**, 351–361.

Stone, M. A., Moore, B. C. J., Alcántara, J. I., and Glasberg, B. R. (**1999**). ''Comparison of different forms of compression using wearable digital hearing aids,'' J. Acoust. Soc. Am. **106**, 3603–3619.

Studebaker, G. (**1985**). ''A 'rationalized' arcsine transform,'' J. Speech Hear. Res. **28**, 455–462.

Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gwaltney, C. A. (**1999**). ''Monosyllabic word recognition at higher-than-normal speech and noise levels,'' J. Acoust. Soc. Am. **105**, 2431–2444.

van Buuren, R. A., Festen, J., and Houtgast, T. (**1999**). ''Compression and expansion of the temporal envelope: Evaluation of speech intelligibility and sound quality,'' J. Acoust. Soc. Am. **105**, 2903–2913.

Van Tasell, D. J., and Trine, T. D. (**1996**). ''Effects of single-band syllabic amplitude compression on temporal speech information in nonsense syllables and in sentences,'' J. Speech Hear. Res. **39**, 912–922.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., and Widin, G. P. (**1987**). ''Speech

waveform envelope cues for consonant recognition,'' J. Acoust. Soc. Am. **82**, 1152–1161.
Villchur, E. (**1973**). ''Signal processing to improve speech intelligibility in perceptive deafness,'' J. Acoust. Soc. Am. **53**, 1646–1657.
Yund, E. W., and Buckles, K. M. (**1995a**). ''Enhanced speech perception at low signal-to-noise ratios with multichannel compression hearing aids,'' J. Acoust. Soc. Am. **97**, 1224–1240.
Yund, E. W., and Buckles, K. M. (**1995b**). ''Multichannel compression hearing aids: Effect of number of channels on speech discrimination in noise,'' J. Acoust. Soc. Am. **97**, 1206–1223.

Measuring the neck frequency response function of laryngectomy patients: Implications for the design of electrolarynx devices

Geoffrey S. Meltzner[a)]
Voice and Speech Laboratory, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114

James B. Kobler
Voice and Speech Laboratory, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114 and Communication Sciences and Disorders, MGH Institute of Health Professions, Boston, Massachusetts 02129

Robert E. Hillman
Voice and Speech Laboratory, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114, Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts 02114, and Communication Sciences and Disorders, MGH Institute of Health Professions, Boston, Massachusetts 02129

(Received 9 May 2002; accepted for publication 17 April 2003)

Measurements of the neck frequency response function (NFRF), defined as the ratio of the spectrum of the estimated volume velocity that excites the vocal tract to the spectrum of the acceleration delivered to the neck wall, were made at three different positions on the necks of nine laryngectomized subjects (five males and four females) and four normal laryngeal speakers (two males and two females). A minishaker driven by broadband noise provided excitation to the necks of subjects as they configured their vocal tracts to mimic the production of the vowels /a/, /æ/, and /I/. The sound pressure at the lips was measured with a microphone and an impedance head mounted on the shaker measured the acceleration. The neck wall passed low-frequency sound energy better than high-frequency sound energy, and thus the NFRF was accurately modeled as a low-pass filter. The NFRFs of the different subject groups (female laryngeal, male laryngeal speakers, laryngectomized males, and laryngectomized females) differed from each other in terms of corner frequency and gain, with both types of male subjects presenting NFRFs with larger overall gains. In addition, there was a notable amount of intersubject variability within groups. Because the NFRF is an estimate of how sound energy passes through the neck wall, these results should aid in the design of improved neck-type electrolarynx devices.
[DOI: 10.1121/1.1582440]

PACS numbers: 43.70.Aj, 43.70.Dn, 43.70.Jt [DOS]

I. INTRODUCTION

Each year, thousands of people lose the ability to speak normally because they have the larynx removed (laryngectomy) or suffer laryngeal trauma. Many of these individuals must rely on a hand-held electrolarynx (EL) to produce speech (Hillman *et al.*, 1998). The most common electrolarynx devices are electro-mechanical vibrators that are typically held against the neck to excite the vocal tract acoustically. EL speech generally provides a serviceable means of communication, but has several serious shortcomings, including an artificial quality, reduced intelligibility, and poor audibility, especially in noisy environments. Because a better understanding of the sound transmission characteristics of the neck wall appears to be essential in designing a more efficient and natural-sounding electrolarynx, this study sought to characterize the impact of this factor on the acoustics of EL speech.

Despite being developed over 40 years ago (Barney *et al.*, 1959), there have been few scientific efforts to improve electrolaryngeal devices or the resulting speech. Qi and Weinberg (1991) found that enhancing the low-frequency content of EL speech resulted in speech that was judged in formal listening experiments to be preferred more often than the speech without enhancement. In an effort to improve the intelligibility of EL speech, Espy-Wilson *et al.* (1998) used an adaptive filtering algorithm to remove the contribution of the directly radiated sound (produced by the EL) to the perceived speech signal. This processing resulted in increased intelligibility for stop consonants but degraded intelligibility of nasal consonants.

Norton and Bernstein (1993) tried to improve EL speech by changing the driving signal of the vibration source. They measured the frequency response function (FRF) of the neck, which they defined as the ratio of the spectrum of the pressure signal measured at the lips (after the formants had been removed) to the spectrum of the input signal of a shaker which vibrated against the neck. They used the neck FRF to modify the output of an electrolarynx such that the spectrum of the sound that excited the vocal tract resembled that of a natural human glottal source. Listeners informally judged the speech produced using their modified electrolarynx as sounding more natural.

While the results of the Norton and Bernstein study were

a)Electronic mail: geoff@mit.edu

promising, there remain unresolved issues, including the fact that Norton and Bernstein measured the neck frequency response for a single normal subject, whose anatomy significantly differed from that of laryngectomized patients. The necks of laryngectomy patients lack the larynx and its surrounding tissue, and additionally, laryngectomy patients are often subjected to radiation treatment, which can change the properties of the neck tissue.

Norton and Bernstein's work also does not address the variations in the FRF that might exist between individuals. For example, the type and extent of laryngectomy is primarily dependent on tumor size and the spread of the disease to other neck structures. Differences in surgical techniques could potentially affect the transmission properties of the neck wall.

Finally, Norton and Bernstein only measured a single place on the neck. Because of the potentially asymmetric nature of a laryngectomy surgery, especially if the surgery is accompanied by a radical neck dissection, the characteristics of the tissue at different locations on the neck may differ. Furthermore, EL users often report that there is one location on their necks that produces an optimal speech output level. These reports of an optimal sound-source location suggest that the neck frequency response function may vary with neck position in laryngectomy patients.

Thus, the goal of this research was to measure the neck frequency response functions of both laryngectomized and nonlaryngectomized subjects at different locations on the neck. This information should assist the design of an improved electrolarynx driving signal that would excite the vocal tract in a manner that is more similar to a natural glottal sound source. Furthermore, if a significant degree of variation between the neck frequency response functions of different subjects was discovered, then there would be a rationale for tailoring the output of an electrolarynx to its user.

II. METHODS

The following experimental procedures were used to collect the necessary data to determine the neck frequency response function (NFRF). The NFRF is defined as the ratio of the spectrum of the estimated volume velocity that excites the vocal tract to the spectrum of the acceleration measured at the neck. The procedures are based on those used by Norton and Bernstein (1993) and Fujimura and Lindqvist (1971), with some additional modifications to improve the accuracy of the measurements.

A. Subjects

The laryngectomy subject group consisted of ten laryngectomized subjects (five males and five females) all of whom had been laryngectomized at least 1 year prior to the date of the experiment and had received postsurgical radiation treatment. The subjects ranged between 50 and 76 years in age (mean of 65.6 ± 7.8 years). All of the laryngectomized subjects were patients at the Voice and Speech Laboratory at the Massachusetts Eye and Ear Infirmary. All but two laryngectomized subjects were part-time or full-time electrolarynx users. In addition to the laryngectomized subjects, four normal subjects (i.e., nonlaryngectomized, or laryngeal speakers), two male and two female, were used. The ages of the normal laryngeal speakers ranged from 22 to 31 (mean 25.8 ± 3.9 years). Although none of the normal subjects was an everyday electrolarynx user, all of them were familiar with how to use the device.

B. Experimental procedures

Two separate experiments were performed in a sound-treated experimental chamber. The main experiment was performed on all of the laryngectomized and normal subjects. A second experiment was performed on a single normal male subject to estimate the near-field lip radiation characteristic.

1. Collection of neck frequency response function data

Each subject was seated in a clinical exam chair and a head-mounted directional microphone (Sennheiser model K3-U) was positioned 1 cm from the subject's lips and oriented so that it faced the lips. The microphone was calibrated against a Brüel & Kjaer $\frac{1}{4}$ in. microphone which had a known, flat response between 60 and 4000 Hz by simultaneously recording the output of a broadband noise source.

A small Brüel & Kjaer minishaker (model 3081) was attached by a 2 in. shaft to an impedance head (a transducer that simultaneously measures force and acceleration—PCB model 288D01/788D01). The other side of the impedance head was attached to a 2.5-cm-diameter metal disk that was placed on the subject's neck. The size of the disk was the same as that found on a Servox electrolarynx (a popular, commercially available EL produced by Siemens Corp.). The shaker was enclosed in a large plastic cylinder that was packed with sound-attenuating foam to reduce the directly radiated sound (into the environment) produced by the shaker. The silencer reduced the shaker self-noise, especially for frequencies between 600 and 4000 Hz, but there were occasions in which the levels of radiated sound at other frequencies confounded the measurements.

The shaker was driven with broadband noise (1 to 12 500 Hz), and in separate trials, each subject was asked to configure the vocal tract as if producing the vowels /a/ (as in bot), /æ/ (as in bat), and /I/ (as in bit). In two additional trials, the subjects were asked to: (1) change the vocal-tract configuration to shift between the production of /a/, /æ/, and /I/ in continuous succession and (2) keep the mouth closed while the shaker was being driven. All of the nonlaryngectomized subjects were instructed to maintain the glottis in a closed position (i.e., to hold their breath) to simulate the laryngectomized condition in which the vocal tract is closed at the level of the hypopharynx and thus is decoupled from the subglottal system.

These steps were repeated three times at three locations on the neck. The first position was always the location where each laryngectomized subject placed the electrolarynx during everyday use. For those subjects who did not use an electrolarynx, position 1 was arbitrarily chosen to be a spot on the upper right neck, with position 2 placed 2 cm below position 1, and position 3 at a spot identical to position 1 on the opposite side of the neck. For some laryngectomized sub-

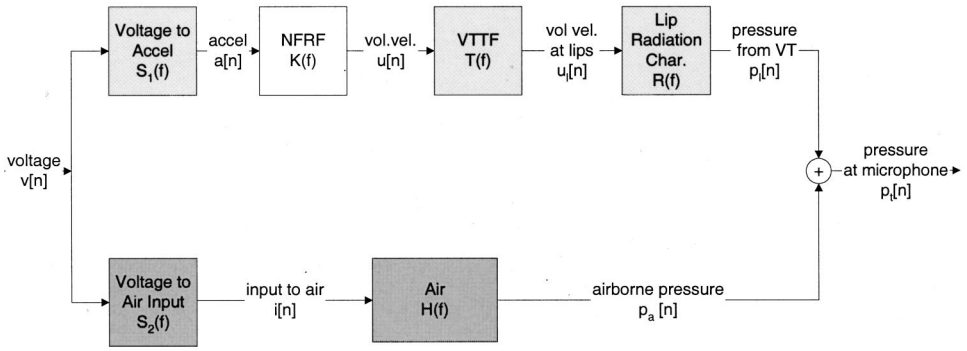


FIG. 1. Block diagram of experimental system. The microphone at the lips measures the sound that is the sum of the sound transmitted through the neck wall and vocal tract, and the sound transmitted via a direct path through the air. Signals were digitized, and therefore the discrete signal notation $x[n]$ is employed where $x[n]$ is the discrete counterpart to $x(t)$, the continuous, measured signal.

jects, measurements at all three positions could not be made because at some locations the signal transmission was too attenuated.

A data acquisition system (Axon Instruments Digidata acquisition board with accompanying AXOSCOPE software) simultaneously recorded the pressure signal from the microphone, the force and acceleration signals produced by the impedance head, and the voltage of the driving signal to the shaker. All of the signals were sampled at 20 kHz after being low-pass filtered at 6 kHz by 4-pole Bessel filters (Axon Instruments Cyberamp).

2. Collection of lip radiation data

For this experiment, an additional microphone was suspended from the ceiling at a distance of 45 cm from the lips of one subject. The subject was asked to sustain the vowels /a/, /æ/, and /I/, for a duration of 5 s.

C. Data analysis

First, the waveforms were downsampled from 20 to 14 kHz to decrease the analysis bandwidth and reduce the order of LPC filters needed to estimate the vocal-tract transfer function. Although it would have been preferable to downsample the data further, a sampling rate of 14 kHz was chosen as a compromise between the need to keep higher formants in the signal (to achieve a better vocal-tract transfer function estimate) and the need to reduce the likelihood of fitting poles to nonvocal tract resonances.

The neck frequency response function is defined as the ratio of the spectrum of the signal that excites the vocal tract (i.e., volume velocity) to the spectrum of the input acceleration. The acceleration was measured directly, but some signal processing was necessary to obtain an estimate of the volume velocity from the pressure signal measured at the mouth. A schematic of the analysis algorithm is shown in Fig. 1.

In this diagram, the system consists of two additive paths containing several linear, time-invariant (LTI) systems. In the primary path (upper path in the figure), the driving voltage signal (broadband noise with a 12.5-kHz bandwidth) is transformed into an acceleration signal by the shaker; this transformation is characterized by the primary shaker transfer function, $S_1(f)$. The acceleration is then filtered by the neck wall, $K(f)$, vocal tract, $T(f)$, and the near-field lip radiation characteristic, $R(f)$.

The output of the primary path is represented by $p_l[n]$, which is the pressure that would be measured by the microphone if this path could be isolated from the rest of the system. In the lower path, the voltage signal is transformed into a signal that excites the air by a secondary shaker transfer function, $S_2(f)$. The input to the air is then filtered by an unknown air transfer function, $H(f)$, and the resulting pressure is $p_a[n]$. $H(f)$ also includes any filter characteristics due to the directionality of the microphone. While it is not possible to measure $p_l[n]$ without contamination by $p_a[n]$, it is possible to estimate $p_a[n]$ under conditions where $p_l[n]$ is greatly attenuated by simply having the subject keep the mouth closed. Thus, the four signals that were measured are $v[n]$ (the voltage driving the shaker), $a[n]$ (the acceleration signal from the accelerometer), $p_t[n]$ the pressure signal from the microphone when the lips are open, and $p_a[n]$, the pressure signal from the microphone when the lips are closed, where

$$p_t[n]=p_l[n]+p_a[n], \tag{1}$$

and the transfer function to be estimated is $K(f)$ where

$$K(f)=\frac{\Phi_{au}(f)}{\Phi_{aa}(f)}, \tag{2}$$

where $\Phi_{au}(f)$ and $\Phi_{aa}(f)$ are the cross-spectral density of $u[n]$ and $a[n]$ and the power spectral density of $a[n]$, respectively. The cross-spectral density, $\Phi_{xy}(f)$, of two signals, $x[n]$ and $y[n]$, is the Fourier transform of their cross correlation, $\phi_{xy}[n]$, defined as

$$\phi_{xy}[n]=\sum_{m=-\infty}^{\infty} x[m+n]y[m]. \tag{3}$$

The autocorrelation of a signal is merely the cross correlation of the signal with itself. The power-spectral density is the Fourier transform of the auto correlation.

Since $v[n]$, $a[n]$, $p_t[n]$, and $p_a[n]$ are the signals that were measured, $K(f)$ cannot be computed directly. Therefore, an alternative method for computing $K(f)$ needed to be developed. The resulting algorithm is discussed below.

1. Step one—removing the airborne path

Given the relationship described in Eq. (1), the following equation is also valid:

$$\phi_{vp_t}[n]=\phi_{vp_l}[n]+\phi_{vp_a}[n], \tag{4}$$

where $\phi_{vp_t}[n]$, $\phi_{vp_l}[n]$, and $\phi_{vp_a}[n]$ are the cross correlations of $v[n]$ with $p_t[n]$, $p_l[n]$, and $p_a[n]$, respectively.

Furthermore, because each system block is assumed to be an LTI system

$$\phi_{vp_l}[n] = \phi_{vv}[n] * (s_1[n] * k[n] * t[n] * r[n]), \tag{5a}$$

and

$$\phi_{vp_a}[n] = \phi_{vv}[n] * (s_2[n] * h[n]), \tag{5b}$$

where $\phi_{vv}[n]$ is the autocorrelation of $v[n]$, $k[n]$ is the impulse response of the neck frequency response function, $h[n]$ is the acoustic impulse response of the environment, and $r[n]$ is the impulse response of the near-field lip radiation characteristic. Combining Eqs. (3) and (5a) and (5b) results in

$$\phi_{vp_t}[n] = \phi_{vv}[n] * (s_1[n] * t[n] * k[n] * r[n] + s_2[n] * h[n]). \tag{6}$$

Performing a Fourier transform on Eq. (6), dividing by $\Phi_{vv}(f)$ (the power spectral density of $v[n]$), and rearranging terms leads to

$$\frac{\Phi_{vp_t}(f)}{\Phi_{vv}(f)} - S_2(f) \cdot H(f) = S_1(f) \cdot T(f) \cdot K(f) \cdot R(f), \tag{7}$$

where $\Phi_{vp_t}(f)$ is the cross-spectral density of $p_t[n]$ and $v[n]$. But, transforming Eq. (5b) into the frequency domain produces

$$\frac{\Phi_{vp_a}(f)}{\Phi_{vv}(f)} = S_2(f) \cdot H(f), \tag{8}$$

and therefore

$$\left(\frac{\Phi_{vp_t}(f)}{\Phi_{vv}(f)} - \frac{\Phi_{vp_a}(f)}{\Phi_{vv}(f)} \right) \cdot \frac{1}{S_1(f) \cdot T(f) \cdot R(f)} = K(f). \tag{9}$$

Thus, if $\Phi_{vp_a}(f)$, $S_1(f)$, $T(f)$, and $R(f)$ can be computed, then Eq. (9) can be used to determine the neck frequency response function, $K(f)$.

Therefore, $\Phi_{vp_a}(f)$ and $\Phi_{vv}(f)$ are computed from $p_a[n]$ (estimated from the closed-mouth measurements) and $v[n]$. The effectiveness of this method for removing the airborne path is explored in the Discussion section. The remaining steps in the analysis algorithm involve computing $S_1(f)$, $T(f)$, and $R(f)$ in order to isolate $K(f)$ from the rest of the system. It should be noted that this step assumes that both the primary shaker transfer function, $S_1(f)$, and the unknown air transfer function, $H(f)$, are independent of whether the mouth is open or closed.

2. Step two—Calculating the primary shaker transfer function

According to the block diagram in Fig. 1, the input to the primary shaker transfer function, $S_1(f)$, is $v[n]$ and its output is $a[n]$; therefore, $S_1(f)$ can be computed as

$$S_1(f) = \frac{\Phi_{va}(f)}{\Phi_{vv}(f)}, \tag{10}$$

where $\Phi_{av}(f)$ is the cross-spectral density of $a[n]$ and $v[n]$. Because both $a[n]$ and $v[n]$ are measured during the experiments, $S_1(f)$ can be determined from measured signals. The primary shaker transfer function was found to be relatively flat, rolling off at -2 dB/octave within the frequency range of interest (100 to 4000 Hz). The unwrapped phase of this transfer function decreased at -0.5 rad/octave over the same frequency range.

3. Step three—Estimating the near-field lip radiation characteristic

As shown in the block diagram of Fig. 1, the output of the vocal tract (at the lips) is a volume velocity, $u_l[n]$. However, the microphone located 1 cm from the lips produces a voltage that is proportional to pressure. Therefore, to accurately estimate the neck transfer function, it is necessary to determine the relationship between the volume velocity at the lips, $u_l[n]$, and the pressure at the microphone, $p_l[n]$. This relationship is called the *near-field lip radiation characteristic*, $R(f)$. Similarly, in the second experiment, the *far-field lip radiation characteristic*, $R_f(f)$ is the relationship between $u_l[n]$, and the pressure measured by the microphone in the far field. Specifically

$$P_n(f) = U_l(f) \cdot R(f), \tag{11a}$$

$$P_f(f) = U_l(f) \cdot R_f(f), \tag{11b}$$

where $P_n(f)$ and $P_f(f)$ are the spectra of the pressures measured at 1 cm and 45 cm away from the lips, respectively, and $R_f(f)$ is the far-field lip radiation characteristic. Because $P_n(f)$ and $P_f(f)$ are measured simultaneously in the second experiment, we can combine Eqs. (11a) and (11b) to compute $R(f)$ as

$$R(f) = \frac{P_n(f) \cdot R_f(f)}{P_f(f)}. \tag{12}$$

For distances, r, which are greater than a few centimeters (in this case, $r = 45$ cm), the open mouth can be considered a simple source radiating in all directions. In such cases, the radiation characteristic can be approximated as

$$R_f(f) = \frac{j2\pi f\rho}{4\pi r} \cdot e^{-j(2\pi fr/c)}, \tag{13}$$

where c is the speed of sound, and ρ is the density of air. Equation (13) is an accurate approximation (within a few decibels) for frequencies up to 4000 Hz. (Stevens, 1998). Thus, combining Eqs. (12) and (13), $R(f)$ can be determined as

$$R(f) = \frac{P_n(f)}{P_f(f)} \cdot \frac{j2\pi f\rho}{4\pi r} \cdot e^{-j(2\pi fr/c)}. \tag{14}$$

The radiation characteristics were computed using $c = 331.6$ m/s and $\rho = 1.293$ kg/m^3 (Kinsler *et al.*, 1982). It was found that $R(f)$ increased in magnitude at a rate of 6 dB/octave between 100 and 2000 Hz, and then flattened out between 2000 and 4000 Hz.

4. Step four—Estimating the vocal-tract transfer function

Based on measurements of other body tissues, one would expect $K(f)$ to be slowly varying with frequency, with few, if any, sharp resonances (Wodicka *et al.*, 1993). The vocal-tract transfer function, $T(f)$, on the other hand, can be effectively modeled as an all-pole system (at least for the production of non-nasalized vowels), with the location of the resonances and bandwidths of the resonances being dependent on the vowel being produced (Stevens, 1998). If the center frequencies and bandwidths of the poles of $T(f)$ can be estimated, it should be possible to separate $T(f)$ from $K(f)$. The validity of this assumption is investigated in the Discussion section.

The vocal-tract transfer function, $T(z)$, was estimated using linear prediction, which assumes that this function can be estimated using an all-pole filter of order p (Rabiner and Schafer, 1978). When using LP analysis, the order of the filter must be specified. Typically, an LP filter of order 12 can accurately represent a speech segment sampled at 10 kHz (Atal and Hanauer, 1971). However, in this study, the signals were sampled at 14 kHz and did not represent typical speech segments because they contained both the neck frequency response function and the vocal-tract transfer function. Therefore, an order of 20 was chosen so that the LP coefficients could properly account for at least six formants, any poles in the neck frequency response function, as well as any zeros present in either the vocal-tract transfer function or the neck frequency response function.

This LP estimate was the first step in the vocal-tract frequency response function estimation process. Because the conditions of the measured system were not ideal, the LP estimate needed to be modified to improve the accuracy of the vocal-tract transfer function estimate.

First, all poles that were not associated with formants were removed from the LP estimate. Some of the removed poles had similar frequencies and bandwidths as those used to model the NFRF (see Sec. III D), but others were most likely the result of using a high filter order in the initial LP estimate. In the cases where the shaker was placed at locations far from the terminal end of the vocal tract, some of the extra poles were present to compensate for the zeros present in the vocal-tract transfer function. Since it was known that nonformant poles would be removed from the estimate, using an LP order that produced extra poles was not considered problematic. The average values of formant frequencies and bandwidths for different vowels (Peterson and Barney, 1952) were used as guides for deciding between formant and nonformant poles. It was also taken into consideration that a laryngectomy patient's vocal tract has been truncated and thus the formant frequencies tend to be shifted higher in frequency, usually by 50 to 100 Hz (Sisty and Weinberg, 1972). Moreover, the formant bandwidths would be expected to be narrower since the vocal tract is decoupled from the subglottal system (Fujimura and Lindqvist, 1971; House and Stevens, 1958). Therefore, some correction was made in the criteria used to separate formant poles from nonformant poles. As a further check, the continuous vowel transition task allowed the formant transitions from vowel to vowel to be tracked using spectrograms. This formant tracking aided in identifying the formant frequencies and bandwidths for each vowel.

Second, while linear prediction assumes an all-pole model, the vocal-tract transfer function in this experiment is not strictly all-pole. In almost every case, the excitation source was placed in a location other than the terminal end of the vocal tract (i.e., the glottis in normal subjects), thus producing a back cavity. Zeros in the vocal-tract transfer function occur at frequencies where the back cavity acts as a short circuit, and if the back cavity is modeled as a rigidly terminated uniform tube, these zeros occur at its quarter-wave frequencies. It was observed that placing the shaker higher on the neck produced lower-frequency zeros, consistent with this model. The LP estimate often produces formant frequencies and bandwidths that are somewhat inaccurate in order to compensate for the presence of these zeros. Therefore, the estimated formant frequencies and bandwidths were manually altered so that the resulting modified LP spectrum more accurately approximated the measured data. In many cases, adding a conjugate zero pair to the spectral estimate improved its accuracy. The frequencies and bandwidths of these zeros were chosen by visually inspecting the measured spectra. The frequencies of the zeros ranged from 900 to 4200 Hz with bandwidths that ranged from 50 to 100 Hz. Additionally, the narrow bandwidths of the zeros were not surprising given that the zeros are not affected by losses from the lip radiation. It should be noted that even though the presence of zeros no longer meets the strict definition of an LP-based estimate of the vocal-tract transfer function, these estimates will still be referred to as such throughout the rest of this paper.

Once the LP spectrum was corrected, a new inverse filter, $A(z) = 1/[T(z)]$, was derived. The impulse response of the neck frequency response function, $k[n]$, is determined by filtering the impulse response associated with $T(f) \cdot K(f)$ by $A(z)$, and from $k[n]$, $K(f)$ can easily be computed.

Theoretically, the NFRF should be independent of the vowel used in the experimental task, and thus comparisons between vowels were used to verify the validity of the experiment and analysis. Section III B discusses the intervowel variability of the NFRF.

The measured NFRFs from each trial (nine trials per position) were averaged at each position for each subject. In addition, average NFRFs were computed for each of the four subject groups: normal males, normal females, laryngectomized males, and laryngectomized females.

D. Coherence and the LTI assumption

The analysis algorithm assumes that each block in Fig. 1 represents an LTI system. Subjects were asked to maintain a static vocal-tract configuration for 5 s, and from this sample, a 1- to 2-s section of constant amplitude was chosen for analysis. To confirm the time invariance of the entire system during this time period, the coherence, $C_{ap}(f)$, was computed using the measured voltage and pressure signals. The coherence is defined as

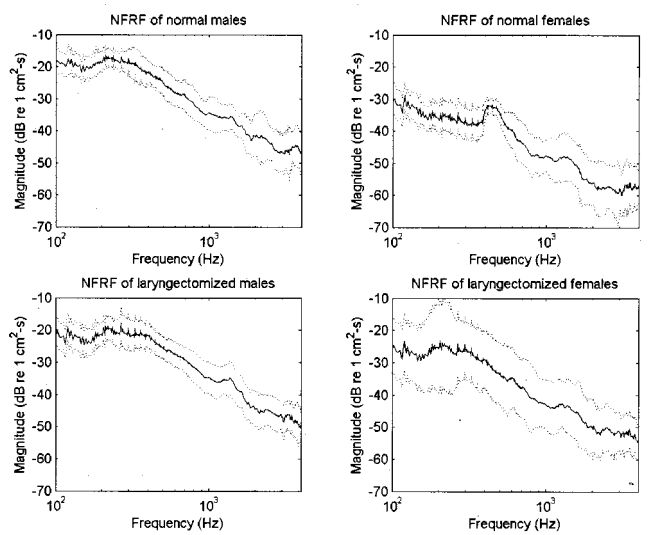


FIG. 2. The mean (solid line) and standard deviation (dotted line) of the log magnitude of the NFRFs for each of the four subject groups.

$$C_{ap_t}(f) = \frac{|\Phi_{ap_t}(f)|^2}{\Phi_{aa}(f) \cdot \Phi_{p_t p_t}(f)}, \tag{15}$$

where $\Phi_{ap_t}(f)$ is the cross-spectral density of $a[n]$ and $p_t[n]$, and $\Phi_{aa}(f)$ and $\Phi_{p_t p_t}(f)$ are the power spectral densities of $a[n]$ and $p_t[n]$, respectively.

Only data for which the coherence was above 0.8 for $100 < f < 4000$ Hz were analyzed. For frequencies greater than 4000 Hz, the coherence always significantly deviated from 1.

E. Minimum signal quality criteria

In addition to meeting the coherence criterion, the spectrum of the measured open-mouth pressure signal had to display at least two prominent formant peaks to be accepted for further analysis. If the pressure signal spectrum does not have enough formant energy then the LPC algorithm would be unable to estimate the vocal-tract transfer function. Those sets of signals that did not meet these minimum criteria were not analyzed, and as a result, one female laryngectomized subject's entire data set was discarded. Not surprisingly, this particular subject was unable to use a neck-type electrolarynx because it was unable to transmit enough sound through her neck to produce audible, intelligible speech.

III. RESULTS

A. General description

The results are based on data from nine laryngectomy patients (five male, four female) and four normal subjects (two male, two female). Figure 2 plots the mean NFRF and corresponding standard deviation for each of the four subject groups. In all four subject groups, the NFRF resembled a low-pass filter, with a constant maximum gain between 100 and 320 ± 80 Hz, rolling off at a slope of -8.8 ± 2.0 dB/octave between 320 and 3000 Hz, and then flattening out between 3000 and 4000 Hz. There was some variability in the NFRFs of the individuals, as reflected by the average standard deviation across frequency of between 1.9 and 5.2 dB for each group. In addition, the NFRF of some individuals did not flatten out at 3000 Hz but instead continued to roll off at the same rate.

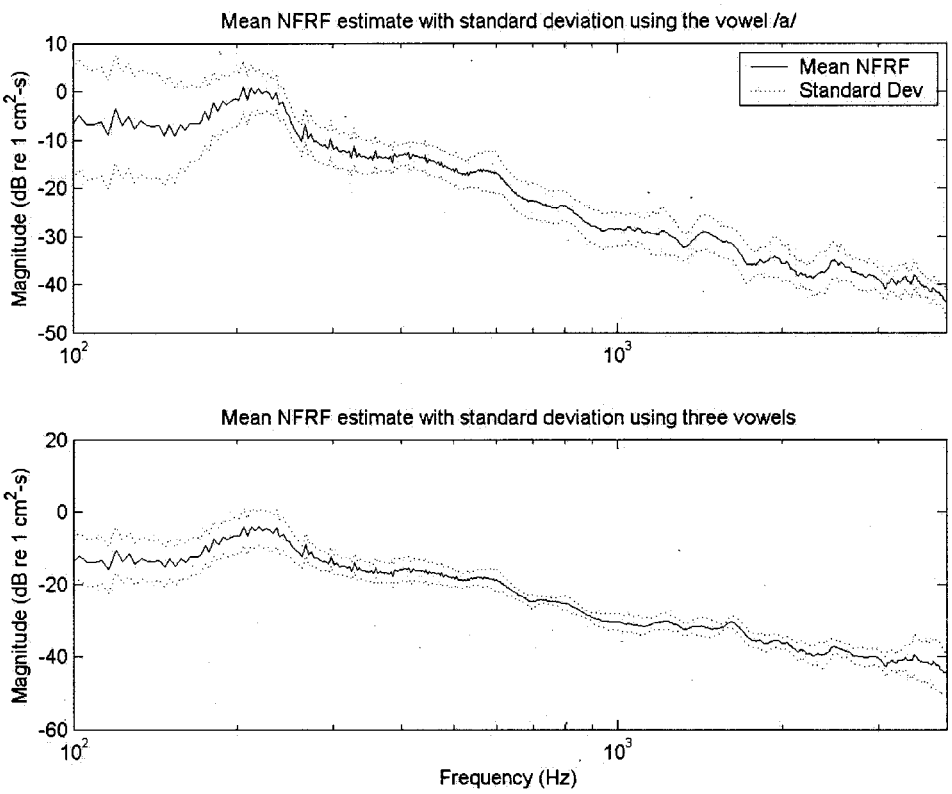


FIG. 3. *Top*: The mean of the NFRF estimation for a female laryngectomy patient producing the vowel /a/. The solid line represents the average transfer function and the dotted lines represent the standard deviation at each frequency. These values were computed by averaging over the three trials using the vowel /a/ at one neck position. The standard deviation ranged from 2.02 dB at 3725 Hz to 11.4 dB at 102 Hz with a mean of 3.30 dB. *Bottom*: The average (solid line) and standard deviation (dashed lines) of the NFRF at a single position on the neck of the same female laryngectomy patient. The average NFRF was computed by averaging over all of the trials for a single position. This particular case produced a mean standard deviation of 2.95 dB with a maximum of 6.94 dB at 3860 Hz and a minimum of 0.88 dB at 1576 Hz.

B. Reproducibility of results

The top plot in Fig. 3 shows a representative example of the ''intravowel variation'' (i.e., the variation across three repetitions of the same vowel for a male laryngectomy subject). In general, the intravowel repeatability was quite good; the average (over frequency) standard deviation across subjects was 4.34 ± 0.32 dB. There was some variation from vowel to vowel, with the vowel /I/ demonstrating more variability than the other vowels: the average standard deviation for /I/ was 4.64 dB, while for /ae/ and /a/ it was 3.99 and 4.39 dB, respectively.

''Intervowel'' (between-vowel) variation was also examined. Since the NFRF is measuring the sound transmission properties of the neck tissue, it should be independent of the vocal-tract configuration. Figure 3 shows that the intervowel variability was small and on the same order as the intravowel variability. Averaged over the frequency range 100–4000 Hz, the standard deviation of the NFRF ranged from 0.91 to 5.88 dB with a mean value of 3.76 dB. The small intervowel variability confirms that the NFRF is independent of vocal-tract configuration and helps validate the experimental procedure and analysis algorithm.

C. Variations in the NFRF

1. Variation between subject groups

As shown in Fig. 2, all four subject groups produced similar average NFRFs with some differences in corner frequency and gain. The corner frequency was measured by constructing ideal Bode plots that approximated the measured NFRF. The corner frequency is defined as the frequency where the slope of the log magnitude of the idealized NFRF begins to deviate from zero. The corner frequency ranged from 256.7 Hz for the normal males to 430.8 Hz for the normal females, with values of 314.7 and 324.1 Hz for the laryngectomized males and females, respectively. The gain of the NFRF avaried from group to group, with the male NFRFs (both normal and laryngectomized) having larger gains than those displayed by the female NFRFs.

2. Variation with different shaker locations

The effect of shaker location on NFRF was highly subject dependent. Location dependence was classified into three different types: (I) no dependence; (II) gain dependence only; and (III) gain and shape dependence. Examples of each type of location dependence are shown in Fig. 4. Type I (*no dependence*): All of the normal male subjects and one laryngectomized male subject displayed little or no change in NFRF with change in neck location (see the top panel in Fig. 4). Type II (*gain dependence*): For three male and two female laryngectomy patients, the shape of the NFRF remained constant, but the gain changed with changes in shaker location (see the middle panel in Fig. 4). Type III (*gain and shape dependence*): One male and two female laryngectomy patients displayed changes in both the gain and shape of the NFRF with changes in the location of the sound source on the neck (see the bottom panel in Fig. 4). An inspection of the data revealed that for almost all the subjects, the greatest neck location-related differences in mean NFRFs occurred at lower frequencies.

D. NFRF approximation

The results show that the neck wall acts as a low-pass filter when coupling EL vibration to vocal-tract excitation, and could attenuate useful high-frequency energy that normally contributes to speech intelligibility. Compensation for this filtering could be accomplished by appropriate ''inverse filtering'' of the EL driving signal. To facilitate the design of this inverse filter, the NFRFs have been approximated by discrete linear filters of the form

$$K(z) = \frac{\sum_{m=0}^{M} b_m z^{-m}}{1 - \sum_{n=1}^{N} a_n z^{-n}}, \tag{16}$$

where b_m are the numerator coefficients, a_n are denominator coefficients, and M and N are the number of coefficients in the numerator and denominator, respectively. Figure 5 shows the mean NFRFs for each subject group with the corresponding discrete linear filter.

The filters were designed by using a least-squares fit to the frequency response data. The orders of both the numerator and the denominator were determined by iterating through all combinations (up to an order of 6) and choosing the combination that produced a frequency response that minimized the energy difference between the measured NFRF data and the discrete filter frequency response. Four distinct $K(z)$'s were computed for the mean NFRFs of the laryngeal and laryngectomized males and females. In all four cases, the rational function provided an excellent fit, with a mean deviation from the measured NFRF of less than 2 dB and maximum deviation of about 5 dB.

IV. DISCUSSION

The goal of this study was to understand the role that the neck tissue plays in determining the acoustic properties of EL speech, with the underlying motivation of using this information to design an improved EL device. This study has shown that on average, the neck wall acts as a low-pass filter, and that the NFRF can be approximated by a discrete linear filter.

The inverses of the discrete filters can be used to inverse filter a natural glottal sound source to produce an electrolarynx driving signal that compensates for the NFRF. While the discrete filters were stable infinite impulse response (IIR) filters, they were not minimum phase, meaning that their inverses would not be causal and stable. However, since any stable filter can be separated into a minimum-phase filter and an all-pass filter (Oppenheim and Schafer, 1989), a minimum-phase filter version of these discrete filters can be easily computed, especially since phase is not important in this instance.

Another possible use of the NFRF is to design a circuit that has the same frequency response as $1/K(z)$ and to incorporate this circuit into the electrolarynx itself. With such a circuit in place, the various components could be adjusted to compensate for the variability in the NFRFs associated with

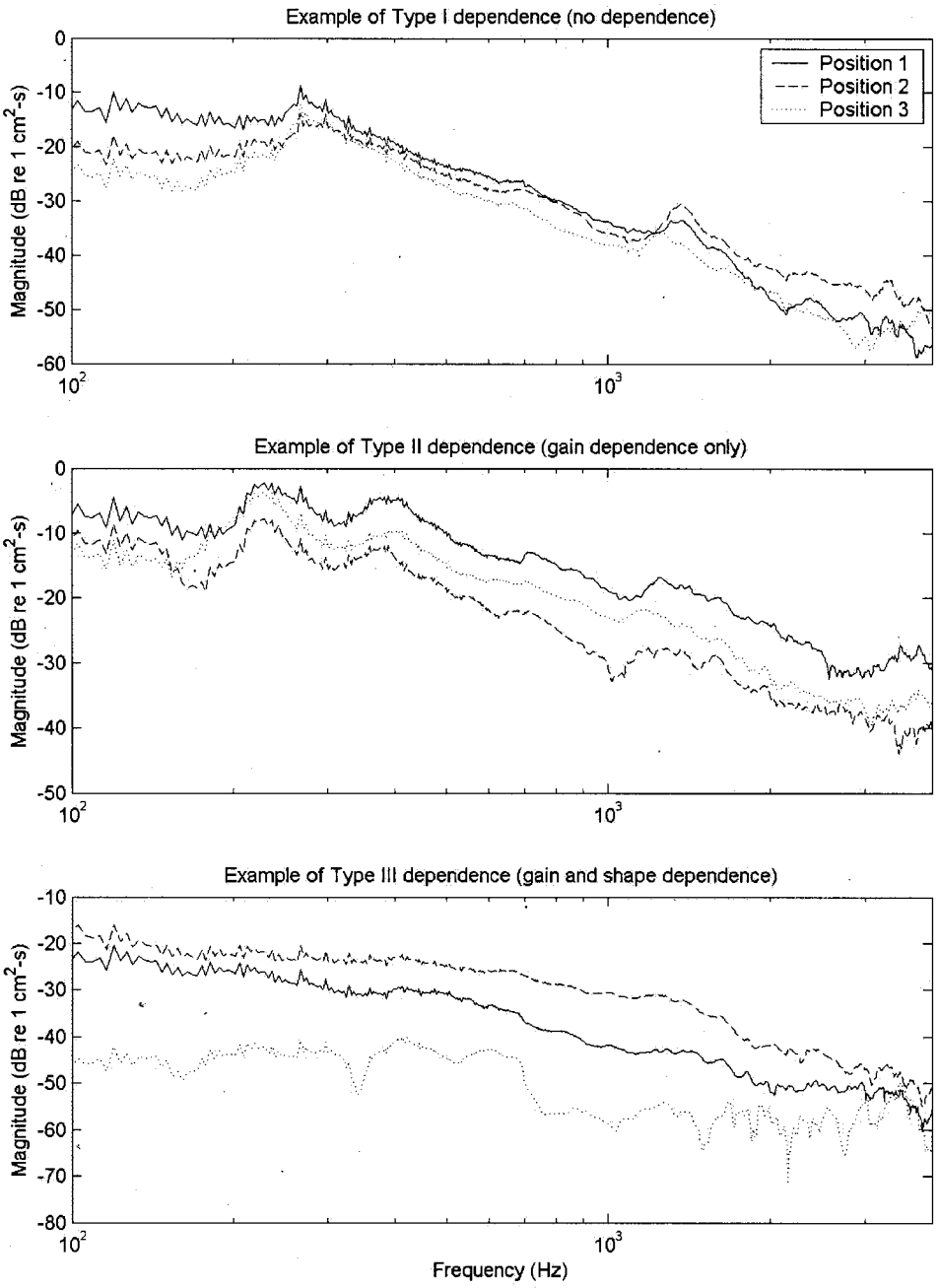


FIG. 4. Plots of the three different types of neck location dependence. The NFRF at position 1 (solid line), position 2 (dashed line), and position 3 (dotted line) are plotted for three subjects. The top plot shows an example of type I dependence where there is little or no change in NFRF with change in neck location. The middle plot is an example of a type II dependence where only the gain of the NFRF changes with neck location. Type III dependence, where both the gain and shape change with neck location, is shown in the bottom plot.

each individual user. As shown in Fig. 2, the corner frequencies of the mean NFRFs of each group differed from each other, producing NFRFs that have somewhat different shapes. Every laryngectomized subject had undergone a total laryngectomy and radiation, and thus it was not unexpected to find notable differences in their NFRFs as compared to normals.

How the transfer function will be used in the electrolarynx design will ultimately depend on which method is simpler and produces a more natural-sounding voice. It should be noted that currently available EL devices do not contain linear transducers that can reproduce arbitrary waveform driving signals.

It is likely that the corner frequency of the NFRF is dependent on the physical properties of the neck tissue. Specifically, as the stiffness of the neck tissue increases, one would expect the corner frequency to increase as well. This means that the normal female group possessed the stiffest neck tissue, the normal male group possessed the most compliant, and the stiffness of the neck tissue of both male and female laryngectomy groups fell in between. On average, the female vocal tract is 15% shorter that that of males and most of this difference is in the pharynx (Klatt and Klatt, 1990). Thus, it is possible that most of the measurement positions on the female necks were located over some part of the thyroid cartilage, while for the male subjects (and laryngectomized subjects) more of the measurements were made over soft tissue. However, only two normal female subjects were tested, and therefore the conclusions that can be drawn from these differences are limited. Indeed, it is possible that the NFRFs of the two subjects that were measured did not rep-

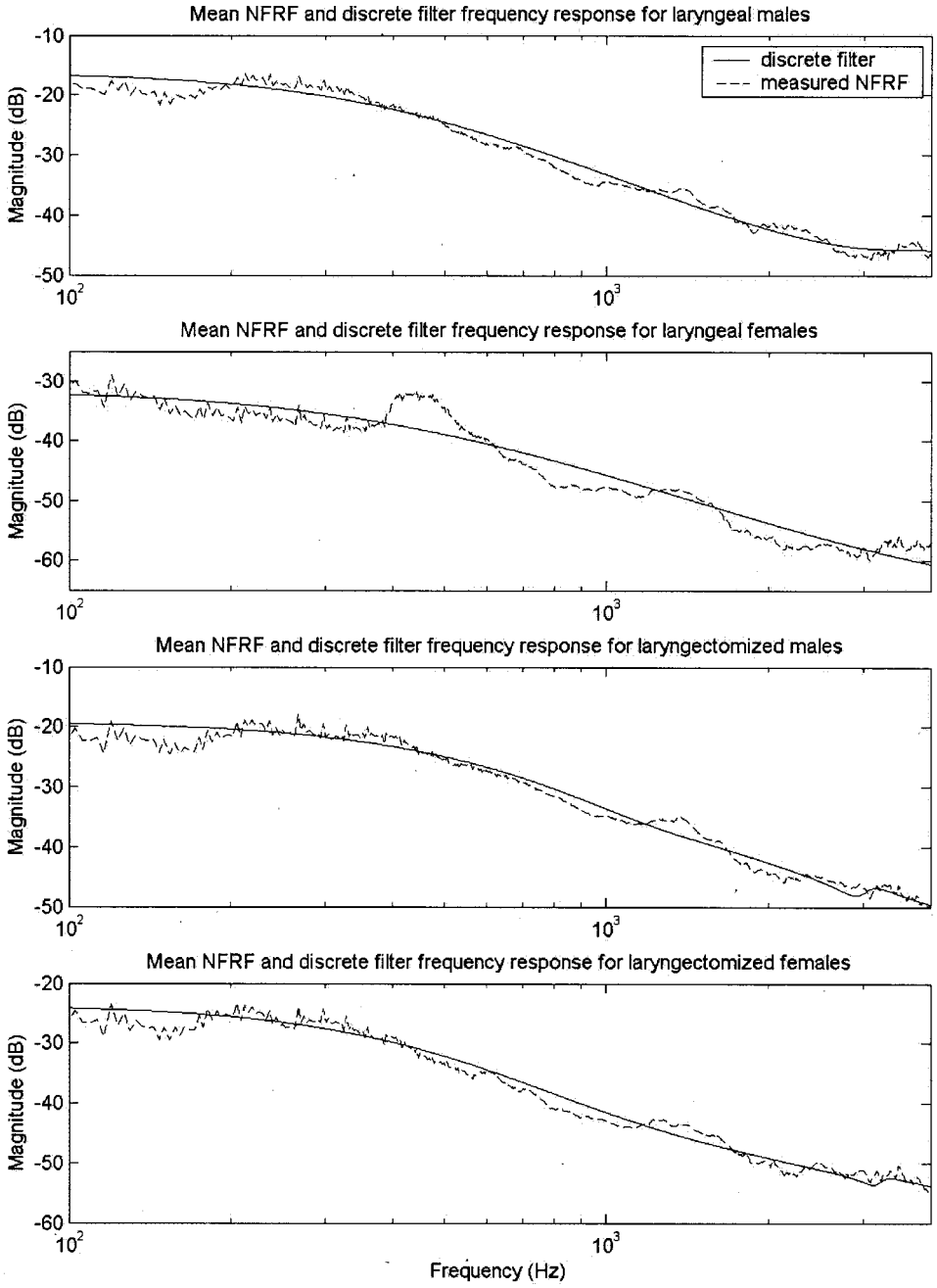


FIG. 5. The mean NFRFs for each subject group (dashed line) with the corresponding frequency response function of the discrete filter estimate (solid line). In all cases, the discrete filter frequency response provided a good fit, usually producing a maximum deviation of less than 5 dB.

resent the average characteristics of the normal female population, i.e., sampling error.

The results indicate that there are three types of neck location dependence. Almost every laryngectomized subject demonstrated either a type II (gain only) or type III (gain and shape) dependence. The location dependence of the NFRFs in laryngectomized subjects was not unexpected since the tissue properties of their necks do vary with location. In many cases, certain parts of the neck are quite rigid due to scarring and the effects of radiation therapy, while other parts are soft and pliable. In addition, a laryngectomy operation tends to produce anatomical asymmetries in the neck. Accordingly, an EL user typically places the device at the spot on the neck that produces the loudest and clearest EL speech. Therefore, in terms of designing an improved EL, one only needs to consider the neck location that produces the NFRF with the largest overall gain.

The group mean NFRFs (and their corresponding linear filter) did not closely approximate the NFRFs of all individual subjects. To illustrate individual variability, Fig. 6 displays the average NFRFs for three laryngectomized males along with the mean NFRF for the male laryngectomized subject group.

While the NFRF of the subject shown in the top panel of Fig. 6 is well fit by the mean NFRF, the NFRFs of the other subjects in the middle panel, and especially the subject in the bottom panel, show significant deviations. The NFRF of the individual in the middle panel of Fig. 6 demonstrates a low corner frequency and a slightly steeper roll-off than does the group mean NFRF. The NFRF of the subject in the bottom panel of Fig. 6 shows increased low-frequency attenuation when compared to the group mean. Both of these latter subjects would probably gain more benefit from an improved EL

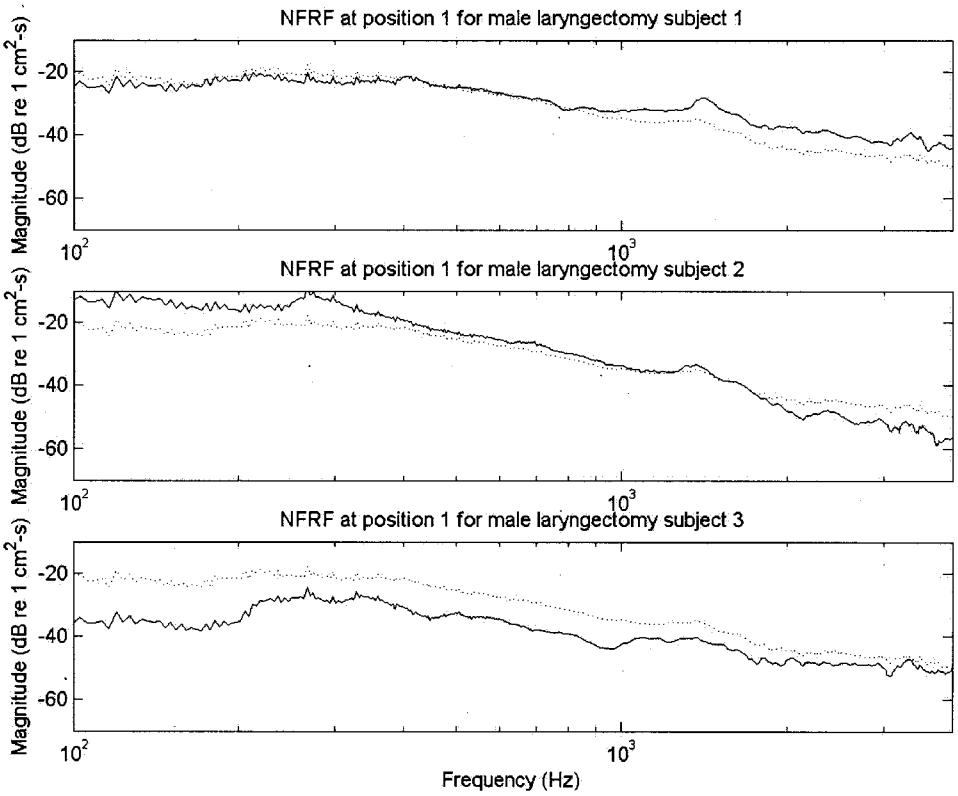


FIG. 6. Plots of individual NFRFs (solid line) and group mean NFRFs (dotted line) for three male laryngectomy patients. The figure demonstrates that there can be a noticeable difference between the NFRFs of individual subjects, despite those subjects being members of the same subject group.

if the associated inverse filters were tailored to their individual neck properties.

It has been well documented (Weiss *et al.*, 1979; Qi and Weinberg, 1991) that electrolaryngeal speech (specifically using a Servox EL) contains a low-frequency deficit. For example, in some cases, the energy in EL speech below 500 Hz can be as much as 35 dB less than that found in normal speech. From the results of this research, it is obvious that the neck does not contribute to the low-frequency deficit but in fact helps (to some degree) to equalize the signal that is ultimately delivered to the vocal tract by devices like the Servox.

A. Comparison with other research

Norton and Bernstein (1993) computed the neck frequency response function (FRF) and found an attenuation of frequencies below 500 Hz and four sharp resonances between 600 and 3000 Hz. Norton and Bernstein followed an experimental protocol that was similar to the one described here, although they did not make closed-mouth measurements in an attempt to remove the directly radiated noise, and they did not estimate and remove the lip radiation characteristic. Furthermore, their frequency response function is the ratio of the spectrum of the estimated volume velocity to spectrum of the input voltage of the shaker, thus making the FRF dependent on the shaker response. However, Norton and Bernstein used the same model B&K shaker that was used in this study, and the shaker response (i.e., the primary shaker transfer function) is relatively flat (a change of about 6 dB over a frequency range of 3000 Hz). In addition, the lip radiation characteristic is also slowly varying in frequency, increasing at about 6 dB/octave to a first-order approximation. Multiplying these two functions produces an all-pass filter over the frequency range in question. Thus, the inclusion of these two functions in the FRF cannot explain why the Norton and Bernstein FRF is quite different from the NFRFs of normal males reported here.

The question arises as to why the FRF and NFRF are drastically different, particularly with respect to the sharp resonant peaks in the FRF that were missing from the NFRF. The first three resonances found by Norton and Bernstein have center frequencies of 650, 1100, and 2450 Hz, which are also very close to the average formant frequencies of the vowel /a/ spoken by a male (Peterson and Barney, 1952). Comparing the FRF resonances with the formants of their experimental subject saying the vowel /a/ shows that they are located at virtually the same frequencies. Norton and Bernstein used cepstral deconvolution instead of linear prediction and inverse filtering to remove the vocal-tract transfer function from the pressure signal measured at the lips. However, in cases of unvoiced speech (e.g., producing a vowel with a shaker driven by noise), it is not obvious what the cutoff points of the frequency-invariant high-pass filter should be, and thus Norton and Bernstein performed the same frequency-invariant filtering as they did when the same subject voiced the same vowel. This technique presents two problems. First, no corrections were made in the cepstral deconvolution to prevent the removal of any spectral features that are part of the FRF. Second, as House and Stevens (1958) discovered, the bandwidths of the vocal-tract formants are significantly narrower when the glottis is closed than when the glottis is open. The speech signal used to estimate the FRF was produced by having the subject hold his breath by closing his glottis, yet the frequency-invariant high-pass filtering location was determined using an open

glottis speech signal. Therefore, the slowly varying part of the spectrum that is being removed has larger bandwidths than the slowly varying part of the speech spectrum measured during shaker excitation. Thus, one likely possibility is that the vocal-tract resonances were not properly deconvolved using the Norton and Bernstein approach, leaving sharp peaks in the remaining signal spectrum.

Fujimura and Linqvist (1971) used a similar experimental protocol to measure vocal-tract characteristics, but instead of trying to remove the vocal-tract transfer function from their measurements, they attempted to remove the neck transfer function. They derived a ''correction function'' to account for the transmission properties of the neck tissue which is similar to the NFRF reported here, specifically in that it contains a low-frequency maximum and a magnitude spectrum that decreases with frequency.

The transmission characteristics of the neck found here are similar to those reported for other body tissues. Wodicka and Shannon (1990) measured the transfer function of the subglottal human respiratory system using eight male subjects. They found that this transfer function is also low pass, with a peak near 100 Hz and a minimum at 600 Hz (which was the limit of the frequency range they were able to measure). Furthermore, the roll-off of the subglottal transfer function was found to be -10 ± 4 dB/octave and -17 ± 5 dB/octave (between 300–600 Hz) (depending on the site on the chest), while the average roll-off of the NFRFs of the laryngeal subjects was measured to be -8.8 ± 2.1 dB/octave (between f_c and 3000 Hz). Although these slopes (for at least at one position on the chest wall) are similar, the roll-off for the chest-wall transfer function continues up to higher frequencies than the NFRF (where the slope becomes flatter). It is not surprising to find a larger degree of high-frequency attenuation in the subglottal transfer function due to the greater thickness of the body wall, which would attenuate higher frequencies to a greater degree than would the neck tissue.

B. Limitations of this research

The experimental protocol that was employed in this study was chosen mainly because of its noninvasive nature. However, a weakness of this approach was the corruption of the pressure signal measured by the microphone at the lips by the sound pressure transmitted through the air from the shaker. This is illustrated in the block diagram in Fig. 1. A cylindrical enclosure for the shaker was used to help reduce the intensity of the directly radiated shaker sound but it was not effective at all frequencies. For certain frequencies, the directly radiated sound was almost at the same level (at times, only about 3 dB lower) as the sound output from the subjects' mouths. When the direct radiation is of similar magnitude to, and is subtracted from, the open-mouth pressure signal, the remaining signal to be analyzed will be small, thus limiting the accuracy of the estimated NFRF.

To verify the effectiveness of the algorithm's ability to remove the direct noise (and to justify the assumptions about the primary shaker and unknown air transfer functions), another experiment was performed which employed an additional physical method to isolate the direct noise from the speech at the lips. A 1.5-in.-thick plywood door was built such that it fit tightly into the doorframe of the acoustic chamber. A 3×6 in. hole was cut into the door at mouth level and a facemask surrounded by sound-insulating foam was cemented into the opening. A subject using a neck-placed sound source (e.g., shaker or EL) could then speak into the acoustic chamber so that the speech from the lips is isolated from the directly radiated sound from the shaker.

In this experiment, one of the normal male subjects was asked to sustain ten different vowels configurations while the shaker was activated. The speech at the lips was recorded by a microphone located in the sound-insulated booth and the signals were analyzed using methods comparable to those already described (see Sec. II). In this case, however, there was no need to subtract an estimate of the direct noise. As shown in Fig. 7, there is good agreement between the NFRF calculated using the door and the NFRF calculated using the original method, with the largest discrepancies occurring between 100 and 200 Hz. The mean difference between the two NFRFs was 1.7 dB, with a maximum difference of 7.8 dB occurring at 106 Hz. Despite the low-frequency difference, these data confirm the validity of the original noise estimation and removal algorithm, the assumptions made about the independence of the primary shaker and the unknown air transfer functions, as well as the validity of these NFRF estimates.

The approach used to estimate the vocal-tract transfer function involves computing a modified LPC estimate on a signal that is essentially a convolution of the impulse responses of the vocal tract and neck wall. Although the LPC estimate is modified to only include the poles of the vocal-tract transfer function, it is still possible that the estimate of the vocal-tract transfer function was corrupted by the presence of the neck impulse response. Unfortunately, because *a priori* knowledge of the shape of the NFRF was unavailable, it was impossible to remove it prior to estimating the vocal-tract transfer function in a manner similar to the one used in Alku (1992). However, to determine the amount of corruption introduced to the vocal-tract estimate by the presence of the neck impulse response, the data collected from one male laryngectomy subject were reanalyzed so that the signal used in the LPC estimate was prefiltered with the inverse of the NFRF approximation described earlier. It was found that the formant frequency estimates only differed by $1.4\%\pm0.3\%$, thus demonstrating that the vocal-tract estimation technique used in this research was not greatly affected by the neck impulse response.

The vocal-tract transfer function often contains zeros, whose presence makes it difficult to estimate the phase of the NFRF. By definition, the linear predictive filter is minimum phase because it is an all-pole filter (Dreller *et al.*, 1993). For non-nasal phonemes (e.g., vowels), the vocal tract can be assumed to be an all-pole filter and hence, minimum phase. However, the presence of the back cavity (posterior to the neck location of the sound source) adds a phase component to the spectrum that cannot be accurately captured by the LP filter, and therefore an accurate estimate of the phase of the NFRF is impossible using the methods described here. While there was an attempt to correct the LPC estimate, the correc-

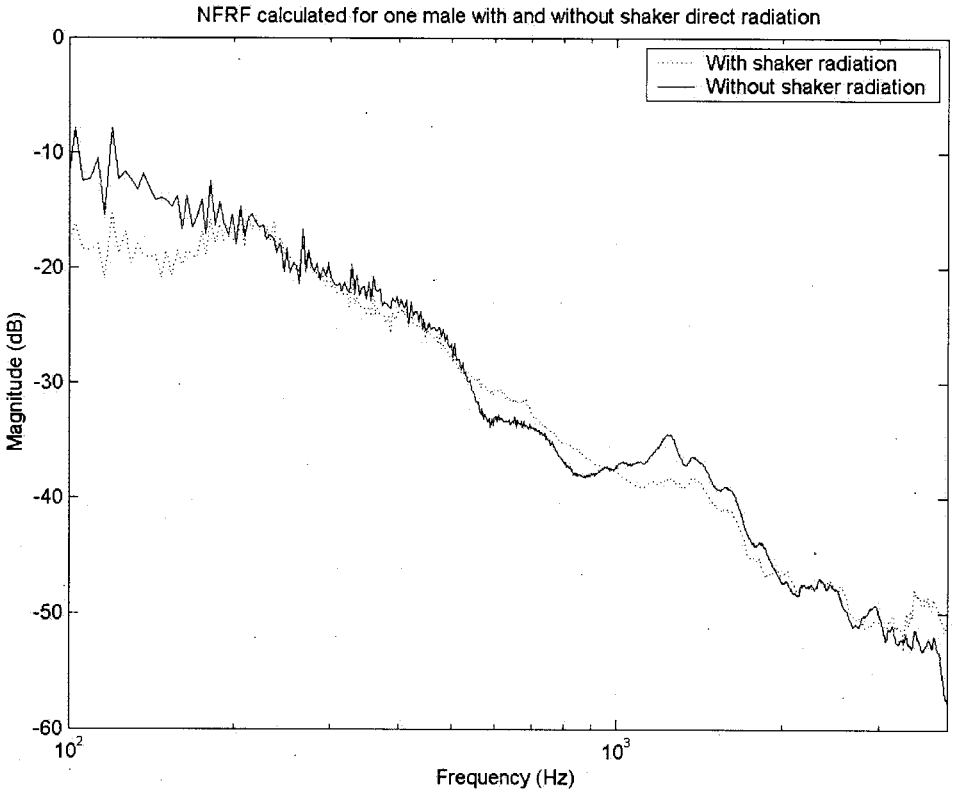


FIG. 7. Plots of the NFRF of one normal male subject estimated using the original method (dotted line) and using the sound-isolating door (solid line). While the two NFRF curves are very similar, especially between 200 and 3500 Hz, there are some differences between 100 and 200 Hz. It should be noted that the NFRF curve calculated with the data gathered using the door was scaled to match the gain of the original NFRF curve in order to emphasize the similarity of the shapes of the NFRFs.

tions were based on producing a better fit to the magnitude of the vocal-tract transfer function, regardless of the phase. Therefore, it is not certain that these corrections improved the phase estimate of the NFRF, and thus phase data are not reported here. It is important to note that the human auditory system is not very sensitive to phase, so as long as the magnitude of the NFRF is available, one can potentially design an EL driving signal that will produce speech with a more normal spectral magnitude.

V. SUMMARY

As a first step in improving the quality of electrolaryngeal speech, measurements of the neck frequency response function (NFRF) were made on nine laryngectomized and four normal subjects. It was found that, in general, the NFRF has a constant maximum gain between 100 Hz and a corner frequency within the range of 200 and 400 Hz, rolls off with a slope of about -9 dB/octave until 3000 Hz, where the magnitude again becomes constant until 4000 Hz. Some subjects' NFRF did not flatten out at 3000 Hz but continued to roll off at about -9 dB/octave. The mean NFRFs of each subject group could be accurately approximated using a discrete linear filter. There was enough intersubject variability to suggest that it would be useful (at least for some laryngectomees) to customize the EL driving signal to the individual's NFRF.

ACKNOWLEDGMENTS

This study is based on a thesis completed in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering and Computer Science at MIT by the first author. This work was supported by grants from the W. M. Keck Foundation, the Department of Veterans Affairs Division of Rehabilitation Research and Development (Grant No. C3343-2DC), and in part by NIH Grant No. T32DC00038. The authors would like to thank Kenneth Stevens of the Research Laboratory of Electronics at MIT, John Rosowski of the Eaton Peabody Laboratory, and Ken Houston of Draper Laboratory for their extensive help and expertise.

Alku, P. (**1992**). "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering," Speech Commun. **11**, 109–118.

Atal, B. S., and Hanauer, S. L. (**1971**). "Speech analysis and synthesis by linear prediction of the speech wave," J. Acoust. Soc. Am. **50**, 637–655.

Barney, H. L., Haworth, F. E., and Dunn, H. K. (**1959**). "An experimental transistorized artificial larynx," in *Readings in Speech Following Total Laryngectomy*, edited by B. Weinberg (University Park, Baltimore).

Dreller, J. R., Proakis, J. G., and Hansen, J. H. L. (**1993**). *Discrete-Time Processing of Speech Signals* (Macmillan, New York).

Espy-Wilson, C., Chari, V., Huang, C., MacAuslan, J., and Walsh, M. (**1998**). "Improvement of electrolaryngeal speech by adaptive filtering," J. Speech Lang. Hear. Res. **41**, 1253–1264.

Fujimura, O., and Lindqvist, J. (**1971**). "Sweep-tone measurements of vocal tract characteristics," J. Acoust. Soc. Am. **49**(2), 541–557.

Hillman, R. E., Walsh, M. J., Wolf, G. T., Fisher, S. G., and Hong, W. K. (**1998**). "Functional outcomes following treatment for advanced laryngeal cancer," Ann. Otol. Rhinol. Laryngol. **107**, 1–27.

House, A. S., and Stevens, K. N. (**1958**). "Estimation of formant band widths from measurements of transient response of the vocal tract," J. Speech Hear. Res. **1**(4), 309–315.

Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**1982**). *Fundamentals of Acoustics* (Wiley, New York).

Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variation among female and male talkers," J. Acoust. Soc. Am. **87**(2), 820–857.

Norton, R. L., and Bernstein, R. S. (**1993**). "Improved LAboratory Prototype ELectrolarynx (LAPEL): Using inverse filtering of the frequency response function of the human throat," Ann. Biomed. Eng. **21**(2), 163–174.

Oppenheim, A. V., and Schafer, R. W. (**1989**). *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).
Peterson, G. E., and Barney, H. L. (**1952**). ''Control methods used in a study of vowels,'' J. Acoust. Soc. Am. **24**(1), 175–184.
Qi, Y., and Weinberg, B. (**1991**). ''Low-frequency energy deficit in electrolaryngeal Speech,'' J. Speech Hear. Res. **34**, 125–1256.
Rabiner, L. R., and Schafer, R. W. (**1978**). *Digital Processing of Speech Signals* (Prentice-Hall, New York).
Sisty, N. L., and Weinberg, B. (**1972**). ''Formant frequency characteristics of esophageal speech,'' J. Speech Hear. Res. **15**, 439–448.
Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge).
Weiss, M. S., Yeni-Komshian, G. H., and Heinz, J. M. (**1979**). ''Acoustical and perceptual characteristics of speech produced with an electronic artificial larynx,'' J. Acoust. Soc. Am. **65**(5), 1298–1308.
Wodicka, G. R., and Shannon, D. C. (**1990**). ''Transfer function of sound transmission in subglottal human respiratory system at low frequencies,'' J. Appl. Physiol. **69**(6), 2126–2130.
Wodicka, G. R., Lam, A. M., Bhargava, V., and Sunkel, D. (**1993**). ''Acoustic impedance of the maternal abdomen,'' J. Acoust. Soc. Am. **94**(1), 13–18.

Audio-vocal responses to repetitive pitch-shift stimulation during a sustained vocalization: Improvements in methodology for the pitch-shifting technique[a)]

Jay J. Bauer[b)] and Charles R. Larson
Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, Illinois 60208

(Received 27 February 2002; revised 4 May 2003; accepted 20 May 2003)

The pitch-shift reflex is a sophisticated system that produces a ''compensatory'' response in voice F_0 that is opposite in direction to a change in voice pitch feedback (pitch-shift stimulus), thus correcting for the discrepancy between the intended voice F_0 and the feedback pitch. In order to more fully exploit the pitch-shift reflex as a tool for studying the influence of sensory feedback mechanisms underlying voice control, the optimal characteristics of the pitch-shift stimulus must be understood. The present study was undertaken to assess the effects of altering the duration of the interstimulus interval (ISI) and the number of trials comprising an average on measures of the pitch-shift reflex. Pitch-shift stimuli were presented to vocalizing subjects with ISI of 5.0, 2.5, 1.0, and 0.5 s to determine if an increase in ISI altered response properties. With each ISI, measures of event-related averages of the first 10, 15, 20, or 30 pitch-shift reflex responses were compared to see if increases in the number of responses comprising an event-related average altered response properties. Measures of response latency, peak time, magnitude, and prevalence were obtained for all ISI and average conditions. While quantitative measures were similar across ISI and averaging conditions, we observed more instances of ''non-responses'' with averages of ten trials as well as at an ISI of 0.5 s. These findings suggest an ISI of 1.0 s and an average consisting of at least 15 trials produce optimal results. Future studies using these stimulus parameters may produce more reliable data due to the fivefold decrease in subject participation time and a concomitant decrease in fatigue, boredom, and inattention. [DOI: 10.1121/1.1592161]

PACS numbers: 43.70.Aj, 43.70.Jt [AL]

I. INTRODUCTION

Neural mechanisms controlling vocalization are poorly understood. Recent years have seen increased study of mechanisms of vocal control in primates and birds (Jürgens, 2002; Luthe *et al.*, 2000; Solls *et al.*, 2000; Suthers *et al.*, 2002). While these studies have increased our understanding of voice control, their invasive nature has precluded parallel studies in humans. An alternative approach for studying vocal control mechanisms in human subjects is to analyze real-time physiologic responses to unanticipated perturbations in sensory feedback during on-going motor tasks. Such perturbations can mimic naturally occurring sensory events that arise from execution of controlled motor tasks, and can reveal important properties of the underlying neural control mechanisms. For example, in the study of mechanisms controlling orofacial musculature during speech, systematic application of mechanical loads to lips during speech demonstrates that the nervous system uses sensory information from the lips for predictive, feed-forward purposes (Abbs and Gracco, 1984; Gracco, 1995; Saltzman *et al.*, 1998; Shaiman and Gracco, 2002).

The perturbation approach is also used to study the properties of the audio-vocal system. These investigations typically involve presenting pitch modulated auditory feedback to a vocalizing subject (Burnett *et al.*, 1998). During sustained vowel phonations and glissandos, the audio-vocal system operates in a negative feedback mode that serves to stabilize voice fundamental frequency (F_0) around the intended pitch (Burnett *et al.*, 1998; Burnett and Larson, 2002; Hain *et al.*, 2000; Larson *et al.*, 2000). Thus, an automatic ''compensatory'' change in voice F_0 corrects for a discrepancy between the intended voice F_0 and the feedback pitch. Because of the automatic nature of the response and lack of habituation, the audio-vocal response to the unexpected change in pitch is termed the pitch-shift reflex (PSR) (Burnett *et al.*, 1998). The latency of the response to altered pitch feedback is ~100 to 150 ms (Hain *et al.*, 2001; Larson *et al.*, 2000), which suggests that underlying neural mechanisms are more complex than a simple multisynaptic reflex loop contained in the medulla. Rather, the pathways may involve higher levels of the brainstem, such as the midbrain periaqueductal gray (PAG) and inferior colliculus (IC), or even cortical and cerebellar pathways. Recent research indicates that the pitch-shift reflex may also be used in control of F_0 related to suprasegmental aspects of speech (Donath *et al.*, 2002; Natke and Kalveram, 2001). Exploring the characteristics of the pitch-shift reflex may provide important information about how the nervous system uses auditory feedback in the regulation of voice F_0 during speech and singing.

If the pitch-shift reflex is to be useful for the study of the

[a)]Material originally presented in ''Improvements in methodology for the pitch-shifting technique,'' proceedings of The Acoustical Society of America, Chicago, IL, June 2001.
[b)]Electronic mail: j-bauer2@northwestern.edu

influence of auditory feedback on voice F_0 control, many parameters of the reflex must be understood. One parameter that has not been addressed in previous studies is interstimulus interval (ISI). Previous investigations of the pitch-shift reflex randomly presented one pitch-shift stimulus during the first few seconds of a sustained vowel repeated for 15–20 consecutive vocalizations. This paradigm corresponded to an interstimulus interval (ISI) greater than 5 s (Burnett *et al.*, 1998; Hain *et al.*, 2000; Larson *et al.*, 2000), which likely provided more than adequate recovery time between successive pitch perturbations. However, it remains unclear whether the audio-vocal system can respond to multiple pitch perturbations presented across the duration of a single vocalization. Stimulation at high rates may result in an overlap of successive responses if response latency and duration exceed the interval between consecutive stimuli. Such an overlap can introduce error, and alter the magnitude (Nelson and Lassman, 1968; Wikström *et al.*, 1996) of the observed response. Complex systems such as the pitch-shift reflex require time to process and transmit neural potentials from one location to another. Thus, temporal processing constraints limit the frequency at which the system can respond to repetitive stimulation without a reduction in response magnitude. We hypothesize that the audio-vocal system continuously monitors auditory feedback pitch throughout the duration of an utterance, and responds to repetitive pitch-shift stimulation with successive compensatory pitch-shift reflexes. We tested this hypothesis by comparing PSR responses elicited under several interstimulus intervals (ISI).

Event-related averaging techniques have been widely used to investigate small biological signals in a noisy environment. In studies such as auditory evoked potentials, thousands of averages are obtained with relative ease because subjects do not need to attend to the stimuli or actively control muscular movements (Abbas and Brown, 1991; Dawson, 1954; Nakamura *et al.*, 1989; Nelson and Lassman, 1968). However, in studies of motor systems, subjects must maintain attention, and control muscles within rather strict limits (Barlow and Bradford, 1996; Burnett *et al.*, 1998; Sapir and McClean, 1981). These demands can be taxing on subjects, but such limitations are partially offset by the fact that motor responses are typically large and reliable. Thus, averaged motor responses can be obtained from a noisy background with far fewer stimulus presentations than those needed to discern evoked brain responses. Studies of speech motor systems have utilized event-related averaging techniques to study the role of sensory feedback on perioral, jaw, and laryngeal muscles (Barlow and Bradford, 1996; Burnett *et al.*, 1998; Sapir and McClean, 1981; Smith *et al.*, 1987). However, in these studies, the number of trials comprising each average varied from 15 to 80 trials. If an objective assessment of the number of averages is not known, experimenters may average insufficient trials to obtain an event-related average, or they may have used too many trials, which confers no greater response reliability and yet may excessively tire the subject. Nevertheless, it is important to understand interactions between stimulus and subject-dependent variables in the study of motor responses to sensory stimulation. Therefore, a secondary objective of the present study was to determine the optimal number of trials required for reliable averaged measurements of the pitch-shift reflex.

Determining the optimal stimulus parameters needed to elicit a pitch-shift reflex has the potential to produce more reliable data. A concern with our previous technique of presenting a single stimulus for each 5-s vocalization is that a lengthy period of vocalization was required in order to study several manipulations of an independent variable. Such extended vocalization time could lead to fatigue, boredom, and more variable data. In this study, we presented pitch-shift stimuli to vocalizing subjects with interstimulus intervals of 5.0, 2.5, 1.0, and 0.5 s. These intervals correspond to stimulation rates of 0.2, 0.5, 1.0, and 2.0 pitch-shifts per second. We then compared dependent measures across event-related averages consisting of the first 10, 15, 20, and 30 trials obtained at each ISI.

II. METHODS

Thirty-one normal-hearing young adult subjects (8 males and 23 females) with no history of speech-language disorders or neurological deficits were tested under the experimental conditions described below. In each condition, subjects repeatedly produced /u/ vowel sounds for ~5 s into an AKG boom-set microphone (model HSC 200) at a conversational voice F_0 at 70 dB SPL at a microphone-to-mouth distance of 5 cm, while seated comfortably in a sound attenuated booth. The microphone (voice) signal was amplified with a Mackie Mixer (model 1202), then processed for pitch-shifting with an Eventide Ultraharmonizer (SE 3000), mixed (Mackie Mixer model 1202-VLZ) with pink masking noise (Goldline Audio Noise Source, model PN2, spectral frequencies 1 to 5000 Hz), and presented to the subject over AKG headphones (model HSC 200) after amplification by a Crown amplifier (D75-A). Voice signals were amplified to 80 dB SPL and pink noise was amplified to 60 dB SPL to partially mask the detection of bone-conducted signals. The Ultraharmonizer was controlled by MIDI software to repeatedly insert pitch-shift stimuli of 0.1-s duration and 100 cents (100 cents=1 semitone) greater than speaking F_0 into the subject's vocal auditory feedback during each vocalization. These stimulus parameters were chosen because they have been shown to elicit reliable reflexes (Burnett *et al.*, 1998). Moreover, stimuli of 0.1-s duration do not elicit secondary, voluntary responses (Burnett *et al.*, 1998). Stimuli were presented at an average ISI of approximately 2.5, 1.0, and 0.5 s. Each ISI varied somewhat (~100 ms) to reduce the ability of subjects to predict the exact time of stimulus onset. Thirty stimuli were presented under each ISI condition. Thus, the number of vocalizations needed under each ISI condition differed. We collected 15 vocalizations for the 2.5-s ISI condition, 6 vocalizations for the 1.0-s ISI condition, and 3 vocalizations for the 0.5-s ISI condition. The order of presentation of the experimental conditions was randomized across subjects.

We tested ten of the original subjects (three males and seven females) under one additional ISI condition. Under this additional condition, pitch-shift stimuli were presented approximately every 5.0 s resulting in only one stimulus per vocalization across a total of 30 vocalizations. This ISI con-

dition was included *posthoc* so that we could directly compare PSR responses elicited using repetitive pitch-shifts per vocalization with those elicited by a single pitch-shift stimulus per vocalization, as in previous studies (Burnett *et al.*, 1998; Hain *et al.*, 2001, 2000; Larson *et al.*, 2000).

Each voice signal, feedback signal, and TTL pulse indicating the timing of the MIDI signal to the Ultraharmonizer and resultant pitch-shift was digitized on-line at 10 kHz (5 kHz anti-aliasing filter) onto a laboratory computer. In off-line analysis in preparation for frequency extraction, the voice signal was low-pass filtered at 200 Hz for females and 100 Hz for males, differentiated, and then smoothed with a five-point binomial, sliding window (Larson *et al.*, 2000). Low-pass filtering of the voice signal removed most of the energy present at harmonic frequencies, allowing error-free triggering of the F_0 in subsequent stages of signal processing. A software algorithm then detected positive-going threshold-voltage crossings, interpolated the time fraction between each pair of sample points that constituted a crossing, and calculated the reciprocal of the period defined by the center points to signify the voice F_0. The resulting F_0 signal was then low-pass filtered at 10 Hz to remove sharp discontinuities associated with each glottal cycle. For each rate condition, the F_0 signal was time aligned to stimulus onset (TTL pulse) and averaged. Event-related averages were calculated using a 0.2-s pretrigger baseline and 0.5-s posttrigger window. For each subject, separate averages were made for voice F_0 responses for the first 10, 15, 20, and 30 stimuli presented under each ISI condition.

A software algorithm was used to extract poststimulus responses that exceeded two SDs of the prestimulus baseline mean of the event-related average. The algorithm measured reflex latency (s), duration (s), magnitude (cents), and peak time (s). Response latency was the time of F_0 departure from the baseline mean by more than 2 SDs, while reflex duration reflected the total time a response remained above 2 SDs. Reflex magnitude (cents) was the maximum F_0 fluctuation from the mean prestimulus baseline F_0 beyond 2 SDs, and response peak time (s) reflected the poststimulus time corresponding to response magnitude. A minimum duration of at least 50 ms, a minimum peak time of 120 ms, a minimum magnitude of 5 cents, and a maximum latency no greater than 400 ms were required for a response to be considered valid. Previous experiments utilized similar duration criteria (Burnett *et al.*, 1998; Burnett and Larson, 2002; Hain *et al.*, 2000; Larson *et al.*, 2001, 2000) to reduce invalid responses. Peak time, magnitude, and latency criteria were used to eliminate extreme data outliers. If no valid responses were observed for a given event-related average, then dependent measures were assigned a value of 0 and classified as a ''non-response.'' If more than one valid response occurred per event-related average, the response with the shortest latency was recorded.

While ''non-responses'' would be discarded from statistical analyses in most studies, we included them in our analyses so as to obtain a more accurate measure of overall variability. Given the relatively small number of these ''non-responses,'' their inclusion in our analysis was not expected to alter averaged data. Nevertheless we felt it was important to maintain the spirit of the data distribution given that the study was designed to test methodological adjustments. Subsequently, each ''non-response'' was transformed to the cell mean to avoid problems due to missing data in the statistical analyses of the dependent measures.

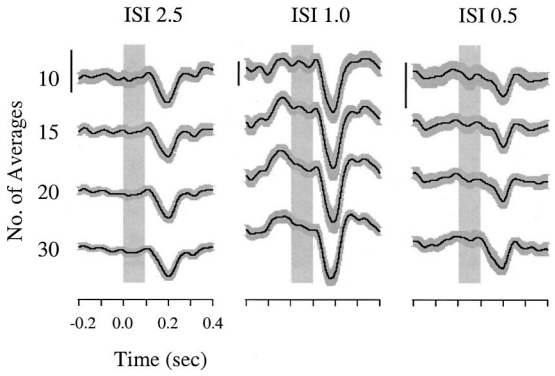


FIG. 1. Event-related averages of voice F_0 for a representative subject tested with ISI of 2.5, 1.0, and 0.5 s. Under each rate, averages of 10, 15, 20, and 30 responses were calculated (vertically stacked traces). Black lines represent the overall average response. Gray bars represent standard error of the mean for all responses comprising an average on a point-by-point basis. Shaded boxes represent pitch-shift stimulus onset and duration (0.1 s) beginning at time 0 s. Vertical line scales=20 cents.

Logarithmic transformations were performed on the adjusted dependent measures to meet assumptions for parametric statistical comparison in repeated measures (within-subject) designs (i.e., equal samples per cell, minimal outliers, normal distribution, homogeneity of variance, compound symmetry, and sphericity). Transformed latency, peak time, and magnitude measurements were tested using repeated-measures ANOVAs (RM ANOVA) with a Bonferoni corrected alpha set at $p=0.01$. *Posthoc* Sheffé tests were used for follow-up analyses when needed. Response prevalence was tested with a nonparametric Cochran's Q test across experimental conditions.

III. RESULTS

Figure 1 illustrates data for a representative subject across three of the four ISI conditions (2.5, 1.0, and 0.5 s). The four vertically stacked traces associated with each experimental condition represent the average responses obtained from 10, 15, 20, and 30 trials. Thin black lines represent average response and thick gray bars represent SE of the mean of all trials comprising the average. The vertical gray box represents the onset and duration of the pitch-shift stimuli. The overall morphology of responses is consistent across conditions and number of averaged trials. While the largest responses for this subject were observed for the 1.0-s ISI condition, we did not observe this trend across all subjects. On average the overall median reflex latency, peak time, and magnitude calculated across all subjects and collapsed across all conditions were 0.135 s, 0.219 s, and 14.34 cents, respectively.

The following RM ANOVA analyses were used to assess the response properties of the PSR across ISI conditions (2.5, 1.0, and 0.5) and averaging conditions (10, 15, 20, and 30). Results of the analysis of response latency did not reveal any

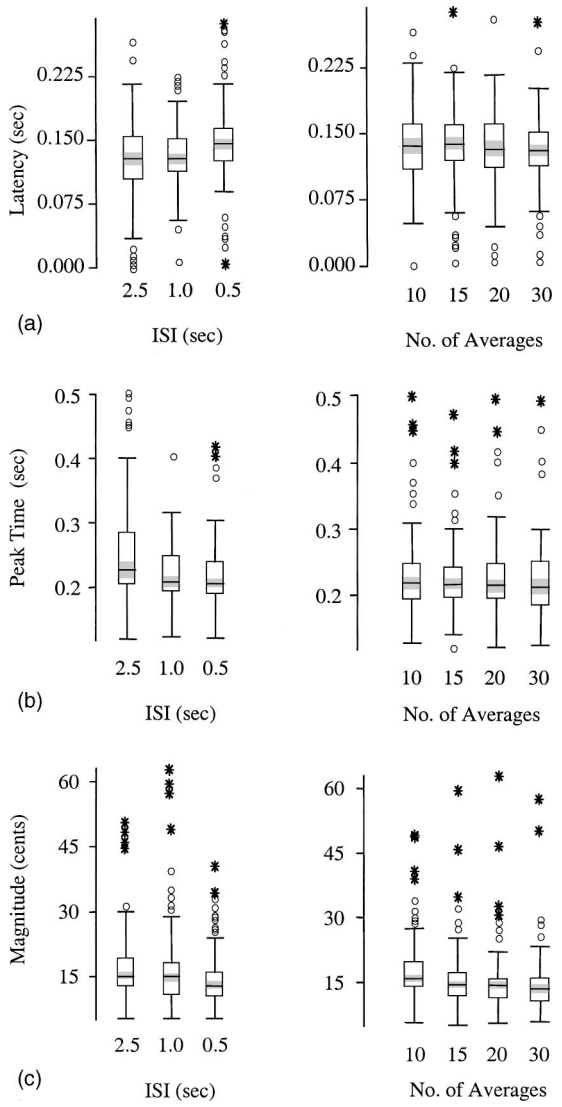


FIG. 2. (a) Median box-plots of response latency (s) by ISI and number of averages ($N=31$). (b) Median box-plots of responses peak time (s). (c) Median box-plots of response magnitude (cents). Box-plot definitions: The horizontal line through a box is the median. The shaded region surrounding the median is the 95% confidence interval. The upper and lower limits of the box represent the 75th and 25th percentiles, respectively. Whiskers extend to upper and lower limits of the main body of data. Points depicted by a circle are considered to be extreme data values, while very extreme values are plotted as asterisks.

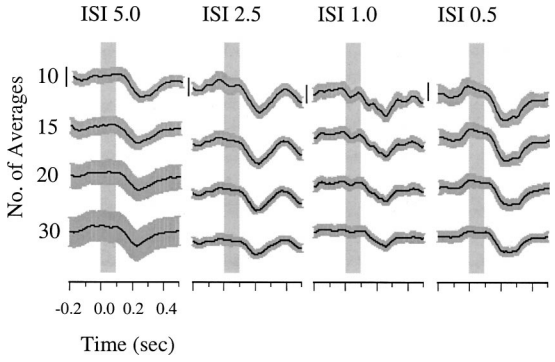


FIG. 3. Event-related averages of voice F_0 for a representative subject tested with ISI of 5.0, 2.5, 1.0, and 0.5 s. Under each rate, averages of 10, 15, 20, and 30 responses were calculated (vertically stacked traces). Black lines represent the overall average response. Gray bars represent standard error of the mean for all responses comprising an average on a point-by-point basis. Shaded boxes represent pitch-shift stimulus onset and duration (0.1 s) beginning at time 0 s. Vertical line scales=20 cents.

statistically significant effects as a function of the ISI condition [$F(2,60)=2.45$, $p>0.05$] or the number of averaged responses [$F(3,90)=2.72$, $p>0.05$]. Likewise, peak-time analysis also did not indicate any statistically significant effects across ISI conditions [$F(2,60)=4.88$, $p>0.01$] or number of averages [$F(3,90)=0.85$, $p>0.05$]. Thus, temporal measures of latency and peak time were not affected by ISI rates as fast as 0.5 s or averages consisting of ten trials [Figs. 2(a) and (b)]. However, analysis of response magnitude revealed a significant main effect across the number of average responses [$F(3,90)=11.30$, $p<0.0001$], but not across ISI conditions [$F(2,60)=3.9$, $p>0.01$]. *Posthoc* Sheffé testing revealed larger magnitude responses for averages of 10 stimuli compared to averages of 15, 20, or 30 stimuli. Thus, larger magnitude responses were the result of averaging too few trials [Fig. 2(c)]. Interaction effects were not statistically significant.

In order to assess the effects of presentation of multiple pitch-shift stimuli compared to single stimulus presentation, data were also compared across the group of ten subjects exposed to four ISI conditions (5.0, 2.5, 1.0, and 0.5). Figure 3 displays data from a representative subject. Overall, there were no observable differences in reflex latency as a function of ISI [$F(3,27)=0.60$, $p>0.01$] or averaging condition [$F(3,27)=0.62$, $p>0.05$] as tested with a RM ANOVA ($N=10$). Likewise, no main effects or interactions were observed for measures of peak time across ISI [$F(3,27)=0.54$, $p>0.05$] or averaging conditions [$F(3,27)=0.73$, $p>0.05$]. In comparison, an increase in response magnitude occurred with an apparent decrease in the number of averages [$F(3,27)=14.14$, $p<0.0001$], but there were no statistically significant differences in magnitude as a function of ISI condition [$F(3,27)=0.31$, $p>0.05$]. *Posthoc* Sheffé analyses revealed that averages of 10 trials were greater in magnitude than averages of 15, 20, or 30 trials. Thus, PSR magnitude appears to be inversely related to the number of trials comprising the average. However, increasing the rate of stimulation from one per vocalization up to ten per vocalization (ISI 0.5 s) did not impact the pitch-shift reflex parameters.

Across all subjects there were 412 possible averaged responses [(21 subjects×3 ISI×4 averages)+(10 subjects×4 ISI×4 averages)]. Overall, 341 responses (83%) decreased in F_0 (opposing responses), 17 responses (4%) increased in F_0 (''following'' responses), and 54 (13%) were considered ''non-responses.'' The percentage of ''following'' responses and ''non-responses'' appeared to increase as a result of the

TABLE I. Percentage (%) of ''following'' (FOLL), opposing (OPP), and ''non-responses'' (NR) across ISI condition.

ISI condition	5.0	2.5	1.0	0.5	Total
FOLL	2	2	3	8	4
OPP	90	88	86	73	83
NR	8	10	11	19	13

TABLE II. Percentage (%) of ''following'' (FOLL), opposing (OPP), and ''non-responses'' (NR) across number of averages.

No. of averages	10	15	20	30	Total
FOLL	3	3	6	5	4
OPP	80	84	82	84	83
NR	17	13	12	11	13

decrease in interstimulus interval (Table I) with the highest incidence observed in the 0.5-s ISI condition. Similarly, an increase in the percentage of ''non-responses'' was observed as a result of a decrease in the number of averages (Table II) with the highest incidence observed for averages consisting of only 10 trials. Table III displays response prevalence as a percentage of the total responses across number of averages and ISI conditions. In general, fewer responses were observed for the 0.5-s ISI condition regardless of the number of responses comprising the average. Similarly, averaging of 10 responses produced a lower percentage of responses than averages of 15, 20, or 30 trials across each of the ISI conditions. However, nonparametric statistical analysis of these response prevalence trends did not reach statistical significance (Cochran's $Q=11.22$, $p>0.05$) given the relatively small subject size.

IV. DISCUSSION

The primary objective of this study was to determine if the pitch-shift reflex could be elicited repetitively at relatively short ISIs during the course of a single vocalization. The secondary objective was to determine the number of trials needed to yield reliable averaged responses. A by-product of both objectives was to determine if studies of the pitch-shift reflex could be conducted such that the time required for testing each subject could be reduced and thereby minimize potential fatiguing factors. It was reasoned that reduction of vocal fatigue would produce less variability in voice F_0 and hence more reliable data.

Determination of the optimal ISI for the study of event-related potentials is important for two reasons. The first is that studying the effects of ISI on response measures can provide indirect evidence of response processing duration. The second is that knowing the optimal ISI allows experiments to be conducted more rapidly than if an ISI is too long. Interstimulus intervals that are longer than the response processing time do not affect response measures, but increase the amount of experimentation time. Interstimulus intervals that are too short in duration may cause a reduction in response magnitude (Nelson and Lassman, 1968; Wikström *et al.*, 1996) regardless of the faster experimentation time. In the present study, reduction of the ISI from 5.0 to 0.5 s had no statistically significant effect on latency, peak time, or magnitude of the response, and thus we believe that the central processing mechanisms of this reflex are less than 0.5 s in duration. However, we could not assess the nature of the pitch-shift reflex with an ISI of less than 0.5 s because of limitations inherent in our averaging technique. Specifically, a prestimulus baseline period of 0.2 s is needed to obtain a stable baseline F_0 (approximately 20–50 vocal cycles). Further, a 0.5-s poststimulus window is needed to verify that the pitch-shift response returns to the baseline F_0 level. Nevertheless, our findings indicate the audio-vocal system adequately compensates for pitch fluctuations every 0.5 s during a sustained vocalization, which corresponds to corrective control at a rate of at least 2 Hz.

Even though the response measures were not affected by the reduction in the ISI, the shortest ISI (0.5 s) did elicit a greater percentage of ''non-responses'' and ''following'' responses (Tables I and II). We speculate that the decrease in response prevalence may have been caused by an overlap between the refractory period of one response and the prestimulus baseline of the next response. To illustrate this point for one set of data from a representative subject (Fig. 4), the poststimulus averaging window was extended to 1.0 s to encompass two sequential responses. Here it can be seen that the termination of the first reflex occurred at the same time as the prestimulus baseline period of the next reflex. The decreased magnitude of the second response suggests there was not a sufficient recovery period between successive stimuli presented with an ISI of 0.5 s. Thus, the overlap between the refractory period following one response and the prestimulus baseline of the next response may have lead to an increase in overall variability and a subsequent decrease in response prevalence in the average waveform. These data suggest that the ISI should be longer than 0.5 s so as to minimize potential response degradation.

The other goal of the present study was to determine the

TABLE III. Response prevalence displayed as a percentage (%) of valid responses per cell (total number of opposing plus following responses divided by the total number of possible responses) across number of averages and ISI conditions. Weighted row and column means are based on raw data.

	No. of averages				
ISI	10 % Response	15 % Response	20 % Response	30 % Response	Weighted row mean
1.0	90	90	100	90	93
2.5	81	94	90	94	90
1.0	84	87	90	94	89
0.5	81	81	81	81	81
Weighted column mean	84	87	88	89	

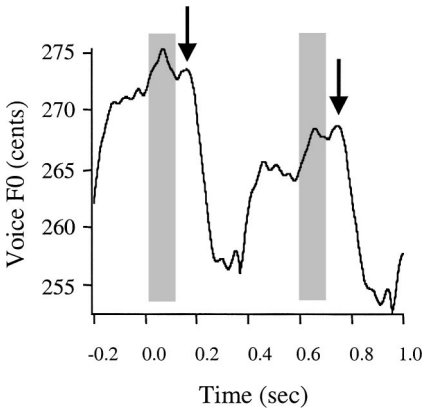


FIG. 4. Event-related average for a representative subject displaying an overlap of successive pitch-shift reflex responses due to the short ISI at 0.5 s. Black lines represent the overall average voice F_0 and arrows denote response onset. Duration of the post-stimulus averaging window was extended to 1.0 s. Shaded boxes represent pitch-shift stimulus onset and duration (0.1 s) beginning at time 0.0 and 0.6 s.

optimal number of trials needed to yield reliable and consistent averaged responses. Statistical analysis showed no difference in response latency or peak time regardless of the number of trials included in the averaged response. However, we observed larger magnitude responses and greater instances of ''non-responses'' when only 10 responses were averaged compared to 15, 20, or 30. It is difficult to explain why responses should have been larger with an average of only ten trials. However, one possibility is that with relatively few responses comprising the average waveform, there was larger variability in the background noise level that added to the average response waveform yielding a larger response magnitude [Fig. 2(c)]. With a greater number of averages, a greater amount of noise cancellation occurs, and the measured response magnitude may be a more accurate estimate of the actual response as it regresses to the mean. Pursuing this logic, the greater incidence of ''non-responses'' with ten averages is most likely due to the fact that the ''noise'' in the background signal was not adequately averaged out. In other words, the response was no larger than the noise (variability) in the averaged signal. For these reasons, we conclude that at least 15 responses should be averaged per condition. Although our results did not show increased improvement with averages of 20 or 30 responses, it may be suggested that in some cases of extreme noise a greater number of averages would be warranted to reduce variability.

Previous studies in our lab using the pitch-shift paradigm have presented stimuli at an ISI of 5.0 s (Burnett *et al.*, 1998; Hain *et al.*, 2001; Kiran and Larson, 2001; Larson *et al.*, 2000) or once per vocalization. Subjects were instructed to vocalize at least 15–20 times per experimental condition to obtain an averaged response comprised of 15–20 trials. The response measures (latency, peak time, and magnitude) from the present study are very similar to those previously obtained. Therefore, it is suggested that similar data could have been obtained using only three vocalizations (five stimuli per vocalization). Although the median response magnitude (~14 cents) measured in the present study was less than values reported in the above studies, response magnitude is in part dependent on stimulus duration. The duration in this study (0.1 s) was substantially less than in most of the above studies, and thus leads to smaller magnitude responses than those previously reported. It must be emphasized, however, that if longer stimulus durations are used, the ISI would have to be increased to accommodate a longer response duration, and this could in turn reduce the number of stimuli per each vocalization, with a further increase in the number of vocalizations required per condition. Although short stimulus durations with short ISIs and repetitive trials per vocalization are appropriate for investigating reflexive properties of the audio-vocal system (Burnett *et al.*, 1998; Kiran and Larson, 2001; Larson *et al.*, 2000), longer durations with a single stimulus per vocalization are still preferable for other studies interested in more voluntary mechanisms of audio-vocal control (Hain *et al.*, 2001, 2000), or possibly studies of speech and prosody (Donath *et al.*, 2002; Natke and Kalveram, 2001).

V. CONCLUSION

Measuring physiologic responses to unanticipated perturbations in sensory feedback provides important information about neural control mechanisms for a particular motor task. Inferences regarding the neural mechanisms of motor planning, initiation, and corrective control can be made based on the timing, magnitude, and form of responses to perturbed sensory feedback in various tasks. Using this technique for the study of voice, it has been shown that auditory feedback is used to help stabilize voice F_0 around an intended level. In order to more fully appreciate the role of auditory feedback on voice control, a detailed understanding of interactions between stimulus variables and vocal responses is necessary. The present study has shown that the audio-vocal system can respond to repetitive pitch perturbations during a sustained vocalization. Furthermore, presenting pitch-shift stimuli with an ISI of at least 1.0 s and averaging at least 15 trials yields measures of response latency, peak time, magnitude, and prevalence comparable to those obtained at longer ISI and with greater number of averaged stimuli. Shorter ISIs and lower number of averages elicit more ''non-responses'' due to potential increased variability and overlap of successive averaging windows. Thus a fivefold reduction in subject participation time compared to previous studies in our lab may be obtained by reducing the ISI from 5.0 to 1.0 s. This reduction has the benefit of decreasing the total number of vocalizations, and thus reduces confounding effects from subject fatigue, boredom, lapse of attention, and voice F_0 variability across vocalizations.

ACKNOWLEDGMENTS

Research partially supported by NIH Grant No. DC02764-01 awarded to Dr. Charles R. Larson, Northwestern University, Evanston, IL. We would like to gratefully acknowledge the assistance of Ciara Leydon for suggestions regarding manuscript preparation, as well as invaluable feedback provided by two reviewers on an earlier version of the manuscript.

Abbas, P. J., and Brown, C. J. (**1991**). ''Electrically evoked auditory brainstem response: Refractory properties and strength-duration functions,'' Hear. Res. **51**, 139–147.

Abbs, J. H., and Gracco, V. L. (**1984**). ''Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech,'' J. Neurophysiol. **51**, 705–723.

Barlow, S. M., and Bradford, P. T. (**1996**). ''Comparison of perioral reflex modulation in the upper and lower lip,'' J. Speech Hear. Res. **39**, 55–75.

Burnett, T. A., and Larson, C. R. (**2002**). ''Early pitch shift response is active in both steady and dynamic voice pitch control,'' J. Acoust. Soc. Am. **112**, 1058–1063.

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (**1998**). ''Voice F_0 responses to manipulations in pitch feedback,'' J. Acoust. Soc. Am. **103**, 3153–3161.

Dawson, G. D. (**1954**). ''A summation technique for the detection of small evoked potentials,'' Electroencephalogr. Clin. Neurophysiol. **6**, 65–84.

Donath, T. M., Natke, U., and Kalveram, K. T. (**2002**). ''Effects of frequency-shifted auditory feedback on voice F_0 contours in syllables,'' J. Acoust. Soc. Am. **111**, 357–366.

Gracco, V. L. (**1995**). ''Central and peripheral components in the control of speech movements,'' in *Producing Speech: Contemporary Issues*, edited by F. Bell-Berti (American Institute of Physics, Woodbury, NY), pp. 417–431.

Hain, T. C., Burnett, T. A., Larson, C. R., and Kiran, S. (**2001**). ''Effects of delayed auditory feedback (DAF) on the pitch-shift reflex,'' J. Acoust. Soc. Am. **109**, 2146–2152.

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., and Kenney, M. K. (**2000**). ''Instructing subjects to make a voluntary response reveals the presence ot two components to the audio-vocal reflex,'' Exp. Brain Res. **130**, 133–141.

Jürgens, U. (**2002**). ''A study of the central control of vocalization using the squirrel monkey,'' Med. Eng. Phys. **24**, 473–477.

Kiran, S., and Larson, C. R. (**2001**). ''Effect of duration of pitch-shifted feedback on vocal responses in Parkinson's disease patients and normal controls,'' J. Speech Lang. Hear. Res. **44**, 975–987.

Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., and Hain, T. C. (**2001**). ''Comparisons of voice F_0 responses to pitch-shift onset and offset conditions,'' J. Acoust. Soc. Am. **110**, 2845–2848.

Larson, C. R., Burnett, T. A., Kiran, S., and Hain, T. C. (**2000**). ''Effects of pitch-shift onset velocity on voice F_0 responses,'' J. Acoust. Soc. Am. **107**, 559–564.

Luthe, L., Hausler, U., and Jürgens, U. (**2000**). ''Neuronal activity in the medulla oblongata during vocalization. A single-unit recording study in the squirrel monkey,'' Behav. Brain Res. **116**, 197–210.

Nakamura, M., Nishida, S., and Shibasaki, H. (**1989**). ''Evaluation of the signal-to-noise ratio for average evoked potentials: Determination of interstimulus interval and averaging number,'' Front Med. Biol. Eng. **1**, 341–349.

Natke, U., and Kalveram, K. T. (**2001**). ''Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables,'' J. Speech Lang. Hear. Res. **44**, 577–584.

Nelson, D. A., and Lassman, F. M. (**1968**). ''Effects of intersignal interval on the human auditory evoked response,'' J. Acoust. Soc. Am. **44**, 1529–1532.

Saltzman, E., Löfqvist, A., Kay, B., Kinsella-Shaw, J., and Rubin, P. (**1998**). ''Dynamics of intergestural timing: A perturbation study of lip-larynx coordination,'' Exp. Brain Res. **123**, 412–424.

Sapir, S., and McClean, M. D. (**1981**). ''Evidence for acoustico-laryngeal and trigemino-laryngeal reflexes in man,'' American Speech-Language-Hearing Association, Los Angeles, CA.

Shaiman, S., and Gracco, V. L. (**2002**). ''Task-specific sensorimotor interactions in speech production,'' Exp. Brain Res. **146**, 411–418.

Smith, A., McFarland, D. H., Weber, C. M., and Moore, C. A. (**1987**). ''Spatial organization of human perioral reflexes,'' Exp. Neurol. **98**, 233–248.

Solls, M. M., Brainard, M. S., Hessler, N. A., and Doupe, A. J. (**2000**). ''Song selectivity and sensorimotor signals in vocal learning and production,'' Proc. Natl. Acad. Sci. U.S.A. **97**, 11836–11842.

Suthers, R. A., Goller, F., and Wild, J. M. (**2002**). ''Somatosensory feedback modulates the respiratory motor program of crystallized birdsong,'' Proc. Natl. Acad. Sci. U.S.A. **99**, 5680–5685.

Wikström, H., Huttunen, J., Korvenoja, A., Virtanen, J., Salonen, O., Aronen, H., and Ilmoniemi, R. J. (**1996**). ''Effects of interstimulus interval on somatosensory evoked magnetic fields (SEFs): A hypothesis concerning SEF generation at the primary sensorimotor cortex,'' Electroencephalogr. Clin. Neurophysiol. **100**, 479–487.

Acoustic analysis of monophthong and diphthong production in acquired severe to profound hearing loss

Sallyanne Palethorpe and Catherine I. Watson
Speech, Hearing and Language Research Centre, Department of Linguistics, Macquarie University, Sydney, Australia 2109 and Macquarie Centre for Cognitive Science, Macquarie University, Sydney, Australia 2109

Rosalind Barker
Speech, Hearing and Language Research Centre, Department of Linguistics, Macquarie University, Sydney, Australia 2109

(Received 23 September 2002; revised 31 May 2003; accepted 2 June 2003)

The effect of diminished auditory feedback on monophthong and diphthong production was examined in postlingually deafened Australian-English speaking adults. The participants were 4 female and 3 male speakers with severe to profound hearing loss, who were compared to 11 age- and accent-matched normally hearing speakers. The test materials were 5 repetitions of hVd words containing 18 vowels. Acoustic measures that were studied included $F1$, $F2$, discrete cosine transform coefficients (DCTs), and vowel duration information. The durational analyses revealed increased total vowel durations with a maintenance of the tense/lax vowel distinctions in the deafened speakers. The deafened speakers preserved a differentiated vowel space, although there were some gender-specific differences seen. For example, there was a retraction of $F2$ in the front vowels for the female speakers that did not occur in the males. However, all deafened speakers showed a close correspondence between the monophthong and diphthong formant movements that did occur. Gaussian classification highlighted vowel confusions resulting from changes in the deafened vowel space. The results support the view that postlingually deafened speakers maintain reasonably good speech intelligibility, in part by employing production strategies designed to bolster auditory feedback. [DOI: 10.1121/1.1593059]

PACS numbers: 43.70.Dn [AL]

I. INTRODUCTION

This study was designed to investigate the effects of severe to profound postlingual deafness on monophthong and diphthong vowel production. Postlingual deafness is a hearing deficit that has occurred after language has been firmly established, usually after age 5, and can also be referred to as adventitious or acquired deafness. Speakers who have postlingual deafness can be called deafened. When the hearing deficit has occurred early enough to have a major effect on language acquisition, the speakers are termed (congenitally) deaf. Numerous studies have investigated vowel production in deaf speakers (e.g., Bakkum *et al.*, 1993; Ertmer *et al.*, 1997; Fourakis *et al.*, 1993; McCaffrey and Sussman, 1994; Monsen, 1976). The majority of these researchers have studied vowels in children and adolescents and report reduced vowels spaces, overlap between vowel categories, a reduction in $F1-F2$ range, tense–lax substitution, diphthongization, and neutralization. However, there is a vast difference between investigating how vowel production has developed in the absence of hearing, and establishing what happens to production when hearing is lost or distorted over a prolonged period of time.

The investigation of vowel production in severe to profound deafness can assist in an understanding of the long-term role of auditory feedback in speech production. Monophthongs and diphthongs represent a valuable field of investigation as they rely less on tactile feedback than on motor-kinesthetic feedback that may be more vulnerable to the effects of long-term hearing deprivation. Waldstein (1990) noted that vowel production was affected by acquired deafness, stating that ''postlingual deafness affects the production of all classes of speech sounds, suggesting that auditory feedback is implicated in regulating the phonetic precision of consonants, [and] vowels'' (p. 2099). Although researchers are in agreement about the importance of auditory feedback in the acquisition of speech production skills (Tobey, 1993), there has been considerable debate in the literature regarding its role in speech maintenance. Clarification of the role of auditory feedback in speech maintenance is important for the construction of a comprehensive theory of the speech production system (Perkell *et al.*, 1992).

Studies examining the role of auditory feedback can be divided into several categories. The first investigates the effect of altering feedback in normally hearing adults, thus providing evidence regarding the short-term effects of feedback deprivation. For example, effects such as increased intensity, segment duration, and spectral changes have been shown to occur when speaking in increased background noise (Lane and Tranel, 1971; van Summers *et al.*, 1988). Compensatory production mechanisms have been shown to result from alterations in the auditory feedback frequencies (Natke and Kalverum, 2001; Houde and Jordan, 1998, 2002). Similar studies have been carried out in deaf and deafened speakers by observing the changes in speech production that occur as a consequence of removal of auditory aids for short

periods, or that occur pre- and postimplantation of cochlear implants (e.g., Economou *et al.*, 1992; Lane *et al.*, 2001; Kishon-Rabin *et al.*, 1999; Perkell *et al.*, 1992). By illustrating the long-term effects of feedback deprivation after skilled speech has been acquired, studies of the differential deterioration in the speech production skills of adults with profound acquired deafness (deafened speakers) ''can illuminate the underlying control mechanism of speech'' (Lane and Webster, 1991; p. 860). In addition, research into the effects of acquired profound deafness in early childhood and adolescence provides additional information about the developmental role of auditory feedback.

The results of investigations into the role of auditory feedback have led to a number of theories regarding the nature of the speech production system. Older theories advocated that speech production is a closed-loop system, relying on auditory feedback for continuous monitoring of output to prevent production errors (Fairbanks, 1954; Siegel and Pick, 1974). However, Perkell *et al.* (2000) argue that ''it is unlikely that auditory feedback is used for closed-loop error correction in the intra-segmental control of individual articulatory movements, because the feedback delay is too large'' (p. 238). Rather, speech motor control is predominantly an open-loop system, where auditory feedback, in conjunction with proprioceptive and visual feedback, enables the learning of a robust internal model that is hypothesized to be a neural representation of spatial, kinematic, tactile and/or proprioceptive features of movements used to predict a desired acoustic consequence. Once speech has been acquired there is less need for peripheral feedback, and the role of auditory feedback is more in tuning the settings of this robust internal model (Perkell *et al.*, 1992). However, recent studies of fundamental frequency perturbation have shown that auditory feedback can be used in a closed-loop fashion in conjunction with internal representations (Donath *et al.*, 2002; Larson *et al.*, 2000) leading to suggestions that the mechanisms involved in suprasegmental control may be more labile than those involved in segmental control (Lane *et al.*, 1997; Perkell *et al.*, 2000; Svirsky *et al.*, 1992).

The evidence for a robust mapping of vocal gestures to their acoustic consequences can be seen in maintenance of reasonably good intelligibility in the speech of people who have been deafened for a number of years (Lane and Webster, 1991). While discernible changes in suprasegmental features of deafened speech have been reported (Cowie and Douglas-Cowie, 1992; Leder and Spitzer, 1993), the segmental features appear to be less affected by the reduction of auditory feedback. However, alterations in vowel production with a reduction in auditory feedback have been shown to occur in acoustic studies of deafened speakers. The results show a pattern of distortion from normal production that varies between and within studies, often as a consequence of differences in auditory capability between speakers. Several studies have shown some reduction of vowel space either along the $F1$ and/or $F2$ dimensions (Economou *et al.*, 1992; Smyth *et al.*, 1991; Waldstein, 1990), although this effect is not always present (Tartter *et al.*, 1989; Lane *et al.*, 2001). For example, while Svirsky and Tobey (1991) reported no significant changes for the American English point vowels /i, a, u/, Langereis *et al.* (1999) found small changes to these vowels, and Barker (1995) and Cowie and Douglas-Cowie (1992) reported the centralization of /i/. Richardson *et al.* (1993) found that two of the five Australian English (AE) speakers they studied showed reduction in /ɪ/ and /ɛ/, while $F1$ change varied according to speaker. Langereis *et al.* (1999) and Waters (1986) reported that monophthongs were replaced with diphthongs, while Plant and Hammarberg (1983) found that diphthongs were reduced to monophthongs. Both Waldstein (1990) and Economou *et al.* (1992) noted that their deafened speakers maintained tense lax vowel distinctions, as did two of the three speakers in the study by Svirsky and Tobey (1991). Waldstein (1990) also showed that her deafened speakers showed increased within-speaker variability, particularly in the increased standard deviations for $F2$ of vowel means.

Vowel production has a complex dependency on audition: speakers need to use auditory feedback to achieve vowel contrasts that are perceptible by a listener. Little tactile feedback is available in vowel production, as, with the exception of high vowels (Fitzpatrick and Ni Chasaide, 2002), the tongue does not generally make direct contact with other articulators. However, other proprioceptive information is available: for example, cutaneous receptors in the vocal tract, sensors transmitting information about joint location and sensors in the muscles providing information on force and movement of articulators (Gracco, 1995). McCaffrey and Sussman (1994) felt that the important question was whether severely and profoundly deaf speakers can produce vowels with significant discriminability and with critical distances between formants for contrasting vowel height and place. Deafened speakers have diminished audibility despite the assistance of clarification (via hearing aids or cochlear implants). For speakers with significant losses, the fundamental frequency ($F0$) and vowel $F1$ are most likely to be audible as they are lower in frequency (less than 2 kHz) than $F2$ and $F3$. ''Thus, the resulting vowel inventory may achieve contrast in auditory space along fewer formant dimensions than those employed by speakers with normal hearing (McCaffrey and Sussman, 1994; p. 939).

The aim of this study is to look at vowel production in deafened speakers. In contrast to the majority of research on deafened speech that generally focuses on monophthongs, we shall be extending our acoustic studies to look in depth at diphthong production, on the premise that the moving formant tracks of diphthongs may be harder to produce in deafened speakers due to the absence of much tactile feedback. In addition, we shall look at male and female speakers separately given the influence of gender on formant frequency ranges coupled with the relatively restricted range of audible sounds in deafened speakers. The acoustic measures to be examined are vowel duration, target timing, formant position, and formant trajectories. In the discussion of our results we shall interpret our findings in the light of the various auditory feedback issues that have been raised in other research.

TABLE I. Deafened speaker characteristics.

Speaker number	Sex	Age	Accent type	Type of hearing loss	Severity (3-frequency average loss in better ear -dB)	Duration of deafness (years)	Hearing aid use at time of recording
1	M	71	General	Acquired—gradual loss	83	40	Bilateral
2	M	72	General	Acquired—gradual loss	80	32	Left ear
3	M	60	Broad	Acquired—gradual then viral infection	105+	18	None—awaiting cochlear implant
4	F	52	Upper General	Sudden onset (viral) then gradual deterioration	103	26	Bilateral
5	F	65	General	Acquired—gradual loss	85	28	Bilateral
6	F	61	General	Sudden onset (viral and drug reaction)	83	48	Bilateral
7	F	50	General	Acquired—gradual then sudden	90	4	Bilateral

II. METHOD

All participants were native Australian English speakers, who used speech as their main means of communication. Seven deafened adults (four females and three males) and 11 normally hearing (NH) adults (six females and five males) took part in the study. All of the seven deafened (DF) speakers were deafened postlingually. Duration of deafness ranged from 6 to 48 years. All of the deafened speakers were classified as severely to profoundly deaf (three-frequency pure-tone average minimum of 80 dB HL in better ear). Table I provides more detailed information on the deafened speakers. All deafened speakers used hearing aids during the task, except one male speaker who was awaiting a cochlear implant; however, this speaker's vowel space was not significantly different from that of the other male speakers. All NH speakers were aged from 32 to 68 years and satisfied our criteria for normal hearing (passing an unaided three-frequency pure-tone average of 25 dB HL or less bilaterally).

It was important to select NH speakers who matched the age range and accent characteristics of the deafened speaker. Australian English has traditionally been described as having three accent types (''broad,'' ''general,'' and ''cultivated'') based on the pronunciation of certain vowels (Bernard, 1970a; Harrington *et al.*, 1997). Although it is convenient to refer to these three varieties as separate entities, there is considerable phonetic overlap between them, and they are perhaps better described as socio-phonetic variation along a broadness continuum. In addition, accent changes over the last 30 years in Australian English have been documented (Cox, 1999), particularly in the monophthongs /æ/, /u/, /ɒ/, and /ɜ/ and the second targets of /oʊ/ and /aʊ/, and also seen in the monopthongisation of the diphthongs /ɪə/ and /eə/. Two trained phoneticians listened to the data from all speakers to determine accent type in order to appropriately match speakers between the deafened and normally hearing groups.

In addition, to ensure that the vowel space of the normally hearing speakers was representative of a larger population of Australian speakers, all the monophthong tokens were checked acoustically against appropriate tokens from the Australian National Database of Spoken English (ANDOSL) (Millar *et al.*, 1994). This database contains both citation form and sentence material from over 250 speakers of ages ranging from 18 to 45+, representing the Australian accent continuum from broad to cultivated. The tokens of the normally hearing speakers were found to fall within the 95%-confidence intervals of the means of the citation form vowel tokens from the ANDOSL database.

The speech material consisted of five repetitions of 18 monophthongs and diphthongs presented in a citation-form /hVd/ context (Table II). The IPA representation of the Australian English vowels is phonemic (Mitchell and Delbridge, 1965). The stimulus words were presented in a different randomized order for each of the five repetitions. All speakers were recorded in the Macquarie University Speech Pathology Clinic. The acoustic profile of these rooms conforms to the Australian Hearing Services (AHS) requirements for

TABLE II. Target words and IPA symbols.

Word	IPA	Word	IPA
HEED	/i/	WHO'D	/u/
HID	/ɪ/	HERD	/ɜ/
HEAD	/e/	HAYED	/aɪ/
HAD	/æ/	HIDE	/eɪ/
HARD	/a/	HOID	/ɔɪ/
HUD	/ʌ/	HOED	/oʊ/
HOD	/ɒ/	HOWED	/aʊ/
HOARD	/ɔ/	HEARED	/ɪə/
HOOD	/ʊ/	HARED	/eə/

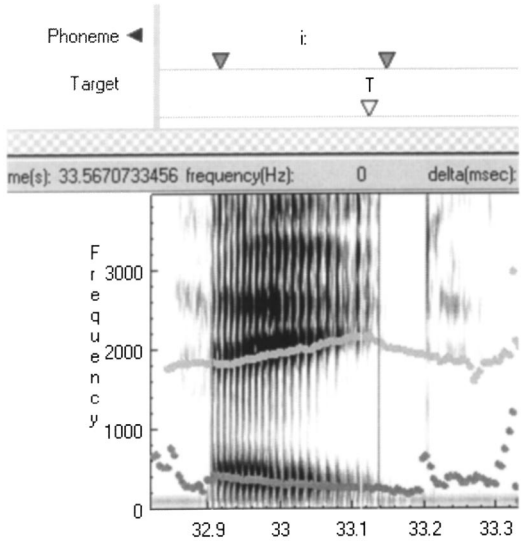


FIG. 1. The display window of the EMU labeler showing a spectrogram of the vowel /i/ with formant tracks superimposed.

sound sensitive spaces. The recordings were made onto digital audio tape (DAT) using a Sennheiser ME80 condenser microphone placed at approximately 45 cm from the speaker's face. The majority of the speakers also wore a Nasometer mask during the recording, as data were also being collected for another experiment; however, speech samples used from four of the male normally hearing speakers were recorded without a Nasometer mask. In order to test whether the mask compromised articulatory movement, in particular jaw movement, a comparison was made between the vowel formant data from two NH speakers (one male and one female) recorded both with and without the Nasometer mask. Statistical analysis showed no significant difference in the first two formant values at the targets of all relevant vowels.

The speech data were digitized at 20 kHz with a 16-bit resolution, and the first two formants and their bandwidths were automatically tracked using ESPS/WAVES (12th-order LPC analysis, cosine window, 49-ms frame size, 5-ms frame shift). All automatically tracked formants were checked for accuracy and, where the formant tracking was unreliable due to discontinuities, the formants were smoothed interactively using the spectrogram and DFT spectrum. All the labeling was done in EMU, a hierarchical speech data management system (Harrington and Cassidy, 1999). Using both the speech waveform and spectrogram as guides, the acoustic onset of the vowel was marked at the first complete pitch period and the offset marked at the closure for the /d/ (see Fig. 1). A single acoustic vowel target was marked in the monophthongs and two targets were marked in the diphthongs. The targets of the monophthongs and the first targets of the diphthongs were marked at a point where there was the least movement in the formant tracks. For the high vowels, this point occurred where $F2$ reached a peak (see Fig. 1 showing the target placement for the vowel /i/); for open vowels the target was marked where $F1$ was at a maximum; for back vowels the target was marked where $F2$ reached a trough. If none of the above criteria was satisfied for a given vowel, the target was marked at the point of maximum amplitude in the waveform. The second target of the rising and falling diphthongs was marked according to the same criteria as the first target. The second target of the centring diphthongs was marked at the right boundary of the vowel because a second target cannot be measured reliably.

The formant frequencies of the vowels were normalized using the Lobanov method (Lobanov, 1971). This was done to reduce speaker-specific aspects of the acoustic signal in order to examine more effectively the group characteristics of the deafened and normally hearing speakers, in particular, the size of the vowel space. In addition, as vowel classification studies will be carried out, normalization is useful in lowering classification errors by reducing intervowel variability while maintaining the appropriate distances between the vowel centroids (Disner, 1980). The normalization was carried out separately for each of the four populations of speakers (NH male/female and DF male/female). For each speaker, the Lobanov normalization of a vowel target formant frequency was carried out by subtracting the mean of the five tokens from each target point and dividing the result by the standard deviation, using the formula

$$F_i = (F_i - \overline{F_i})/\mathrm{SD}_i,$$

where F_i is a given formant, $\overline{F_i}$ is the mean of $\underline{F}$ across all vowels, and SD_i is the standard deviation of F_i about its mean for all vowels. In order to present the data graphically in Hz values comparable to unnormalized values, the data were rescaled using the mean and standard deviation (Disner, 1980; p. 260).

The time-varying nature of the formant tracks was modeled with the coefficients of the discrete cosine transform (DCT) (Zahorian and Jaghardi, 1993; for the mathematical modeling equations, see Watson and Harrington, 1999). Pilot vowel classification studies were carried out on the present vowel data to see which DCT coefficients should be used to encode the formant tracks. As in Watson and Harrington (1999), we found that only the first and second DCT coefficients played a significant role in separating the vowels: the first coefficient is proportional to the mean of the original formant track; the second models both the direction and tilt of the original formant track. Therefore, in the results reported herein, we have represented the first two formant tracks with the first two discrete cosine transformation (DCT) coefficients.

Statistical analysis was carried out in SPSS using two-way analyses of variance (ANOVA) with vowel type (VT) and hearing type (HT) as main factors using a p value of 0.05. Due to the inherent nature of the individual vowels studied, VT was always a significant main effect and will not be mentioned further except when there is a significant VT–HT interaction. *Post hoc* two-sample t-test analyses had a significance level of 0.01. In carrying out these statistical tests, we recognize that while there is generally thought to be less homogeneity in the speech parameters of deafened speakers, it is important to establish group patterns of behavior. In general, the group statistics reported reflect the performance of the individual members of the group; however, if statistical results were to be biased by a part of the group of speakers, it would be noted in the text.

III. RESULTS

A. Timing

For the monophthongs, analysis of variance shows a significant main effect for deafness in female speakers [$F(1,21)=48.23$, $p=0.000$] and, in male speakers, a significant two-way interaction between HT and VT [$F(10,21)=2.51$, $p=0.006$]: *post hoc t*-tests show a significantly longer duration for all DF monophthongs except /ɒ/, where the p value (0.013) is slightly above the 0.01 cutoff limit. The total mean durations (standard deviations) averaged over all vowels for female speakers are: NH: 262 (96) ms.; DF: 285 (101) ms.; those for male speakers are: NH: 221 (83) ms; DF: 278 (106) ms, bearing in mind that the duration for /ɒ/ is longer for DF compared to NH speakers, but not significantly so.

In addition, for all speakers the tense vowels are always longer than the lax vowels, and there is little difference in the lax/tense duration ratio between NH and DF speakers [female: NH(0.53), DF(0.52); male: NH(0.52), DF(0.52)]. This value is very similar to other tense/lax ratios of vowels in similar phonetic contexts found in Australian English: 0.56 in a study of male adolescents (Bernard, 1970b), and 0.53 (females) and 0.55 (males) from the ANDOSL database (Millar *et al.*, 1994). Therefore, the tense/lax distinction is preserved in the deafened speakers, as was found by Waldstein (1990). The preservation of the tense/lax distinction is important for phonemic contrast in Australian English. Australian English has two tense/lax pairs /a/ and /ʌ/, and /i/ and /ɪ/ (although the former can also be distinguished by an onglide) (Harrington *et al.*, 1997).

In order to examine whether there are differences between NH and DF speakers in the time to attain a monophthong vowel target, the more accurate onset-target timing information provided by the /i/ on-glide that is generally present in Australian English was examined. We calculated the ratio of the duration between the vowel onset and vowel target to the total vowel duration and then arc-sine transformed the result. However, there was no significant difference between DF and NH male and female speakers for this target-time duration.

In examining the total duration of the diphthongs, analysis of variance shows a significant main effect for deafness with no significant interaction [female: $F(1,13)=56.896$, $p=0.000$; male: $F(1,13)=79.601$, $p=0.000$]. The mean total duration averaged over all diphthongs is longer for the deafened speakers compared with the NH speakers for females [NH: 370 (56) ms.; DF: 410 (51) ms] and males [NH: 315 (56) ms.; DF: 377 (59) ms]. To investigate whether the DF speakers took longer to reach the second target, we looked at the arc-sined ratio of the time between the first and second targets to the total duration; however, no consistent significant differences between the NH and DF speakers are seen.

The gender difference noted here in total vowel duration, where female speakers generally have longer duration than male speakers, has been commented on previously (Hillenbrand *et al.*, 1995; Simpson, 2001, 2002). To test the significance of such a difference, analyses of variance were carried out on each population of NH and DF speaker separately. For monophthong total durations, there was no significant gender difference for the DF speakers and a significant gender/vowel interaction for the NH speakers [$F(10,583)=2.404$, $p=0.008$]: *post hoc t*-tests showed that the total duration for female speakers' monophthong productions was longer than that for males', but not significantly so for /ɒ/ and /ʊ/. The lack of a significant gender difference for the deafened speakers was due primarily to the increased duration of some of the tense monophthongs in one male speaker. The total duration for diphthongs was significantly longer for females than for males in both groups of speakers [NH: $F(1,371)=102.283$, $p=0.000$; DF: $F(1,231)=24.324$, $p=0.000$].

Waldstein (1990) observed an increase in the standard deviations for total duration and formant frequencies in her study of monophthongs in deafened speakers, but did not carry out any statistical analyses of these differences in variability. In order to see whether this increased variability was also found in the present data, we examined the variability in total duration values across all vowels for male and female DF and NH speakers using a Levene's test for equality of variance (Levene, 1960). Contrary to expectations, there are no consistently significant increases in total duration variability for the deafened speakers for either monophthongs or for diphthongs.

B. Formants

1. Monophthongs

Table III shows the means and standard deviations of the first two formant values for the male and female speakers. Figure 2(a) shows the $F1$ and $F2$ centroids at the vowel targets for each of the monophthongs from the female speakers: the individual deafened speakers are superimposed on the average for the NH speakers. For both formants in the female data there are significant two-way VT–HT interactions [formant 1: $F(10,21)=10.17$, $p=0.000$; formant 2: $F(10,21)=20.50$, $p=0.000$]. *Post hoc t*-tests show significantly lower formant values in the deafened speakers for the $F1$ of /a/ and /ʌ/ (there is a trend only ($p=0.017$) in /ɛ/ due to the lowered $F1$ in three speakers only). There is also significantly retracted $F2$ in /i/, /ɪ/, /ɛ/, /æ/, /u/, and /ɜ/. This retraction can be seen in Fig. 2(a), where the deafened vowel space is retracted in the $F2$ plane for the front vowels, and reduced in the $F1$ mainly in the open vowels. Figure 2(b) shows the data for the deafened male speakers and the averaged normally hearing speakers. Analysis of variance shows a significant VT–HT interaction for both [formant 1: $F(10,21)=5.49$, $p=0.000$; formant 2: $F(10,21)=2.91$, $p=0.002$]. *Post hoc t*-tests reveal a significant decrease in $F1$ of the vowels /a/, /ʌ/, and /ɒ/ and retraction in $F2$ of /æ/ for the deafened speakers: resulting in a reduction in vowel space due to a raising of the open vowels and retraction of /æ/ with no change in the other front vowels.

We investigated the possibility of an increase of variability around the vowel mean in the deafened speakers as might be expected if there was instability in vowel production. Three different methods were used: Levene's test for homogeneity of variance (Levene, 1960), coefficient of variation

TABLE III. Average monophthong $F1$ and $F2$ values in Hz (standard deviation) for female (left) and male (right) normally hearing (NH) and deafened (DF) speakers.

	Female				Male			
	Formant one		Formant two		Formant one		Formant two	
	NH	DF	NH	DF	NH	DF	NH	DF
HEED	314(41)	321(32)	2850(105)	2512(138)	288(25)	306(27)	2282(58)	2247(55)
HID	367(42)	377(29)	2777(92)	2417(133)	324(22)	332(22)	2163(46)	2158(67)
HEAD	476(64)	435(47)	2623(76)	2212(215)	409(32)	412(34)	2020(51)	1982(96)
HAD	789(112)	817(55)	2118(229)	1750(112)	667(59)	635(73)	1726(61)	1649(81)
HARD	1044(67)	869(70)	1533(85)	1465(91)	764(42)	711(53)	1287(67)	1301(63)
HUD	1015(96)	918(43)	1565(86)	1544(72)	770(51)	721(40)	1317(41)	1344(87)
HOD	742(84)	712(85)	1125(98)	1092(87)	633(56)	579(52)	991(46)	991(52)
HOARD	426(56)	414(47)	792(90)	758(66)	382(26)	401(42)	765(61)	808(43)
HOOD	383(68)	387(53)	869(90)	798(100)	352(25)	366(27)	922(129)	914(49)
WHO'D	351(42)	330(24)	1946(169)	1796(122)	310(23)	328(22)	1566(72)	1656(240)
HERD	454(47)	424(57)	1800(102)	1650(79)	433(34)	434(45)	1493(80)	1517(53)

(standard deviation/mean), and the ratio of within to between vowel variability per speaker. In general, there is no consistent increase in variability for either $F1$ or $F2$ in the deafened population, and in some cases the variability in the formants of the normally hearing speakers is greater.

Previous research has suggested a relationship between distortion of segmental parameters (including the vowel space) and age of deafness onset (Cowie *et al.*, 1982; Plant and Hammarberg, 1983; Waldstein, 1990). In order to test this, the degree of $F2$ retraction of the front vowels in the female deafened speakers was examined because this was the most significant change noted in the deafened speakers' vowel space. The distances from each token of the front vowels /i/, /ɪ/, /ɛ/, /æ/ from the deafened speakers was compared with the $F1/F2$ centroids of the corresponding vowels for the NH speakers using a Bayesian distance metric (Harrington and Cassidy, 1999; pp. 247–254). No consistent relationship between the calculated distances and duration of deafness was found in these particular deafened speakers.

2. Diphthongs

The diphthong formant track data are shown for both male and female speakers in Figs. 3(a) and (b). The figures depict the movement of the formant tracks from vowel onset to offset. Each formant track has been time normalized to 20 equidistant values; then, each of these 20 values is frequency normalized as described in Sec. II (Method). The normally hearing data are a solid line and each deafened speakers a dashed line. Although overall the $F1$ and $F2$ formant tracks for the NH and DF speakers are fairly similar, there are differences nonetheless. In particular, when the $F2$ tracks are above about 2 kHz in the NH speakers, they are noticeable lower in the DF speakers.

Table IV gives the means and standard deviations for $F1$ and $F2$ at the first and second target for the normally hearing and deafened female speakers and Table V show the equivalent information for the male speakers. Analysis of variance reveals that there is a significant VT–HT interaction for $F1$ and $F2$ of target one in the female speakers [$T1F1{:}F(6,13)=4.86$, $p=0.000$; $T1F2{:}F(6,13)=24.82$, $p=0.000$]. The results of *post hoc t*-tests indicate that for target one, $F1$ is significantly reduced for the deafened female speakers for /eɪ/, /aɪ/, /oɪ/, /oʊ/, and /aʊ/, and $F2$ is significantly retracted in /eɪ/, /oɪ/, /ɪə/, and /eə/. There is also a significant VT–HT interaction for the first and second formants of target one in the male data [$T1F1{:}F(6,13)=3.11$, $p=0.006$; $T1F2{:}F(6,13)=4.57$, $p=0.000$]. *Post hoc t*-tests show that $F1$ at target one is significantly reduced for /aɪ/, and $F2$ is significantly reduced for /eɪ/ and /aɪ/.

There is a significant VT–HT interaction for $F2$ of target two in the female data [$F(6,13)=15.56$, $p=0.000$]: *post hoc t*-tests show that all words spoken by the deafened group except for /aʊ/ have a significantly lower $F2$ than the normally hearing group. There is a similar significant interaction for the second target $F2$ in the male data [$F(6,13)=3.02$, $p=0.007$], and *post hoc t*-tests reveal that only for /eɪ/ and /oɪ/ is there a significantly lower $F2$ in the deafened compared with the normally hearing group.

Figures 3(a) and (b) also suggest that for the NH and DF there may be differences between the slope of the formant track between T1 and T2. In order to investigate any differences between the deafened and normally hearing speakers in the dynamic nature of the diphthongs, we calculated the DCT coefficients of the formant tracks. We modeled the formant track slope with the second coefficient of the DCT as it models both the direction and tilt of the original trajectory. The second coefficient indicated that the slopes of the formant trajectories for the NH and DF speakers were in the same direction, as can be seen in Figs. 3(a) and (b): however, there were some differences in the degree of slope. Analyses of variance were carried out on values of DCTs of the $F1$ and $F2$ formant tracks for each of the diphthongs for the NH and DF speakers. In the male speakers there is no significant difference in the slope of the formant tracks between the DF and NH speakers. In the female speakers there are significant interactions both for $F1$ and $F2$ [$F1{:}F(6,13)=2.84$, $p=0.010$; $F2{:}F(6,13)=6.99$, $p=0.000$]. *Post hoc* analyses show that in the DF speakers there is significantly less slope for the $F1$ of /eɪ/, /oɪ/, /oʊ/, and /aʊ/ due to a lower $F1$ at the first target; there is also a significantly reduced slope for the $F2$ of /eɪ/, /oɪ/, /oʊ/, with a similar trend for /aɪ/ and /aʊ/ ($p=0.013$), caused by a retraction of $F2$ at the second tar-

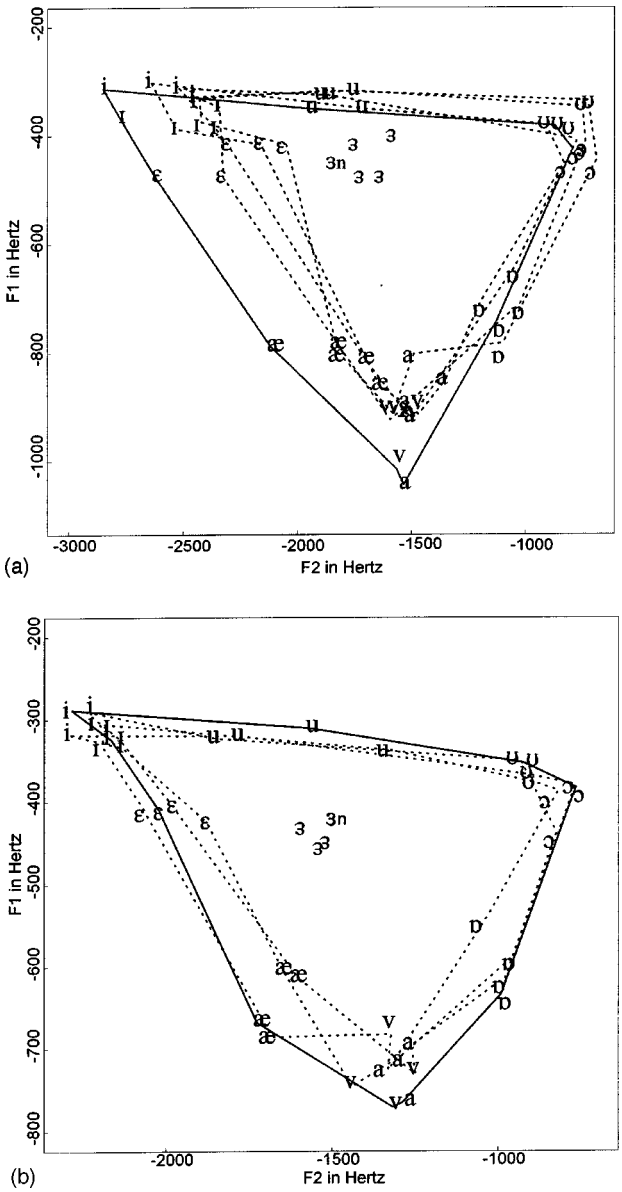


FIG. 2. (a) Female $F1-F2$ vowel space in Hz for the individual deafened speakers (dashed line) superimposed on the average for the normally hearing speakers (solid line). The label "n" for /ɜ/ indicates normally hearing speaker mean. (b) Male $F1-F2$ vowel space in Hz for the individual deafened speakers (dashed line) superimposed on the average for the normally hearing speakers (solid line). The label "n" for /ɜ/ indicates normally hearing speaker mean.

get. For the centering diphthong /ɪə/, *post hoc* analysis shows that there is a significantly greater $F2$ slope.

The above results are not unexpected. Since the monophthongal space of the DF speakers is reduced compared to the NH speakers, we would also expect the absolute formant trajectory difference between T1 and T2 to be reduced. The increased $F2$ slope in /ɪə/ is also compatible with this viewpoint in that the result is a more centralized second target. Other studies on normally hearing speakers have suggested the monophthong vowel space is strongly linked to the position of the first and second vowel targets for the diphthongs, a vowel shift in the monophthongs space results in a vowel shift in the diphthongs (Cox, 1996; Cox and Palethorpe, 2001).

The strong link between the monophthongs and diphthongs for the NH and DF speakers can be observed in Figs. 4 and 5. These figures show the movement of the time- and frequency-normalized formant tracks from T1 to T2 in the $F1/F2$ plane, henceforth termed the formant trajectory. The trajectories for the deafened speakers (dashed line) and normally hearing speakers (solid line) have been superimposed on the outline of the normalized monophthongal vowel space of the NH speakers and the DF speakers. The results for the female speakers are shown in Fig. 4(a) (the vowels /eɪ/, /aɪ/, and /oɪ/) and Fig. 4(b) (the vowels /oʊ/, /aʊ/, /ɪə/, and /eə/). Figures 5(a) and (b) present the equivalent information for the male speakers. It can be seen that for both the female and male speakers, the difference in the monophthongal vowel space for both normal-hearing and deafened speakers is also reflected in the altered trajectories of the diphthongs from the two speaker groups. For the female speakers there is both a retraction of the targets in the front region of the vowel space and a reduction for targets in the open region. However, for the male speakers the change is mainly in the reduction of the $F1$ at target one in the more open vowels.

These results reflect an inter-relationship between the realizations of monophthongs and diphthongs within the vowel space. The position of the first targets for all the diphthongs relative to the closest monophthong is very similar regardless of whether it is the deafened or normally hearing data (e.g., the first target of /ɪə/ with /i/ and the first target of /aɪ/ with /a/); this relationship can also be seen in the second targets. These results are also reflected in the statistical analysis. In the front region of the vowel space there are far more significant differences between the female deafened and normally hearing speakers than for the male speakers in terms of $F2$ values for either vowel target.

C. Classification experiments

In the first part of this study we established ways in which the deafened speech differed acoustically from the speech of normally hearing. There were differences in duration, and formant values at the vowel target, and consequently in the curvature of the formant tracks. The next question is whether these differences are sufficient to cause problems in perceiving the correct vowels.

Zahorian and Jagharghi (1993) found there is a high correlation between vowel confusions using formant-based automatic classifiers and perceptual experiments. In addition, other acoustic classification studies have found that vowels of different quality are more effectively distinguished when the acoustic parameters are based on spectral information sampled at multiple time points rather than from just at the vowel target (Harrington and Cassidy, 1994; Hillenbrand *et al.*, 1995; Watson and Harrington, 1999).

In this final experiment to investigate the types of vowel confusions that potentially may be made on the deafened data by a normally hearing speaker we performed an automatic Gaussian classification based on modeling the formant tracks of $F1$ and $F2$ with the coefficients of the discrete cosine transform (DCT), as described in the Methods section. Vowel duration was also used as a classification parameter because it is known to be an essential feature in the

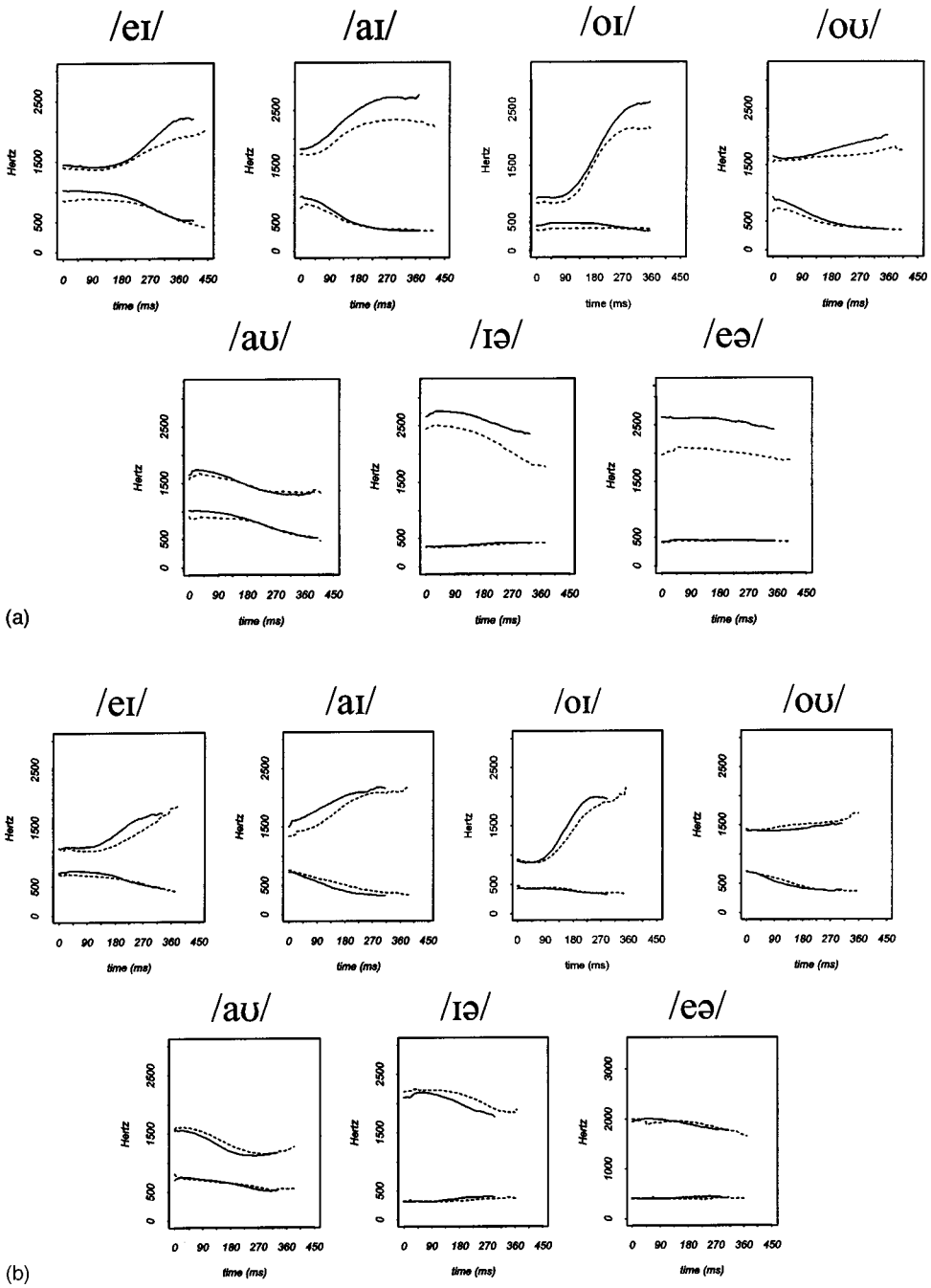


FIG. 3. (a) Female $F1$ and $F2$ tracks time aligned to vowel target one for normally hearing (solid line) and deafened (dashed line) speakers. (b) Male $F1$ and $F2$ tracks time aligned to vowel target one for normally hearing (solid line) and deafened (dashed line) speakers.

separation of Australian English vowels (Cox, 1996; Watson and Harrington, 1999). In the Gaussian classification, the centroid and covariance matrix of the training set (the tokens of the normally hearing speakers) were estimated for each vowel class. Tokens from the test set (those of the deafened speakers) were then classified based on their Bayesian distances to each of the training class centroids. A ''round-robin'' procedure was used whereby the tokens for each deafened speaker were used in turn as a test set. This resulted in a classification score for each deafened speaker, and the overall classification score was a summation of the individual results.

Vowel classification was carried out on the monophthongs and diphthongs together. The purpose of using the dynamic information contained in the formant track for classification of monophthongs is that, among other studies, a previous study of Australian English (Watson and Harrington, 1999) showed a differential between lax and tense monophthongs in the timing from vowel onset to target. This suggests that the shape of the formant contour may be different in these pairs of vowels and that this difference can be encoded by the DCT information and used to separate the vowel pairs in classification studies. This is relevant in the case of the Australian high vowels, particularly /i/, which in many speakers is realized with a delayed target (see Fig. 1). In addition, for many speakers of Australian English the

TABLE IV. Average diphthong $F1$ and $F2$ for target one (left) and target two (right) in Hz (standard deviation) for normally hearing (NH) and deafened (DF) female speakers.

	Target one				Target two			
	Formant one		Formant two		Formant one		Formant two	
	NH	DF	NH	DF	NH	DF	NH	DF
HAYED	929(86)	829(68)	1837(135)	1710(85)	376(34)	365(52)	2739(89)	2351(238)
HIDE	1016(71)	899(76)	1422(111)	1384(131)	459(59)	456(66)	2336(140)	2063(242)
HOID	488(62)	376(69)	934(80)	843(107)	372(32)	390(61)	2592(122)	2187(112)
HOED	867(115)	727(85)	1615(121)	1577(84)	373(39)	342(87)	1987(173)	1807(108)
HOWED	1021(97)	912(66)	1733(121)	1660(128)	603(81)	664(140)	1236(146)	1223(163)
HEARED	352(70)	333(46)	2754(98)	2496(113)	389(44)	389(78)	2320(160)	1849(162)
HARED	457(44)	438(65)	2614(75)	2098(242)	408(49)	381(62)	2344(101)	1845(160)

diphthongs /ɪə/ and /eə/ are realized as monophthongonized variants when preceding /d/ (Cox, 1999).

The confusion matrix from the female data is given in Table VI, and that for the male data is given in Table VII. For the deafened female data overall 70% were correctly classified, and for the male data 71.5% were correctly classified. Watson and Harrington (1999) performed a similar vowel classification of monophthongs and diphthongs together using only normally hearing speakers from the ANDOSL database, speakers who were different from the normally hearing speakers in the present research. They found 90% of the female vowels and 94.3% of the male vowels were correctly identified. These values are clearly a lot higher than those from the deafened speakers, suggesting that normally hearing speakers might have more trouble differentiating certain vowels spoken by a deafened speaker compared with those spoken by a normally hearing speaker.

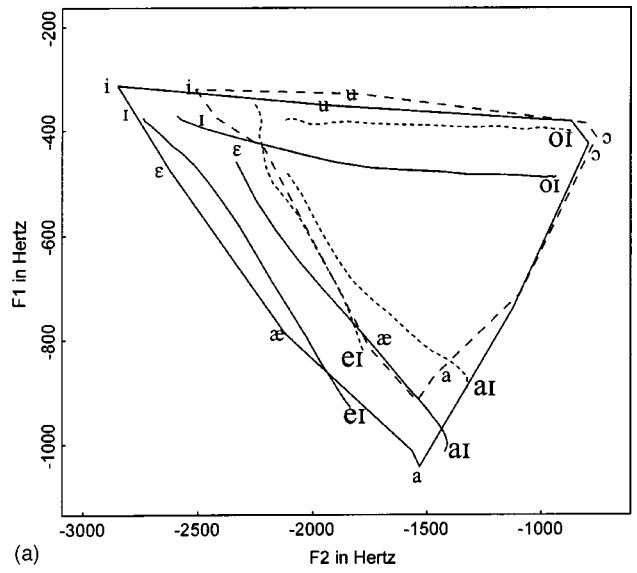
The confusions occurred generally either as a result of increased vowel duration or a retracted vowel space. For the females there is some confusion between the high front vowels due to a combination of retracted $F2$ and slightly lowered $F1$. Also, both /aɪ/ and /eɪ/ were misclassified as /oʊ/; this result is not unexpected in that there would be a tendency in reduced vowel space in these female deafened speakers for these more peripheral diphthongs to be identified as the more central /oʊ/. In the male speakers, no /ʌ/ vowel was correctly classified; it was confused with /a/, /æ/, and /ɒ/: in Australian English /ʌ/ and /a/ are distinguished by length so an increased duration for /ʌ/ will produce more confusion with /a/; the confusion with the other two vowels was due presumably to a lower $F1$ in /ʌ/. There was some confusion between /ɪə/ and some monophthongs. This is not an unexpected result given the changing nature of /ɪə/ in Australian English, where it can be heard in some speakers as a lengthened monophthong (Cox, 1999). However, most incorrectly classified deafened monophthongs were classified as another monophthong, and most incorrectly classified diphthongs were classified as another diphthong.

When we separated the data from the deafened speakers into either a diphthong or monophthong group and repeated the classification experiment, we got 93.6% accuracy in the separation of diphthongs and monophthongs for the deafened female speakers and 90% accuracy for the males. In their study using normally hearing speakers from the ANDOSL database, Watson and Harrington (1999) also separated the vowels into monophthong and diphthong groups, and carried out a vowel classification using these two parameters. They found a similar high accuracy of separation between monophthongs and diphthongs of 94.3% and 93.2% for female and male data, respectively.

In order to further clarify the effect of the retraction of $F2$ on the classification scores of the deafened female speakers, the vowel classification was repeated for both males and females using just the monophthong data, separated into two vowel sets: A front set containing the vowels /i/, /ɪ/, /ɛ/, and /æ/, and a central/back set containing the vowels /a/, /ʌ/, /ɒ/, /ɔ/, /ʊ/, /u/, and /ɜ/. The correct classification scores for the deafened male speakers were 90% for the front and 80% for the back sets. In contrast, the correct score for the deafened female speakers was 87% for the back set, and only 59% for the front set of vowels.

TABLE V. Average diphthong $F1$ and $F2$ for target one (left) and target two (right) in Hz (standard deviation) for normally hearing (NH) and deafened (DF) male speakers.

	Target one				Target two			
	Formant one		Formant two		Formant one		Formant two	
	NH	DF	NH	DF	NH	DF	NH	DF
HAYED	715(83)	696(62)	1604(155)	1448(174)	337(27)	374(51)	2160(68)	2178(162)
HIDE	769(45)	688(65)	1157(49)	1102(50)	441(50)	445(52)	1790(87)	1718(108)
HOID	422(39)	441(54)	896(44)	874(71)	340(16)	339(30)	2025(84)	1922(91)
HOED	677(53)	668(87)	1405(55)	1414(69)	367(45)	387(27)	1513(121)	1568(136)
HOWED	749(59)	718(67)	1561(110)	1598(141)	571(66)	566(54)	1051(72)	1042(98)
HEARED	303(29)	308(28)	2199(52)	2229(64)	571(66)	345(49)	1730(103)	1754(101)
HARED	406(43)	395(127)	2015(56)	1949(181)	378(53)	384(44)	1740(78)	1691(113)



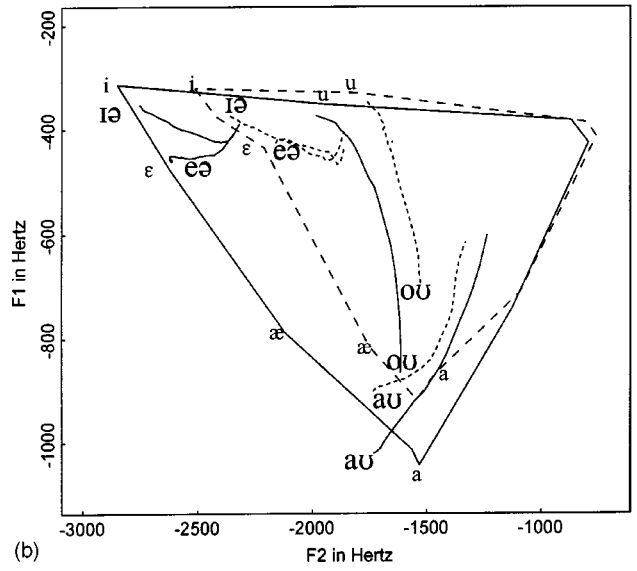
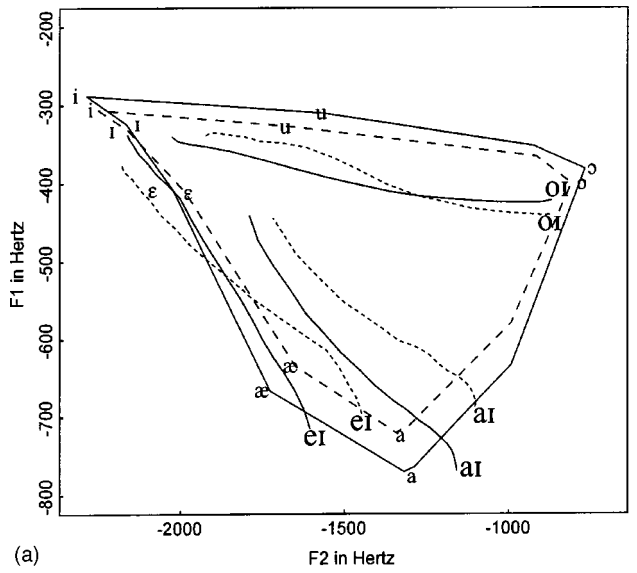


FIG. 4. (a) Female diphthong trajectories of vowels /eɪ/, /aɪ/, /ɔɪ/ for normally hearing (dashed line) and deafened (solid line) speakers superimposed on monophthong NH and DF vowel spaces. The trajectory is drawn from T1 to T2 with the diphthong label placed near T1. (b) Female diphthong trajectories of vowels /oʊ/, /aʊ/, /ɪə/, /eə/ for normally hearing (dashed line) and deafened (solid line) speakers superimposed on monophthong NH and DF vowel spaces. The trajectory is drawn from T1 to T2 with the diphthong label placed near T1.



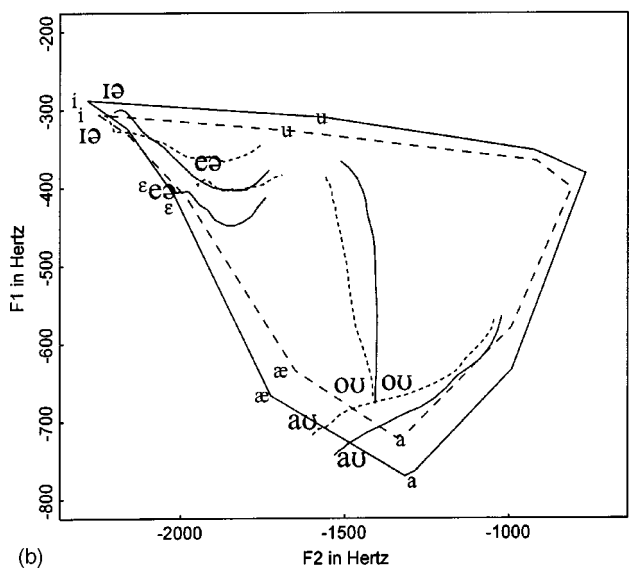


FIG. 5. (a) Male diphthong trajectories of vowels /eɪ/, /aɪ/, /ɔɪ/ for normally hearing (dashed line) and deafened (solid line) speakers superimposed on monophthong NH and DF vowel spaces. The trajectory is drawn from T1 to T2 with the diphthong label placed near T1. (b) Male diphthong trajectories of vowels /oʊ/, /aʊ/, /ɪə/, /eə/ for normally hearing (dashed line) and deafened (solid line) speakers superimposed on monophthong NH and DF vowel spaces. The trajectory is drawn from T1 to T2 with the diphthong label placed near T1.

IV. DISCUSSION

The results of this study show that the deafened speakers generally preserve a differentiated vowel space for Australian English. However, there are some patterns in the vowel changes that do occur that are gender specific. The deafened females demonstrate a vowel space retraction in the $F2$ plane for the front monophthongs and a reduction in $F1$ for the open monophthongs. By contrast, the deafened males demonstrate a lowering of $F1$ for the open monophthongs and a retraction of /æ/, but no change for the other front monophthongs. The average values of $F2$ for the front vowels (/i/, /ɪ/, /ɛ/, and /æ/) for the normally hearing female speakers were all much greater than 2 kHz, whereas for the normally hearing male only /i/ had an $F2$ greater than 2 kHz. Since the deafened speakers have some hearing loss over 2 kHz, the differences in the vowel spaces between the female and male deafened speakers suggest that the pattern of front vowel retraction in the female speakers may be due to an attempt to bring these vowels into a more perceptible range. This change is less necessary for the deafened males, as the $F2$ of front vowels for males is generally lower than that of female speakers. A similar pattern of reduction in vowel space can be seen in the data of Waldstein (1990). She comments that her deafened speakers had a reduced $F1$ and $F2$ range, and it can be seen in her data that the reduction for $F2$ occurs mainly in the front vowels where the frequency value tends to be reduced in the male and female deafened speakers to approximately 2 kHz. The deafened female speaker in

TABLE VI. Classification experiment confusion matrix for female speakers. Diagonal shows numbers correct out of 20.

	i	ɪ	e	æ	a	ʌ	ɒ	ɔ	ʊ	u	ɜ	eɪ	aɪ	ɔɪ	oʊ	aʊ	ɪə	eə
i	**6**		1							7	1						2	3
ɪ		**4**	12	2														2
e		1	**8**	6						2	2							1
æ				**17**		3												
a					**15**						1					4		
ʌ				4		**15**	1											
ɒ							**20**											
ɔ								**20**										
ʊ								2	**18**									
u										**8**	10				2			
ɜ										4	**15**				1			
eɪ												**12**			8			
aɪ													**14**		5	1		
ɔɪ														**20**				
oʊ											1				**19**			
aʊ															1	**19**		
ɪə															2		**15**	3
eə										2	5				3		3	**7**

a study by Plant and Hammarberg (1983) also had a reduced $F2$ range where the reduction was mainly in the front vowels to about 2 kHz. It is also interesting to note that while changes in the point vowels /i/ and /a/ have been seen in our deafened data and that of Waldstein (1990), the high back vowel /ɔ/ in our data, and the /u/ in Waldstein's American English deafened data remain relatively unchanged. Therefore, it may not be all the point vowels that remain stable under changes auditory conditions, as suggested by Svirsky and Tobey (1991), but the high back vowels that are more stable due to their stronger orosensory attribution.

The deafened speakers' production of the diphthongs showed the same relationship to monophthongs as for the normally hearing speakers. The position in $F1/F2$ space of the first target for all the diphthongs relative to the nearest monophthong is similar regardless of hearing type (cf. /a/

TABLE VII. Classification experiment confusion matrix for male speakers. Diagonal shows numbers correct out of 15.

	i	ɪ	e	æ	a	ʌ	ɒ	ɔ	ʊ	u	ɜ	eɪ	aɪ	ɔɪ	oʊ	aʊ	ɪə	eə
i	**11**																4	
ɪ	1	**11**											1					2
e			**8**										2				2	3
æ				**15**														
a					**14**											1		
ʌ				5	5		4									1		
ɒ					4		**11**											
ɔ								**15**										
ʊ					1			1	**13**									
u										**4**					3		2	6
ɜ										1	**8**				6			
eɪ												**10**	3		2			
aɪ												1	**14**					
ɔɪ	1												7	**7**				
oʊ											2				**13**			
aʊ															1	**14**		
ɪə	2																**13**	
eə											**1**	1					1	**12**

compared to the first target of /aɪ/ in Figs. 4 and 5). The changes in the deafened monophthong space were also observed in the diphthong space. The most noticeable was a retracted $F2$ for both the front monophthongs and the diphthong targets which fell in the front region of the vowel space of the deafened females. These results suggest the deafened speakers have a built-in model of the vowel space, and that their ability to relate diphthongs to monophthongs is not lost despite a reduction in auditory feedback. Further, that for deafened speakers, a loss of formant movement in diphthongs is not necessarily a tendency towards steady state, as suggested by Plant and Hammarberg (1983), but rather it maybe a consequence of the smaller difference between the first and second target due to a reduced monophthong space.

We found no consistent relationship between the degree of retraction of the front vowels and degree or duration of deafness in the deafened female speakers, although this has been reported elsewhere (Cowie *et al.*, 1982; Plant and Hammarberg, 1983; Waldstein, 1990). Much research in this area has shown idiosyncratic speaker responses to a reduction of auditory feedback and this may be the case here. For example, the female speaker who had been deafened for the longest time had a position involving some public speaking and this may have played a role in the maintenance of her speech intelligibility.

We found no reliable increases in the formant frequency variability around the vowel mean in the deafened speakers compared to normally hearing speakers, contrary to other studies (Langereis *et al.*, 1997; Waldstein, 1990). This result may be due to more careful speech on the part of the deafened subjects, particularly as the total durations for the vowels of the deafened speakers showed no greater increase in variability compared with those of the normally hearing speakers. In addition, the consistency in the pattern of vowel space change in our deafened speakers is unusual when compared with other studies on deafened speech. However, most of these studies were concerned with the pre- and postcochlear implantation effects on the vowel formant space. Their results have shown that speakers with a disordered vowel space respond differently to the reintroduction of hearing (Economou *et al.*, 1992; Kishon-Rabin *et al.*, 1999; Perkell *et al.*, 1992; Lane *et al.*, 2001), although this variability may be, in part, a consequence of differing acoustic cues presented to the speakers by the particular speech processors used.

The variability in the mean formant frequencies of some vowels between the NH speakers merits further comment. Languages do undergo vowel change: in particular, Australian English has been known to have undergone a vowel change over the last 30 years (Cox, 1999). For example, note the large standard deviation in the $F1$ and $F2$ of /æ/ for the normally hearing female speakers seen in Table III; this vowel is unstable in Australian English as a consequence both of accent difference along the broad-general-cultivated continuum and of an observed change in the /æ/ vowel, which has become a more open retracted vowel in the last 30 years (Cox, 1999). The /u/ vowel is also unstable in Australian English and the fronting of this vowel in recent years has been noted in Australian as well as other English dialects (e.g., Cox and Palethorpe, 2001; Harrington *et al.*, 2000; Watson *et al.*, 1998). Therefore, it could well be that the increased $F2$ range for this vowel in the deafened male speakers may be a consequence of vowel shift rather than of hearing impairment, with two speakers showing fronting and one a retraction compared with the normally hearing speakers. These results highlight the need to have an accurate picture of the speech of a normally hearing population that has been recorded at a similar time to that of the target population in order to remove, as far as possible, the confounding effects of accent change.

Two aspects of timing in vowel production for deafened speakers were investigated in this study; total duration and the timing of the vowel target. We found that durations for both monophthongs and diphthongs for the deafened speakers were significantly longer than for the normally hearing, with the exception of /ɒ/ in the male speakers. This increased duration is likely to reflect a more careful and precise vowel production in the deafened speakers. It may be an attempt to get more orosensory feedback or to enhance intelligibility, as increased duration has been shown by many studies to occur in clear speech that is more intelligible (see Hargus-Ferguson and Kewley-Port, 2002, for a summary). Also, the deafened speakers do preserve the duration distinctions between tense and lax vowels, as was also found by Economou *et al.* (1992) and Waldstein (1990). There were no significant differences in the time taken from the first to the second targets in diphthongs. However, since the formant space of the diphthong is reduced, the result may be in fact a slower transition between targets since the formant distance between the first and second targets will be less. In the present data, the gender differences in duration were compatible with gender differences found in normally hearing data (Simpson, 2001, 2002).

The classification experiment suggested that although the vowel quadrilateral remains intact, the reduced deafened vowel space would have an effect on vowel intelligibility in these speakers. The confusions mainly occur between neighboring vowels and can generally be predicted by the change in vowel space. However, unlike the intelligibility results in Plant (1984), the deafened speakers do preserve the monophthong/diphthong distinction. Further research is necessary to see if a listener panel can confirm these results.

Many of our monophthong findings are very similar to that of Waldstein (1990). In that study she argues that the errors produced by the deafened speakers are consistent with the claim that the speech production system is under some form of closed-loop feedback control. However, Perkell *et al.* (2000), among others, suggest that due to a long feedback delay, it is unlikely the auditory signal provides immediate feedback for the control of segmental features in speech.

Perkell *et al.* (2000) have developed a speech motor control theory that states that speech movements are programed to achieve auditory/acoustic goals. The programming of articulatory movements to achieve auditory goals uses an internal mapping of the relationship between articulatory profiles and their acoustic consequences. The auditory/

acoustic targets of vowels, in particular, are described as depending on the consequences of economy of effort and sufficient perceptual contrast (cf. Lindblom and Engstrand, 1989). Once the internal model is acquired, it is maintained through auditory feedback. However, it is not closed-loop auditory feedback, for there is no need for continuous direct auditory feedback with an internal model. With the maturation of the speaker, the model becomes robust and peripheral feedback is only used intermittently. As a consequence, they propose, deafened speakers should have reasonably good intelligibility many years after the onset of profound deafness, in spite of a reduction in auditory feedback.

The findings in this study would also lend some support to this speech motor control model. Overall, the deafened speakers maintain relatively differentiated vowels in spite of some changes in the vowel space, together with a clear separation in the classification of monophthongs and diphthongs, and a tense/lax distinction between vowels. They are able to achieve this distinction after a long time with reduced and/or distorted auditory feedback, suggesting they have a reasonably robust internal mapping between articulatory configurations and their acoustic consequences. Perkell *et al.* (2000) argue that auditory feedback is used to refine the internal model parameters. It is feasible then that a reduction in auditory feedback may result in some adaptation of the acoustic/auditory goals of the internal model to bolster the effectiveness of the remaining auditory feedback.

Given the interesting findings in the current study, we are continuing research in this area of deafened speech. The relatively small sample of deafened speakers needs to be expanded to include speakers with a wider degree and duration of deafness. We are interested in studying segmental changes in deafened speakers on a long-term basis, paying particular attention to inter- and intraspeaker variability and using both acoustic and articulatory measures. We also acknowledge the need to document more audiological information about the speakers. Information on hearing loss above 2 kHz would seem to be especially important, given the present results. This research topic is relevant as it not only can give descriptions of segmental distortions in postlingual deafness that can be related to possible audiological remediation, but also for the opportunity it provides for studying control strategies in speech production.

ACKNOWLEDGMENTS

This work was supported in part by a Macquarie University Research Grant. Our thanks to Felicity Cox for the use of some acoustic data from normal-hearing speakers collected by herself and Sallyanne Palethorpe in the course of another research project. We are grateful to Libby Harricks for her support of this project. We also thank Jonathan Harrington, Anders Löfqvist, and two anonymous reviewers for their helpful comments.

Bakkum, M. J., Plomp, R., and Pols, A. C. W. (**1993**). ''Objective analysis versus subjective assessment of vowels pronounced by native, non-native, and deaf male speakers of Dutch,'' J. Acoust. Soc. Am. **94**, 1989–2004.

Barker, R. (**1995**). ''Qualitative and quantitative analysis of profoundly deafened speech hearing aid versus cochlear implant users,'' Unpublished Ph.D. thesis, Macquarie University.

Bernard, J. R. L. (**1970a**). ''Toward the acoustic specification of Australian English,'' Zeitschrift für Phonetik **2/3**, 113–128.

Bernard, J. R. L. (**1970b**). ''On nucleus component durations,'' Lang. Speech **13**, 89–101.

Cowie, R., and Douglas-Cowie, E. (**1992**). *Postlingually Acquired Deafness* (Mouton de Gruyter, Berlin).

Cowie, R., Douglas-Cowie, E., and Kerr, A. (**1982**). ''A study of speech deterioration in postlingually deafened adults,'' J. Laryngol. Otol. **96**, 101–112.

Cox, F. (**1996**). ''An Acoustic Study of Vowel Variation in Australian English,'' unpublished Ph.D. thesis, Macquarie University.

Cox, F. (**1999**). ''Vowel change in Australian English,'' Phonetica **56**, 1–27.

Cox, F., and Palethorpe, S. (**2001**). ''Vowel change: Synchronic and diachronic evidence,'' in *English in Australia*, edited by D. Blair and P. Collins (John Benjamins, Amsterdam), pp. 17–45.

Disner, S. F. (**1980**). ''Evaluation of vowel normalization procedures,'' J. Acoust. Soc. Am. **67**, 253–261.

Donath, T. M., Natke, U., and Kalveram, K. T. (**2002**). ''Effects of frequency-shifter auditory feedback on voice F_0 contours in syllables,'' J. Acoust. Soc. Am. **111**, 357–366.

Economou, A., Tartter, V., Chute, P. M., and Hellman, S. A. (**1992**). ''Speech changes following reimplantation from a single-channel to a multichannel cochlear implant,'' J. Acoust. Soc. Am. **92**, 1310–1323.

Ertmer, D. J., Kirk, K. I., Sehgal, S. T., Riley, A. I., and Osberger, M. J. (**1997**). ''A comparison of vowel production by children with multichannel cochlear implants or tactile aids: Perceptual evidence,'' Ear Hear. **18**, 307–315.

Fairbanks, G. (**1954**). ''Systematic research in experimental phonetics. I. A theory of the speech mechanism as a servosystem,'' J. Speech Hear. Disord. **19**, 133–139.

Fitzpatick, L., and Ni Chasaide, A. (**2002**). ''Estimating lingual constriction location in high vowels: A comparison of EMA- and EPG-based measures,'' J. Phonetics **30**, 397–415.

Fourakis, M., Geers, A., and Tobey, E. A. (**1993**). ''An acoustic metric for assessing change in vowel production by profoundly hearing-impaired children,'' J. Speech Hear. Disord. **94**, 2544–2552.

Gracco, V. (**1995**). ''Central and peripheral components in the control of speech movements,'' in *Producing Speech: Contemporary Issues*, edited by F. Bell-Berti and L. J. Raphael (American Institute of Physics, New York), pp. 418–432.

Hargus-Ferguson, S., and Kewley-Port, D. (**2002**). ''Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **112**, 259–271.

Harrington, J., and Cassidy, S. (**1994**). ''Dynamic and target theories of vowel classification: Evidence from monophthongs and diphthongs in Australian English,'' Lang. Speech **37**, 357–373.

Harrington, J., and Cassidy, S. (**1999**). *Techniques in Speech Acoustics* (Kluwer Academic, Dordrecht).

Harrington, J., Cox, F., and Evans, Z. (**1997**). ''An acoustic phonetic study of broad, general and cultivated Australian English vowels,'' Aust. J. Ling. **17**, 155–184.

Harrington, J., Palethorpe, S., and Watson, C. I. (**2000**). ''Monophthongal vowel changes in received pronunciation: A acoustic analysis of the Queen's Christmas Broadcasts,'' J. Int. Phonetic Assoc. **30**, 63–78.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.

Houde, J. F., and Jordan, M. I. (**1998**). ''Sensorimotor adaptation in speech production,'' Science **279**, 1213–1216.

Houde, J. F., and Jordan, M. I. (**2002**). ''Sensorimotor adaptation of speech. I. Compensation and adaptation,'' J. Speech Lang. Hear. Res. **45**, 295–311.

Kishon-Rabin, L., Taitelbaum, R., Tobin, Y., and Hildesheimer, M. (**1999**). ''The effect of partially restored hearing on speech production of postlingually deafened adults with multichannel cochlear implants,'' J. Acoust. Soc. Am. **106**, 2843–2957.

Lane, H., Matthies, M., Perkell, J., Vick, J., and Zandipour, M. (**2001**). ''The effects of changes in hearing status in cochlear implant users on the acoustic vowel space and cv coarticulation,'' J. Speech Lang. Hear. Res. **44**, 552–563.

Lane, H., and Tranel, B. (**1971**). ''The Lombard sign and the role of hearing in speech,'' J. Speech Hear. Res. **14**, 677–709.

Lane, H., and Webster, J. W. (**1991**). ''Speech deterioration in postlingually deafened adults,'' J. Acoust. Soc. Am. **89**, 859–866.

Lane, H., Wozniak, J., Matthies, M., Svirsky, M., Perkell, J., O'Connell, M., and Manzella, J. (**1997**). ''Changes in sound pressure and fundamental frequency contours following changes in hearing status,'' J. Acoust. Soc. Am. **101**, 2244–2252.

Langereis, M. C., Bosman, A. J., van Olphen, A. F., and Smoorenburg, G. F. (**1997**). ''Changes in vowel quality in postlingually deafened cochlear implant users,'' Audiology **36**, 279–297.

Langereis, M. C., Bosman, A. J., van Olphen, A. F., and Smoorenburg, G. F. (**1999**). ''Intelligibility of vowel produced by postlingually deafened cochlear implant users,'' Audiology **38**, 206–224.

Larson, C. R., Burnett, T. A., Kiran, S., and Hain, T. (**2000**). ''Effects of pitch-shift velocity on voice F_0 responses,'' J. Acoust. Soc. Am. **107**, 559–564.

Leder, S. B., and Spitzer, J. B. (**1993**). ''Speaking fundamental frequency, intensity and rate of adventitiously profoundly hearing-impaired adult women,'' J. Acoust. Soc. Am. **93**, 2146–2151.

Levene, H. (**1960**). ''Robust test for equality of variance,'' in *Contributions to Probability and Statistics*, edited by I. Olkin, S. G. Ghuyre, W. Hoffding, W. G. Madow, and H. B. Mann (Stanford University Press, Stanford, CA), pp. 278–292.

Lindblom, B., and Engstrand, O. (**1989**). ''In what sense is speech quantal?'' J. Phonetics **17**, 107–121.

Lobanov, B. M. (**1971**). ''Classification of Russian vowels spoken by different speakers,'' J. Acoust. Soc. Am. **49**, 606–608.

McCaffrey, H. A., and Sussman, H. M. (**1994**). ''An investigation of vowel organization in speakers with severe and profound hearing loss,'' J. Speech Hear. Res. **37**, 938–951.

Millar, J., Vonwiller, J., Harrington, J., and Dermody, P. (**1994**). ''The Australian National Database of Spoken Language,'' in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing* (Institute of Electrical and Electronic Engineers, Adelaide), Vol. 2, pp. 67–100.

Mitchell, A. G., and Delbridge, A. (**1965**). *The Pronunciation of English in Australia* (Angus and Robertson, Sydney).

Monsen, R. (**1976**). ''Normal and reduced phonological space: The production of English vowels by deaf adolescents,'' J. Phonetics **4**, 189–198.

Natke, U., and Kalverum, K. T. (**2001**). ''Effects of frequency shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables,'' J. Speech Lang. Hear. Res. **44**, 577–584.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., Wilhelms-Tricarico, R., and Zandipour, M. (**2000**). ''A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss,'' J. Phonetics **28**, 233–272.

Perkell, J., Lane, H., Svirsky, M. A., and Webster, J. (**1992**). ''Speech of cochlear implant patients: A longitudinal study of vowel production,'' J. Acoust. Soc. Am. **91**, 2961–2978.

Plant, G. (**1984**). ''The effects of an acquired profound hearing loss on speech production,'' Br. J. Audiol. **18**, 39–48.

Plant, G., and Hammarberg, B. (**1983**). ''Acoustic and perceptual analysis of the speech of the deafened,'' STL-QPSR, 2–3, 85–107.

Richardson, L. M., Busby, P. A., Blamey, P. J., Dowell, R. C., and Clark, G. M. (**1993**). ''The effects of auditory feedback from the nucleus cochlear implant on the vowel formant frequencies produced by children and adults,'' Ear Hear. **14**, 339–349.

Siegel, G. M., and Pick, Jr., H. L. (**1974**). ''Auditory feedback in the regulation of voice,'' J. Acoust. Soc. Am. **56**(5), 1618–1624.

Simpson, A. P. (**2001**). ''Dynamic consequences of differences in male and female vocal tract dimensions,'' J. Acoust. Soc. Am. **109**, 2153–2164.

Simpson, A. P. (**2002**). ''Gender-specific articulatory-acoustic relations in vowel sequences,'' J. Phonetics **30**, 417–435.

Smyth, V., Murdoch, B. E., McCormack, P., and Marshall, I. (**1991**). ''Objective and subjective evaluation of subjects fitted with the cochlear multichannel cochlear prostheses: 3 studies,'' Aust. J. Human Commun. Disord. **19**, 31–52.

Svirsky, M., Lane, H. L., Perkell, J. S., and Wozniak, J. (**1992**). ''Effects of short-term auditory deprivation on speech production in adult cochlear implant users,'' J. Acoust. Soc. Am. **92**, 1284–1300.

Svirsky, M. A., and Tobey, E. A. (**1991**). ''Effect of different types of auditory stimulation on vowel formant frequencies in multichannel cochlear implant users,'' J. Acoust. Soc. Am. **89**, 2895–2904.

Tartter, V. C., Chute, P. M., and Helman, S. A. (**1989**). ''The speech of a postlingually deafened teenager during the first year of use of a multichannel cochlear implant,'' J. Acoust. Soc. Am. **86**, 2113–2121.

Tobey, E. A. (**1993**). ''Speech Production,'' in *Cochlear Implants: Audiological Foundations*, edited by R. S. Tyler (Singular, San Diego), pp. 257–316.

van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (**1988**). ''Effects of noise on speech production: Acoustic and perceptual analyses,'' J. Acoust. Soc. Am. **84**, 917–928.

Waldstein, R. S. (**1990**). ''Effects of postlingual deafness on speech production: Implications for the role of auditory feedback,'' J. Acoust. Soc. Am. **88**, 2099–2114.

Waters, T. (**1986**). ''Speech therapy with cochlear implant wearers,'' Br. J. Audiol. **20**, 35–43.

Watson, C. I., and Harrington, J. (**1999**). ''Acoustic evidence for dynamic formant trajectories in Australian English vowels,'' J. Acoust. Soc. Am. **106**, 458–468.

Watson, C. I., Harrington, J., and Evans, Z. (**1998**). ''An acoustic comparison between New Zealand and Australian English vowels,'' Aust. J. Ling. **18**, 185–207.

Zahorian, S. A., and Jagharghi, A. J. (**1993**). ''Spectral-shape features versus formants as acoustic correlates for vowels,'' J. Acoust. Soc. Am. **94**, 1966–1982.

Speech production in noise with and without hearing protection[a)]

Jennifer B. Tufts[b)] and Tom Frank
Department of Communication Disorders, The Pennsylvania State University, University Park, Pennsylvania 16802

(Received 23 August 2002; revised 25 April 2003; accepted 19 May 2003)

People working in noisy environments often complain of difficulty communicating when they wear hearing protection. It was hypothesized that part of the workers' communication difficulties stem from changes in speech production that occur when hearing protectors are worn. To address this possibility, overall and one-third-octave-band SPL measurements were obtained for 16 men and 16 women as they produced connected speech while wearing foam, flange, or no earplugs (open ears) in quiet and in pink noise at 60, 70, 80, 90, and 100 dB SPL. The attenuation and the occlusion effect produced by the earplugs were measured. The Speech Intelligibility Index (SII) was also calculated for each condition. The talkers produced lower overall speech levels, speech-to-noise ratios, and SII values, and less high-frequency speech energy, when they wore earplugs compared with the open-ear condition. Small differences in the speech measures between the talkers wearing foam and flange earplugs were observed. Overall, the results of the study indicate that talkers wearing earplugs (and consequently their listeners) are at a disadvantage when communicating in noise. [DOI: 10.1121/1.1592165]

PACS numbers: 43.70.Gr, 43.50.Hg, 43.66.Vt [DKW]

I. INTRODUCTION

Among the many hazards that workers face on their jobs, exposure to excessive noise is one of the most pervasive. An estimated 30 million American workers are exposed to hazardous noise levels on their jobs (NIOSH, 1996) and are at risk for developing noise-induced hearing loss. Hearing protection devices (HPDs), such as earplugs or earmuffs, can reduce the danger of incurring noise-induced hearing loss if worn correctly and consistently. Wearing HPDs does not impair the ability of normal-hearing people to understand speech in noise levels above 85 dBA (Berger, 1980; Abel *et al.*, 1982; Nixon and Berger, 1991), and may even improve it (Kryter, 1946; Chung and Gannon, 1979; Rink, 1979). However, workers often resist the use of HPDs because they claim that they have difficulty understanding speech while wearing them (Royster and Holder, 1982; Helmkamp, 1986; Suter, 1992). A possible reason for this contradiction is that some workers may have pre-existing hearing loss, which, when combined with the attenuation of their HPDs, may make parts of the speech signal inaudible. In fact, speech understanding when wearing HPDs is often degraded for people with sensorineural hearing loss (Lindeman, 1976; Chung and Gannon, 1979; Abel *et al.*, 1982; Bauman and Marston, 1986).

Another possible reason is that speech may be less intelligible if the talker is wearing HPDs, because wearing HPDs tends to induce changes in speech production (Kryter, 1946; Howell and Martin, 1975; Martin, Howell, and Lower, 1976). Kryter (1946) reported that the intelligibility of male talkers reading monosyllabic word lists in a background of simulated engine room noise at levels of 75 to 105 dB SPL was slightly lower when they wore earplugs, compared with an open-ear condition. Further, the talkers lowered the level of their speech by 1 to 2 dB in all of the noise levels when they wore earplugs.

Howell and Martin (1975) measured the intelligibility of four male talkers reading monosyllabic word lists in a background of broadband noise at levels of 93 and 54 dB SPL while wearing earmuffs, earplugs, or no HPDs. In noise at 93 dB SPL, the percentage of phonemes correctly heard by listeners was 56% when the talkers were not wearing HPDs. This percentage decreased by 15% when the talkers wore earmuffs and by 26% when the talkers wore earplugs. In addition, the talkers lowered their speech levels by an average of 2.7 dB when they wore earmuffs and 4.2 dB when they wore earplugs. In noise at 54 dB SPL, all of the intelligibility scores approached 100%, although they were slightly lower when the talkers wore earplugs. At this noise level, the difference in the talkers' speech levels was negligible (0.3 dB) when the talkers wore earmuffs and was 1.8 dB lower with earplugs, compared with the no-HPD condition.

In a follow-up study, Martin *et al.* (1976) recorded overall and octave-band speech levels and measured the intelligibility of eight male talkers reading monosyllabic word lists while wearing earplugs, earmuffs, or no HPDs in a background of broadband noise presented at 67, 77, 87, and 95 dBA. They found that the percentage of phonemes correctly understood by listeners decreased when the talkers wore HPDs, particularly if the listeners also wore HPDs. Further, the overall speech level of the talkers was 2 to 3 dB lower

[a)]Portions of this work were presented at the annual meeting of the National Hearing Conservation Association, Dallas, TX, February 2002, and at the annual meeting of the American Academy of Audiology, Philadelphia, PA, April 2002.

[b)]Current affiliation: Army Audiology & Speech Center, Walter Reed Army Medical Center, 6900 Georgia Ave. NW, Washington, DC 20307-5001. Electronic mail: jennifer.tufts@na.amedd.army.mil

when they wore HPDs compared with the open-ear condition. The octave-band analysis revealed no significant changes in the talkers' speech spectra when wearing HPDs other than would be expected from the change in their speech level. Martin *et al.* (1976) surmised that when the talkers wore HPDs, changes in their voice quality might have occurred that were not reflected in the octave-band speech spectrum but that might have been detrimental to intelligibility. To evaluate this possibility, the recordings of the eight male talkers reading the word lists with and without earplugs in the background noise at 87 dBA were mixed with noise at a +3-dB speech-to-noise ratio (SNR) and presented to four male listeners at approximately 70 dB SPL. Any differences in intelligibility between speech produced with and without HPDs would therefore be due to factors other than speech level and SNR of the presentation. For seven of the eight talkers, no differences were observed in the intelligibility of their speech. Thus, the difference in intelligibility between talkers wearing and not wearing HPDs largely disappeared when speech level and SNR were controlled.

Previous studies, including those just described, have shown that talkers lower the level of their voices in noise by 1 to 6 dB on average when they wear HPDs compared with an open-ear condition (Kryter, 1946; Howell and Martin, 1975; Martin *et al.*, 1976; Hormann *et al.*, 1984; Casali, Horylev, and Grenell, 1987). Two of these studies (Kryter, 1946; Casali *et al.*, 1987) also reported that talkers raise their voices by approximately 4 dB when they speak in quiet while wearing HPDs, perhaps to compensate for the attenuation of the air-conduction component of their speech.

Two explanations have been offered as to why people speak more softly when wearing HPDs in noise. First, complete occlusion of the ear canal improves bone-conduction hearing at frequencies below 2000 Hz. Due to the occlusion effect created by wearing HPDs, talkers' voices will seem louder to them, and they may lower the level of their voices in response (Howell and Martin, 1975; Martin *et al.*, 1976; Casali *et al.*, 1987; Suter, 1992). Second, since HPDs attenuate ambient noise, talkers wearing HPDs may lower their voices in response to the decrease in the perceived ambient noise level (Kryter, 1946; Lindeman, 1976; Foreshaw and Cruchley, 1982).

Aside from the changes in overall level, little is known about the speech of talkers wearing HPDs. In general, when speech is produced in a background of noise, without HPDs, several of its acoustic and prosodic characteristics (e.g., overall level, fundamental frequency, formant frequencies, and spectral composition) change relative to speech produced at a normal conversational level in quiet. For example, for every 10-dB increase in the background noise level, talkers raise their voices by 1 to 6 dB, a phenomenon known as the Lombard effect (Lane and Tranel, 1971). In addition, talkers produce higher fundamental frequencies when they speak in a background of noise (Summers *et al.*, 1988; Bond, Moore, and Gable, 1989; Junqua, 1993; Letowski, Frank, and Caravella, 1993). The frequencies of the first two formants also change, but in less consistent ways. (Summers *et al.*, 1988; Bond *et al.*, 1989; Junqua, 1993; Tartter, Gomes, and Litwin, 1993). In general, relatively more high-frequency energy is contained in speech that is produced in noise versus speech that is produced in quiet (Webster and Klumpp, 1962; Summers *et al.*, 1988; Junqua, 1993; Tartter *et al.*, 1993; Pittman and Wiley, 2001). The spectral and prosodic differences that occur when talkers produce speech in noise and when they produce speech at greater vocal efforts in quiet are similar (Bond *et al.*, 1989). However, Letowski *et al.* (1993) found that speech produced in noise contains more high-frequency energy than loud speech produced in quiet.

Perhaps due in part to these changes, several investigators have found that speech produced in noise is more intelligible than speech produced in quiet, when both types of speech are presented in noise at the same SNR (Dreher and O'Neill, 1957; Summers *et al.*, 1988; Pittman and Wiley, 2001). Summers *et al.* (1988) also reported that the relative improvement in the intelligibility of speech produced in noise increases as the SNR decreases.

Most of the research concerning the speech of talkers wearing HPDs in noise has been limited to measurements of the overall SPLs, as described previously. Further, these studies were typically carried out in background noise levels that were restricted in either the number of discrete levels presented or the range they encompassed. In most cases, the speech stimuli were monosyllabic words, which are not representative of most communication situations. Further information about the speech of talkers wearing HPDs in noise may provide a basis for evaluating the validity of workers' complaints and for selecting HPDs for particular work environments.

The main purpose of the present study was to obtain overall and one-third-octave-band SPL measurements of connected speech produced by men and women wearing two types of earplugs and wearing no earplugs (open ears). Measurements were made in quiet and in several noise levels. Differences between the two types of earplugs in terms of their effect on speech production were also assessed. Additionally, the Speech Intelligibility Index (SII) (ANSI, 1997) was used to determine if the amount of speech information available to a hypothetical listener changed as a function of background noise level and whether or not the talkers were wearing earplugs. The SII is especially useful for evaluating communication systems when it is not practical to use actual listeners for that purpose, and it has previously been used as a means of ranking HPDs for maximum speech communication ability in noise (Williams and Michael, 1991).

II. METHOD

A. Subjects, HPDs, and test environment

Thirty-two subjects (16 males, 16 females) ranging in age from 19 to 35 years (mean=25.4 years, s.d.=4.9 years) were enrolled in the study and received $20.00 for their participation. Each subject was a fluent speaker of American English, and had normal air-conduction thresholds ($\leqslant$20 dB HL from 0.25 to 8 kHz *re*: ANSI S3.6-1996) and a normal tympanogram (Margolis and Heller, 1987) in each ear. For part of the test session, each subject wore one of two types of insert HPDs commonly used in industry. One type was a

user-molded foam earplug [EAR Classic] and the other was a premolded triple flange earplug [Bilsom 556]. All testing was conducted in an 18×18×6-ft. sound-treated room having ambient noise levels suitable for ears-open testing from 0.125 to 8 kHz (*re*: ANSI S3.1-1991).

B. Stimulus characteristics

1. Speech passages

During the test session, each subject was required to read twelve speech passages from the Speech Intelligibility Rating (SIR) test (McDaniel and Cox, 1992). The 12 passages were chosen from the 20 passages of the SIR test by measuring the average one-third-octave-band spectra of the commercially available CD recordings of all 20 passages, computing the slopes of regression lines fitted to the spectra, and then selecting those passages whose slopes fell within 1 standard deviation of the mean. Each passage consisted of approximately 110 words on a familiar topic, and all were equated on the basis of reading level, vocabulary, and sentence structure.

2. Background noise

Background noise was presented to the subjects via TDH-49 earphones mounted in circumaural enclosures as they read the speech passages. The background noise was a CD recording of pink noise having equal power within each one-third-octave band, as measured in an IEC 318 artificial ear [Bruel & Kjaer, 4153]. Pink noise was chosen because its spectral level decreases with increasing frequency, similar to many industrial noises.

To create a flat one-third octave-band spectrum (i.e., pink noise) as measured in the artificial ear, the following steps were taken. Initially, a CD recording of pink noise was created by looping a pink-noise wave file and writing the wave file to a CD using an editing program [SOUND FORGE, v4.5]. The pink-noise CD was then played from a CD player [Optimus, CD-1850] to the right channel input of an audiometer [Madsen, OB 822]. The audiometer output was split and directed to two TDH-49 earphones, located in the test booth. Each TDH-49 earphone was mounted in a circumaural enclosure having an oval cushion [David Clark, H10–40]. Next, each earphone was placed on a flat plate adaptor [Bruel & Kjaer type 1, DB0843] and loaded with a static force of 10 N. The flat plate adaptor was seated on the IEC 318 artificial ear, which was equipped with a $\frac{1}{2}$-in. microphone [GRAS, 40AF] and preamplifier [01 dB, PRE 12H] that was connected to a real-time spectrum analyzer [01 dB Symphonie]. The audiometer output was adjusted to 100 dB SPL as measured in the IEC 318 coupler. Once this was done, the one-third-octave-band spectrum of the pink noise was averaged over a 1-min period, with samples taken once every 10 ms. Using the measured spectrum as a guide, the spectrum of the original pink-noise wave file was modified using the one-third-octave-band graphic equalizer in the SOUND FORGE editing program, and then rerecorded onto a new CD. This procedure was repeated until the one-third-octave-band spectrum measured in the coupler was essentially flat (±1 dB from 0.16 to 2.5 kHz; ±3 dB from 3.15 to 5 kHz; ±1 dB from 6.3 to 8 kHz). During data collection, the flat-spectrum pink-noise CD was played from the CD player to the audiometer and then to the circumaural earphones at 60, 70, 80, 90, and 100 dB SPL as measured in the IEC 318 coupler. The coupler SPLs remained stable throughout data collection (±0.5 dB), as did the linearity of the audiometer output from 60 to 100 dB SPL (±0.5 dB). The difference in the coupler output between the two earphones was <1 dB at each one-third-octave-band from 0.16 to 8 kHz.

C. Measurements

1. Speech

As each subject read the SIR test passages, his or her overall and one-third-octave-band speech levels were measured with a type 1 sound-level meter (SLM) [Casella CEL, 593] mounted on an adjustable tripod and equipped with a $\frac{1}{2}$-in. omnidirectional sound-field microphone [Casella CE, 3122] having a flat response (±1 dB) from 0.05 to 12 kHz. The SLM was interfaced with a laboratory computer [Dell, Dimension 4100] in an adjacent control room. A commercial software program [Casella CEL, 6695] was used for remotely controlling the SLM and for downloading and storing each subject's speech measurements. The SLM was programmed to average each subject's overall and one-third-octave-band speech levels from 0.08 to 20 kHz, using a fast response time (0.125 ms), a sampling rate of 75.6 kHz, and a dynamic range of 25–100 dB SPL. The SLM was calibrated with an acoustical calibrator [Casella CEL, 284/2] and remained stable throughout data collection (±0.1 dB).

2. Earplug attenuation

The attenuation provided by the earplugs was measured with a Fitcheck measurement system [Michael & Associates, 700] consisting of both hardware and software. The hardware consisted of a 40-dB fixed attenuator and a 70-dB variable attenuator, a subject response switch, and circumaural earphones, which were the same earphones used to transduce the pink noise. The software was used to control the attenuator settings and the test signals, record the subject's responses, and perform calculations. The Fitcheck unit was connected to the laboratory computer located in the control room, while the response switch and the circumaural earphones were located in the test booth. The software instructed the sound card of the computer to generate one-third-octave-bands of noise centered at 0.25, 0.5, 1, 2, and 4 kHz, which were directed to the Fitcheck unit and then to the circumaural earphones. The software also determined the status of the response switch, increasing the attenuation of the one-third-octave-band noise test signals when the response switch was pressed, and decreasing the attenuation when the response switch was released. The variable attenuator had a step size of 1.5 dB and an attenuation rate of 3 dB/second.

Each subject's one-third-octave-band noise (1/3 OBN) thresholds were obtained for each center frequency while he or she wore the circumaural earphones without the earplugs (open-ear 1/3-OBN thresholds) and with the earplugs (closed-ear 1/3-OBN thresholds) using the Fitcheck automated threshold-tracing procedure, which was controlled by

the software. This was done by having the subject continuously press the response switch whenever the 1/3-OBN test signal was audible and release the switch whenever the signal was inaudible. For each 1/3-OBN test signal, eight threshold reversals (four peaks and four valleys) were obtained. The subject's threshold for each 1/3-OBN test signal was calculated automatically by the software as the midpoint between the average of the peaks and the average of the valleys, with the first two reversals ignored.

The output of the Fitcheck measurement system was calibrated by adjusting its variable attenuator to an arbitrary level at each 1/3-OBN center frequency, and measuring the output voltage at the earphone terminals (with the earphones in line) using a Hewlett-Packard 400GL voltmeter. The output voltage remained stable (± 1.3 dB) throughout data collection.

3. Occlusion effect

Pure-tone bone-conduction thresholds (BCTs) at frequencies of 0.25, 0.5, and 1 kHz were obtained for each subject. The BCTs were measured using forehead placement of a Radioear B-71 bone vibrator. The pure tones were generated by the audiometer [Madsen, OB 822] and directed to a 20-dB fixed-pad attenuator in the test booth and then to the bone vibrator. The bone vibrator was fitted in the center of each subject's forehead and held in place with a headband adjusted to produce >300 grams of force as measured by a strain gauge [Abbott, 4580].

The output of the bone vibrator was calibrated by setting the audiometer output to an arbitrary level at each test frequency, and measuring the output voltage at the terminal of the bone vibrator (with the bone vibrator in line) using a Hewlett-Packard 400 GL voltmeter. The output voltage remained stable (± 0.7 dB) throughout data collection.

During the test session, BCTs were measured in several conditions: with no earplugs or earphones worn, with the earphones only, with the earplugs only, and with the earplugs and earphones worn. This was done to determine the magnitude of the occlusion effect due to the earphones alone, the earplugs alone, the earplugs and earphones together, and the earplugs and earphones relative to the earphone-only condition. Of particular interest was the latter measurement. Since the subjects were wearing earphones in the control condition and earplugs and earphones in the experimental condition, the difference between the BCTs at each test frequency in these two conditions represented the magnitude of the occlusion effect experienced by the subjects when they wore earplugs relative to the control condition.

D. Procedure

Sixteen subjects were tested with the foam earplugs and 16 subjects were tested with the flange earplugs. Within each earplug group (foam or flange), speech measurements were obtained for eight subjects (four males, four females) first while wearing the earplugs and then without the earplugs. The other eight subjects (four males, four females) were tested first without the earplugs and then while wearing the earplugs. Each subject received a randomized order for the 12 reading passages, a randomized order for the six noise conditions (quiet, 60, 70, 80, 90, and 100 dB SPL) for testing with the earplugs, and a different randomized order for the six noise conditions for testing without the earplugs. Prior to the start of each subject's test session, the subject was seated in the test booth and the microphone of the SLM was adjusted to be level with and 1 m away from his or her mouth at 0° azimuth. The face of the microphone was positioned at 90° relative to the subject's lips so that his or her voice reached the microphone with a grazing incidence.

The procedures carried out during data collection are described below. The order in which they are listed corresponds to the order of procedures for the 16 subjects who were tested first with no earplugs, then with earplugs.

(1) Without earplugs, speech in quiet and in five levels of pink noise: The subject was required to read several of the speech passages while different levels of noise (60, 70, 80, 90, or 100 dB SPL), or quiet, were presented over the circumaural earphones. One noise level or quiet was presented per passage. The subject was instructed to read each passage in such a way that a conversation partner listening in noise at the same level would be able to understand what was said. Further, the subject was instructed to read continuously until told to stop, even if he or she made a mistake or had to begin reading the passage again. Once told to stop, the subject was instructed to turn to the next reading passage and wait for the next noise level to be presented. The subject was allowed to look over each new passage and, if necessary, take a drink of water before continuing.
For each speech passage, the investigator adjusted the audiometer to provide the appropriate level of pink noise, or quiet, according to the randomized order for that subject. As soon as the subject began reading each passage, the SLM was activated to record his or her overall and one-third-octave-band speech levels for a period of 45 s. This procedure was repeated until all of the noise level conditions were completed.
(2) Open-ear 1/3-OBN thresholds: The subject's binaural open-ear 1/3-OBN thresholds were measured with the Fitcheck system at the center frequencies of 0.25, 0.5, 1, 2, and 4 kHz, in ascending order by frequency.
(3) Insertion of the earplugs: Following a demonstration by the investigator, the subject inserted the earplugs. If the earplugs appeared to be poorly inserted (e.g., most of the earplug was visible outside the ear canal), the investigator asked the subject to remove and reinsert them.
(4) With earplugs, speech in quiet and in five levels of pink noise: The instructions for reading the speech passages were repeated. The speech measurements were obtained in quiet and in noise levels of 60, 70, 80, 90, and 100 dB, with the subject wearing the earplugs in addition to the circumaural earphones.
(5) Closed-ear 1/3-OBN thresholds: The subject's closed-ear 1/3-OBN thresholds were obtained with the Fitcheck system.
(6) BCTs: The subject's BCTs were measured at 1, 0.5, and 0.25 kHz, in descending order by frequency. Threshold

at each test frequency was defined as the mean of four ascending thresholds using a 2-dB step. The BCTs were measured in four conditions: while wearing both the circumaural earphones and earplugs, while wearing the earplugs without the earphones, while wearing neither the earplugs nor the earphones, and while wearing the earphones only.

The order of procedures was reversed for the 16 subjects who were tested first with earplugs, then without earplugs, except that this group of subjects inserted the earplugs halfway through the first procedure (i.e., step 6).

1. Dependent measures

Three types of dependent measures were obtained or calculated for each talker in each background noise level and in each ear condition (with versus without earplugs). These included (1) two measures of speech level; (2) one measure that described the speech spectrum; and (3) one measure of predicted intelligibility.

The first measure of speech level was the overall speech level (in dB SPL). The second was the speech-to-noise ratio (SNR; in dB). The SNR was defined as the overall speech level minus the overall noise level presented to the subject. Since the speech and the noise were measured under different conditions (i.e., the speech was measured in the sound field, and the noise was measured in an IEC 318 artificial ear), the difference between the two is not, strictly speaking, a true SNR. However, due to the necessity of measuring the speech without contamination by the background noise, the SNR was calculated in this way.

The measure that described the speech spectrum was the spectral center of gravity (SCoG; in Hz). This was defined as the center frequency of the one-third-octave-band between 0.16 and 8 kHz, inclusive, containing the frequency above and below which there were equal amounts of power in the speech signal. The SCoG was determined by successively adding the power in each consecutive one-third-octave-band, in order of increasing center frequency, and then choosing the one-third octave band whose addition caused the running sum to first exceed one-half of the total power in the signal. Last, the SII was calculated assuming a hypothetical listener with open ears (i.e., without HPDs) and normal hearing threshold levels, and using the one-third-octave-band procedure and the frequency importance function for continuous discourse described in ANSI S3.5-1997. In essence, an SNR was calculated in each one-third-octave-band and then weighted by the frequency importance function for continuous discourse. These weighted SNRs were summed to produce a number giving the proportion of speech information available to a hypothetical listener. The SII ranges from a minimum of 0.0 to a maximum of 1.0.

III. RESULTS

Three separate analyses were conducted on the data. The first analysis was done to examine the differences between the ear conditions (with versus without earplugs). For this analysis, the data were collapsed over earplug type (foam and flange). A four-way mixed-factors ANOVA was conducted on each dependent measure, with background noise level and ear condition as within-subjects factors, and testing order and sex as between-subjects factors. The Huynh–Feldt epsilon was used to correct the probabilities for the main effect of background noise level and its interaction with ear condition, due to the violation of the sphericity assumption (Huynh and Feldt, 1976).

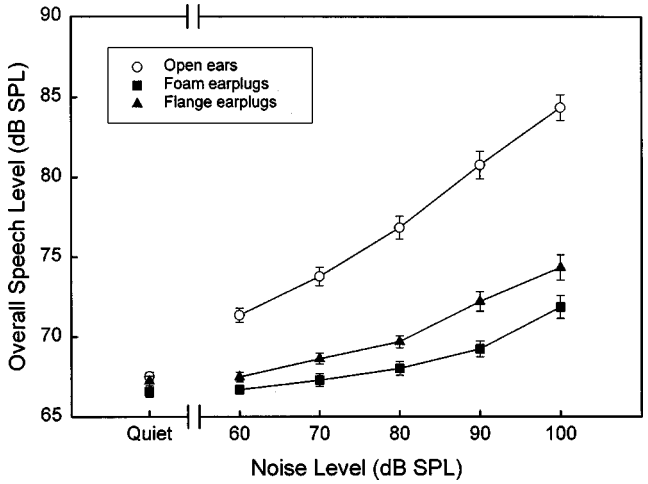


FIG. 1. Mean overall speech levels (in dB SPL) and standard errors as a function of noise level, with open ears and with foam and flange earplugs.

The second analysis examined the differences between the earplug groups (foam versus flange). Using only the data obtained when the earplugs were worn, a four-way mixed-factors ANOVA was conducted on each dependent measure, with background noise level as the within-subjects factor, and earplug group, testing order, and sex as the between-subjects factors. As before and where applicable, the probabilities were corrected for the violation of sphericity using the Huynh–Feldt epsilon.

The third analysis was conducted to ensure that the two earplug groups were similar with respect to the dependent measures when the earplugs were not worn. For each dependent measure, a four-way mixed-factors ANOVA was conducted on the data obtained in the open-ear condition. For each ANOVA, the main effect of earplug group and all of its interactions were nonsignificant ($p > 0.05$). Thus, any differences between the earplug groups when the earplugs were worn could be attributed to the earplug type.

A. Overall speech level and SNR

Figures 1 and 2 illustrate the mean overall speech levels and SNRs, respectively, and their standard errors, for the talkers with and without earplugs in quiet (overall speech level only) and in each background noise level. In quiet, the mean overall speech levels were very similar between ear conditions (67.5 dB SPL without earplugs versus 66.6 and 67.2 dB SPL for the foam and flange earplug groups, respectively). In the presence of background noise, the talkers raised the level of their voices. However, the increase in the overall speech level did not match the increase in the background noise level. That is, although the background noise level increased 40 dB from 60 to 100 dB SPL, the overall

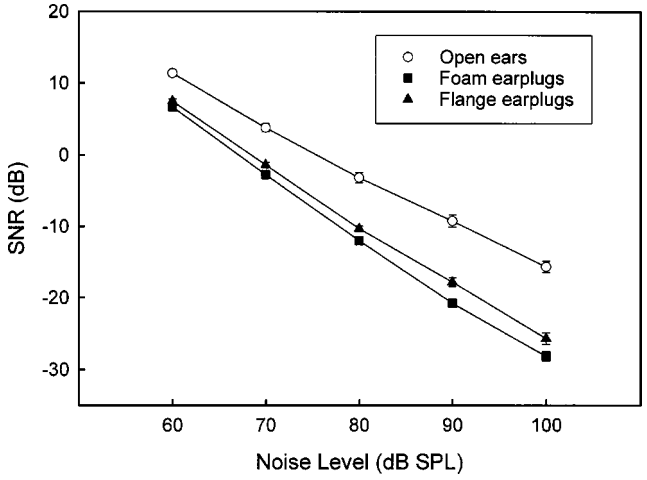


FIG. 2. Mean SNRs (in dB) and standard errors as a function of noise level, with open ears and with foam and flange earplugs.

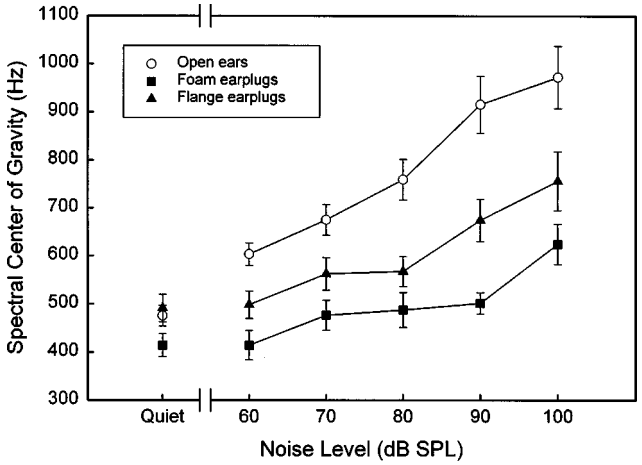


FIG. 3. Mean spectral centers of gravity (in Hz) and standard errors as a function of noise level, with open ears and with foam and flange earplugs.

speech level only increased by 5 to 13 dB, depending on the ear condition. Therefore, the SNR decreased as the background noise level increased.

When the talkers wore earplugs, the mean overall speech level and SNR were consistently lower in each background noise level compared with the open-ear condition. In fact, the mean differences in both the overall speech levels and the SNRs between the two ear conditions increased from 4 to 11 dB as the background noise level increased from 60 to 100 dB. In background noise at 100 dB SPL, the overall speech levels were 84.4 dB for the talkers with open ears, and 71.9 and 74.3 dB for the foam and flange earplug groups, respectively. The corresponding SNRs were -15.6, -28.1, and -25.7 dB.

For the first analysis on the overall speech level (with versus without earplugs), the main effects of background noise level and ear condition, and their interaction, were statistically significant [noise level: $F(5,140)=326.517$, $p<0.001$; ear condition: $F(1,28)=226.956$, $p<0.001$; interaction: $F(5,140)=106.826$, $p<0.001$]. In addition, the interaction of background noise level and testing order was significant [$F(5,140)=4.327$, $p=0.015$]. For the second analysis (foam versus flange earplug groups), the main effects of background noise level and earplug group, and their interaction, were significant [noise level: $F(5,120)=119.726$, $p<0.001$; earplug group: $F(1,24)=10.757$, $p=0.003$; interaction: $F(5,120)=4.389$, $p=0.008$]. The interaction of background noise level and testing order was also significant [$F(5,120)=7.398$, $p<0.001$].

Similarly, for the first analysis on the SNR, the main effects of background noise level and ear condition, and their interaction, were significant [noise level: $F(4,112)=3041.646$, $p<0.001$; ear condition: $F(1,28)=4965.188$, $p<0.001$; interaction: $F(4,112)=60.568$, $p<0.001$]. Further, the three-way interaction of background noise level, ear condition, and sex was significant [$F(4,112)=2.776$, $p=0.034$]. For the second analysis, the main effects of background noise level and earplug group, and their interaction, were significant [noise level: $F(4,96)=3478.527$, $p<0.001$; earplug group: $F(1,24)=11.620$, $p=0.002$; interaction: $F(4,96)=3.642$, $p=0.023$]. In addition, the interaction of background noise level and testing order was significant [$F(4,96)=5.487$, $p=0.003$].

These results indicate that the level of a talker's voice (and consequently the SNR) is significantly lower on average when earplugs are worn in the presence of background noise compared with an open-ear condition. In addition, the foam earplug group had slightly lower overall speech levels and SNRs than the flange earplug group. However, it should be noted that the increases in overall speech level in the presence of background noise were not sufficient to maintain a favorable SNR either with or without earplugs.

B. Spectral center of gravity

Figure 3 illustrates the mean frequencies of the spectral center of gravity (SCoG) and their standard errors, for the talkers with and without earplugs, in quiet and in each background noise level. In quiet, the SCoGs were very similar between ear conditions (475 Hz for the talkers without earplugs versus 414 and 491 Hz for the talkers with foam and flange earplugs, respectively). In the presence of background noise, the SCoG increased in frequency. However, the frequency of the SCoG was consistently lower in each background noise level when the talkers wore earplugs. The differences between the two ear conditions increased as the noise level increased. In background noise at 100 dB SPL, the mean SCoGs were 972 Hz for the talkers without earplugs and 624 and 756 Hz for the talkers with foam and flange earplugs, respectively.

The effect on the frequency spectrum of an increase in the SCoG is shown more clearly in Fig. 4, which illustrates the one-third-octave-band speech spectra for (A) men and (B) women, respectively, in quiet and in background noise at 100 dB SPL, with and without earplugs. Inspection of Figs. 4(A) and (B) reveals that the speech spectra in quiet were very similar in level and shape between ear conditions. In background noise at 100 dB SPL, the speech spectra changed in both level and shape, reflecting the increase in the overall speech level as well as the shift of speech energy to the

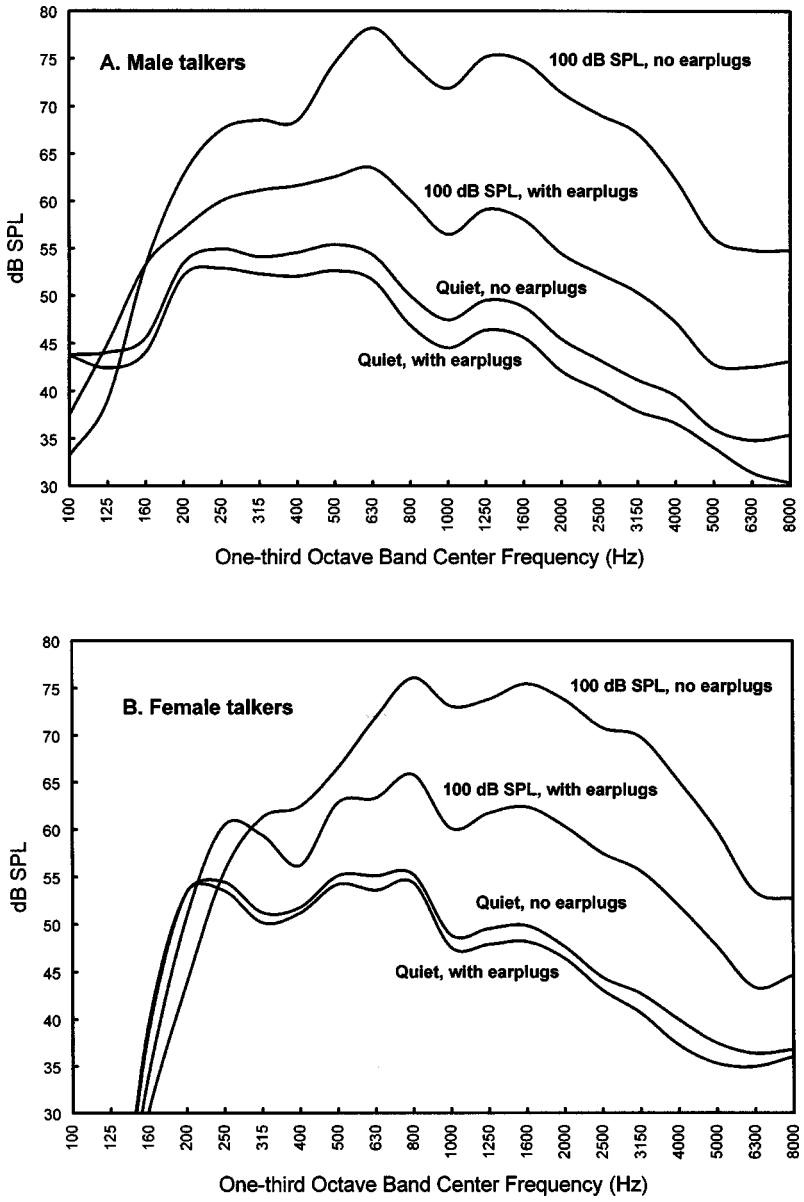


FIG. 4. The one-third-octave-band speech spectra of (A) male and (B) female talkers with and without earplugs in quiet and in background noise at 100 dB SPL. The parameter is background noise level.

higher frequencies. However, these changes were very different in magnitude between the ear conditions. For both men and women, the average speech spectrum of the talkers without earplugs was higher in level and showed a greater shift of energy to the higher frequencies than the average speech spectrum obtained when the talkers wore earplugs.

In general, women had more high-frequency energy in their speech than men. For example, the SCoGs in quiet, without earplugs, were 520 Hz for the women and 430 Hz for the men. In background noise at 100 dB SPL, without earplugs, the corresponding SCoGs were 1200 and 743 Hz. This finding was not unexpected, given that women have generally higher fundamental and formant frequencies than men.

For the first data analysis (with versus without earplugs), the main effects on the SCoG of background noise level and ear condition, and their interaction, were significant [noise level: $F(5,140)=73.438$, $p<0.001$; ear condition: $F(1,28)=64.652$, $p<0.001$; interaction: $F(5,140)=9.407$, $p<0.001$]. The main effect of sex and its interaction with background noise level were also significant [sex: $F(1,28)=51.221$, $p<0.001$; interaction: $F(5,140)=11.156$, $p<0.001$]. For the second analysis (foam versus flange earplug groups), the main effects of background noise level and earplug group were significant [noise level: $F(5,120)=33.853$, $p<0.001$; earplug group: $F(1,24)=17.866$, $p<0.001$]. In addition, the main effect of sex, its two-way interaction with background noise level, and its three-way interaction with background noise level and earplug group were significant [$F(1,24)=44.994$, $p<0.001$; two-way interaction: $F(5,120)=6.006$, $p<0.001$; three-way interaction: $F(5,120)=2.370$, $p=0.043$]. The two-way interaction of background noise level and testing order was also significant [$F(5,120)=4.746$, $p=0.001$].

These results indicate that the talkers produced relatively more high-frequency energy in their speech as the background noise level increased. However, the increase in high-frequency energy was not as pronounced when the talkers wore earplugs, particularly for the foam earplug group. In

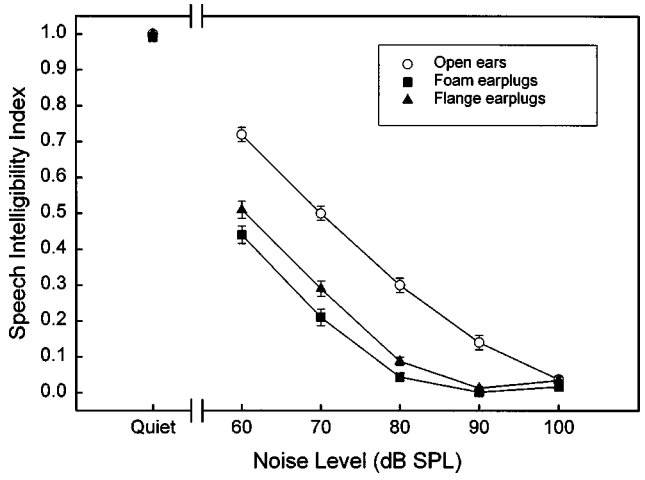


FIG. 5. Mean Speech Intelligibility Index values and standard errors as a function of noise level, with open ears and with foam and flange earplugs. Note that the three symbols overlap in the "quiet" condition.

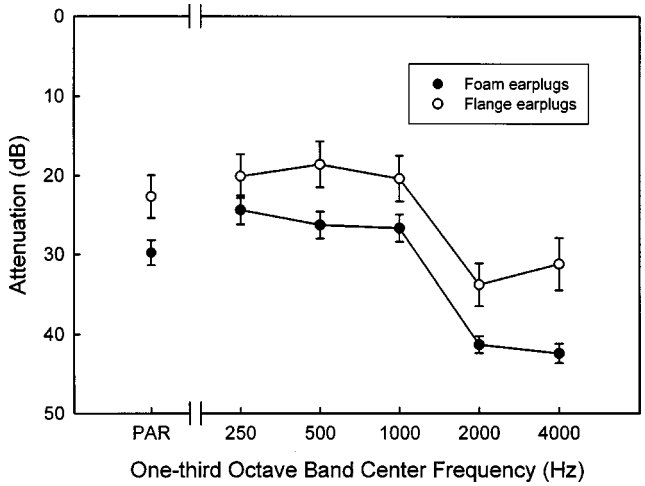


FIG. 6. Mean attenuation values (in dB) and standard errors as a function of one-third octave-band center frequency for the foam and flange earplugs. The mean PAR is also shown for each earplug type.

quiet and in background noise, women produced more high-frequency energy in their speech than men.

C. Speech Intelligibility Index (SII)

Figure 5 shows the mean SII values and standard errors for the talkers with and without earplugs, in quiet and in each background noise level. In quiet, the talkers achieved a mean SII of 1.00 both with and without earplugs (0.99 and 1.00 for the talkers in the foam and flange earplug groups, respectively). The SII decreased to nearly zero in the higher background noise levels both with and without earplugs. In the background noise levels of 60, 70, 80, and 90 dB, the SII was consistently lower when the talkers wore earplugs.

Because the SII data were less variable at each end of their range than in the middle, an arcsin transformation was applied before the analyses were conducted. For the first analysis (with versus without earplugs), the main effects of background noise level and ear condition, and their interaction, were significant [noise level: $F(5,140)=2546.759$, $p<0.001$; ear condition: $F(1,28)=212.351$, $p<0.001$; interaction: $F(5,140)=91.338$, $p<0.001$]. In addition, the two-way interaction of background noise level and sex was significant [$F(5,140)=3.966$; $p=0.002$], as well as the three-way interaction of background noise level, ear condition, and testing order [$F(5,140)=2.776$, $p=0.029$]. For the second analysis (foam versus flange earplug groups), the main effects of background noise level and earplug group were significant [noise level: $F(5,120)=1945.291$, $p<0.001$; earplug group: $F(1,24)=10.690$, $p=0.003$]. The two-way interactions of background noise level with testing order and with sex were also significant [noise level and testing order: $F(5,120)=3.822$, $p=0.004$; noise level and sex: $F(5,120)=3.626$, $p=0.006$].

These results indicate that less and less speech information was available from the talkers as the background noise level increased. Further, when the talkers wore earplugs, the amount of available speech information was reduced relative to the open-ear condition, particularly in moderate to high background noise levels (60 to 90 dB SPL). The SII was consistently lower in each background noise level for the foam earplug group compared with the flange earplug group, particularly in moderate background noise levels (60 to 80 dB SPL).

The differences in the SII values between the ear conditions and between the earplug groups were smaller in the higher background noise levels than in the more moderate background noise levels. This occurred because in the calculation of the SII, it is assumed that no speech information is received by a listener in a particular frequency band if the SNR in that band is less than -15 dB. Thus, in the higher background noise levels, where the SNRs were extremely poor, the differences between the groups were effectively ignored.

D. Attenuation and occlusion effect

Figure 6 shows the mean attenuation values and their standard errors at each one-third-octave-band center frequency for the two earplug types. The mean attenuation characteristics of the two earplug types were typical of passive insert-type HPDs, with greater attenuation of the higher frequencies (>2 kHz) than of the lower frequencies (<1 kHz). Also shown in the figure is the mean Personal Attenuation Rating (PAR; Michael, 1999) of each earplug type and the associated standard errors. The PAR is a single-number rating of the attenuation of HPDs similar in derivation to the Noise Reduction Rating (NRR). The mean PARs were 22.7 dB for the flange earplugs and 29.8 dB for the foam earplugs. (The published NRRs for these earplugs are 27 and 29 dB, respectively.)

Collapsed over earplug type, the mean values of the occlusion effect experienced by the subjects when they wore earplugs relative to the control (earphone-only) condition at 0.25, 0.5, and 1 kHz were 2.8, 7.4, and 4.6 dB, respectively.[1] The magnitude of the occlusion effect was similar between the two earplug types ($\leq$1.5-dB mean difference at each test frequency).

IV. DISCUSSION

A. The effects of wearing earplugs on speech production

Speech communication is a two-way process involving a talker and a listener. To understand how HPDs affect that process, it is important to assess their effects on both the talker and the listener. However, much of the research to date on the effects of HPDs on speech communication has focused on the listener. These studies have generally involved a comparison of the ability to understand speech in a background of noise with open ears and with ears occluded by HPDs. In contrast, the purpose of this study was to assess the effects of HPDs on speech production, *viz.*, the overall level and spectral characteristics of speech. Several speech measures were obtained for talkers with and without earplugs in quiet and in a background of pink noise presented at several levels.

In quiet, occluding the talkers' ear canals with earplugs resulted in a slight decrease of 0.6 dB in the mean overall speech level. Navarro (1996) reported a similar decrease of 0.7 dB for talkers wearing EAR foam earplugs in quiet. These two findings suggest that talkers do not significantly alter the level of their voices when they wear earplugs in quiet. This conclusion contrasts with that reached by Kryter (1946) and Casali *et al.* (1987), who reported that talkers raise the level of their voices by approximately 4 dB when they wear HPDs in quiet. It is possible that, for the talkers in this study [and perhaps the Navarro (1996) study], the effects of the attenuation of the air-conduction component and the enhancement of the bone-conduction component of their speech offset one another. This may have occurred if the earplugs provided an acoustic seal but were not deeply inserted.[2]

In the presence of background noise, the talkers raised the level of their voices. However, depending on the noise level, the mean overall speech levels of the talkers in background noise were 4 to 11 dB lower when they wore earplugs compared with the open-ear condition. These differences were generally larger than the differences of 1 to 6 dB reported in previous research (Kryter, 1946; Howell and Martin, 1975; Martin *et al.*, 1976; Hormann *et al.*, 1984; Casali *et al.*, 1987). The variation across studies may be due to different amounts of HPD attenuation, different test room conditions (e.g., reverberant vs anechoic), and different methods of measuring the speech level. Differences in the background noises presented to talkers may have also affected the speech levels. Casali *et al.* (1987) showed that talkers wearing earmuffs in background noise having a high-frequency bias produced lower speech levels than talkers in background noise having a low-frequency bias. This may have occurred because the earmuffs attenuated the high-frequency noise to a greater extent than the low-frequency noise, and/or because the high-frequency noise provided less masking of the speech signal. The pink noise used in this study likely resulted in speech measures more representative of an environment with low-frequency, steady-state noise than other kinds of noise (e.g., intermittent noise or noise with a high-frequency emphasis). Despite the methodological variations, the talkers in this and previous studies reliably lowered the level of their voices while wearing HPDs in noise, although the magnitude of the effect varied across studies.

As the background noise level increased, the amount of high-frequency energy in the talkers' speech increased, as evidenced by the upward shift in the SCoG. This spectral change is consistent with previous research on speech produced in noise (Webster and Klumpp, 1962; Summers *et al.*, 1988; Junqua, 1993; Tartter *et al.*, 1993; Pittman and Wiley, 2001) and speech produced with greater vocal effort in quiet (Lazarus, 1990; Traunmüller and Eriksson, 2000). Further, the SCoGs of the male talkers were lower in frequency than those of the female talkers in quiet and in each background noise level, both with and without earplugs. This finding was expected, given that men produce more low-frequency speech energy on average due to their lower fundamental frequencies. No other consistent differences between men and women were found for any other speech measure. When the talkers wore earplugs, the mean SCoG decreased in frequency by 147 to 327 Hz compared with the open-ear condition, depending on the background noise level. Thus, in a given background noise level, the talkers produced less high-frequency speech energy when they wore earplugs.

In general, the increase in high-frequency energy that occurs when speech is produced in a background of noise may be beneficial for (normal-hearing) listeners. Several investigators have found that speech produced in noise is more intelligible than speech produced in quiet, when both types of speech are mixed with noise at the same SNR (Dreher and O'Neill, 1957; Summers *et al.*, 1988; Pittman and Wiley, 2001). Apparently, the changes that occur in the frequency spectrum serve to increase the intelligibility of speech produced in noise relative to speech produced in quiet, when level and SNR are controlled. Thus, talkers wearing earplugs (and consequently their listeners) may be at an additional disadvantage, since the shift of speech energy to the higher frequencies is not as pronounced for them.

Although the talkers in this study raised the level of their voices and produced more high-frequency speech energy in background noise, any benefit to intelligibility that might have occurred as a result of these changes was more than offset by the decrease in the SNR. This conclusion is supported by the marked decrease in the SII as the background noise level increased. Further, these effects were even more pronounced when the talkers wore earplugs. Since the SNR and SII are important predictors of intelligibility (Bradley, 1986), these data suggest that the intelligibility of the talkers would decrease as the background noise level increased, and that intelligibility would be further compromised when the talkers wore earplugs. Indeed, previous research has shown that talkers wearing HPDs are less intelligible in noise than talkers without HPDs (Kryter, 1946; Howell and Martin, 1975; Martin *et al.*, 1976).

Despite the differences in the speech spectra between talkers with and without earplugs in the same background noise level, the shape of a talker's speech spectrum does not appear to depend on whether the talker is wearing earplugs,

but on the level of the talker's voice. As an illustration, the talkers without earplugs in background noise at 70 dB SPL and the talkers with earplugs in background noise at 100 dB SPL had overall speech levels of 73.8 and 73.1 dB SPL, respectively, and corresponding SCoGs of 675 and 690 Hz. These measures were similar despite the fact that they were obtained in different ear conditions. Previously, Martin *et al.* (1976) found no differences in the octave-band speech spectra of talkers with and without HPDs (V51-R earplugs and Amplivox Sonogard earmuffs). These data suggest that earplug attenuation and the occlusion effect, in and of themselves, do not cause a talker to alter his or her speech spectrum. Instead, the changes in the speech spectrum are driven by the changes in the overall level of the talker's voice.

Thus, talkers with and without earplugs should be equally intelligible in background noise, provided their overall speech levels are the same. Unfortunately, the desired effect of wearing earplugs, the attenuation of ambient noise, creates an undesirable effect, the reduction of overall speech levels. This lowers the SNR and reduces the potentially beneficial changes in spectral shape that occur with an increase in the overall speech level. As a result, talkers wearing earplugs in background noise provide less speech information to listeners than talkers without earplugs.

Small but consistent differences in the speech measures were observed between the earplug groups. The mean overall speech levels, SNRs, SCoGs, and SII values of the foam earplug group were slightly lower than those of the flange earplug group. These data suggest that, within the constraints of this study, talkers wearing foam earplugs may be slightly less intelligible than talkers wearing flange earplugs. However, the small differences between the earplug groups in this study would probably be of little or no consequence in a real-world setting, in which many other variables impact the communication process.

Although it was not a significant main effect in any of the analyses that were conducted, the factor of testing order appeared in a small number of significant interactions with the factor of background noise level. Because its appearance was not consistent across dependent measures, and because the pattern of each dependent measure across background noise levels was very similar for both testing orders, it did not affect the interpretation of the results. The interactions involving the factor of sex for the SII were unexpected. However, the largest differences between the SII values for men and women, which occurred in the background noise levels of 60, 70, and 80 dB, were on the order of several hundredths and thus are not predictive of substantial differences in intelligibility between men and women. The significant main effect of sex and its interactions for the SCoG were expected due to the inherent differences in the speech spectra of men and women.

The relationship between the attenuation and occlusion effect caused by HPDs on the one hand, and speech production on the other, is often mentioned in passing in the literature. In order to examine this relationship more thoroughly, the speech measurements were reanalyzed with attenuation and occlusion effect included as covariates. If, after the effects of attenuation and occlusion were removed, the main effect of ear condition, its interaction with background noise level, or both were no longer significant, this indicated that one or both of the covariates accounted for the differences in the dependent measures between the ear conditions. In that case, follow-up ANOVAs were then conducted with each covariate separately. This reanalysis revealed that the combination of the earplug attenuation and occlusion effect mediated the differences in the overall speech levels, SNRs, and SII values between the ear conditions. Earplug attenuation alone mediated the differences in the SCoG between the ear conditions. However, these results do not imply a causal relationship. Instead, it is reasonable to assume that the earplug attenuation and the occlusion effect changed the talkers' perception of the ambient noise level, and perhaps the loudness of their own voices relative to the ambient noise level. As a result, the talkers adjusted the level of their voices, which in turn drove the differences in the SNR, SCoG, and SII between the ear conditions.

Previously, it was shown that the mean overall speech levels, SNRs, SCoGs, and SII values of the talkers wearing foam earplugs were slightly lower than those of the talkers wearing flange earplugs. It is tempting to ascribe these differences to the greater attenuation provided on average by the foam earplugs. However, a similar reanalysis of the data between the earplug groups revealed that neither the earplug attenuation nor the occlusion effect accounted for the differences in any of the speech measures between the foam and flange earplug groups. It is not clear whether other factors were responsible for the differences in the speech measures between the earplug groups, or if perhaps the PAR was not a sensitive enough measure to account for the differences. Further research is necessary to evaluate the effects of attenuation and the occlusion effect on the overall level of talkers' speech when HPDs are worn.

B. Communication in the workplace

Communicating in a noisy environment can be frustrating even for people with normal hearing who are not wearing HPDs. One of the greatest challenges in hearing conservation is to protect workers' hearing without further impacting their ability to communicate. Unfortunately, these two goals sometimes appear to be at odds.

The results of this study suggest that talkers wearing earplugs alter their speech in ways that are deleterious to intelligibility. However, several variables in addition to those manipulated in this study may also contribute to the success or failure of speech communication in a given situation. For this reason, several caveats should be kept in mind. First, the SII was calculated under the assumption that potential listeners would have normal hearing and would not be wearing HPDs. In a noisy work environment, listeners may have noise-induced hearing loss and/or may wear HPDs. The combined effects of cochlear distortion and audibility loss in a listener would probably further reduce the intelligibility of the talker. Second, the contributions of talker motivation, feedback from a conversation partner, and visual and contextual cues, were not assessed in this study. However, these factors would probably improve the intelligibility of talkers

in a real-world setting. Third, the speech measures obtained from talkers producing connected speech while wearing foam or flange earplugs may not be generalizable to studies using other types of speech, such as single words or short phrases, or to other types of HPDs, such as earmuffs or flat-attenuation earplugs. In addition, the speech measures may have been affected by the characteristics of the background noise presented to the talkers, the size and absorptive characteristics of the test room, and the talker–listener distance implied by the distance of the microphone from the talker. Further research is necessary to evaluate the effects of these additional variables on speech production. Therefore, in generalizing the results of this study to the real world, the speech measures are best viewed in terms of the trends they reveal across background noise levels and ear conditions, not in terms of their absolute values.

A thorough assessment of communication problems in the workplace should consider not only the listener, but also the talker. One way to improve workplace communication may be to train workers to raise their voices when wearing HPDs. However, because changes in one's speech level occur unconsciously in response to changes in the perceived ambient noise level, this would be very difficult (Casali *et al.*, 1987). HPDs that provide different attenuation characteristics may offer another solution. For example, in environments with sufficiently low noise levels, the use of HPDs having minimal attenuation may induce talkers to raise their voices while also eliminating the need to remove the HPDs to hear what others are saying. More generally, a research strategy that combines qualitative and quantitative methodologies may be the best way to pinpoint and evaluate workplace communication problems. Examining the conditions in particular work environments (e.g., noise characteristics, HPD types, talker–listener distance, available visual cues, etc.) could lead to the development of unique strategies for improving communication in these settings.

V. CONCLUSIONS

In quiet, the talkers' mean overall speech levels, SCoGs, and SII values were very similar with and without earplugs. In the presence of background noise, talkers automatically raised the level of their voices, as expected. Further, the relative amount of high-frequency energy in their speech increased, as evidenced by the upward shift of the SCoG. However, the mean SNR and SII decreased markedly, indicating that less and less speech information was available to a potential listener as the background noise level increased. These trends occurred regardless of whether or not the talkers were wearing earplugs, and are consistent with previous research on the effects of noise on speech production and intelligibility.

Although the trends in the speech measures were similar with and without earplugs, the differences that were observed would be of potential significance for communication. The overall speech levels and SNRs were 4 to 11 dB lower when the talkers wore earplugs, compared with the open-ear condition. In addition, the upward shift of the SCoG was much smaller. Finally, in moderate to high background noise levels (60 to 90 dB SPL), less speech information was available from the talkers when they wore earplugs, as quantified by the SII. The differences that were observed in the speech measures between the foam and flange earplug groups were statistically significant, but small in magnitude and probably of little real-world significance. No consistent differences in the speech measures were noted between men and women with the exception of the SCoG.

The results of the study indicate that talkers wearing earplugs (and consequently their listeners) are at a disadvantage when communicating in noise. The challenge to hearing conservationists is to reconcile the twin goals of protecting workers' hearing and preserving their ability to communicate.

ACKNOWLEDGMENTS

This research was conducted as part of a doctoral dissertation under the supervision of Tom Frank. It was supported by a grant from the Grants and Contracts Office of the College of Health and Human Development at Penn State University, by matching funds from the Department of Communication Disorders at Penn State University, and by HHS/CDC U60/CCU315855-02. The authors are grateful for the advice and assistance offered by Ingrid Blood, Marjorie Leek, Kevin Michael, Toby Mordkoff, and Robert Prosek. The authors would also like to thank the staff of the Research Section of the Army Audiology & Speech Center at the Walter Reed Army Medical Center and three anonymous reviewers for their helpful comments on an earlier version of this manuscript. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

[1]The mean values of the occlusion effect created by the earphones at 0.25, 0.5, and 1 kHz were 20.2, 4.2, and 2.1 dB, respectively. At the same test frequencies, the mean values of the occlusion effect created when the earplugs were worn without the earphones were 22, 10.8, and 6.9 dB, respectively (collapsed over earplug type). When both earplugs and earphones were worn, the mean values of the occlusion effect were 23.2, 11.8, and 6.6 dB, respectively. The occlusion effect created by wearing both earphones and earplugs was similar in magnitude to the occlusion effect created by the earplugs alone. Each of the occlusion effects mentioned above was calculated relative to a completely unoccluded ear canal, with no earplugs or earphones worn. However, the occlusion effect of interest here was that created by wearing the earphones and earplugs together relative to the earphone-only condition. These values, reported in the text, were obtained by subtracting, at each test frequency, the BCT with earplugs and earphones from the BCT with earphones only. Alternatively, these values could be obtained by subtracting, at each test frequency, the occlusion effect due to the earphones alone from the occlusion effect due to wearing the earplugs and earphones (with very minor corrections for rounding errors). Additional measurements of the talkers' speech in quiet were made to assess the impact of the occlusion effect due to the earphones themselves on the overall speech levels. These measurements were made while the talkers wore neither the earplugs nor the earphones, and while they wore the earplugs without the earphones. The mean overall speech level when neither earplugs nor earphones were worn was 66.4 dB, compared with a mean overall speech level of 67.5 when earphones only were worn. When the earplugs were worn, the mean overall speech levels were 66.9 and 67.0 dB with and without earphones, respectively. The fact that the talkers were wearing earphones during the test session appears to have had minimal to no impact on their overall speech levels.

[2]This explanation was suggested by Elliott Berger, E-A-R/Aearo Company, Indianapolis, IN.

Abel, S. M., Alberti, P. W., Haythornthwaite, C., and Riko, K. (**1982**). ''Speech intelligibility in noise with and without ear protectors,'' in *Personal Hearing Protection in Industry*, edited by P. W. Alberti (Raven, New York), pp. 371–385.
ANSI (**1991**). ANSI S3.1-1991, ''American National Standard: Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms'' (American National Standards Institute, New York).
ANSI (**1996**). ANSI S3.6-1996, ''American National Standard: Specifications for Audiometers'' (American National Standards Institute, New York).
ANSI (**1997**). ANSI S3.5-1997, ''American National Standard: Methods for Calculation of the Speech Intelligibility Index'' (American National Standards Institute, New York).
Bauman, K. S., and Marston, L. E. (**1986**). ''Effects of hearing protection on speech intelligibility in noise,'' Sound Vib. **20**, 12–14.
Berger, E. H. (**1980**). *EARLog 3: The Effects of Hearing Protectors on Auditory Communications* (Aearo Company, Indianapolis, IN).
Bond, Z. S., Moore, T. J., and Gable, B. (**1989**). ''Acoustic-phonetic characteristics of speech produced in noise and while wearing an oxygen mask,'' J. Acoust. Soc. Am. **85**, 907–912.
Bradley, J. S. (**1986**). ''Predictors of speech intelligibility in rooms,'' J. Acoust. Soc. Am. **80**, 837–845.
Casali, J. G., Horylev, M. J., and Grenell, J. F. (**1987**). ''A pilot study on the effects of hearing protection and ambient noise characteristics on intensity of uttered speech,'' in *Trends in Ergonomics/Human Factors IV*, edited by S. S. Asfour (Elsevier Science, North Holland), pp. 303–310.
Chung, D. Y., and Gannon, R. P. (**1979**). ''The effect of ear protectors on word discrimination in subjects with normal hearing and subjects with noise-induced hearing loss,'' J. Am. Aud. Soc. **5**, 11–16.
Dreher, J. J., and O'Neill, J. J. (**1957**). ''Effects of ambient noise on speaker intelligibility for words and phrases,'' J. Acoust. Soc. Am. **29**, 1320–1323.
Foreshaw, S. E., and Cruchley, J. I. (**1982**). ''Hearing protector problems in military operations,'' in *Personal Hearing Protection in Industry*, edited by P. W. Alberti (Raven, New York), pp. 387–402.
Helmkamp, J. D. (**1986**). ''Why workers do not use hearing protection,'' Occup. Health Saf. **55**, 52.
Hormann, H., Lazarus-Mainka, G., Schubeius, M., and Lazarus, H. (**1984**). ''The effect of noise and the wearing of ear protectors on verbal communication,'' Noise Control Eng. J. **23**, 69–77.
Howell, K., and Martin, A. M. (**1975**). ''An investigation of the effects of hearing protectors on vocal communication in noise,'' J. Sound Vib. **41**, 181–196.
Huynh, H., and Feldt, L. S. (**1976**). ''Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs,'' J. Educ. Stat. **1**, 69–82.
Junqua, J. (**1993**). ''The Lombard reflex and its role on human listeners and automatic speech recognizers,'' J. Acoust. Soc. Am. **93**, 510–524.
Kryter, K. D. (**1946**). ''Effects of ear protective devices on the intelligibility of speech in noise,'' J. Acoust. Soc. Am. **18**, 413–417.
Lane, H., and Tranel, B. (**1971**). ''The Lombard sign and the role of hearing in speech,'' J. Speech Hear. Res. **14**, 677–709.
Lazarus, H. (**1990**). ''New methods for describing and assessing direct speech communication under disturbing conditions,'' Environ. Int. **16**, 373–392.
Letowski, T., Frank, T., and Caravella, J. (**1993**). ''Acoustical properties of speech produced in noise presented through supra-aural earphones,'' Ear Hear. **14**, 332–338.
Lindeman, H. E. (**1976**). ''Speech intelligibility and the use of hearing protectors,'' Audiology **15**, 348–356.
Margolis, R. H., and Heller, J. W. (**1987**). ''Screening tympanometry: Criteria for medical referral,'' Audiology **26**, 197–208.
Martin, A. M., Howell, K., and Lower, M. C. (**1976**). ''Hearing protection and communication in noise,'' in *Disorders of Auditory Function*, edited by S. D. G. Stephens (Academic, New York), pp. 47–62.
McDaniel, D. M., and Cox, R. M. (**1992**). ''Evaluation of the Speech Intelligibility Rating (SIR) test for hearing aid comparisons,'' J. Speech Hear. Res. **35**, 686–693.
Michael, K. (**1999**). ''Measurement of insert-type hearing protector attenuation on the end-user: A practical alternative to rely on the NRR,'' Spectrum **16**, 13–17.
National Institute for Occupational Safety and Health (NIOSH) (**1996**). *Preventing Occupational Hearing Loss—A Practical Guide*, DHHS (NIOSH) Publication No. 96-110 (U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control, National Institute for Occupational Safety and Health, Cincinnati, OH).
Navarro, R. (**1996**). ''Effects of ear canal occlusion and masking on the perception of voice,'' Percept. Mot. Skills **82**, 199–208.
Nixon, C. W., and Berger, E. H. (**1991**). ''Hearing protection devices,'' in *Handbook of Acoustical Measurements and Noise Control*, 3rd ed., edited by C. M. Harris (McGraw-Hill, New York), pp. 21.1–21.24.
Pittman, A. L., and Wiley, T. L. (**2001**). ''Recognition of speech produced in noise,'' J. Speech Lang. Hear. Res. **44**, 487–496.
Rink, T. L. (**1979**). ''Hearing protection and speech discrimination in hearing-impaired persons,'' Sound Vib. **13**, 22–25.
Royster, L. H., and Holder, S. R. (**1982**). ''Personal hearing protection: Problems associated with the hearing protection phase of the hearing conservation program,'' in *Personal Hearing Protection in Industry*, edited by P. W. Alberti (Raven, New York), pp. 447–470.
Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (**1988**). ''Effects of noise on speech production: Acoustic and perceptual analyses,'' J. Acoust. Soc. Am. **84**, 917–928.
Suter, A. H. (**1992**). *Communication and Job Performance in Noise: A Review* (American Speech-Language-Hearing Association, Rockville, MD), Monograph No. 28.
Tartter, V. C., Gomes, H., and Litwin, E. (**1993**). ''Some acoustic effects of listening to noise on speech production,'' J. Acoust. Soc. Am. **94**, 2437–2440.
Traunmüller, H., and Eriksson, A. (**2000**). ''Acoustic effects of variation in vocal effort by men, women, and children,'' J. Acoust. Soc. Am. **107**, 3438–3451.
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.
Williams, D., and Michael, K. (**1991**). ''Maximizing communication ability in the selection of hearing protective devices,'' J. Acoust. Soc. Am. **89**, 1905.

Squared error as a measure of perceived phase distortion[a)]

Harald Pobloth[b)] and W. Bastiaan Kleijn
Department of Signals, Sensors, and Systems KTH (Royal Institute of Technology), 100 44 Stockholm, Sweden

(Received 19 July 2002; accepted for publication 26 May 2003)

Based on two well-known auditory models, it is investigated whether the squared error between an original signal and a phase-distorted signal is a perceptually relevant measure for distortions in the Fourier phase spectrum of periodic signals obtained from speech. Both the performance of phase vector quantizers and the direct relationship between the squared error and two perceptual distortion measures are studied. The results indicate that for small values the squared error correlates well to the perceptual measures. However, for large errors, an increase in squared error does not, on average, lead to an increase in the perceptual measures. Empirical rate-perceptual distortion curves and listening tests confirm that, for low to medium codebook sizes, the average perceived distortion does not decrease with increasing codebook size when the squared error is used as encoding criterion. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1594190]

PACS numbers: 43.72.Gy, 43.66.Nm, 43.71.Es [DOS]

I. INTRODUCTION

This article addresses issues related to the encoding of the Fourier phase spectrum of periodic signals obtained from speech. In particular, the objective is to

(i) Specify how well squared errors predict perceived distortion caused by phase modifications.
(ii) Evaluate and explain the perceptual performance of phase vector quantizers described in the literature, e.g., Gottesman (1999); Jiang and Cuperman (1995); Thomson (1988).

For a discrete time signal $s[n]$, the Fourier coefficients describing a short-time Fourier spectrum at frame v are

$$S^v(\omega)=\sum_{n=-\infty}^{\infty} u[n-vn_0]s[n]e^{-j\omega n}, \tag{1}$$

where $u[n]$ is a window, and n_0 is the frame advance in samples. Equation (1) is commonly referred to as the short-time Fourier transform (STFT). In the present context, a rectangular window with a support of N samples is used, where N coincides with the pitch period of the signal. The amplitude spectrum $a^v(\omega)$ and the phase spectrum $\phi^v(\omega)$ are defined by the relation $S^v(\omega)=a^v(\omega)e^{j\phi^v(\omega)}$. The phase $\phi^v(\omega)$ can be split into linear phase and dispersion phase (Gottesman, 1999; Jiang and Cuperman, 1995). Linear phase represents the circular shift of a signal segment with respect to a reference, and dispersion phase is what remains after subtracting the linear phase. For a periodic signal linear phase corresponds to the window position, while dispersion phase is independent of the window position.

For nearly periodic signals, such as voiced speech, the Fourier domain provides a sparse representation since relatively few coefficients, $S^v(\omega)$, are needed to describe the periodic component. As a result, it is common to use a sequence of short-time Fourier spectra as a basis for encoding speech and audio signals. Such a Fourier spectrum-based representation is used in, e.g., sinusoidal coders (McAulay and Quatieri, 1986b; Quatieri and McAulay, 1986), waveform interpolation (WI) coders (Burnett and Bradley, 1995; Jiang and Cuperman, 1995; Kleijn and Haagen, 1995), and audio coders (Purnhagen, 1999; Purnhagen and Meine, 2000). For STFTs with a window size of one pitch period the amplitude spectrum is generally considered to be perceptually more important than the phase spectrum (Kim, 2001; Ma and O'Shaughnessy, 1994; Moriya and Honda, 1986; Plomp and Steeneken, 1969).

Research on the perceptual relevance of the Fourier phase spectrum started in the 19th century. The first reference usually named is von Helmholtz (1859). In their historical review Plomp and Steeneken (1969) argue that the equipment used in early experiments, e.g., Hermann (1894); König (1881); Lindig (1903) was not sufficiently precise to study phase. Chapin and Firestone (1934) among others associate the audibility of phase distortions with nonlinear overtone effects in the auditory system. Later experiments, e.g., Goldstein (1967); Licklider (1957); Maths and Miller (1947); Patterson (1987); Plomp and Steeneken (1969), leave little doubt that the auditory system is sensitive to phase itself.

From their comprehensive investigation to quantify the effects of phase distortion on the timbre of periodic complex signals, Plomp and Steeneken (1969) conclude that low-pitch signals suffer more from phase distortion than high-pitch signals, that the significance of phase differs between subjects, and that the effect of phase distortion is only weakly dependent on sound-pressure levels. They also conclude that the perceived distortion stemming from phase can be as strong as the difference between perceptually close vowel sounds, but phase distortions do not change the sounds such that one vowel is mistaken for another.

[a)]Part of this work has been published in "Squared Error as a Measure of Phase Distortion," Proceedings Eurospeech, Aalborg, Denmark, September 2001.
[b)]Electronic mail: harald.pobloth@s3.kth.se

To address issues related to phase spectrum encoding, it is important to understand the structure of the relevant coders. In sinusoidal coding, the reconstructed signal in frame v is modeled as (Quatieri and McAulay, 1989)

$$\hat{s}^v(t) = \sum_m a^v(\omega[m],t)\cos(\varphi^v(\omega[m],t)), \tag{2}$$

where $a^v(\omega[m],t)$ denotes a time-dependent amplitude and $\varphi^v(\omega[m],t)$ is a time-dependent phase. In early sinusoidal coding, $\varphi^v(\omega[m],t)$ often was represented using a cubic polynomial that fit the conditions at frame boundaries (McAulay and Quatieri, 1986b). To facilitate modeling and coding, the phase φ^v is commonly split into $\varphi^v(\omega[m],t) = m\omega_v(t-t_0)+\psi^v[m]$, e.g., Quatieri and McAulay (1989) where ω_v is the estimated pitch frequency, t_0 is a time offset, and $\psi^v[m]$ is the so-called phase offset. The time offset results in the linear phase $-m\omega_v t_0$, chosen to align the pitch pulses in voiced speech, and $\psi^v[m]$ is the remaining dispersion phase.

For coding it is most efficient if the transmission of phase information can be avoided by approximating the structure of the phase spectrum with a model. Quatieri and McAulay (1989) use a simple model with $\psi^v[m]=0$ for the voiced speech component and random $\psi^v[m]$ for the unvoiced speech component. Following Makhoul *et al.* (1978) the voiced and unvoiced components are disjoint in frequency, with the signal component above an encoded cutoff frequency being unvoiced and below that voiced. McAulay *et al.* (1990) argue that the coherence of the voiced component's phase spectrum removes the reverberation for voiced speech sounds, which was commonly observed in early sinusoidal coders. However, using zero phase results in an increase in so-called buzziness. This deleterious effect is reduced by replacing the zero-phase model for the voiced component by a minimum-phase model (McAulay and Quatieri, 1991), a phase structure that also can be obtained using linear prediction. The speech quality associated with minimum-phase models is generally considered to be sufficient for low-rate coders.

It is fair to state that no sinusoidal-modeling-based speech coders exist that provide transparent speech quality without the use of explicit information about the STFT phase spectrum. Some authors have argued that phase is a major source of perceived distortion (Hedelin, 1988; Thomson, 1988) in speech coders. This has led to coding systems that encode the phase spectrum, e.g., Gottesman (1999); Jiang and Cuperman (1995); Marques *et al.* (1990); McAulay and Quatieri (1986a); Thomson (1988); Trancoso *et al.* (1988). These systems generally use squared errors as fidelity measures for encoding.

The squared error has the well-known advantages that it is easy to compute, and easy to handle in optimization problems, such as vector-quantizer training. This study considers three different squared errors. The first two are the speech squared error, and the linear-prediction (LP) residual squared error, which are commonly used. The third is the perceptually weighted squared error (Gottesman, 1999). All three are measured between the original and phase-distorted (coded) signals and not directly between the phase spectra.

We study artificial vowels. These were generated by repeating one original or one distorted pitch cycle. The reason for studying these vowels is that they are close to voiced speech while avoiding the problems that occur when voiced speech is used. Voiced speech changes from one pitch cycle to the next. These changes are difficult to handle in experiments since: First, dispersion phase becomes dependent on the window position. Second, for a given window position a change of only the phase spectrum results in a distortion of the signal waveform that for a different window position requires a change of amplitude- and phase spectra. In addition, it is hard to find perceptual distortion measures as the ones in Sec. III for continuous speech.

Kubin *et al.* (1993) show that the LP residual phase spectrum, which is considered throughout this article, is irrelevant for unvoiced speech. This is supported by observations of McAulay *et al.* (1990), who note that the voiced component suffers from audible distortion when zero phase is used, while for unvoiced components no audible distortion is reported when random phase is used. Similarly, for WI coders it is observed that phase is perceptually most important for the so-called slowly evolving waveform. For example, in Gottesman (1999), phase encoding is only considered for the slowly evolving waveform. The slowly evolving waveform can be seen as the nearly periodic component of the speech.

In summary, the artificial vowels are used to overcome practical problems that arise from the random fluctuations in real speech. It can be argued that they correspond to the part of the speech signal for which phase is most relevant. This notion is supported by the fact that our results coincide with commonly observed behavior in phase vector quantization for speech. Similar signals were used in Plomp and Steeneken (1969); Pols *et al.* (1969).

The main question this article tries to address is whether the various flavors of squared error predict the perceived distortion caused by phase distortions. If this prediction is poor, the usage of a perceptually more relevant measure than the squared error may be useful. To determine the relation between squared error and perceived distortion, we need a quantitative measure of perceived distortion. The most reliable and least objectionable method to obtain such a measure is to perform listening tests. Since our experiments involved several million signal pairs, this procedure is impractical. Instead, we utilized two well-known auditory models, both designed to capture temporal effects, to find objective distance measures. The accuracy of the measures was validated by the listening tests of Sec. III C. The models are the auditory image model (AIM) (Patterson *et al.*, 1995), and the quantitative model of the ''effective'' signal processing in the auditory system (Dau *et al.*, 1996a). Two models were used to show that the results obtained are not an artifact of the particular model used.

Section IV B shows that increasing SNR does not necessarily correspond to less perceived distortion. In general, less distortion is perceived when encoding phase for high-pitch signals, a commonly observed phenomenon, e.g., Licklider (1957); Plomp and Steeneken (1969); Skoglund *et al.* (1997). The results in Sec. IV C provide a relation between

perceived distortion and required bit allocation for phase vector quantization. For low-to-medium rate quantizers (0 to 7 bit), the average perceived distortion does not decrease with increasing rate when squared errors are used in the encoding system for low-pitch signals. This result is verified by a listening test in Sec. IV D, and in Sec. V we indicate that differences between the squared error and the perceptual distortion measures in their reaction to time shifts within one auditory channel might be the reason for the observed behavior.

II. FRAMEWORK

A. Signals

As argued in the Introduction, it is reasonable to examine periodic artificial vowels to study phase. To obtain a number of artificial vowels, two databases were built. Both contain voiced pitch cycles; one of low pitch the other of high pitch. The periodic artificial vowel s_k^c was then formed by repeating the kth pitch-cycle's waveform s_k. For the auditory models of Sec. III it was sufficient to have s_k^c of 400 ms length to reach a stable internal representation. For the listening tests in Secs. III C 1 and IV D, s_k^c was chosen slightly longer. The two databases were obtained from TIMIT (Darpa, 1990) sampled at 8-kHz sampling frequency. The speech signals were divided into segments of one pitch cycle length using the algorithm proposed by Eriksson and Kleijn (1999). The algorithm contains a preprocessing that introduced some inaudible changes to the signal but allowed for a segmentation with minimum edge effects. The voiced segments were identified using a correlation criterion operating on the linear prediction (LP) residuals and stored in an intermediate database. To simplify the vector quantization used throughout this article, the pitch cycles from this intermediate database lying in a certain pitch range were normalized to a fixed pitch by means of interpolation (zero padding or truncating in the Fourier domain). The truncation removes some of the phase samples from the database. Otherwise, this interpolation did not change the effective distribution of the phase. The zero-padded components are not effective as they have zero amplitude.

In this manner two databases were obtained. The first contains normalized pitch cycles in the range between 80 and 120 Hz normalized to 100 Hz. The second contains pitch cycles in the range between 180 and 220 Hz normalized to 200 Hz. All pitch cycles in these databases were made zero mean and were energy normalized such that $1/N\Sigma_{n=0}^{N-1}s_k^2[n]=1$, where s_k is the kth pitch cycle in the database and N is the number of samples in one pitch cycle.

The waveform of s_k can be written as a function of the amplitude spectrum a_k and the phase spectrum ϕ_k

$$s_k[n]=\sum_{m=0}^{N-1} a_k[m]e^{j\phi_k[m]}e^{j2\pi(nm/N)}, \quad n\in[0,N-1]. \tag{3}$$

A phase-distorted pitch cycle $\hat{s}_k[n]$ had identical amplitude spectrum, but its phase spectrum was changed to $\hat{\phi}_k$, which introduced a phase distortion

$$\Delta\phi_k=\phi_k-\hat{\phi}_k. \tag{4}$$

The phase-distorted vowel $\hat{s}_k^c[n]$ was then formed by repeating the pitch-cycle waveform $\hat{s}_k[n]$.

B. Squared error

The squared errors used measure the distortion between the original and the phase-distorted signals (not between the original and the distorted phase spectra). We considered three different squared errors, measuring the error in the speech domain, in the linear-prediction (LP) residual domain, and in a weighted domain (Gottesman, 1999), respectively. All three were found as

$$\eta(a_k,\Delta\phi_k)=\sum_{m=0}^{N-1} w_k[m]|S_k[m]-\hat{S}_k[m]|^2, \tag{5}$$

where $S_k[m]$ and $\hat{S}_k[m]$ are the Fourier coefficients of the original and distorted pitch cycle, respectively.

The weighting function $w_k[m]$ distinguishes the three different errors. For the speech squared error η_s, $w_k[m]=1$. For the residual error η_r, $w_k(z)=|A_k(z)|^2$, where $A_k(z)$ is the tenth-order LP analysis filter[1] found for s_k^c via the autocorrelation method, e.g., Makhoul (1975). For the weighted squared error η_w (Gottesman, 1999)

$$w_k(z)=\left|\frac{A_k(z/\gamma_1)}{A_k(z/\gamma_2)}\right|^2, \tag{6}$$

where $\gamma_1=0.9$ and $\gamma_2=0.7$ were selected. The subscript of η indicates which weighting is used. Relations valid for all three squared errors are described using η without subscript.

Since $\hat{s}_k[n]$ only differs in its phase spectrum, Eq. (5) can be rewritten as

$$\begin{aligned}\eta(a_k,\Delta\phi_k)&=\sum_{m=0}^{N-1} w_k[m]a_k^2[m]|e^{j\phi_k[m]}-e^{j\hat{\phi}_k[m]}|^2\\&=4\sum_{m=0}^{N-1} w_k[m]a_k^2[m]\sin^2\left(\frac{\Delta\phi_k[m]}{2}\right).\end{aligned} \tag{7}$$

Considering the symmetry of the Fourier coefficients and assuming that the phases of the DC component ($m=0$) and at half the sampling frequency ($m=N/2$, for even N) are undistorted,[2] Eq. (7) becomes

$$\eta(a_k,\Delta\phi_k)=8\sum_{m=1}^{\lceil N/2\rceil-1} w_k[m]a_k^2[m]\sin^2\left(\frac{\Delta\phi_k[m]}{2}\right), \tag{8}$$

where $\lceil x\rceil$ is the closest integer greater or equal x.

We decompose the phase distortion $\Delta\phi_k$ into $\Delta\phi_k=\phi_l+\phi_d$. The linear phase ϕ_l is a measure for the cyclic time shift of the signal, and the dispersion-phase distortion ϕ_d is what remains after this time shift is removed. It is reasonable to transmit the linear phase minimizing η separately since it can be represented by a scalar. The optimal linear phase for s_k and $\Delta\phi_k$ is

$$\phi_{l_k}^{\text{opt}}=\arg\min_{\phi_l\in\Phi_l}\eta(a_k,\Delta\phi_k-\phi_l), \tag{9}$$

where Φ_l is the set of all linear phase vectors ϕ_l

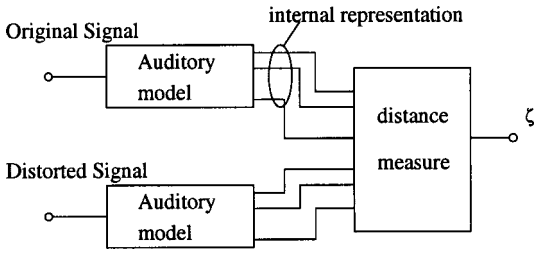


FIG. 1. Block diagram for the calculation of the perceptual distance measures using an auditory model.

$$\Phi_l = \{\phi_l : \phi_l[m] = \lambda m + 2\pi c$$

and (10)

$$\phi_l[N-m] = -\phi_l[m]\},$$

with $\lambda \in \mathcal{R}$, $c \in \mathcal{Z}$, $m \in [0, \lceil N/2 \rceil - 1]$, where $\mathcal{R}$ denotes the set of real numbers and $\mathcal{Z}$ denotes the set of integer numbers. We restricted linear phase to represent integer samples delay of the signal which helped to facilitate methods based on cross correlation (Jiang and Cuperman, 1995) to find $\phi_{l_k}^{\text{opt}}$.

Throughout this work, the phase-distorted signals were generated by replacing the original phase with entries from phase codebooks. All codebooks contain LP residual phase vectors. Thus, the coded phase $\hat{\phi}_{k,i}$ which appears when s_k is encoded with codebook entry ϕ_i^{CB} is

$$\hat{\phi}_{k,i} = \phi_i^{CB} - \phi_{A_k} + \phi_{l_{k,i}}, \tag{11}$$

where ϕ_{A_k} is the phase spectrum of the LP analysis filter $A_k(z)$, and $\phi_{l_{k,i}}$ is the linear phase. The dispersion phase distortion when ϕ_k is encoded with ϕ_i^{CB} is

$$\Delta\phi_{k,i} = \phi_k - \hat{\phi}_{k,i}. \tag{12}$$

III. AUDITORY MODELS AND PERCEPTUAL DISTANCE MEASURES

To evaluate if the squared error is capable of representing the perceived distortion introduced by phase alterations, the perceived distortion itself had to be measured. The most accurate way of doing that is by means of listening tests. For the results in Sec. IV, the perceived distortion for several million cases had to be found. Consequently, listening tests are not a feasible solution; instead, perceptual distance measures based on the signal representation of two auditory models were used as depicted in Fig. 1. One is the auditory image model (AIM) (Patterson *et al.*, 1995) and the other is the quantitative model of the ''effective'' signal processing in the auditory system (Dau *et al.*, 1996a). Consistent with, e.g., Kasper *et al.* (1997) and Tchorz *et al.* (1997), we refer to the latter model as PEMO (perception model). In Fig. 1 the original signal is the concatenation of $s_k[n]$ with amplitude a_k and phase ϕ_k, and the distorted signal is the concatenation of $\hat{s}_k$ with amplitude a_k and phase $\hat{\phi}_k$.

Both auditory models use a linear filterbank. Even though suggestions towards filters that give a better description of the phase response of the basilar membrane have been published recently, e.g., Oxenham and Dau (2001), we used the simpler linear filters. This is motivated by the successful prediction of the effect of time-domain changes in a signal and of the effect of target phase in a noise masker (Dau *et al.*, 1996b).

TABLE I. Parameters for the AIM function gennap (left) and the PEMO (right). The second and third row represent the minimum and maximum center frequencies of the gamma-tone filterbanks, respectively. The fourth row gives the number of channels used for the AIM, and the fifth row gives the filter spacing in the PEMO in ERB. MaxLevel controls to which level in dB SPL the maximum input sample is normalized.

AIM parameter		PEMO parameter	
Name	Value	Name	Value
samplerate	8000	Sampling frequency	8000
mincf_afb	96	MinCF	30
maxcf_afb	3696	MaxCF	3800
channels_afb	24		
		BW	1.0
		MaxLevel	80 dB SPL

The distance measures used were chosen from a set of candidate measures as the ones that gave the best fit to a series of preliminary listening tests. This led to different distance measures for the AIM and PEMO. In Sec. III C we describe a listening test evaluating the performance of the selected candidates. It should be stressed that it was not intended to investigate which of the two models is better suited to predict phase distortions. We used two models solely to show that the results in Sec. IV are independent of the model used.

A. Auditory image model (AIM)

This section describes how we found the distance measure ζ_A between an original signal s_k^c and a distorted signal $\hat{s}_k^c$ using the AIM (Patterson *et al.*, 1995). First, the autocorrelation within the neural activity pattern (NAP) was found for both s_k^c and $\hat{s}_k^c$. The NAP represents the activity in regions of the inner hair cells along the basilar membrane and was generated using the software described by Patterson *et al.* (1995). Except for the gamma-tone filter parameters, which had to be adapted for the sampling frequency of 8 kHz, the default parameters were used. The gamma-tone filter parameters shown in Table I closely resemble the setup used in Patterson (1987).

Patterson *et al.* (1995) suggest two alternative methods to model the higher level (brain) processing of the signal. One is strobed temporal integration and the other is autocorrelation leading to correlograms. We used the latter since correlograms are insensitive to cyclic shifts of the input signal. This renders the distance measure insensitive to time delays between s_k^c and $\hat{s}_k^c$ (i.e., it is insensitive to linear phase). The autocorrelation of the NAPs in each of the 24 channels was found and normalized such that it is 1 at zero delay in each channel. Since the signals were periodic, the autocorrelation was similar throughout the entire NAP. Thus, we chose the autocorrelation for an N sample period in the center of the 400-ms NAP representation to calculate ζ_A, where N is the pitch period. The set of all NAP autocorrelation functions in the 24 channels is called a correlogram.

The perceptual distortion measure ζ_A is

$$\zeta_A(a_k,\phi_k,\hat{\phi}_k)=\frac{1}{N}\sum_{\iota=3}^{24}\left(\sum_{n=0}^{N-1}(c_{k,\iota}[n]-\hat{c}_{k,\iota}[n])^4\right)^{1/4}. \tag{13}$$

ζ_A is the L4 distance between the original NAP correlation in channel ι, $c_{k,\iota}[n]$, and the distorted NAP correlation in channel ι, $\hat{c}_{k,\iota}[n]$ summed over channel 3 to 24.

Due to the time dependency of the adaptive thresholding done in the NAP algorithm, the NAP of channels 1 and 2 could have a period that corresponds to double the pitch. Thus, it can happen that for the original signal a different half of the $2N$ sample period of the NAP is considered when finding the autocorrelation $c_{k,\iota}$ than for the distorted signal when finding $\hat{c}_{k,\iota}$. This type of misalignment led to an increase in ζ_A where it was not appropriate. Instead of aligning the NAPs in these channels, we omitted the first two channels. This did neither show any influence on the measure ζ_A nor on the correlation between ζ_A and the listening test in Sec. III C but is computationally more efficient than the alignment. This notion is confirmed by Kim (2001), who describes and proves that the phase of low-frequency components is irrelevant for the perception of periodic signals.

B. Perception model (PEMO)

The PEMO measure ζ_P was found from $p_{k,\iota}[n]$ and $\hat{p}_{k,\iota}[n]$, the internal representation of the PEMO (Dau *et al.*, 1996a) in channel ι of the original signal and the distorted signal, respectively. The PEMO was realized using software provided by Universität Oldenburg. Using the parameters (shown in Table I) suggested by the authors of this software, the internal representation consists of 26 auditory channels. It was not down-sampled after the final low-pass filtering. ζ_P is defined as

$$\zeta_P(a_k,\phi_k,\hat{\phi}_k)=26-\sum_{\iota=1}^{26}\max_l(\rho_{k,\iota}[l]),\quad 0\leqslant l<N. \tag{14}$$

With $N_m=\lceil 2/3N_t\rceil$ and N_t the number of samples in the artificial vowel, $\rho_{k,\iota}[l]$ is the Pearson correlation between $(p_{k,\iota}[N_m],p_{k,\iota}[N_m+1],\ldots,p_{k,\iota}[N_m+N-1])^T$ and $(\hat{p}_{k,\iota}[N_m+l],\ldots,\hat{p}_{k,\iota}[N_m+l+N-1])^T$.

Since the delay l can be different for each channel, the measure explicitly follows the common assumption, e.g., Kim (2000); Patterson (1987) that within-channel distortions are of higher importance than across-channel distortions. In fact, we neglect that strong across-channel distortions may contribute to the perceived distortion (Patterson, 1987, 1988). ζ_P is, as ζ_A, not sensitive to linear phase.

The similarity measure introduced in Hansen and Kollmeier (2000) is based on correlation as well. However, due to its operation on a block-by-block basis, it requires prealignment even for the periodic signals used here. With ζ_P no prealignment is necessary for the periodic signals. In addition, ζ_P gives better performance for the preliminary tests as well as for the listening test in Sec. III C than the measure in Hansen and Kollmeier (2000).

C. Evaluation

1. Listening test

To evaluate how well the perceptual measures ζ_A and ζ_P perform in predicting the perceived distortion for phase-distorted artificial vowels, a listening test was performed. A method-of-constants test (Levitt, 1971) was performed double blind with an AXB forced-choice procedure. The probabilities of positive response found are plotted as a function of the perceptual measures.

A triplet of signals A, X, B was presented to the listener, where A and X were random choices of either the original or the distorted signal, and B was the distorted signal if A was the original and vice versa. The listeners were forced to choose A=X or X=B. The signals were generated on a standard PC with a Creative AWE 64-GOLD sound card connected digitally to a digital audio tape (DAT) deck serving as A/D converter. The analog signal was then presented monaurally via Sennheiser HD 250 linear II headphones. The level for all listening tests was 79 dB SPL measured using an artificial ear (IEC 60318). The sound intervals A, X, and B were of 700 ms duration, separated by 1 s of silence.

The original signals were artificial vowels s_k^c generated by repeating randomly selected s_k. One pitch cycle with high pitch (211 Hz) and one with low pitch (94 Hz) were selected. They are from the databases of Sec. II A before pitch normalization was performed. The distorted $\hat{s}_k$ had an additive distortion phase $\Delta\phi$ and $\hat{s}_k^c$ was generated by repeating $\hat{s}_k$. The distortion was composed as $\Delta\phi=\alpha d$, where the factor $\alpha\geqslant 0$ controlled the distortion level, and d is a unit norm phase vector with random elements.

We define a run as all responses gathered for one original vowel with one d. The test consisted of three runs for the selected high-pitch vowel and two for the selected low-pitch vowel. Each run involved seven listeners. Two listeners were the same expert listeners throughout all runs. They acquired their expertise by performing a long series of preliminary tests. The other listeners were mainly employees at the department with no particular training for this type of distortion. Each of the inexperienced listeners was only used for one run of the test.

Each listener started with a short training sequence of five AXB triplets during which the listener was informed if his/her decision was correct. This was followed by a 1up–1down (Levitt, 1971) method. The results of the up–down method are not used in this work. Then, the method of constants with five predefined distortion levels α was started. During pilot tests performed by one of the experienced listeners, $\alpha_{\min}$ and $\alpha_{\max}$ that are likely to cover the entire range of possible correct-response probabilities from 50% to 100% were determined for each d. The triplet AXB for each αd was repeated $N_{\text{trial}}=24$ times in random order.

For a run involving L listeners, each correct response contributed with $\Delta p=1/(N_{\text{trial}}L)$ to the approximated probability of positive response $\bar{p}$. Figure 2 shows the probability of positive response $\bar{p}$ for all seven listeners from one run for the 94-Hz vowel. The vertical bars are the intervals at 95% confidence level as described in Srednicki (1988). Individual

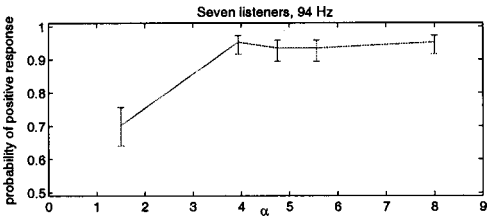


FIG. 2. $\overline{p}$ averaged over listeners as a function of distortion strength α for one run of the listening test with the 94-Hz vowel. The bars show the intervals at 95%-confidence level.

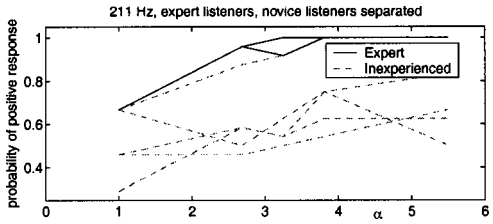


FIG. 3. $\overline{p}$ for the 211-Hz vowel for individual listeners. Solid lines: the two expert listeners; dashed lines: the five inexperienced listeners.

listeners' intervals are about $\sqrt{L}$ larger than the intervals across all listeners shown in the figure.

The experienced listeners gave more consistent responses than the inexperienced ones for the 211-Hz vowel, as seen in Fig. 3, a trend observed for all runs with the 211-Hz vowel. For all runs with the 211-Hz vowel, the $\overline{p}$ from the group of experienced listeners is higher than the $\overline{p}$ for the group of inexperienced listeners, and in 67% of all cases the confidence intervals of the two groups do not overlap. For the 94-Hz vowels experienced and inexperienced listeners were more consistent. Plomp and Steeneken (1969) used mainly high-pitch signals (292.4 Hz) and note that sensitivity to phase distortion varied among subjects. The difference in our results between untrained and trained subjects suggests that this variation can be reduced and a high sensitivity to phase distortions can be achieved by training. Nelson (1994) also reports increasing ability to detect phase distortions with training.

2. Predicting listening tests from perceptual distance measures

To compare the subjective results of the listening tests to the two perceptual measures ζ_A and ζ_P, the probabilities of positive response $\overline{p}$ are plotted as a function of the perceptual measures in Figs. 4 and 5. Since the difference due to experience observed in Sec. III C 1 leads to different dependencies between ζ and $\overline{p}$, the correlation between ζ and $\overline{p}$ is examined for experienced and inexperienced listeners separately.

In addition, we observed that the relation between ζ and $\overline{p}$ differs between high-pitch and low-pitch vowels. Since the perceptual measures are used with either high-pitch or low-pitch vowels within one experiment in Sec. IV, the experiments in Sec. IV do not require consistency of the measures between different pitch values. Thus, the sensitivity of the measures to pitch is not an obstacle in the framework of this article. It should be noted that this sensitivity is not necessarily a shortcoming of the PEMO or AIM; we can only conclude that the described combination of auditory models and perceptual distance measures is sensitive to pitch.

Accordingly, we split the data of the listening test into four parts: low pitch with inexperienced subjects, high pitch with inexperienced subjects, low pitch with experienced subjects, and high pitch with experienced subjects. This leads to eight plots (shown in Figs. 4 and 5) since each of the four parts is related to the AIM measure and the PEMO measure separately.

In practice, the difference between the original and distorted signal can become sufficiently audible to be detected with 100% certainty, but still different strengths of distortions are perceived. Thus, $\overline{p}$ saturates at 1, while ζ can increase further. The further analysis only takes pairs s of

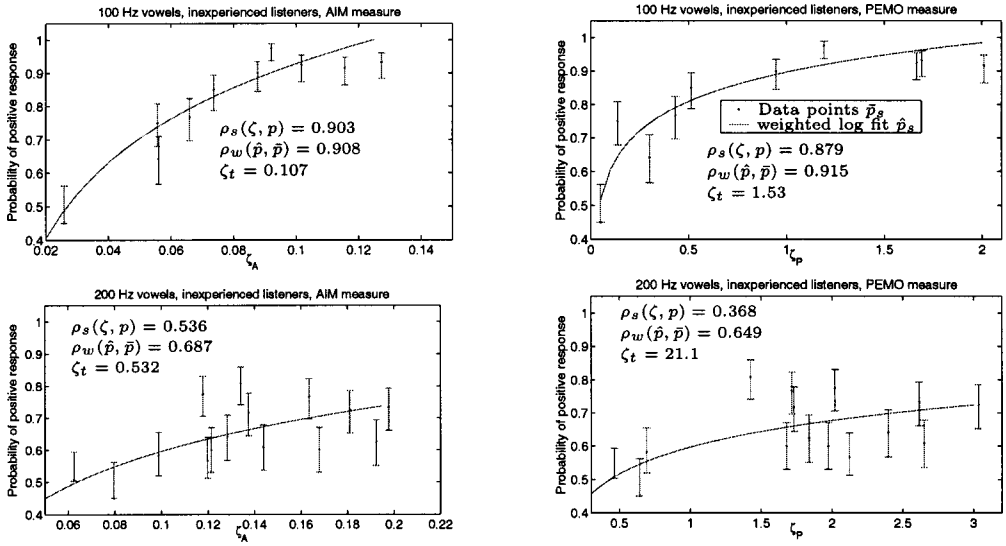


FIG. 4. Probability of positive response $\overline{p}$ for inexperienced listeners as a function of the AIM measure ζ_A (left) and the PEMO measure ζ_P (right). Upper graphs low-pitch vowels, lower graphs high-pitch vowels. ρ_S is the Spearman rank correlation coefficient. $\rho_w(\hat{p},\overline{p})$ denotes the weighted correlation coefficient with weights as in Eq. (17). ζ_t gives the threshold for which $\hat{p}(\zeta)=0.95$ [Eq. (15)].

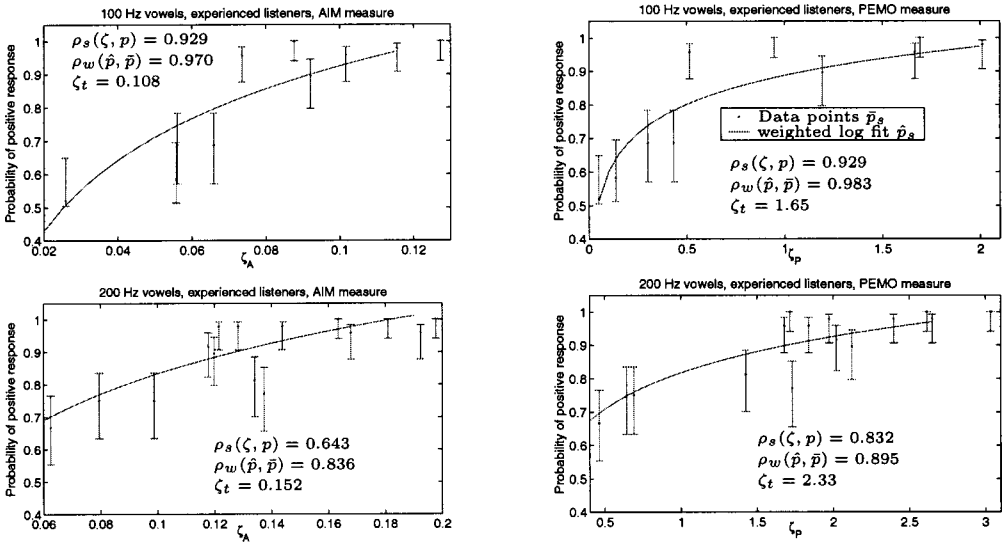


FIG. 5. Probability of positive response $\bar{p}$ for experienced listeners as a function of the AIM measure ζ_A (left) and the PEMO measure ζ_P (right). Upper graphs low-pitch vowels, lower graphs high-pitch vowels. ρ_S is the Spearman rank correlation coefficient. $\rho_w(\hat{p},\bar{p})$ denotes the weighted correlation coefficient with weights as in Eq. (17). ζ_t gives the threshold for which $\hat{p}(\zeta)=0.95$ [Eq. (15)].

original and distorted signals into account for which $\bar{p}_s < 1.0$, where $\bar{p}_s$ is $\bar{p}$ for pair s.

Given the saturation of $\bar{p}$ at 100%, a logarithmic relation is assumed between $\bar{p}$ and ζ and the prediction $\hat{p}_s$ of $\bar{p}_s$ is

$$\hat{p}_s = a + b \log(\zeta_s), \tag{15}$$

where a and b are regression coefficients, and ζ_s is the auditory measure for signal pair s.

Figures 4 and 5 show the logarithmic regressions minimizing the error

$$J_w = \sum_{s=1}^{N_s} (\bar{p}_s - \hat{p}_s)^2 w_s, \tag{16}$$

where N_s is the number of pairs in one of the four parts of the listening test for which $\bar{p}_s < 1.0$. The weights w_s are the square of the inverse length l_s of the intervals at 95% confidence level on $\bar{p}$ found as in Srednicki (1988)

$$w_s = \frac{1}{l_s^2}. \tag{17}$$

To check how well the results of the listening test can be predicted from the perceptual measures, Figs. 4 and 5 show the Spearman rank correlation coefficients between ζ_s and $\bar{p}_s$ and the weighted correlation coefficient between p and $\hat{p}$. The weighted correlation coefficient is found as

$$\rho_w = \frac{C(\bar{p},\hat{p})}{\sqrt{C(\bar{p},\bar{p})C(\hat{p},\hat{p})}},$$
$$C(\bar{p},\hat{p}) = \frac{\sum_{s=1}^{N_s}(w_s(\bar{p}_s - \tilde{p})(\hat{p}_s - \tilde{\hat{p}}))}{\sum_{s=1}^{N_s} w_s}, \tag{18}$$

where $\tilde{p} = (\sum_{s=1}^{N_s} w_s \bar{p}_s)/(\sum_{s=1}^{N_s} w_s)$. $\tilde{\hat{p}}$ is found as $\tilde{p}$, with $\bar{p}_s$ replaced by $\hat{p}_s$.

The correlations between $\bar{p}$ and $\hat{p}$ for the two auditory models are very similar. In general, the correlation between $\bar{p}$ and $\hat{p}$ is higher for low-pitch vowels than it is for high-pitch vowels, especially for the inexperienced listeners. This is consistent with the behavior of human subjects. Their decisions for high-pitch vowels were more random, especially for the inexperienced subjects.

To compare predictions from the two models and to compare the performance of experienced and inexperienced subjects, we arbitrarily choose 95% probability of positive response as threshold. In Figs. 4 and 5 ζ_A and ζ_P at $\hat{p} = 95\%$, are given. For high-pitch vowels and inexperienced listeners this common reference point of 95% probability of positive response is not reached by any observed $\bar{p}_s$ or predicted $\hat{p}_s$. Thus, we extrapolated $\hat{p}(\zeta)$ to 95%. The exact value of the extrapolated threshold might be unreliable since it is remote from the highest ζ_s observed and the variation among subjects is large.

The 95% thresholds for high-pitch vowels are higher than the threshold for the low-pitch vowels. They differ by a factor of 1.4 (both AIM and PEMO) for the experienced listeners and by factors of 5 (AIM) and 13.8 (PEMO) for the inexperienced listeners. Even though the exact factors for the inexperienced listeners are as unreliable as the extrapolation used to obtain them, there is a clear indication that inexperienced listeners have more difficulties to detect phase distortion in high-pitch signals. For low-pitch signals the threshold for experienced and inexperienced listeners is almost identical. For experienced subjects the difference between high pitch and low pitch is less for the monaural listening tests used here than what was reported earlier for binaural listening (Pobloth and Kleijn, 1999, 2001).

IV. EXPERIMENTS

In this section, the perceptual distortion measures ζ_A of Eq. (13) and ζ_P of Eq. (14) are used to investigate the relation between the squared error of Eq. (8) and the perceived

distortion. If the perceived distortion can be determined from the squared error, it is not necessary to use computationally expensive perceptual measures.

In Sec. IV B the average perceptual measures are shown as a function of the squared error. In Sec. IV C, the perceptual performance of phase vector quantizers (VQs) similar to the ones used in Gottesman (1999); Jiang and Cuperman (1995); Thomson (1988); Trancoso *et al.* (1988) is shown as a function of bit rate. The perceptual performance of VQs trained to minimize, respectively, the squared error and the perceptual measures is evaluated. Some of these results are verified by a listening test in Sec. IV D.

A. Codebook training

Since all experiments in Secs. IV B through IV D require trained codebooks, this section explains the training of dispersion-phase codebooks. It was possible to use a modified version of the generalized Lloyd algorithm (GLA) (Linde *et al.*, 1980) since constrained resolution VQs were used. Modifications are necessary since the error minimized is not the squared error between the phase vectors ϕ, but an error between the signals s_k and $\hat{s}_k$.

The aim is to find a codebook C containing a set of $N_{CB}=2^B$ LP residual dispersion-phase vectors ϕ^{CB}, that minimizes the mean error between the original speech and the encoded speech given a set of training data S. The error can be either one of the squared errors (η_s, η_r, η_w) or one of the perceptual distance measures (ζ_A, ζ_P). A codebook containing 2^B vectors is referred to as B-bit codebook.

The standard GLA iterates through two steps. It first associates each vector in the training set S to one codebook vector ϕ_i^{CB}. All training vectors associated to ϕ_i^{CB} build the cell S_i. In the second step the codebook vector ϕ_i^{CB} minimizing the error contribution of each cell S_i is found. The difference between the GLA used here and the conventional GLA (Linde *et al.*, 1980) is in the expression for the optimal codebook vector ϕ_i^{CB}.

1. Codebook training for the squared error

This section gives an expression for the codebook vector minimizing the squared error in cell S_i. The elements of the optimal codebook vector are

$$\phi_i^{CB}[m]=\arg\min_{\varphi[m]}\sum_{k\in\{k:s_k\in S_i\}}\eta(a_k[m],\Delta\phi_{k,i}[m]), \tag{19}$$

with $\Delta\phi_{k,i}[m]=\phi_{r_k}[m]-\varphi[m]-\phi_{l_{k,i}}[m]$, where $\phi_{l_{k,i}}[m]$ is the linear phase found to be optimal in the partitioning step of the GLA. $\phi_{r_k}[m]$ is the LP residual phase for pitch cycle s_k. The contribution of the phase distortion in dimension m to the squared error is

$$\eta(a_k[m],\Delta\phi_{k,i}[m])=w_k[m]a_k^2[m]\sin^2\left(\frac{\phi_{r_k}[m]-\varphi[m]-\phi_{l_{k,i}}[m]}{2}\right). \tag{20}$$

The sum over all s_k in S_i: $\Sigma_{k\in\{k:s_k\in S_i\}}\eta(a_k[m],\Delta\phi_{k,i}[m])$ has extrema at

$$\tan(\varphi[m])=B_i[m], \tag{21}$$

with

$$B_i[m]=\frac{\Sigma_{k\in\{k:s_k\in S_i\}}w_k[m]a_k^2[m]\sin(\phi_{r_k}[m]-\phi_{l_{k,i}}[m])}{\Sigma_{k\in\{k:s_k\in S_i\}}w_k[m]a_k^2[m]\cos(\phi_{r_k}[m]-\phi_{l_{k,i}}[m])}.$$

For a given $B_i[m]$ Eq. (21) has two solutions

$$\varphi_1[m]=\arctan(B_i[m])$$

$$\varphi_2[m]=\begin{cases}\arctan(B_i[m])+\pi & \text{for } \arctan(B_i[m])<0\\ \arctan(B_i[m])-\pi & \text{for } \arctan(B_i[m])\geqslant 0\end{cases}.$$

Thus, $\phi_i^{CB}[m]$ is selected as either $\varphi_1[m]$ or $\varphi_2[m]$, whichever leads to the minimum in Eq. (19).

The modified GLA converges due to the same reason as the original GLA. After both finding the cells S_i for given codebook vectors, and the optimal codebook vectors ϕ_i^{CB} for given cells, the average squared error between the encoded signals and the original signals is either reduced or remains unchanged. When it remains unchanged the algorithm has converged.

The size of the training sets S was $N_t=\max(6400,100N_{CB})$. Codebooks of sizes 0–10 bit were trained for each of the three squared errors and for 100-Hz and 200-Hz pitch cycles, giving a total of 66 codebooks.

2. Codebook training for the perceptual distance measures

The two iterative steps of the GLA were used to train resolution-constrained codebooks minimizing perceptual distance measures ζ_A or ζ_P. In the perceptual case, there is no analytical expression for the optimal codebook vector given a cell S_i. Instead, the Powell numerical minimization method as described in Press *et al.* (1992) was used to find the codebook vector $\phi_i^{CB}=\text{argmin}_\varphi\Sigma_{k\in\{k:s_k\in S_i\}}\zeta(a_k,\phi_k,\varphi-\phi_{A_k})$, where ϕ_{A_k} is the linear-prediction analysis-filter phase. Either ζ_A of Eq. (13) or ζ_P of Eq. (14) can be used.

In the perceptual case the GLA approaches to minimize the average perceptual distance

$$\bar{\zeta}=\frac{1}{N_t}\sum_{k\in S}\zeta(a_k,\phi_k,\hat{\phi}_{k,i(k)}), \tag{22}$$

where $\hat{\phi}_{k,i(k)}$ is given in Eq. (11). Here, $i(k)$ is the codebook index i minimizing $\zeta(a_k,\phi_k,\hat{\phi}_{k,i})$. The $\phi_{l_{k,i}}$ of Eq. (11) can be chosen to be an arbitrary linear phase since the perceptual measures are insensitive to linear phase for the artificial vowels. N_t is the size of the training set S.

The convergence of the perceptual training is seen by the same argument as the convergence of the squared-error training. It should be noted however, that, due to numerical minimization, the training algorithm for the perceptual distortion measures is more suboptimal than the squared-error GLA since global and local minima cannot be distinguished when finding ϕ_i^{CB}.

The perceptual measures and the numerical procedure increase the complexity of the algorithm. Thus, the size of S was always chosen as $N_t=100N_{CB}$. For the same reason,

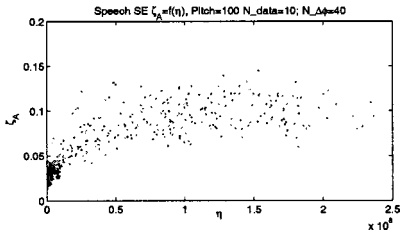


FIG. 6. Scatter plot of speech squared-error η_s and AIM distortion measure ζ_A using 100-Hz vowels. To make it possible to distinguish individual data points, only 400 randomly chosen signal pairs out of 204 600 are shown.

perceptual codebooks of sizes 0 to 5 bit (instead of up to 10 bit) were trained for both ζ_A and ζ_P and the 100-Hz and 200-Hz pitch cycles. This makes a total of 24 perceptual codebooks.

B. Squared error versus perceptual distance

In this section, we investigate how the average of the perceptual measures ζ_A and ζ_P evolves as a function of the average of the squared errors. To find the perceptual measures and squared errors, we defined a set of original and sets of distorted signals. For the original signal $N_p=200$ random pitch cycles s_k were selected for each of the two pitch frequencies. All of the original signals were distorted by replacing their phase by $\hat{\phi}_{k,i}$ according to Eq. (11) using all ϕ^{CB} stored in squared-error trained codebooks. Then, $\eta(a_k,\Delta\phi_{k,i})$ [Eq. (8)] and $\zeta(a_k,\phi_k,\hat{\phi}_{k,i})$ [Eq. (13) for the AIM and Eq. (14) for the PEMO] were found for all resulting signal pairs.

The codebooks were always used consistently. Thus, when measuring, e.g., the residual squared-error η_r, only codebooks trained for the residual squared error were used. We used trained codebooks to investigate distortions which appear during the exhaustive search in phase vector quantization. If η and ζ correlate well, it is likely that the same phase vector in a codebook gives the minimum for both η and ζ during such a search; in this case encoders minimizing a squared error and encoders minimizing a perceptual measure select the same codebook vector.

The size of the codebooks used was 0 to 7 bit. We observed that the squared errors for these codebooks are almost all in the saturated region of the graphs in Fig. 7. In this region the average of the perceptual measures does not decrease with decreasing squared error. Even when using larger codebooks (up to 10 bit) we observed most of the pairs in the saturated region. Thus, instead of using larger codebooks, the phase distortions $\Delta\phi$ obtained when using the 7-bit codebooks were scaled by the six arbitrary factors $\{0.5,0.3,0.2,0.1,0.05,0.03\}$. This resulted in a total of $N_p(6\cdot 2^7+\Sigma_{i=0}^{7}2^i)=204\,600$ signal pairs for 100 Hz and for 200 Hz, each. This number is a compromise between a high number of data points and computational complexity.

Both perceptual measures (ζ_A, and ζ_P) can be related to any of the three squared errors (η_s, η_r, and η_w). Figure 6 shows an example of a η, ζ scatter plot for 100-Hz pitch cycles. Figure 6 suggests that the average perceptual measure saturates for high squared errors. This was investigated by finding the average perceptual measure as a function of the squared error using the group average discussed in Ezekiel and Fox (1959). For the group average the set of all signal pairs was sorted with increasing η and then partitioned into groups with $N_g=3500$ signal pairs in each group. Then, the average η and ζ in each group r is found as $\tilde{\eta}[r]$ and $\tilde{\zeta}[r]$. The groups for high η do not contain pairs from all s_k. Finally, $\tilde{\eta}$ was normalized with

$$\eta_0=\frac{1}{N_p}\sum_k 4Na_k^Ta_k, \tag{23}$$

to be able to draw the results from all three different squared errors in one graph. η_0 normalizes the average of the maximum possible squared error, which appears for $\hat{s}_k=-s_k$, to 1.

1. Discussion of the results

Figure 7 shows that the average perceptual distance $\tilde{\zeta}[r]$ saturates in all cases; a further increase of the squared error does not, on average, lead to a higher perceptual distance. Thus, within codebooks providing vectors only in the saturated range of $\tilde{\zeta}[r]$ the vector minimizing the squared error cannot be expected to be the perceptual optimal choice. This was further investigated and confirmed in Sec. IV C.

To judge how audible the phase distortions from the codebooks are on average, the approximate values of the

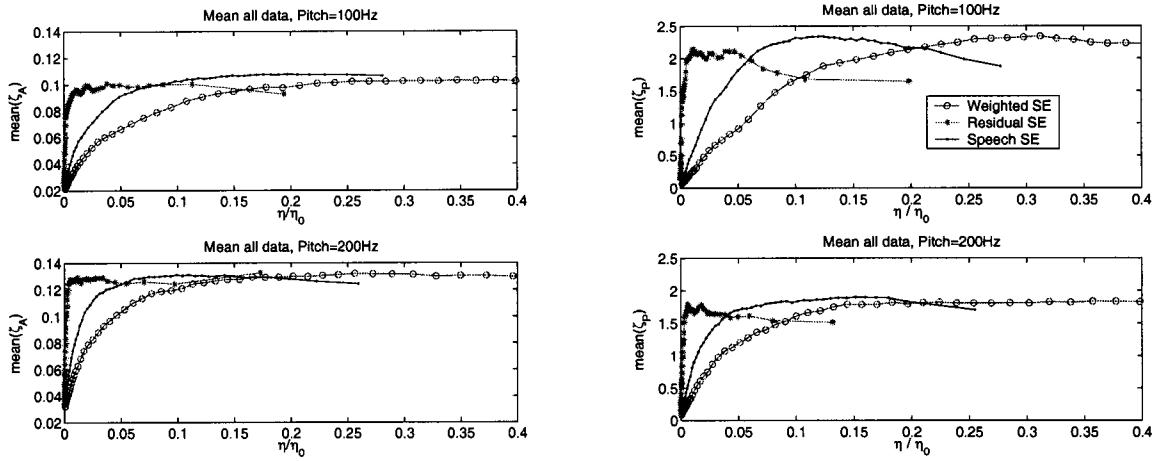


FIG. 7. Group average $\tilde{\zeta}[r]$ as a function of $\tilde{\eta}[r]/\eta_0$. AIM results left, PEMO results right, 100-Hz upper graphs, and 200-Hz lower graphs.

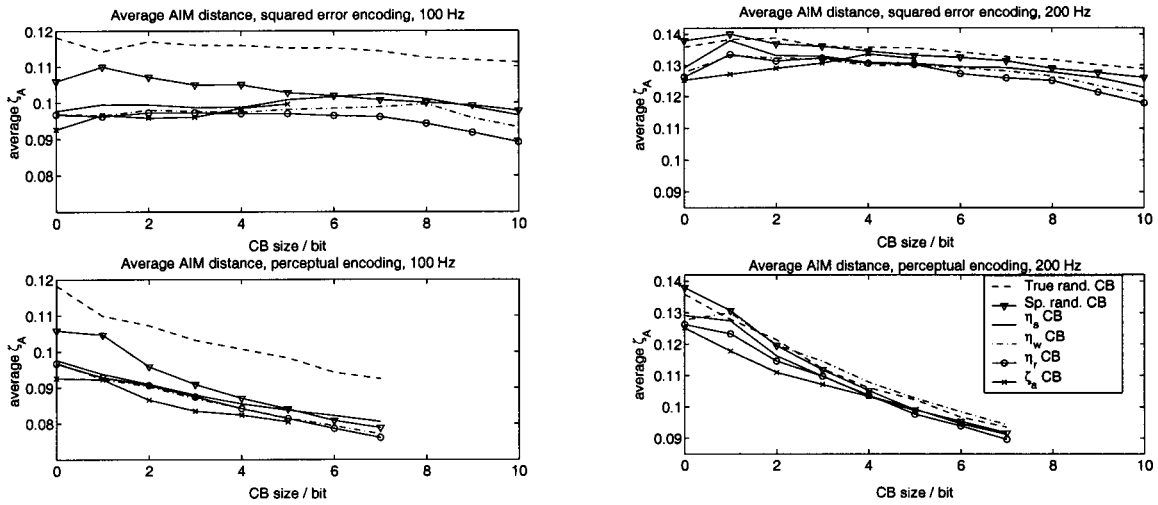


FIG. 8. Average AIM distance for 100-Hz vowels (left) and 200-Hz vowels (right). Upper graph: squared-error encoding. The true random CB is encoded using η_s, the speech-random CB using η_r, and the perceptual CB using η_r for 100 Hz and η_s for 200 Hz. Lower graph: perceptual encoding.

saturated average perceptual measures in Fig. 7 were related to the threshold ζ_t for 95% correct responses found in Sec. III C 2. Using the threshold for the experienced listeners given in Figs. 4 and 5, it is seen that for 100 Hz the saturated $\tilde{\zeta}/\zeta_t$ is 93% (133%) for the AIM (PEMO) measure, and for 200 Hz the saturated $\tilde{\zeta}/\zeta_t$ is 86% (81%) for the AIM (PEMO). Since the relation between the probability of positive response and the perceptual measures is nonlinear, and differs for ζ_A and ζ_P, the observed differences when averaging ζ_A or ζ_P are not unexpected. Still, the average audibility of phase distortions caused by the codebooks is slightly higher for 100 Hz.

Finding the same relations for the inexperienced listeners gives different results only for the 200-Hz signals, where $\tilde{\zeta}_A$ and $\tilde{\zeta}_P$ saturate far below the thresholds. Here, the saturated $\tilde{\zeta}/\zeta_t$ is 24.4% (9%) for the AIM (PEMO). The differences between AIM and PEMO are now due to both the different nonlinear relations between $\overline{p}$ and ζ, and the fact that the extrapolation towards the threshold values is likely to be imprecise. Still, the average perceptual distance for 200-Hz vowels is so far away from the threshold that the phase distortions can be considered inaudible for the inexperienced listeners, while phase distortions in the 100-Hz signals are likely to be audible.

The standard deviation around $\tilde{\zeta}$ in Fig. 7 saturates at around 18% (30%) of $\tilde{\zeta}$ for the AIM (PEMO). From the results in this section, it is not obvious which of the squared-error measures is more suitable to measure the perceived distortion. The weighted squared error leads to higher $\tilde{\eta}/\eta_0$ than η_r or η_s, which should compensate for the fact that it saturates at higher $\tilde{\eta}/\eta_0$.

C. Vector quantizers (VQs)

We evaluate the perceptual performance of phase VQs, similar to the quantizers used in, e.g., Gottesman (1999); Jiang and Cuperman (1995); Thomson (1988). For the evaluation we found empirical rate-perceptual distortion curves. They show the average perceptual distance of Eq. (22) when the dispersion phase of sets of s_k are encoded with codebooks of different size. The sets S consisted of pitch cycles s_k chosen randomly from the pitch-cycle databases of Sec. II A and were disjoint from the ones used for training. They were of size $N_t = \max(6400; 100N_{CB})$, where N_{CB} is the codebook size.

Using different codebook types we found how much improvement slow and difficult perceptual codebook (CB) training gains compared to fast squared-error CB training and compared to random codebooks. Two different random codebooks were used, true random with phase vectors containing independent uniformly distributed phase values between $[-\pi,\pi)$, and speech random with LP residual phase vectors ϕ_{r_k} from randomly selected s_k. The trained codebooks were trained for the three different squared errors η_s, η_r, and η_w, and the perceptual measures ζ_A and ζ_P. All codebooks contained phase vectors stored in the LP residual domain.

By encoding using the perceptual measures as encoding criterion or the squared error as encoding criterion, it was found how suboptimal the one method is compared to the other. The upper plots in Figs. 8 and 9 show the average perceptual distance as a function of CB size when ϕ_i^{CB} is chosen to minimize the squared error (squared-error encoding). The measures (η_s, η_r, and η_w) are chosen consistently for the squared-error trained CBs, e.g., for the speech squared-error trained CB the speech squared error is used as encoding criterion. For the random, speech-random, and perceptual CBs there is no consistent choice for the squared-error encoding criteria; the curves which for most CB sizes lead to the minimum average perceptual distance are shown.

The lower plots show the empirical rate-perceptual distortion when choosing ϕ_i^{CB} to minimize the perceptual measures ζ_A or ζ_P (perceptual encoding). The perceptual measures are used consistently, e.g., the PEMO measure trained CB is only used in conjunction with the PEMO measure for evaluation and/or encoding.

1. Discussion of results

For the 200-Hz pitch the average perceptual distance $\overline{\zeta}$ in Figs. 8 and 9 is always below the 95% thresholds found in

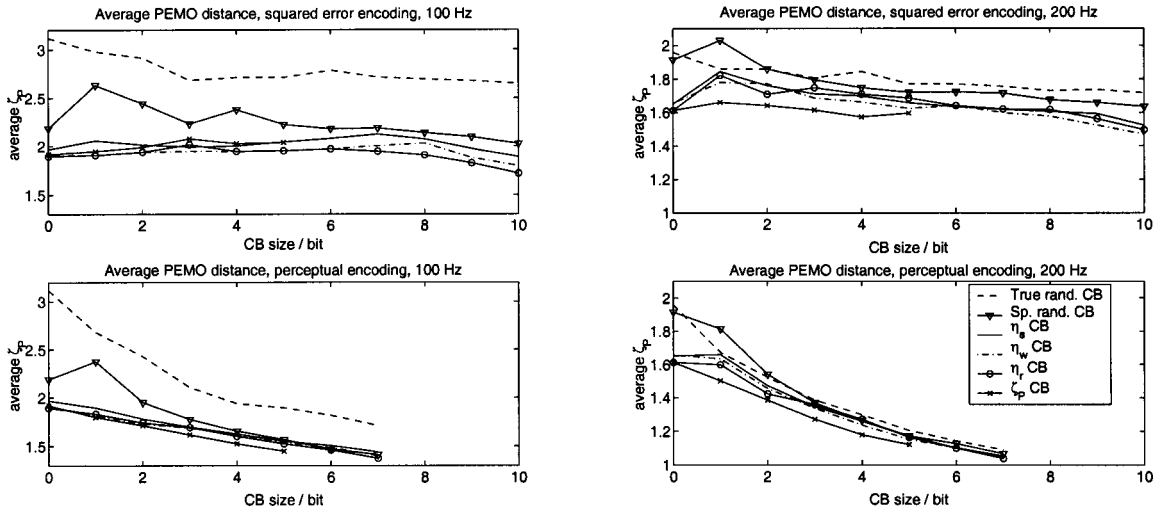


FIG. 9. Average PEMO distance for 100-Hz vowels (left), 200-Hz vowels (right). Upper graph: squared-error encoding. The true random CB is encoded using η_s, the speech-random CB using η_r for 100 Hz and η_w for 200 Hz, and the perceptual CB using η_r for 100 Hz and η_s for 200 Hz. Lower graph: perceptual encoding.

Sec. III C 2, while for 100 Hz it is close to the 95% threshold (just below for the AIM and just above for the PEMO). This confirms the conclusion drawn in Sec. IV B that phase distortion caused by VQ is more audible for 100-Hz vowels than for 200-Hz vowels.

In Sec. IV B, it was anticipated that squared-error encoding cannot be used to select the perceptually optimal codebook vectors. This is confirmed by the fact that empirical rate-distortion curves for squared-error encoding are flat for low to moderate rates. For low pitch (100 Hz), ζ for the AIM (Fig. 8) as well as for the PEMO (Fig. 9) is almost constant for rates up to 7–8 bit, and trained CBs. We verified this behavior with the listening test described in Sec. IV D. For 200 Hz the empirical rate-distortion curves for squared-error encoding show a decrease in ζ with increasing rate already for low rates. The earlier decrease of ζ for the 200-Hz vowels is reasonable since the bit/(component $[m]$) ratio is higher than for 100-Hz vowels. However, relating the decrease of ζ to the 95% thresholds of Sec. III C 2, the improvement in the investigated rate range is probably not audible. In summary, increasing the size of phase VQ codebooks does not lead to strongly audible improvements for squared-error encoding in the range between 0 to 10 bit.

The difference between codebooks trained for different measures, including the perceptual measures, is small in all graphs of Figs. 8 and 9. The most significant improvement gained from perceptual training compared to squared-error training is observed in the upper right graph of Fig. 9 (squared-error encoding, PEMO measure evaluation, 200-Hz pitch cycles). However, considering the graphs of Figs. 4 and 5 and that the improvement is of the order of $0.2\,\zeta_P$ units, the improvement is not likely to be perceivable. Thus, perceptual training does not significantly improve the performance of phase VQs for the artificial vowels.

Comparing squared-error encoding to perceptual encoding in Figs. 8 and 9, we see that the latter always leads to improved average perceptual distance with increasing rate. The improvement compared to squared-error encoding is small and generally will not justify the additional computational effort required. The fact that the structures of phase spectra do not have an obvious connection to physiological or psycho-acoustic properties is likely to be the reason for this negligible improvement. In addition, the importance of the phase at dimension m is dependent on the amplitude $a[m]$. Thus, the dispersion-phase CBs are a compromise for all amplitude spectra, which makes them inefficient.

Random CBs perform worse than trained CBs for 100-Hz pitch cycles, both for the squared-error encoding and for the perceptual encoding (Figs. 8 and 9). Thus, training is useful for the 100-Hz vowels, even though the advantage is not large. For 200-Hz pitch cycles, the difference between trained and untrained CBs is smaller.

D. Listening test

To verify that for low-pitch artificial vowels there is no perceptual difference between squared-error encoding with small and midsize dispersion-phase CBs, a listening test was performed. In addition, we ensured that listeners perceive the phase distortion introduced by quantization. If the distortion is not perceived, it is obvious that there is no preference for high-rate or low-rate encoding since they are both indistinguishable from the original.

The codebooks optimized for the speech squared error, the residual squared error, and the weighted squared error were used in separate tests. The listening tests were AOB forced-choice tests and the signals were presented monaurally via Sennheiser HD 250 linear II headphones connected to a SoundBlaster Live sound card in a standard PC. The AOB signal triplets consisted of three artificial vowels, each of 500 ms duration separated by 500 ms silence. A total of seven untrained subjects was used.

Two tests were run simultaneously. The first was intended to verify the results of Figs. 8 and 9. The signals AOB were: A: random choice between an artificial vowel from a random 100-Hz s_k either encoded with a 1-bit or a 7-bit

TABLE II. Percentage that listeners judged: The 1-bit or 7-bit encoded signals to be closer to the original (second and third column). The original or 1-bit encoded signals to be closer to the original (fourth and fifth column). Columns three and five include the interval at 95%-confidence level (Srednicki, 1988).

	1-bit	7-bit	Orig.	1-bit
η_s	43.5	56.5±4.1	76.3	23.7±3.5
η_w	50	50±4.1	81.5	18.5±3.2
η_r	52	48±4.1	82.8	17.2±3.1

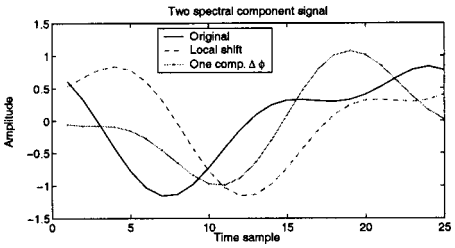


FIG. 10. Signal with two spectral components illustrating local time shift.

codebook; O: vowel from the same s_k with original phase; and B: vowel from the 7-bit encoded s_k if A was from the 1-bit encoded s_k and vice versa. The codebooks used were trained to minimize the squared error used as encoding criterion. The listeners were forced to choose whether A was closer to O or B was closer to O. Based on the results shown in Figs. 8 and 9, one expects that there should be no preference for either 1-bit or 7-bit encoding on average.[3]

The second test was to verify whether the distortion introduced by the encoding is perceived at all. In this second test, A was a random choice between an artificial vowel from a random s_k with original or 1-bit encoded phase, O was the vowel from s_k with original phase, and B was the complement of A. The tests were run double blind such that subjects were not aware if they were confronted with a triplet of the first or the second test.

The discrimination of the original and the 1-bit encoded signal did not exceed 60% for two of the subjects. Consequently, these two subjects cannot be expected to be able to judge whether the 1-bit or the 7-bit encoded signal is closer to the original, since it is to be expected that these subjects perceive both as equal to the original. Thus, the results of these two subjects are not considered. It should be stressed that this does not influence the results of the 1-bit vs 7-bit listening tests in a way which helps to confirm Figs. 8 and 9. It shows, however, that there are inexperienced subjects who are not sensitive to 100-Hz phase distortion. This phenomenon was not observed throughout the listening tests in Sec. III C 1.

For each of the three different squared errors, four out of the five remaining listeners listened to 100 triplets for each of the two test types. This makes 400 triplets for each test type and for each of the three squared errors.

Table II shows the results of the tests. Columns two and three show the percentage that the listeners judged the original to be closer to the 1-bit or 7-bit encoded signal, respectively. The fourth and fifth columns show the percentage that the listeners judged the original or 1-bit encoded signal to be closer to the original.

The results confirm that, in general, there was no preference for the 7-bit squared-error encoded signals over the 1-bit encoded signals, even though the results for η_s indicate a very minor preference for the 7-bit encodings. The detection of the encoded signal in the second test type was around 80% for all three squared errors. Even though the average perceptual distance for the 1–7-bit encodings of 100-Hz pitch cycles in Figs. 8 and 9 is close to the 95% threshold ζ_t, it is reasonable that the percentage of positive responses in the fourth column of Table II is lower than 95%. Averaging over ζ leads to different results than averaging over the probability of positive response, since the two are related nonlinearly.

V. DISCUSSION

In this discussion we show that using a squared error that is insensitive to time shifts in the individual auditory channels gives similar results as the empirical rate-perceptual distortion curves of Figs. 8 and 9. This suggests that the difference between the perceptual measures and the squared error in the sensitivity to time shifts is one of the main reasons for the failure of the squared error to select the codebook vectors that lead to the minimum perceived distortions.

The filterbanks of the auditory models split the signal into auditory channels and the time shift minimizing the squared error in the signal (found as described in Sec. II B) is not necessarily equivalent to the time shifts minimizing the squared error in each of the auditory channels individually. When neglecting that the channels overlap, the squared error for the signal is the sum of the squared errors in each channel. With this approximation, one difference between the squared error and the perceptual measures becomes evident: while the perceptual measures are insensitive to the shifts in the individual channels (ζ_A is found from autocorrelations, which are not sensitive to shifts; for ζ_P the l in Eq. (14) is found individually for each channel), the squared error is sensitive to these shifts. We refer to the time shifts in the individual auditory channels as local shifts.

Figure 10 illustrates the difference between the squared error and the perceptual measures. It depicts the waveform after one single auditory filter, containing two spectral components in this example. A phase distortion leading to a shift of the signal in the auditory channel but preserving its waveform in this and all other auditory channels causes a local time shift as illustrated by the dashed curve. The squared error increases while the perceptual distortion measures ζ remain zero. In the third curve of Fig. 10 the squared error is reduced by correctly encoding one phase component, while the perceptual measures increase since the consistency between the two phases is destroyed.

Introducing a local-time-shift insensitive squared error removes the above-mentioned inconsistency between squared errors and perceptual measures. The local-time-shift insensitive squared error η_l is found by first dividing the spectra of s_k, and $\hat{s}_k$ into discontinuous segments of one equivalent rectangular bandwidth (ERB) found according to Glasberg and Moore (1990). Then, the circular shifts that

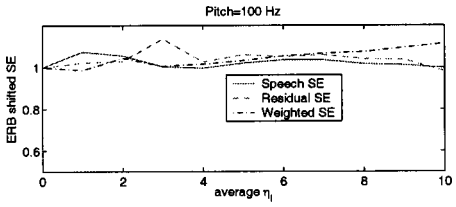


FIG. 11. Average local-time-shift insensitive error as a function of bit rate for 100-Hz vowels. The graph is normalized such that the average error for 0 bit equals 1.

minimize the squared error are found in each ERB band individually. The squared error remaining in band ι, after the circular shift, is η_{l_ι}. Thus, the local-time-shift insensitive squared error η_l is the sum of η_{l_ι} over all bands ι, $\eta_l = \Sigma_\iota \eta_{l_\iota}$.

Figure 11 shows the empirical rate-distortion curves of the squared-error encodings for 100-Hz vowels with the average η_l as distortion criterion on the ordinate previously shown with ζ on the ordinate in Figs. 8 and 9 (the data sets in both figures are equivalent). Figure 11 shows that η_l remains almost unchanged with varying codebook size. Plotting the same data with the average squared error on the ordinate, we observed that the average squared error decreases with increasing rate. The behavior of η_l is consistent with the average perceptual distance in Figs. 8 and 9. For the 200-Hz pitch cycles, η_l is constant only between 0 and 2–3 bit and then shows a slight decrease with increasing rate, which is consistent with the slight tilt of the upper graphs in Figs. 8 and 9.

Given this result, one can speculate that the (compared to the perceptual measures computationally simple) local-time-shift insensitive error as encoding criterion might lead to results similar to the empirical rate-perceptual distortion curves shown for perceptual encoding in Figs. 8 and 9. Since the perceptual encoding did not lead to strong improvements with increasing rates, we did not investigate if this speculation is true.

VI. CONCLUSIONS

The listening tests in Secs. III C 1 and IV D show that dispersion-phase distortion introduced by vector quantizers (VQs) is perceived for low-pitch vowels. In particular for untrained subjects, phase distortion is easier perceived for low-pitch frequencies than for high-pitch frequencies.

The relations between the squared errors and the perceptual measures found in Sec. IV B indicate that squared-error based dispersion-phase VQs as used in Gottesman (1999); Jiang and Cuperman (1995); Thomson (1988) do not always lead to improved perceptual performance with increasing rate for the periodic signals used. This is found to be true in the empirical rate-distortion curves of Sec. IV C and is verified by the listening test in Sec. IV D. As explained in the Introduction, it can be speculated that the results for the periodic signals carry over to natural speech. The discussion in Sec. V suggests that the squared-error VQs do not gain perceptual performance with increasing rate since their sensitivity to time shifts local to one auditory channel is not consistent with the human auditory system.

Observing that the usage of perceptual measures in vector quantization only leads to minor improvements (Sec. IV C) shows that phase spectra are difficult to encode. We speculate that the reason is that the perceptual features phase represents are not readily obtained from observing its structure. Further, the dependency of the weight of a phase at a certain frequency on the amplitude at that frequency prohibits the full utilization of all vector-quantization advantages.

ACKNOWLEDGMENTS

The authors would like to thank V. Bidault for running the listening tests of Sec. III C 1. We also are grateful to all subjects participating in the tests of Secs. III C 1 and IV D. In addition, we wish to thank the anonymous reviewers for their detailed comments and suggestions, which helped to improve the original manuscript significantly. This work was supported by the Swedish Research Council (Fund 271-2000-784).

[1]In general, we define $A_k(z=e^{j2\pi m/N}) \triangleq A_k[m]$.

[2]For the dc component this assumption does not have any influence since all pitch cycles s_k are zero mean.

[3]One-bit codebooks are used instead of the 0-bit codebooks since the 0-bit codebooks were not available at the time the tests were performed.

Burnett, I. S., and Bradley, G. J. (**1995**). "New techniques for multiprototype waveform coding at 2.84 kb/s," Proc. IEEE Int. Conf. Acoust., Speech, and Sign. Process., Detroit, pp. 261–264.

Chapin, E. K., and Firestone, F. A. (**1934**). "The influence of phase on tonal quality and loudness; the interference of subjective harmonics," J. Acoust. Soc. Am. **5**, 173–180.

Darpa (**1990**). TIMIT, CD-ROM. NIST Speech Disc 1-1.1.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996a**). "A quantitative model of the 'effective' signal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am. **99**(6), 3615–3622.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996b**). "A quantitative model of the 'effective' signal processing in the auditory system. II. Simulations and measurements," J. Acoust. Soc. Am. **99**(6), 3623–3631.

Eriksson, T., and Kleijn, W. B. (**1999**). "On waveform-interpolation coding with asymptotically perfect reconstruction," IEEE Workshop on Speech Coding, pp. 93–95.

Ezekiel, M., and Fox, K. A. (**1959**). *Methods of Correlation and Regression Analysis*, 3rd ed. (Wiley, New York).

Glasberg, B., and Moore, B. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Goldstein, J. L. (**1967**). "Auditory spectral filtering and monaural phase perception," J. Acoust. Soc. Am. **41**, 458–479.

Gottesman, O. (**1999**). "Dispersion phase vector quantization for enhancement of waveform interpolative coder," Proc. IEEE Int. Conf. Acoust., Speech, and Sign. Process., Phoenix, AZ, pp. 269–272.

Hansen, M., and Kollmeier, B. (**2000**). "Objective modeling of speech quality with a psychoacoustically validated auditory model," J. Audio Eng. Soc. **48**(5), 395–409.

Hedelin, P. (**1988**). "Phase compensation in all-pole speech analysis," Proc. IEEE Int. Conf. Acoust., Speech, and Sign. Process., pp. 339–342.

Hermann, L. (**1894**). "Beiträge zur Lehre von der Klangwahrnehmung," Arch. Ges. Physiol. **56**, 467–499.

Jiang, Y., and Cuperman, V. (**1995**) "Encoding prototype waveforms using a phase codebook," IEEE Workshop on Speech Coding, Annapolis, MD, pp. 21–22.

Kasper, K., Reininger, H., and Wolf, D. (**1997**). "Exploiting the potential of auditory preprocessing for robust speech recognition by locally recurrent neural networks," Proc. IEEE Int. Conf. Acoust., Speech, and Sign. Process., Munich, pp. 1223–1226.

Kim, D. S. (**2000**). ''Perceptual phase redundancy in speech,'' Proc. IEEE Int. Conf. Acoust., Speech, and Sign. Process., Istanbul/Turkey, pp. 1383–1386.
Kim, D. S. (**2001**). ''On the perceptually irrelevant phase information in sinusoidal representation of speech,'' IEEE Trans. Speech Audio Process. **9**(8), 900–905.
Kleijn, W. B., and Haagen, J. (**1995**). ''Waveform interpolation for speech coding and synthesis,'' in *Speech Coding and Synthesis*, edited by W. B. Kleijn and K. K. Paliwal (Elsevier Science, Amsterdam), pp. 175–208.
König, R. (**1881**). ''Bemerkungen über die Klangfarbe,'' Ann. Phys. (Leipzig) **14**, 369–393.
Kubin, G., Atal, B. S., and Kleijn, W. B. (**1993**). ''Performance of noise excitation for unvoiced speech,'' Proc. IEEE Workshop on Speech Coding for Telecomm., Sainte-Adele, Quebec, pp. 35–36.
Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Licklider, J. C. R. (**1957**). ''Effects of changes in the phase pattern upon the sound of a 16-harmonic tone,'' J. Acoust. Soc. Am. **29**, 780(A).
Linde, Y., Buzo, A., and Gray, R. M. (**1980**). ''An algorithm for vector quantizer design,'' IEEE Trans. Comm. **COM-28**, 84–95.
Lindig, F. (**1903**). ''Über den Einfluß der Phasen auf die Klangfarbe,'' Ann. Phys. (Leipzig) **10**, 242–269.
Ma, C., and O'Shaughnessy, D. (**1994**). ''A perceptual study of source coding of Fourier phase and amplitude of the linear predictive coding residual of vowel sounds,'' J. Acoust. Soc. Am. **95**(4), 2231–2239.
Makhoul, J. (**1975**). ''Linear prediction: A tutorial review,'' Proc. IEEE **63**(4), 561–580.
Makhoul, J., Viswanathan, R., Schwartz, R., and Huggins, A. (**1978**). ''A mixed-source model for speech compression and synthesis,'' Proc. IEEE Int. Conf. Acoust., Speech., and Sign. Process., Tulsa, pp. 163–176.
Marques, J. S., Trancoso, I. M., and Tribolet, J. M. (**1990**). ''Harmonic coding at 4.8 kb/s,'' Proc. IEEE Int. Conf. Acoust., Speech, and Sign. Process., Albuquerque, pp. 17–20.
Maths, R. C., and Miller, R. L. (**1947**). ''Phase effects in monaural perception,'' J. Acoust. Soc. Am. **19**, 780–797.
McAulay, R. J., Parks, T., Quatieri, T. F., and Sabin, M. (**1990**). ''Sine-Wave Amplitude Coding at Low Data Rates,'' in *Advances in Speech Coding* (Kluwer, Dordrecht, Netherlands), pp. 203–213.
McAulay, R. J., and Quatieri, T. F. (**1986a**). ''Phase modeling and its application to sinusoidal transform coding,'' Proc. IEEE Int. Conf. Acoust., Speech, and Sign. Process., Tokyo, Japan, pp. 1713–1715.
McAulay, R. J., and Quatieri, T. F. (**1986b**). ''Speech analysis/synthesis based on a sinusoidal representation,'' IEEE Trans. Acoust., Speech, Signal Process. **34**, 744–754.
McAulay, R. J., and Quatieri, T. F. (**1991**). ''Sine-wave phase coding at low data rates,'' Proc. IEEE Int. Conf. Acoust., Speech, and Sign. Process., Toronto, pp. 577–580.
Moriya, T., and Honda, M. (**1986**). ''Speech coder using phase equalization and vector quantization,'' Proc. IEEE Int. Conf. Acoust., Speech, and Sign. Process., Tokyo, pp. 1701–1704.
Nelson, D. A. (**1994**). ''Level-dependent critical bandwidth for phase discrimination,'' J. Acoust. Soc. Am. **95**(3), 1514–1524.
Oxenham, A. J., and Dau, T. (**2001**). ''Towards a measure of auditory-filter phase response,'' J. Acoust. Soc. Am. **110**(6), 3169–3178.
Patterson, R. D. (**1987**). ''A pulse ribbon model of monaural phase perception,'' J. Acoust. Soc. Am. **82**, 1560–1586.
Patterson, R. D. (**1988**). ''Timbre cues in monaural phase perception: Distinguishing within-channel cues and between-channel cues,'' in *Proceedings of the 8th International Symposium on Hearing*, Basic Issues in Hearing, edited by H. Duifhuis, J. Horst, and H. Wit (Academic, London), pp. 351–358.
Patterson, R. D., Allerhand, M. H., and Giguère, C. (**1995**). ''Time-domain modeling of peripheral auditory processing: A modular architecture software platform,'' J. Acoust. Soc. Am. **98**(4), 1890–1894.
Plomp, R., and Steeneken, H. J. M. (**1969**). ''Effect of phase on the timbre of complex tones,'' J. Acoust. Soc. Am. **46**(2), 409–421.
Pobloth, H., and Kleijn, W. B. (**1999**). ''On phase perception in speech,'' Proc. IEEE Int. Conf. Acoust., Speech, and Sign. Process., Phoenix, AZ, pp. 29–32.
Pobloth, H., and Kleijn, W. B. (**2001**). ''Squared error as a measure of phase distortion,'' Proc. Eurospeech, Aalborg, Denmark, pp. 1973–1976.
Pols, L. C. W., van der Kamp, L. J. T., and Plomp, R. (**1969**). ''Perceptual and physical space of vowel sounds,'' J. Acoust. Soc. Am. **46**, 458–467.
Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (**1992**). *Numerical Recipes in C*, 2nd ed. (Cambridge University Press, Cambridge).
Purnhagen, H. (**1999**). ''Advances in parametric audio coding,'' Proc. IEEE Workshop on Appl. of Signal Proc. to Audio and Acoustics, pp. 31–34.
Purnhagen, H., and Meine, N. (**2000**). ''HILN—The MPEG-4 parametric audio coding tools,'' IEEE Int. Symp. on Circuits and Systems, Vol. 3, Geneva, pp. 201–204.
Quatieri, T. F., and McAulay, R. J. (**1989**). ''Phase coherence in speech reconstruction for enhancement and coding applications,'' Proc. IEEE Int. Conf. Acoust., Speech, and Sign. Process., Glasgow, pp. 207–210.
Quatieri, T., and McAulay, R. (**1986**). ''Speech transformations based on a sinusoidal representation,'' IEEE Trans. Acoust., Speech, Signal Process. **34**(6), 1449–1464.
Skoglund, J., Kleijn, W. B., and Hedelin, P. (**1997**). ''Audibility of pitch-synchronously modulated noise,'' Proc. IEEE Speech Coding Workshop, Pocono Manor, PA.
Srednicki, M. (**1988**). ''A Bayesian analysis of A–B listening tests,'' J. Audio Eng. Soc. **36**(3), 143–146.
Tchorz, J., Kasper, K., Reininger, H., and Kollmeier, B. (**1997**). ''On the interplay between auditory-based features and locally recurrent neural networks for robust speech recognition in noise,'' Proc. Eurospeech, Patras, Greece, pp. 2075–2078.
Thomson, D. L. (**1988**). ''Parametric models of the magnitude/phase spectrum for harmonic speech coding,'' Proc. IEEE Int. Conf. Acoust., Speech, and Sign. Process., pp. 378–381.
Trancoso, I. M., Rodrigues, J. S., Almeida, L. B., Marques, J. S., Serralheiro, A., Santos, D. S., and Tribolet, J. M. (**1988**). ''Quantization issues in harmonic coders,'' Proc. IEEE Int. Conf. Acoust., Speech, and Sign. Process., pp. 382–385.
von Helmholtz, H. L. F. (**1859**). ''Über die Klangfarbe der Vocale,'' Ann. Phys. (Leipzig) **18**, 280–290.

The simulation of piano string vibration: From physical models to finite difference schemes and digital waveguides

Julien Bensa, Stefan Bilbao, Richard Kronland-Martinet, and Julius O. Smith III
Center for Computer Research in Music and Acoustics (CCRMA), Department of Music, Stanford University, Stanford, California 94305-8180

(Received 6 September 2002; revised 20 April 2003; accepted 8 May 2003)

A model of transverse piano string vibration, second order in time, which models frequency-dependent loss and dispersion effects is presented here. This model has many desirable properties, in particular that it can be written as a well-posed initial-boundary value problem (permitting stable finite difference schemes) and that it may be directly related to a digital waveguide model, a digital filter-based algorithm which can be used for musical sound synthesis. Techniques for the extraction of model parameters from experimental data over the full range of the grand piano are discussed, as is the link between the model parameters and the filter responses in a digital waveguide. Simulations are performed. Finally, the waveguide model is extended to the case of several coupled strings. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1587146]

PACS numbers: 43.75.Mn, 43.40.Cw, 43.75.Wx [NHF]

I. INTRODUCTION

Several models of transverse wave propagation on a stiff string, of varying degrees of complexity, have appeared in the literature.[1–5] These models, intended for the synthesis of musical tones, are always framed in terms of a partial differential equation (PDE), or system of PDEs; usually, the simplified starting point for such a model is the *one-dimensional wave equation*,[6] and the more realistic features, such as dispersion, frequency-dependent loss and nonlinear hammer excitation (in the case of the piano string), are incorporated through several extra terms. Chaigne and Askenfelt[3] have proposed such a model (see the Appendix for a concise description of this model), and used it as the basis for a synthesis technique, through the use of finite differences—the time waveform on a struck piano string is simulated in this way to a remarkable degree of fidelity.[4] Frequency-dependent loss, the feature of primary interest in this paper, is modeled through the use of a third-order time derivative perturbation to the dispersive wave equation; while a physical justification for the use of such a term is tenuous, it does give rise to perceptually important variations in damping rates.

In Sec. II, we introduce a model of string vibration, which is of second order in time differentiation; frequency-dependent loss is introduced via mixed time–space derivative terms. As it turns out, the model discussed here is a substantial improvement in several different ways. First, the frequency domain analysis of a second-order system is quite straightforward, and it is quite easy to obtain explicit formulas for dispersion and loss curves; this is considerably more complicated for systems which are higher order in time, essentially requiring the factorization of a higher order polynomial dispersion relation. Second, it is easy to prove that our model, when complemented by initial and boundary conditions, is *well posed*.[7,8] Though we do not give a complete description of this condition here, to say that such an initial-boundary value problem is well posed is to say, generally speaking, that solutions may not grow faster than exponentially; for linear and shift-invariant systems such as simple stiff string models, the condition can be conveniently expressed in the frequency domain. We show in the Appendix that the PDE first proposed by Ruiz,[9] and later popularized by Chaigne and Askenfelt[3,4] is in fact not well-posed, and possesses a spurious unstable solution. Third, it becomes easy to develop *finite difference schemes*, for which precise numerical stability conditions may be easily derived. Finite difference schemes are discussed briefly in Sec. III. Fourth, it is possible to extend the model described here to a more realistic representation of dispersion and loss as a function of frequency through additional terms in the PDE, without compromising well-posedness, or the stability of resulting numerical schemes.

Finally, it is possible to identify the model PDE with a *digital waveguide*—this filterlike structure models one-dimensional wave propagation as purely lossless throughout the length of the string, with loss and dispersion lumped in terminating filters. It thus performs a simulation of a *modified* physical system.

We show, in Sec. IV A, how one can relate the PDE model presented in Sec. II to a digital waveguide structure, paying particular attention to the relationship between the lumped filters used to model loss and dispersion and the model parameters which define our PDE. In Sec. V, we perform several numerical simulations in order to compare the finite difference and waveguide approaches for this particular problem. In particular, in Sec. VI, we examine in detail a procedure allowing the resynthesis of natural string vibration. Using experimental data obtained from a grand piano, both the terminating filters of a digital waveguide and the parameters of the physical model are estimated over most of the keyboard range. This leads to a simple description of the variation of some of these parameters (namely loss parameters and string stiffness) over the piano's range which can be used for the convenient synthesis string vibrations at a given excitation point. Finally, in Sec. VII, interstring coupling is discussed and modeled using coupled digital

waveguides. This is a further step towards the design of a realistic piano simulator, which should ultimately also model effects such as nonlinear hammer-string coupling and soundboard radiation phenomena.

II. SECOND-ORDER MODELS OF ONE-DIMENSIONAL WAVE PROPAGATION

A. A family of PDEs

Consider a general linear second-order (in time) wave equation, of the form

$$\frac{\partial^2 y}{\partial t^2} + 2\sum_{k=0}^{M} q_k \frac{\partial^{2k+1} y}{\partial x^{2k} \partial t} + \sum_{k=0}^{N} r_k \frac{\partial^{2k} y}{\partial x^{2k}} = 0. \tag{1}$$

Here, $y(x,t)$, the solution, is a function of position $x \in \mathbb{R}$ and time $t \geqslant 0$, and q_k and r_k are real constants; the solution is not uniquely defined until two initial conditions are supplied. (For the moment, we concentrate on the pure initial value problem and ignore boundary conditions—we will return to this subject in Sec. II B.) Because this equation describes a linear and shift-invariant system, it is possible to analyze it through Fourier techniques. Defining the spatial Fourier transform $\hat{y}(\beta,t)$ of $y(x,t)$ by

$$\hat{y}(\beta,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y(x,t) e^{-j\beta x}\, dx,$$

Eq. (1) can be rewritten as

$$\frac{\partial^2 \hat{y}}{\partial t^2} + 2\sum_{k=0}^{M} (j\beta)^{2k} q_k \frac{\partial \hat{y}}{\partial t} + \sum_{k=0}^{N} (j\beta)^{2k} r_k \hat{y} = 0. \tag{2}$$

This second-order linear ordinary differential equation with real coefficients will have solutions of the form

$$\hat{y}(\beta,t) = \hat{y}_0(\beta) e^{st} \tag{3}$$

for complex frequencies s which satisfy the characteristic polynomial equation

$$s^2 + 2q(\beta)s + r(\beta) = 0 \tag{4}$$

with

$$q(\beta) = \sum_{k=0}^{M} (j\beta)^{2k} q_k, \quad r(\beta) = \sum_{k=0}^{N} (j\beta)^{2k} r_k.$$

Notice that because only even derivatives appear in the family Eq. (1), the functions q and r are real. The characteristic polynomial equation has roots

$$s_{\pm} = -q \pm \sqrt{q^2 - r}.$$

The condition that the initial value problem corresponding to Eq. (1) be *well-posed* is that these roots have real parts which are bounded from above as a function of β; this is in effect saying that solution growth can be no faster than exponential. A more restrictive (and physically relevant) condition is that these roots have nonpositive real part for all β, so that all exponential solutions are nonincreasing. It is simple to show that this will be true for

$$q(\beta), r(\beta) \geqslant 0. \tag{5}$$

For q and r satisfying Eq. (5), the imaginary parts of these roots correspond to oscillation frequencies, and the real parts to loss. Clearly, for real wave numbers β such that $q^2 \leqslant r$, the real parts of $s_{\pm}$ are simply $-q$. This case corresponds to normal damped wave propagation; notice in particular that if q depends on β, then damping rates will be wave number (and thus frequency) dependent. If $q^2 > r$, then both roots are purely real and nonpositive, yielding damped nontraveling solutions.

TABLE I. Physical model parameters for piano tones C2, C4, and C7.

	C2	C4	C7	Units
L	1.23	0.63	0.10	m
c	160.9	329.6	418.6	$\mathrm{m\,s^{-1}}$
κ	0.58	1.25	1.24	$\mathrm{m^2\,s^{-1}}$
b_1	0.25	1.1	9.17	$\mathrm{s^{-1}}$
b_2	7.5×10^{-5}	2.7×10^{-4}	2.1×10^{-3}	$\mathrm{m^2\,s^{-1}}$
F_s	16 000	32 000	96 000	$\mathrm{s^{-1}}$

Consider a member of the family defined by Eq. (1),

$$\frac{\partial^2 y}{\partial t^2} = c^2 \frac{\partial^2 y}{\partial x^2} - \kappa^2 \frac{\partial^4 y}{\partial x^4} - 2b_1 \frac{\partial y}{\partial t} + 2b_2 \frac{\partial^3 y}{\partial x^2 \partial t}. \tag{6}$$

The first term on the right-hand side of the equation, in the absence of the others, gives rise to wavelike motion, with speed c. The second "ideal bar" term[10] introduces dispersion, or frequency-dependent wave velocity, and is parametrized by a stiffness coefficient κ. The third and fourth terms allow for loss, and if $b_2 \neq 0$, decay rates will be frequency dependent. [A complete model, for a piano string, is obtained by including a hammer excitation term, $f(x,t)$, possibly accounting for nonlinear effects, on the right-hand side, and by restricting the spatial domain to a finite interval and supplying a realistic set of boundary conditions.] This model differs from that of Ruiz[9] in only the last term (see the Appendix).

The characteristic equation has the form of Eq. (4), with

$$q(\beta) = b_1 + b_2\beta^2, \quad r(\beta) = c^2\beta^2 + \kappa^2\beta^4.$$

For b_1, $b_2 \geqslant 0$, condition Eq. (5) is satisfied and this PDE obviously possesses exponentially decaying solutions, and what is more, loss increases as a function of wave number. The PDE of Eq. (6) possesses traveling wave solutions when $q^2 \leqslant r$, which, for realistic values of the defining parameters for a piano string, includes the overwhelming part of the audio spectrum. For instance, for a C2 piano string, described by the parameters given in Table I, the lower cutoff wave number for traveling waves will be $\beta = 0.0025$, corresponding to a frequency of 0.080 Hz. There is no upper cutoff.

In order to relate this PDE model with a digital waveguide numerical simulation method, it is useful to write the expressions for dispersion and loss directly. Taking

$$s_{\pm} = \sigma \pm j\omega \tag{7}$$

over the range of β for which traveling wave solutions exist, we obtain

$$\sigma(\beta) = -b_1 - b_2\beta^2, \tag{8}$$

$$\omega(\beta) = \sqrt{-(b_1 + b_2\beta^2)^2 + c^2\beta^2 + \kappa^2\beta^4}. \tag{9}$$

We will discuss digital waveguide models in detail in Sec. IV A.

Finally, we mention that the general form Eq. (1) serves as a useful point of departure for more accurate models of loss and dispersion in a stiff string. The more terms are included, the better these phenomena can be modeled over the entire frequency range of interest. Though it is difficult to associate physical processes directly with the various extra perturbation terms in the equation, it is at least simple to ensure, through condition Eq. (5), that the model is well-posed, an important first step in developing stable numerical methods.

B. Boundary conditions

In this section, we provide a brief analysis of pinned boundary conditions, to show that when coupled to our string model, the same analysis of well-posedness may be applied (i.e., wave numbers of modal solutions are real). Let us now restrict the spatial domain for the problem defined by Eq. (6) to $x \in [0,L]$. As Eq. (6) is of fourth order in the spatial derivatives, we need to supply two boundary conditions at either end, i.e., at $x=0$ and $x=L$. Following Chaigne,[3] we apply pinned boundary conditions,

$$y|_{x=0} = y|_{x=L} = \left.\frac{\partial^2 y}{\partial x^2}\right|_{x=0} = \left.\frac{\partial^2 y}{\partial x^2}\right|_{x=L} = 0. \tag{10}$$

For a solution of the form $y(x,t)=e^{st+j\beta x}$, from dispersion relation Eq. (4), there are thus four solutions for β in terms of s,

$$\beta_+^2(s) = \frac{-\gamma + \sqrt{\gamma^2 - 4\kappa^2(s^2+2b_1 s)}}{2\kappa^2}, \tag{11a}$$

$$\beta_-^2(s) = \frac{-\gamma - \sqrt{\gamma^2 - 4\kappa^2(s^2+2b_1 s)}}{2\kappa^2} \tag{11b}$$

with

$$\gamma = c^2 + 2b_2 s.$$

At frequency s, thus, any linear combination

$$y(x,t) = e^{st}(a_{+,+}e^{j\beta_+ x} + a_{+,-}e^{-j\beta_+ x} + a_{-,+}e^{j\beta_- x} + a_{-,-}e^{-j\beta_- x}) \tag{12}$$

is a solution to Eq. (6). Applying the boundary conditions Eq. (10) to this solution gives the matrix equation

$$\underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 \\ e^{j\beta_+ L} & e^{-j\beta_+ L} & e^{j\beta_- L} & e^{-j\beta_- L} \\ -\beta_+^2 & -\beta_+^2 & -\beta_-^2 & -\beta_-^2 \\ -\beta_+^2 e^{j\beta_+ L} & -\beta_+^2 e^{-j\beta_+ L} & -\beta_-^2 e^{j\beta_- L} & -\beta_-^2 e^{-j\beta_- L} \end{bmatrix}}_{\triangleq A(s)} \times \begin{bmatrix} a_{+,+} \\ a_{+,-} \\ a_{-,+} \\ a_{-,-} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{13}$$

Nontrivial solutions can occur only when $\det(A)=0$, giving the relation

$$(\beta_+^2 - \beta_-^2)^2 \sin(\beta_+ L)\sin(\beta_- L) = 0. \tag{14}$$

Discounting the case $\beta_+^2 = \beta_-^2$ [which yields an identically zero solution $y(x,t)$], then obviously, solutions are of the form $\beta_+ = n\pi/L$, for integer $n \neq 0$ (similarly for β_-), and the modal frequencies s_n are, from the solutions Eq. (8) and Eq. (9) of the dispersion relation (4),

$$s_n = \sigma(n\pi/L) + j\omega(n\pi/L) \tag{15}$$

over wave numbers for which a traveling solution exists (for small b_1 and b_2, this will be true for all n).

III. A FINITE DIFFERENCE SCHEME

In order to solve Eq. (6) numerically, we may approximate its solution over a grid with spacing X, and with time step T. Equation (6) can be written as

$$\delta_t^2 y = c^2\delta_x^2 y - \kappa^2\delta_x^2\delta_x^2 y - 2b_1\delta_{t,0}y + 2b_2\delta_x^2\delta_{t,-}y + O(T,X^2), \tag{16}$$

where the various difference operators are defined by

$$\delta_x^2 y(x,t) = \frac{1}{X^2}(y(x+X,t) - 2y(x,t) + y(x-X,t)),$$

$$\delta_t^2 y(x,t) = \frac{1}{T^2}(y(x,t+T) - 2y(x,t) + y(x,t-T)),$$

$$\delta_{t,0} y(x,t) = \frac{1}{2T}(y(x,t+T) - y(x,t-T)),$$

$$\delta_{t,-} y(x,t) = \frac{1}{T}(y(x,t) - y(x,t-T)).$$

All these operators are "centered" about the point (x,t), except for the backward difference operator $\delta_{t,-}$, which is used in order to obtain an explicit algorithm. The approximation is first-order accurate in the time step T, and second-order accurate in the space step X. [It is worth mentioning, that this is but one among many ways of discretizing (6).] We may now rewrite Eq. (16) as a difference scheme, operating on the grid function y_m^n, indexed by integer m and n, which will serve as an approximation to $y(x,t)$ at the location $x=mX$, $t=nT$:

$$y_m^{n+1} = a_{10}y_m^n + a_{11}(y_{m+1}^n + y_{m-1}^n) + a_{12}(y_{m+2}^n + y_{m-2}^n) + a_{20}y_m^{n-1} + a_{21}(y_{m+1}^{n-1} + y_{m-1}^{n-1}). \tag{17}$$

Here, the difference scheme coefficients are defined by

$$\begin{aligned} a_{10} &= (2 - 2\lambda^2 - 6\mu^2 - 4b_2\mu/\kappa)/(1+b_1 T), \\ a_{11} &= (\lambda^2 + 4\mu^2 + 2b_2\mu/\kappa)/(1+b_1 T), \\ a_{12} &= -\mu^2/(1+b_1 T), \\ a_{20} &= (-1 + 4b_2\mu/\kappa + b_1 T)/(1+b_1 T), \\ a_{21} &= (-2b_2\mu/\kappa)/(1+b_1 T), \end{aligned} \tag{18}$$

where, for brevity, we have introduced the quantities

$$\lambda = cT/X, \quad \mu = \kappa T/X^2.$$

In order to examine the stability of scheme Eq. (17), which is, like its generating PDE, Eq. (6), linear and shift

invariant, we may apply frequency domain techniques—in the finite difference setting, these techniques are referred to as *Von Neumann analysis*.[7,8] This analysis proceeds in a fashion exactly analogous to the analysis applied to the continuous time and space systems of the preceding sections. Shortcutting the process somewhat, we may consider a solution of the form $y_m^n = z^n e^{imX\beta}$, where $z = e^{sT}$ (we could equivalently employ a z transform and a discrete-time Fourier transform). We thus obtain the two-step characteristic, or *amplification equation*

$$z^2 + a_1(\beta)z + a_2(\beta) = 0,$$

where the functions $a_1(\beta)$ and $a_2(\beta)$ are defined in terms of the difference scheme parameters of Eq. (18) by

$$a_1(\beta) = -a_{10} - 2a_{11}\cos(\beta X) - 2a_{12}\cos(2\beta X),$$

$$a_2(\beta) = -a_{20} - 2a_{21}\cos(\beta X).$$

The necessary and sufficient stability conditions for Eq. (17) are that the roots of the amplification equation be of magnitude less than or equal to unity, for all wave numbers β; for this real-valued quadratic, these conditions can be written in terms of $a_1(\beta)$ and $a_2(\beta)$ as

$$|a_1(\beta)| - 1 \leqslant a_2(\beta) \leqslant 1.$$

The right inequality is satisfied for b_1, $b_2 \geqslant 0$, and after some algebra, it can be shown that the left inequality is equivalent to the condition $\lambda^2 + 4\mu^2 + 4b_2\mu/\kappa \leqslant 1$, further implying that

$$T \leqslant X^2 \frac{-4b_2 + \sqrt{16b_2^2 + 4(c^2X^2 + 4\kappa^2)}}{2(c^2X^2 + 4\kappa^2)}.$$

The relative ease with which an exact bound such as the above may be derived is a direct consequence of the use of a two-step scheme and the relative simplicity of the model itself; for schemes involving more steps of lookback (which results from the discretization of a model with higher time derivatives, such as Ruiz's system), this analysis becomes much more involved, though we do approach it nonetheless in the Appendix. This, in addition to reduced memory requirements, is a further advantage of using a second-order model as a starting point.

Equation (16) is but one of many possible discretizations of Eq. (6)—for instance, replacing $\delta_{t,-}$ by $\delta_{t,0}$ yields an *implicit* algorithm,[7] and other implicit schemes such as the θ-forms discussed in the work of Chaigne[5] may be of interest in reducing *numerical dispersion*,[7] and may be of higher formal accuracy (which may be tempered by the stability requirements). To emphasize this point, we have plotted the phase velocities and loss curves for the model system of Eq. (6) versus those of the difference scheme of Eq. (17) in Fig. 1. Notice that this simple difference scheme is a reasonable approximation to the model only for small ω (i.e., for low frequencies). As we will see later in Sec. V, this deviation from the model PDE will account for differences in simulation results obtained from digital waveguide models, which approximates the phase velocity and loss curves directly.

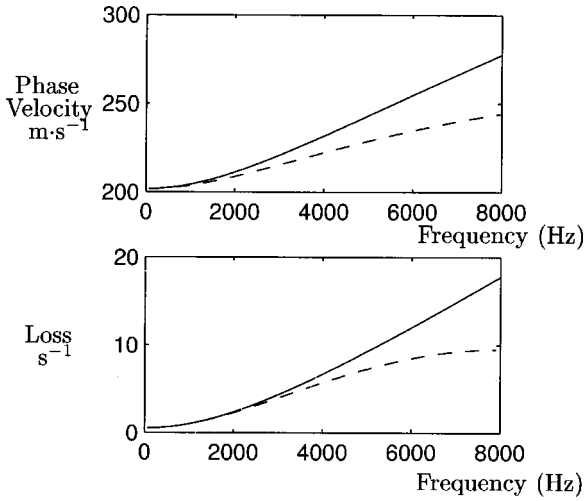


FIG. 1. Phase velocity (top) and loss (bottom) for the model of Eq. (6) (solid line) and for difference scheme Eq. (17) (dashed line) as a function of the frequency. The model parameters are those corresponding to the note C2, as given in Table I.

IV. THE DIGITAL WAVEGUIDE MODEL

The digital waveguide approach provides computational models for musical instruments primarily in the string, wind, and brass families.[11] They have also been developed specifically for piano synthesis.[11–13] This section summarizes the basic ideas of the digital waveguide approach, and relates the parameters of a digital waveguide model to an underlying physical model.

A. Background

As mentioned in Sec. I, to arrive at a PDE modeling the piano string, it is fruitful to start with the ideal wave equation and add perturbation terms to give more realistic frequency-dependent dispersion and losses. The perturbed PDE is then numerically integrated via a finite-difference scheme (or possibly by another approach, such as finite element methods, etc.). The digital waveguide approach interchanges the order of these operations: the ideal wave equation is integrated first using a trivial finite-difference scheme, and the resulting solutions are perturbed using *digital filters* to add frequency-dependent loss and dispersion. In the case of a strongly dissipative and dispersive string, the modulus of these so-called *loop filters* decreases rapidly with frequency, and phase can become strongly nonlinear. For a frequency-domain implementation, this has no effect on computational complexity, but for a time-domain implementation, a larger filter size may be required in order to match the large variations of the phase response.

It has been known since d'Alembert[14] that the ideal one-dimensional wave equation is solved exactly by arbitrary (sufficiently smooth) wave shapes propagating in both directions. The digital waveguide formulation works directly with these traveling wave components. An isolated traveling wave is trivially simulated in practice using a *delay line*. An ideal vibrating string is then modeled as a pair of delay lines, one for each direction of travel.

For digital implementation, the traveling waves are *sampled* at intervals of T seconds. By Shannon's sampling

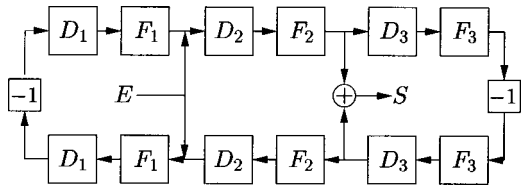


FIG. 2. Digital waveguide model of a rigidly terminated string.

theorem,[15] the solution remains exact, in principle, at all frequencies up to half the sampling rate. To avoid aliasing, all initial conditions and ongoing excitations must be *bandlimited* to less than half the sampling rate $f_s = 1/T$.

Figure 2 shows the simulation diagram for a digital waveguide model of a rigidly terminated string. The string is excited by the signal E and observed via the signal S. Sampled traveling velocity waves propagate to the right along the upper rail, and to the left along the lower rail. The rigid terminations cause inverting reflections (the two -1 scale factors). The delay lines are denoted D_i, $i=1,2,3$, and the F_i blocks are digital filters to be described further below.

Consider a wavelike solution propagating from a point M_1 to a point M_2 along a string (see Fig. 3, top). The distance M_1M_2 will be arbitrarily called l and the propagation time d at the *minimal* phase velocity. At the observation point M_2, the wave will have arrived after having undergone the effects of loss and dispersion. In terms of digital waveguides, the wave will undergo a pure delay [in the frequency domain, a multiplicative phase factor $\exp(-i\omega d)$], times a multiplicative factor $F(\omega)$ representing the loss and the dispersion experienced by the wave during this interval (see Fig. 3, bottom).

Since loss and dispersion are, for this system, linear time-invariant (LTI) phenomena, even when frequency dependent, the perturbations needed for added realism in the digital waveguide string model are LTI *digital filters*. Since LTI filters *commute*, we may *lump all* of the filtering associated with propagation in one direction into a *single* LTI filter. These filters are denoted F_i in Fig. 2, $i=1,2,3$.

For purposes of computing the output signal S from the input signal E, Fig. 2 may be further simplified.

The two filters labeled F_1 can be replaced by a single filter F_1^2 (by commuting one of them with the intervening two delay lines D_1 and -1 gain). A similar simplification is possible for F_3.

In the same way, the two delay lines labeled D_1 can be replaced by a single delay line D_1^2 (having twice the length of D_1), and the two D_3 blocks can be replaced by one D_3^2 block.

In general, any uniform section of a linear vibrating string which is excited and observed only at its endpoints can be accurately modeled (in one vibrational plane) by a pair of digital delay lines, each in series with a digital filter. A discussion of more generalized approaches involving nonuniform string sections, and the relationship with finite difference schemes, is provided in a recent dissertation.[16]

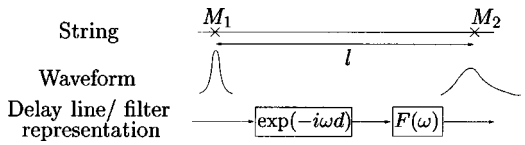


FIG. 3. The physical system and its corresponding delay line/filter.

Since losses and dispersion are relatively *weak* in vibrating strings and acoustic bores, a *low-order* filter can approximate very well the distributed filtering (infinite-order in principle) associated with a particular stretch of string or bore. (We should repeat, however, that the approach outlined above is equally applicable in the case of strongly dispersive or lossy systems, though in these cases, higher-order filters may be necessary.) In practice, the desired loss and dispersion filters are normally derived from measurements such as the decay time of overtones in the freely vibrating string.[17–19] In the next section, the filter will be derived from the stiff-string model of Eq. (16). Interestingly, the filter(s) so designed can be mapped back to an equivalent PDE, including many higher-order terms (which may or may not have a physical interpretation). Lumping of traveling-wave filtering in this way can yield computational savings by orders of magnitude relative to more typical finite difference schemes,[20–22] and this efficiency can be important in applications such as real-time modeling of musical instruments for purposes of automatic sound synthesis or ''virtual acoustic instrument'' performance.

B. Relating digital waveguide parameters to the physical model

We address here the problem of relating digital waveguide filter parameters to the loss and dispersion curves from the physical model discussed in Sec. II. For that, we consider the continuous frequency representation of the loop filter and show its relation with the physical parameters. The digital waveguide parameters can then be obtained by discretization. We do not address here the problem of the time domain implementation of the digital waveguide.

According to Eqs. (3) and (7), the transformations of the wave due to propagation along the string segment can be represented in terms of a digital waveguide filter by a multiplicative phase factor $\exp(sd+j\beta l)$. Ideally, the modulus and phase of this expression are related to the filter F by

$$|F(\omega)| = |e^{sd+j\beta l}| = e^{\sigma d} = e^{-(b_1+b_2\beta^2)d}, \tag{19}$$

$$\arg(F(\omega)) = \arg(e^{sd+j\beta l}) = \omega d + \beta l. \tag{20}$$

In order to write the expressions of the modulus and the phase of the loop filter in terms of the frequency ω, it is necessary to express the wave number β in terms of ω. From Eq. (9), solving for β, one gets

$$\beta^2(\omega) = \frac{-\alpha_1 \pm \sqrt{\alpha_1^2 + 4\alpha_2(b_1^2+\omega^2)}}{2\alpha_2} \tag{21}$$

with

$$\alpha_1 = c^2 - 2b_1b_2, \quad \alpha_2 = \kappa^2 - b_2^2. \tag{22}$$

Since β must be real (see Sec. II B), we keep only the solution for which the term inside the root is positive. Then,

$$\beta(\omega) = \pm \sqrt{\frac{-\alpha_1 + \sqrt{\alpha_1^2 + 4\alpha_2(b_1^2 + \omega^2)}}{2\alpha_2}}. \tag{23}$$

Given that, for realistic piano string modeling, $b_1 \simeq 1$, $b_2 \simeq 10^{-4}$, $c \simeq 200$, $\kappa \simeq 1$, and $\omega \simeq 400$, we make the simplifying assumptions

$$b_1 b_2 \ll c^2, \quad b_2^2 \ll \kappa^2 \tag{24}$$

(the second of which also ensures the realness of β)

$$b_1^2 \ll \omega^2 \tag{25}$$

which permit the following approximation of β:

$$\beta(\omega) \simeq \pm \frac{c}{\kappa\sqrt{2}} \sqrt{-1 + \sqrt{1 + 4\frac{\kappa^2}{c^4}\omega^2}}. \tag{26}$$

In practice, it is helpful to work with more perceptually significant parameters for sound synthesis purposes. For that purpose, we will now suppose that the string is of length L, with perfect reflections at the extremities. The delay D, which corresponds to the propagation time of the minimum phase-velocity wave along the length L can be expressed as

$$D = L/c = \pi/\omega_0, \tag{27}$$

where ω_0 is the fundamental frequency (rad/s) of the ideal string and c is the minimum phase velocity given by

$$c = \omega_0 L/\pi. \tag{28}$$

Thus, D is the propagation time over distance L for a sinusoidal traveling wave tuned to the first resonant mode of the ideal string. This choice of nominal propagation-time D simplifies the frequency-domain computations to follow; however, since phase velocity increases with frequency, the associated propagation filter F will be noncausal in the time domain. This poses no difficulty for frequency-domain implementation.

Next, we may express β as

$$\beta(\omega) \simeq \pm \frac{\pi}{L\sqrt{2B}} \sqrt{\xi} \tag{29}$$

with

$$\xi = -1 + \sqrt{1 + 4B\omega^2/\omega_0^2} \tag{30}$$

in terms of the inharmonicity coefficient[23] B given by

$$B = \kappa^2 \omega_0^2 / c^4, \tag{31}$$

where c denotes $c(\omega_0)$ for notational simplicity.

We now have to choose the sign of β in the expression for the phase. Since we want the output signal to be *delayed* with respect to the input signal, the loop filter/pure delay combination has to be causal. This means that the phase of the whole transfer function must be negative, i.e.,

$$-\omega D + \arg(F(\omega)) < 0. \tag{32}$$

This indicates that we choose the negative solution for β in the phase expression. Finally, using Eq. (29), we arrive at approximate expressions for the modulus and phase of the filter F as a function of frequency ω,

$$|F(\omega)| \simeq \exp\left(-D\left[b_1 + \frac{b_2 \pi^2 \xi}{2BL^2} \right] \right), \tag{33a}$$

$$\arg(F(\omega)) \simeq \omega D - \pi \sqrt{\frac{\xi}{2B}}. \tag{33b}$$

These expressions serve as the link the PDE model of Eq. (6) to the lumped filters of the digital waveguide. For the sake of simplicity, one can choose an "elementary filter" $\delta F = F^{1/D}$ such that

$$|\delta F(\omega)| \simeq \exp\left(-\left[b_1 + \frac{b_2 \pi^2 \xi}{2BL^2} \right] \right), \tag{34a}$$

$$\arg(\delta F(\omega)) \simeq \omega - \omega_0 \sqrt{\frac{\xi}{2B}}. \tag{34b}$$

The filters F of the digital waveguide, which correspond to propagation over a time duration D, can be then easily expressed in terms of δF by

$$F = \delta F^D. \tag{35}$$

The stability of the digital waveguide model is ensured if the modulus of the filter δF is less than one. This condition is here always respected, since the expression in the exponential of Eq. (33a) is negative.

V. NUMERICAL SIMULATIONS

We now address the validity of the analytical approach of the preceding section to determine digital waveguide loop filters. The modulus and phase of the filters corresponding to the digital waveguide model have been directly linked to the parameters of the physical model. For a given set of PDE parameters, thus, we can design a complete digital waveguide model simulating the signal of the vibrating string at a given location, for a predetermined excitation location. For purposes of comparison, we have generated output signals using both the finite difference scheme discussed in Sec. III, and the digital waveguide model for the same set of model parameters. The digital waveguide has been computed in the frequency domain, allowing the use of Eq. (33a). The excitation, a Gaussian function of the form $e^{-[(x-x_0)^2/\sigma^2]}$, simulates an initial velocity distribution of the string at a distance $x_0 = L/8$ from one end of characteristic width $\sigma = K_c L_H$, where K_c is an arbitrary constant and L_H is the hammer width (we do not enter into too much detail here, as hammer modeling is not dealt with in this paper). The signal is observed at the location $9L/10$. We have performed simulations for the notes C2, C4, and C7 using the parameters in Table I (all of which are taken from values provided in the papers by Chaigne and Askenfelt,[3,4] except for the parameter b_2, which comes from the calibration procedure applied subsequently in Sec. VI.

Figure 4 shows the two first periods of the waveforms generated by both approaches. The amplitudes are similar. Nevertheless, there is a slight discrepancy between the two signals, due to the numerical dispersion introduced by the finite difference scheme (see Sec. III).

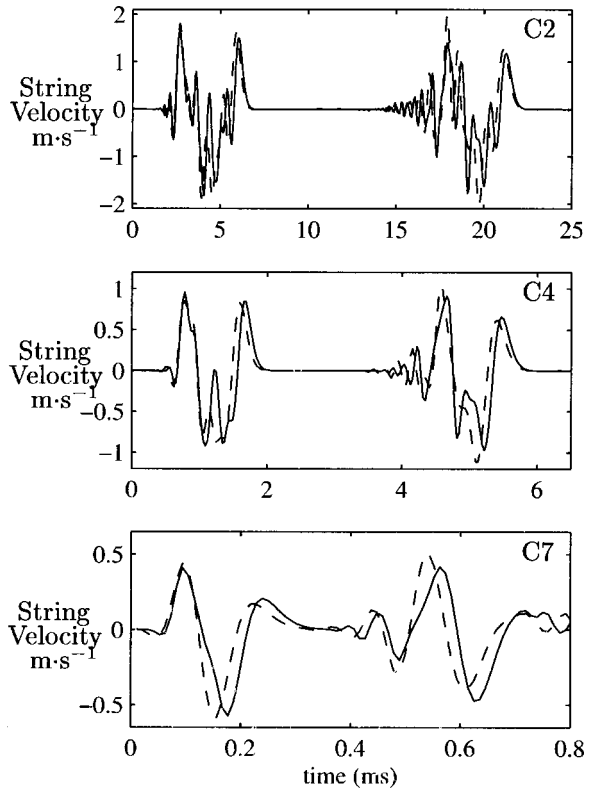


FIG. 4. Velocity signals obtained from the finite difference scheme presented in Sec. III (solid line) and a waveguide model (dotted line), for three different notes and for two periods of sound. The model is excited at distance $L/8$ from one endpoint, and the output signal is measured at distance $9L/10$. Note that the abscissa scales is different for each figure.

This discrepancy can be better seen by comparing the phase velocity of the two signals in Fig. 5. This corresponds to the phase velocity plotted in Fig. 1. The phase velocity of the signal produced by the waveguide is similar to the one of the model.

The long-time behavior of the generated signal is also very similar. In Fig. 6 spectrograms obtained over the whole length of the sound are shown. It is clear that the global damping behavior is similar in the low-frequency range. However, at high frequencies, the finite-difference model suffers from an artificially high propagation gain, as derived in Sec. III. The fundamental frequencies are essentially

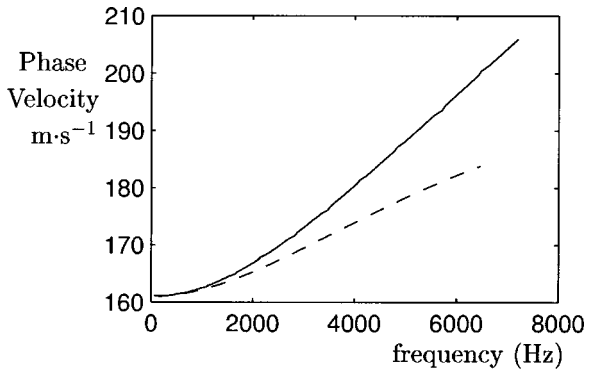


FIG. 5. Phase velocity for the waveguide model (solid line) and for difference scheme as a function of the frequency for the note C2. The analytically obtained phase is not shown, as it is identical (by definition) to the phase response of the waveguide network.

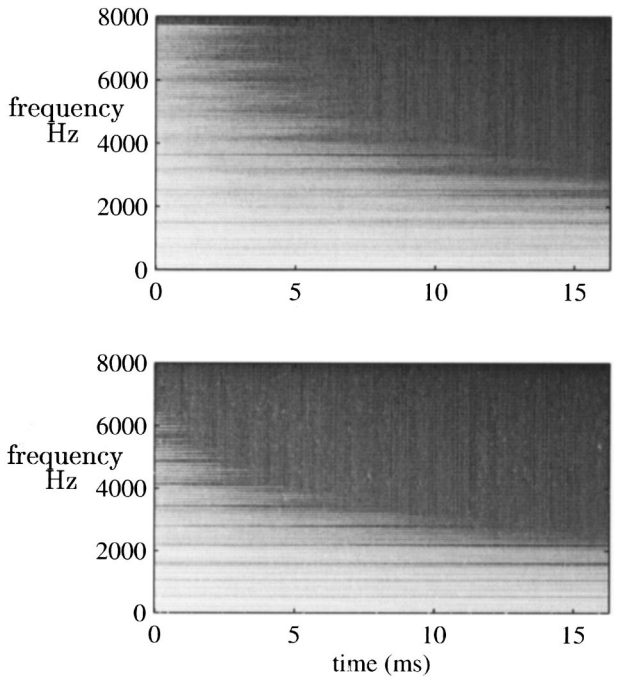


FIG. 6. Spectrograms of the output of the finite difference scheme (at top) and the waveguide model (at bottom) for the note C2.

equal, but the wave dispersion due to string stiffness is different due to the *numerical* dispersion introduced by the difference scheme. (We can see in Fig. 7 a slight tuning difference in the high-frequency partials.) In summary, these figures illustrate the extent to which the waveguide model provides a more accurate digital simulation of stiff, lossy strings with respect to both attenuation and dispersion of wave propagation, when compared with finite difference schemes.

VI. CALIBRATION OF PHYSICAL MODEL PARAMETERS FOR A GRAND PIANO FROM EXPERIMENTAL DATA

A. Experimental setup

In order to calibrate the parameters of the physical model, data were collected on an experimental setup consisting of a Yamaha Disklavier C6 grand piano equipped with sensors (see Fig. 8). The string vibrations were measured

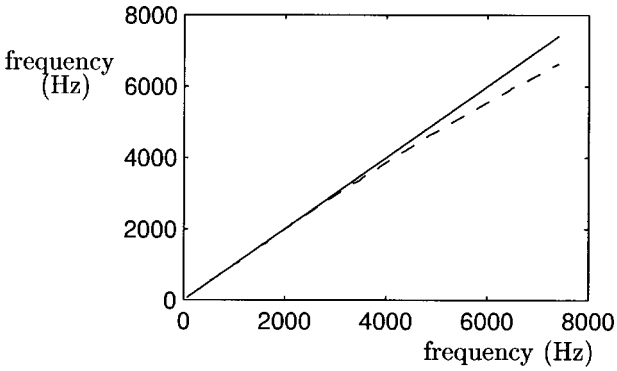


FIG. 7. Partial frequencies of the output of the finite difference scheme (dotted) and of the waveguide model (plain) as function of the theoretical frequencies.

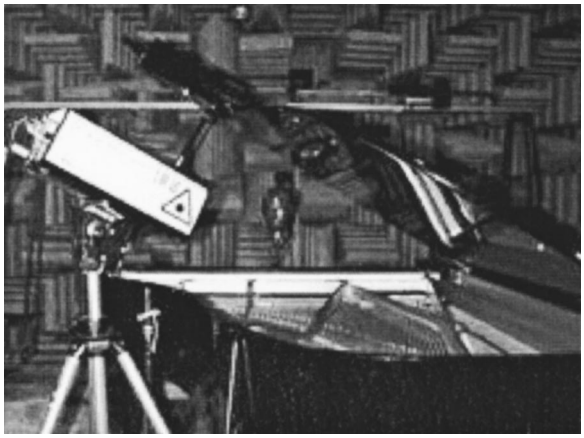


FIG. 8. Experimental setup. The grand piano was isolated in an anechoic room, and both the string vibrations and the acoustical radiated signal were measured. The string vibrations were measured using both an accelerometer located at the bridge level and a laser vibrometer, while the acoustic signal was measured at the head level of the pianist using an artificial headset. Our library of measured data also includes signals corresponding to various hammer velocities [referred to in recent and forthcoming articles concerned with the hammer-string interaction (Ref. 31)], but for this paper, we only need acceleration measurements for each string.

using an accelerometer located at the bridge level. The Disklavier allows the piano to be played under computer control. These measurements were made at the Laboratoire de Mécanique et d'Acoustique, Marseille, France.

In order to ensure that the measurements were taken under similar excitation conditions, we measured the hammer velocity using a photonic sensor (MTI 2000, probe module 2125H) (see Fig. 9).

Since we were interested in exciting a large portion of the frequency spectrum while remaining in the linear domain of vibration, we chose a medium (*mezzo-forte*) hammer velocity of 2.2 $\mathrm{m\,s^{-1}}$, which corresponds roughly to a MIDI value of 80. Such a hammer velocity allows the generation of about 140 spectral components for low frequency tones with a reasonable signal-to-noise ratio (see Fig. 10).

For each note of the grand piano, the optical sensor was placed close to the hammer and the accelerometer (B&K

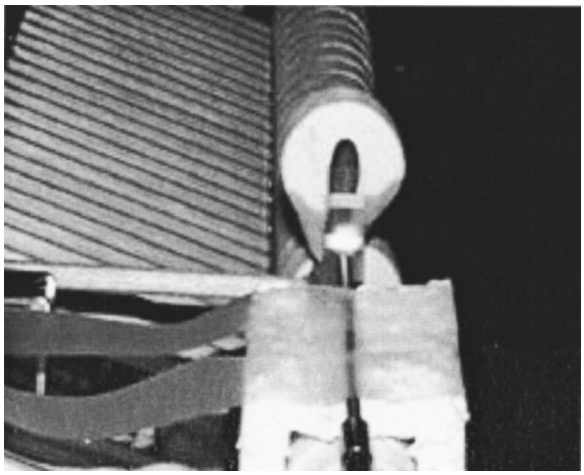


FIG. 9. Optical sensor used to measure the hammer velocity. The velocity is obtained through the duration corresponding to the travel time of the hammer between two reflectors placed on it.

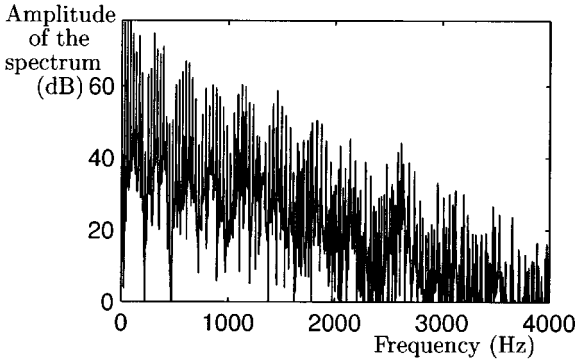


FIG. 10. Spectrum of the note A0.

4374) at bridge level. For notes corresponding to double or triple sets of strings, the accelerometer was placed as close as possible to the strings (see Fig. 11). Due to the imprecision of MIDI coding, several measurements were made, until a target value of the hammer speed was obtained. We have deemed an uncertainty of $\pm 0.1\ \mathrm{m\,s^{-1}}$ for the hammer velocity to be acceptable, as the estimation of modal frequency and decay rates is relatively insensitive to such an error. The acceleration measured at the bridge level was then digitally recorded at 16 bits at a sampling rate of 44.1 kHz, before being entered in the database.

B. Estimation of parameters

Because this model is intended for use in the context of musical sound synthesis, we here discuss the calibration of b_1 and b_2, and the determination of the stiffness parameter for a given string.

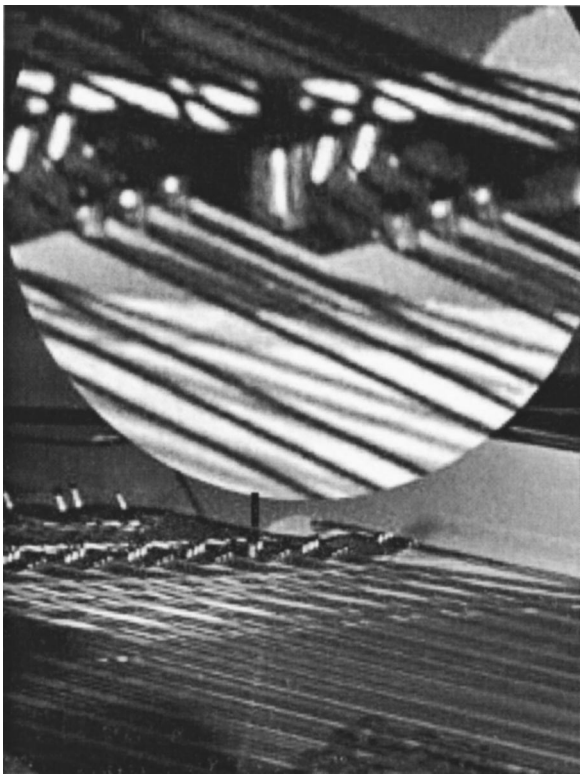


FIG. 11. Accelerometer at the bridge level.

To estimate the damping factor associated with each modal component of the signal, we used a signal processing technique based on the theory of analytic signals.[24,25] The analytic signal representation provides an easy way of extracting both the instantaneous frequency and the damping of each modal component, through band-pass filtering. To isolate each component in frequency, we used a truncated Gaussian window, the frequency bandwidth of which was chosen so as to minimize smoothing effects over the attack duration and to avoid overlapping two successive frequency components. The Gaussian was employed since its time-bandwidth product is minimized. As a consequence, it optimizes the exponential damping support after convolution with a causal single component for a given band pass filtering.[26]

The analytic signal associated with each individual component facilitates the estimation of both the instantaneous frequency and the amplitude modulation law of the component. The frequency dependent damping factor is directly related to the amplitude modulation law of each partial. According to the physics of a single string vibrating in one plane, the amplitude modulation of each component is expected to be exponential. This makes the damping factor easy to estimate by the measurement of the slope of the logarithmic representation of the amplitude of the analytic signal.[19] This technique has advantages over other parametric methods such as Prony's method,[27] mainly due to its ability to extract a coherent mean damping factor when multiple components are present. In fact, the hammer usually strikes not one, but two or three strings simultaneously. The coupling gives rise to perceptually significant phenomena such as beating and two-stage decay;[28] these effects are not accounted for in our model Eq. (6). For these multistring notes, the calculated damping coefficients σ can be thought of describing the global perceived decay of the sound.

For the same reason, the spectral representation of each partial is the result of the summation of several contributions due to each string and each polarization of vibration. The phase of the analytic signal allows an accurate estimation of the mean frequency of the partials. Actually, it permits the calculation of the instantaneous frequency which is a time-dependent function oscillating around a mean value. This mean value coincides with the spectral centroid of the partial[29] and consequently with the more likely perceived frequency. From the mean frequency values estimated this way for each partial, B may be deduced for each note, and consequently ω_0, the fundamental frequency of the corresponding ideal string. The inharmonicity factor B is plotted as a function of the frequency in Fig. 12; B is an increasing function of the note number, except over the bass range, where the strings are double-wrapped (this behavior has also been measured by Conklin[30]).

Combining Eqs. (8) and (29), one obtains

$$\sigma(\omega) = -b_1 - b_2\left(\frac{\pi^2}{2BL^2}\left[-1+\sqrt{1+4B(\omega/\omega_0)^2}\right]\right). \tag{36}$$

Then, from the value of σ obtained for each partial, b_1 and b_2 may be estimated for a given tone. The evolution of these parameters as a function of the fundamental frequency is shown in Fig. 13.

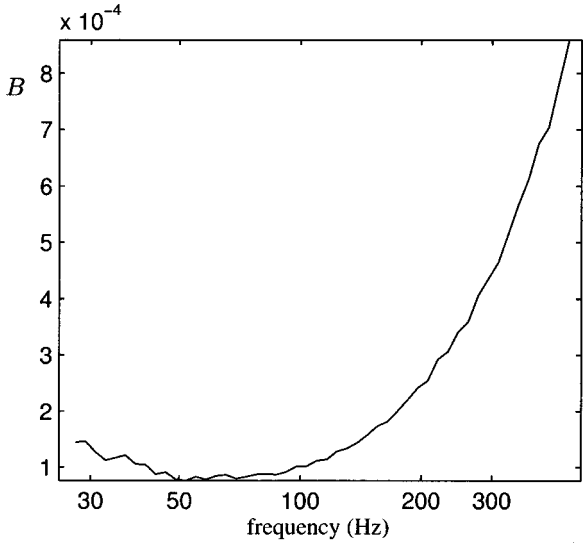


FIG. 12. The measured inharmonicity factor B.

We see that b_1 and b_2 are both increasing functions of MIDI note number, indicating increasing loss as one approaches the treble range. In Fig. 13, we have also fit extremely simple curves to the loss parameter data. The fits are linear as a function of the fundamental frequency, and are given by

$$b_1 = 4.4\times10^{-3} f_0 - 4\times10^{-2}, \tag{37a}$$

$$b_2 = 1.0\times10^{-6} f_0 + 1\times10^{-5}. \tag{37b}$$

These simple empirical descriptions of b_1 and b_2 allow the reproduction of piano tones whose damping will be close to that of the perceived acoustic note. A multistring waveguide

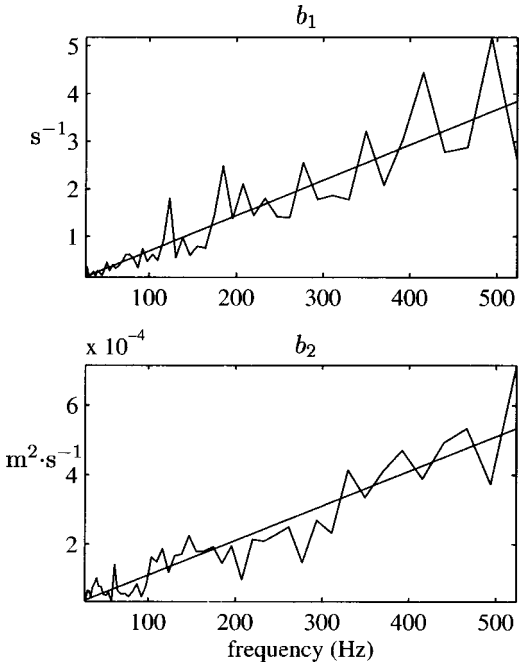


FIG. 13. Values of "equivalent" b_1 and b_2 fitted from measured data as a function of the fundamental frequency.

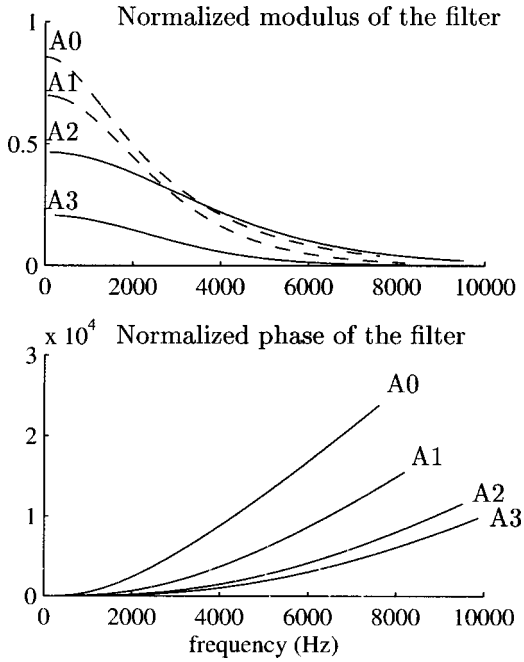


FIG. 14. Normalized modulus and phase of F for selected tones.

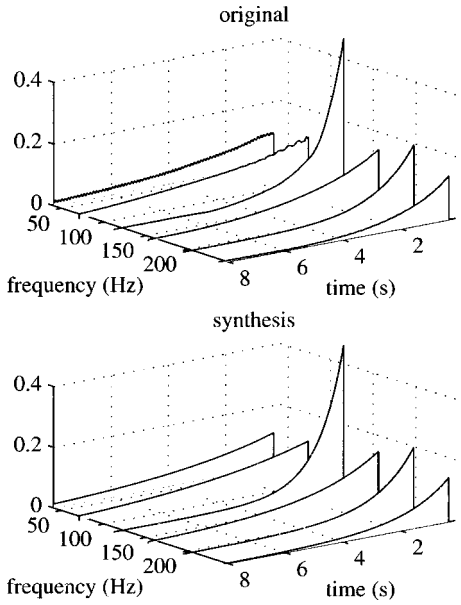


FIG. 15. Amplitude of partials one to six for the note E1 as a function of time and frequency.

model has also been designed,[26,31] allowing for beating and two-stage decay. Its use for synthesis purposes is discussed in Sec. VII. The deviations in the curves (Fig. 13) from the linear fits must be attributed to the impedance ratio between the strings and the soundboard, which varies over the length of the bridge. As a result, the curves should not be interpreted as impedance curves as they do not represent measurements at a single point and data are taken only at string modal frequencies. For a detailed discussion of soundboard impedance and its measurement, we refer to the work of Giordano.[32,33]

VII. WAVEGUIDES AND SOUND SYNTHESIS

The determination of b_1, b_2, B, and ω_0 for each note allows for an explicit expression of the behavior of the filter F as a function of note number. In order to represent the evolution of the elementary loop filter in terms of note value, we show in Fig. 14 the modulus and phase of the elementary filter δF, normalized with respect to the time delay D,

$$\delta F = F^{1/D}. \tag{38}$$

In order to understand the general behavior of the modulus and the phase of the loop filter with respect to the note played, we expand their expressions [Eq. (34a)] for $4B(\omega/\omega_0)$ to third order near zero. We obtain

$$|\delta F(\omega)| \simeq \exp\left(-\left[b_1 + b_2\frac{\omega^2}{c^2}\right]\right), \tag{39a}$$

$$\arg(\delta F) \simeq \frac{\omega^3 B}{2\omega_0^2}. \tag{39b}$$

The modulus (which also accounts for the losses at the endpoints), is decreasing with note number as shown in Fig. 14. This is mainly due to the increase in b_1. But the parameter b_2, which allows for frequency dependent loss is also increasing, leading to a modification of the slope of the modulus versus frequency for different note numbers. We note that this behavior is slightly different for the wrapped strings (A0, A1) than for the other strings (A2, A3). We also note that although B is mainly an increasing function of note number, the phase of the filters δF grows less rapidly for the bass tones than for higher tones. This is due to the fact that the phase of the filter depends not only on B, but also on the fundamental frequency, as evidenced by Eq. (39b). Though this expression is meaningful only for the first few partials, it is clear that phase dispersion decreases as a function of fundamental frequency.

In the case of the piano, the string is struck at a distance of approximately one-eighth to one-sixteenth of its length from the bridge, depending on the note. We are only interested in the vibration generated at the bridge termination, as this is the mechanism by which energy is transmitted to the soundboard. This situation corresponds to a digital waveguide structure identical to the one presented in Fig. 2, except that the loop filters F_2 and F_3 are combined. Using the parameters estimated by experiment, one can reproduce the vibration generated by a single string. Figure 15 shows the evolution of the amplitude of the first six partials of the vibration velocity for the note E1, respectively, measured on the piano and generated by the digital waveguide model. From a perceptual point of view, the two sounds are identical.

If the tones are produced by two or three strings struck simultaneously, the basic digital waveguide model still generates a signal having the same amplitude and damping of the modes. It does not, however, account for the modulations and double decays due to the coupling of the strings at the bridge.[28] Using two or three coupled digital waveguides, and thus allowing for energy transfer between the strings, one can easily reproduce this phenomenon.[20,26,31] Figure 16 shows the time evolution of the amplitude of the first six partials for the tone C2 (two strings), using the coupled-

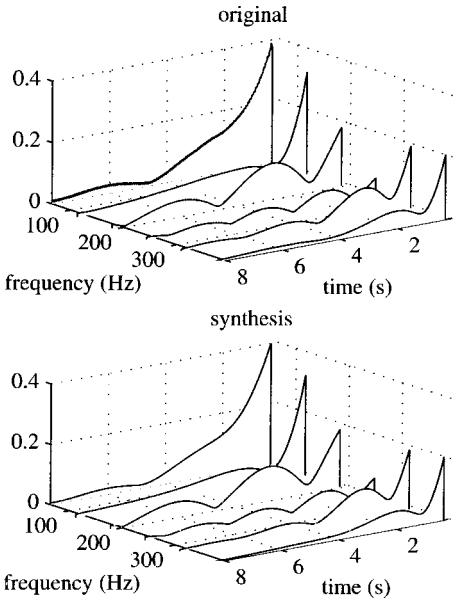


FIG. 16. Amplitude of partials one to six for the note C2 as a function of time and frequency.

strings model described in a recent publication.[26] The modulations are essentially perfectly reproduced. Moreover, this model follows directly from the physics of coupled strings. In fact, the loop filters are again related to the parameters of the physical model, and numerous sound transformations are conceivable. One could, for instance, extend the use of such a waveguide model to practically unrealizable situations involving, for example, widely mistuned strings or coupled strings with differing material properties.

VIII. CONCLUSIONS

We have presented a model of transverse vibrations on a string which includes the effects of stiffness and frequency-dependent loss. This model possesses several advantages over those proposed previously, in particular that it can be framed as a well-posed initial-boundary value problem (leading to stable numerical methods), and also that it can be easily related to digital waveguides. The source of these good properties is the fact that this model allows only two traveling-wave type solutions; the nonphysical third unstable term in the model of Ruiz (which can lead to difficulties both analytically and numerically) is thus eliminated in favor of higher-order spatial terms. It is also simple to write expressions for dispersion and loss as a function of frequency in terms of the model system parameters—such information is critical for the design of the terminating filters in a digital waveguide implementation.

For the sake of comparison, we have performed numerical simulations of the model system, with pianolike parameters, using both finite differences and a digital waveguide; the most notable distinction is the complete lack of numerical dispersion (which leads to mode mistuning) in the waveguide implementation. On the other hand, the waveguide allows the computation of a solution (''sound'') only at preselected points on the string, whereas a finite difference scheme computes the entire string state in (sampled) physical form. This is not a drawback for sound synthesis applications because, only the behavior of the string at the bridge is of interest in most stringed instruments. Moreover, physically accurate outputs from additional points along the string are easily added to a digital-waveguide simulation at the price of one small digital filter each.

A set of experimental data measured from a grand piano was used in order to calibrate the PDE model parameters over the entire keyboard range. String vibration was measured at the bridge through the use of an accelerometer, for each note on the piano, and for an average hammer velocity. The piano employed was equipped with sensors to provide hammer velocity data; from these measurements, all the parameters relating to the relevant PDE model were estimated. Given that the model itself is not completely physical—that is, the various loss mechanisms, interstring coupling, as well as energy transfer to the soundboard are modelled, for simplicity, as internal to the string itself—these parameters must be considered as those describing an ''equivalent'' string, under fixed termination. The equivalent parameters, are, however, sufficient for the resynthesis of piano tones to a high degree of fidelity, when a digital waveguide is employed. The digital waveguide model was also extended in order to directly take into account the effects of interstring coupling, through the use of two or three coupled waveguides.

The modeling of the excitation mechanism for the piano string (i.e., the hammer) is also of great importance, and must be carried out with some care; we have not addressed this issue here. As has been shown in the work of Chaigne and Askenfelt, it is possible to design a nonlinear hammer, which, when applied to a stiff string with frequency-dependent loss, produces signals quite similar to those measured on a real piano. The problem of extracting hammer parameters from measured data is also worthy of future research.

ACKNOWLEDGMENTS

This work was supported in part by the DSP (France) under Contract No. 016060 and by a fellowship from Region PACA (France).

APPENDIX: THE PIANO STRING MODEL OF CHAIGNE AND ASKENFELT

The results in this section have appeared, in a similar form, in the thesis of Ruiz.[9] We have added various comments regarding well-posedness and numerical stability.

The stiff string model in the thesis of Ruiz[9] and in the papers by Chaigne and Askenfelt[3,5] is described by the following equation:

$$\frac{\partial^2 y}{\partial t^2} = c^2 \frac{\partial^2 y}{\partial x^2} - \kappa^2 \frac{\partial^4 y}{\partial x^4} - 2b_1 \frac{\partial y}{\partial t} + 2b_3 \frac{\partial^3 y}{\partial t^3}. \tag{A1}$$

This model differs from Eq. (6) only by the replacement of the term $2b_2(\partial^3 y/\partial x^2 \partial t)$ by $2b_3(\partial^3 y/\partial t^3)$; this model also allows for frequency-dependent loss, but the system itself is of a quite different character, due to the increased degree of the equation with respect to the time variable. We spend a

little time here explaining the significance of the difference, which has a radical effect on the analysis of the system as a whole.

We can examine the well posedness of the system by inserting a solution of the form $y(x,t)=e^{st+j\beta x}$ into Eq. (A1), in order to obtain a dispersion relation,

$$-2b_3s^3+s^2+2b_1s+c^2\beta^2+\kappa^2\beta^4=0. \tag{A2}$$

This is a cubic in the variable s [the quantity on the left-hand side is referred to as the *symbol*[7] of Eq. (A1)], and again, as discussed in Sec. II, a necessary condition for the system of Eq. (A1) to be *well posed* is that the roots of this equation have real part *bounded from above* as a function of spatial frequency β. It is simple to see that the real part of at least one root of Eq. (A2) will be positive and unbounded as a function of wave number. As this is a third-degree polynomial equation with real coefficients, one root will always be real, and the two others occur as a complex conjugate pair (or perhaps as two other real roots). Consider Eq. (A2) as $|\beta|$ becomes large. In this case, the three roots will behave as

$$-2b_3s^3+\kappa^2\beta^4\approx 0$$

and will be evenly distributed over a circle of radius $(\kappa^2\beta^4/(2b_3))^{1/3}$. If $b_3>0$ (as suggested in the numerical experiments in the papers by Chaigne and Askenfelt[4]), then there will be one positive real root of the magnitude mentioned above, clearly unbounded as a function of wave number β. (If $b_3<0$, there will two roots in the right half plane, of this same magnitude, at an angle approaching ± 60 degrees with respect to the positive real axis.) We have thus shown that the initial value problem corresponding to the system of Eq. (A1) is, formally speaking, ill-posed.

We can extract some more detailed information by asking under what conditions the roots of Eq. (A2) have positive real part. A straightforward application of the *Routh–Hurwitz* stability criterion[34] to Eq. (A2) shows that in fact, if $b_3>0$, there is always exactly one real positive root, regardless of the value of the wave number β.

The following question then arises: How can we explain the apparently stable behavior exhibited by simulations[4] of these equations? Indeed, for realistic piano string parameters, the numerical integration routine provided in the paper by Chaigne and Askenfelt[4] is stable, and produces piano sounds of excellent quality. A first guess might be that the above analysis is incomplete due to the neglect of boundary conditions. Using the boundary conditions supplied by Chaigne, however, leads to an analysis identical to that performed in Sec. II B—the modal frequencies for the string system defined by Eq. (A1) will be given by solutions of the dispersion relation Eq. (A2) under the replacement of β by $n\pi/L$ for integer n. For any n, there will be exactly one modal frequency s_n with positive real part. Thus the instability persists even in the presence of boundary conditions.

We must then conclude that discretization has a stabilizing effect on system of Eq. (A1). To explore this idea in more detail, consider the discretization,[4] which can be written as

$$y_m^{n+1}=a_{10}y_m^n+a_{20}y_m^{n-1}+a_{11}(y_{m+1}^n+y_{m-1}^n)+a_{12}(y_{m+2}^n+y_{m-2}^n)+a_{21}(y_{m+1}^{n-1}+y_{m-1}^{n-1})+a_{30}y_m^{n-2}. \tag{A3}$$

This difference scheme involves three steps of lookback, reflecting the degree of the model system of Eq. (A1). Here, the difference scheme coefficients are defined by

$$\begin{aligned} &a_{10}=(2-2\lambda^2-6\mu^2+b_3/T)/D,\\ &a_{20}=(-1+b_1T+2b_3/T)/D, \quad a_{11}=(\lambda^2+4\mu^2)/D,\\ &a_{12}=(b_3/T-\mu^2)/D, \quad a_{21}=a_{30}=(-b_3/T)/D, \end{aligned} \tag{A4}$$

where, for brevity, we have again used

$$\lambda=cT/X, \quad \mu=\kappa T/X^2,$$

as well as

$$D=1+b_1T+2b_3/T.$$

Let us now examine the characteristic polynomial, which can be written as

$$z^3+a_1(\beta)z^2+a_2(\beta)z+a_3(\beta)=0 \tag{A5}$$

with

$$\begin{aligned} &a_1(\beta)=-a_{10}-2a_{11}\cos(\beta X)-2a_{12}\cos(2\beta X),\\ &a_2(\beta)=-a_{20}-2a_{21}\cos(\beta X), \quad a_3(\beta)=-a_{30}. \end{aligned}$$

The solution to the recursion will be bounded and decay if the solutions to this equation are confined to the interior of the unit circle for all $\beta\in[-\pi/X,\pi/X]$. It is simple to show that this is in fact true, for any of the choices of parameters given in the papers by Chaigne and Askenfelt.[4] This does not mean, however, that the difference scheme can be considered to be numerically stable in the Von Neumann sense.[7] This is a rather subtle point, and is worth elaborating.

According to the Lax–Richtmeyer equivalence theorem,[7] if the initial-boundary value problem is well-posed, the solution to a finite difference scheme will converge to the solution of the model problem if it is consistent and stable. In this case, though, the model system is not well posed, and thus no finite difference can possibly converge to a stable solution *in some limit as the time step T and the grid spacing X approach zero*. The difference scheme Eq. (A3) is indeed consistent with (A1) to first-order accuracy (and we note that if one does wish to use this ill-posed model system, it is in fact possible to design second-order accurate explicit methods), but it is possible to show (as we expect) that it cannot be stable in the limit as T becomes small. Because the recursion is of third order, the analysis is somewhat involved, and requires the application of the Schur–Cohn recursive procedure[7] (the discrete time analog of the Routh–Hurwitz stability test, again allowing us to check the stability of a polynomial without explicitly calculating the roots). Nevertheless, it is possible to show in this way that a necessary condition that the roots of the polynomial of Eq. (A5) be inside the unit circle is that

$$b_3/T\leqslant\lambda^2+4\mu^2\leqslant 1.$$

(The second inequality is exactly the necessary stability condition given in the paper by Chaigne and Askenfelt.[3]) Al-

though we have not provided all the details, we note that it suffices to perform the Schur–Cohn test at the wave number $\beta = \pi/X$ in order to arrive at these conditions. Clearly, these two conditions cannot be satisfied if

$$T \leqslant b_3 \Rightarrow \text{Sample rate} \geqslant 1/b_3$$

and thus for a small enough time step, the system poles must cross to the exterior of the unit circle, *regardless of the grid spacing X*.

It should be said, however, that because b_3 is in general quite small for realistic piano string models (on the order of 10^{-9}), for any reasonable sample rate in the audio range, the recursion does not exhibit this unbounded growth. On the other hand, as we have shown in this paper, it is simple enough to dispense with the nonphysical solution and all the concomitant analysis by making use of a simpler second-order model.

[1] L. Hiller and P. Ruiz, "Synthesizing musical sounds by solving the wave equation for vibrating objects," J. Audio Eng. Soc. **19**, 462–551 (1971).

[2] R. Bacon and J. Bowsher, "A discrete model of a struck string," Acustica **41**, 21–27 (1978).

[3] A. Chaigne and A. Askenfelt, "Numerical simulations of struck strings. I. A physical model for a struck string using finite difference methods," J. Acoust. Soc. Am. **95**, 1112–1118 (1994).

[4] A. Chaigne and A. Askenfelt, "Numerical simulations of struck strings. II. Comparisons with measurements and systematic exploration of some hammer-string parameters," J. Acoust. Soc. Am. **95**, 1631–1640 (1994).

[5] A. Chaigne, "On the use of finite differences for musical synthesis. Application to plucked stringed instruments," J. Acoust. **5**, 181–211 (1992).

[6] N. Fletcher and T. Rossing, *The Physics of Musical Instruments* (Springer-Verlag, New York, 1991).

[7] J. Strikwerda, *Finite Difference Schemes and Partial Differential Equations* (Wadsworth and Brooks/Cole Advanced Books and Software, Pacific Grove, CA, 1989).

[8] B. Gustaffson, H.-O. Kreiss, and J. Oliger, *Time Dependent Problems and Difference Methods* (Wiley, New York, 1995).

[9] P. M. Ruiz, "A technique for simulating the vibrations of strings with a digital computer," Ph.D. thesis, University of Illinois, 1970.

[10] K. Graff, *Wave Motion in Elastic Solids*, 3rd ed. (Prentice-Hall, Englewood Cliffs, NJ, 1974).

[11] J. O. Smith III, *Digital Waveguide Modeling of Musical Instruments*, http://www.ccrma.stanford.edu/~jos/waveguide/http://www-ccrma.stanford.edu/~jos/waveguide/, 2002 (printed version forthcoming).

[12] J. O. Smith III and S. A. Van Duyne, "Commuted piano synthesis," *Proceedings of the 1995 International Computer Music Conference, Banff* (Computer Music Association, 1995), pp. 319–326.

[13] S. A. Van Duyne and J. O. Smith III, "Developments for the commuted piano," *Proceedings of the 1995 International Computer Music Conference, Banff* (Computer Music Association, 1995), pp. 335–343.

[14] J. le Rond d'Alembert, "Investigation of the curve formed by a vibrating string, 1747," in *Acoustics: Historical and Philosophical Development*, edited by R. Bruce Lindsay (Dowden, Hutchinson & Ross, Stroudsburg, 1973), pp. 119–123.

[15] J. O. Smith III and P. Gossett, "A flexible sampling-rate conversion method," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, San Diego* (IEEE Press, New York, 1984), Vol. 2, pp. 19.4.1–19.4.2. An expanded tutorial based on this paper and associated free software are available online at The Digital Audio Resampling Home Page: http://www.ccrma.stanford.edu/~jos/resample/http://www-ccrma.stanford.edu/~jos/resample/

[16] S. Bilbao, Wave and scattering methods for the numerical integration of partial differential equations, Ph.D. thesis, Stanford University, 2001.

[17] J. O. Smith III, "Techniques for digital filter design and system identification with applicaltion to the violin," Ph.D. thesis, Stanford University (CCRMA), 1983. Available as CCRMA Technical Report STAN-M-14. Portions available online at http://www.ccrma.stanford.edu/~jos/http://www-ccrma.stanford.edu/~jos/

[18] J. Laroche and J. L. Meillier, "Multichannel excitation/filter modeling of percussive sounds with application to the piano," IEEE Trans. Speech Audio Process. **2**, 329–344 (1994).

[19] V. Välimäki, J. Huopaniemi, M. Karjalainen, and Zoltan Jánosy, "Physical modeling of plucked string instruments with application to real-time sound synthesis," J. Audio Eng. Soc. **44**, 331–353 (1996).

[20] J. O. Smith III, "Efficient synthesis of stringed musical instruments," *Proceedings of the 1993 International Computer Music Conference, Tokyo* (Computer Music Association, 1993), pp. 64–71. Available online at http://www.ccrma.stanford.edu/~jos/cs/

[21] M. Karjalainen, V. Välimäki, and Z. Jánosy, "Towards high-quality sound synthesis of the guitar and string instruments," *Proceedings of the 1993 International Computer Music Conference, Tokyo* (Computer Music Association, 1993), pp. 56–63. Available online at http://www.acoustics.hut.fi/~vpv/publications/icmc93-guitar.htm

[22] T. Hikichi and N. Osaka, "An approach to sound morphing based on physical modeling," *Proceedings of the International Computer Music Conference* (Beijing, Computer Music Association, 1999), pp. 108–111.

[23] H. Fletcher, E. D. Blackham, and R. Stratton, "Quality of piano tones," J. Acoust. Soc. Am. **34**, 749–761 (1961).

[24] D. Gabor, "Theory of communication," J. Inst. Electr. Eng., Part 1 **93**, 429–457 (1946).

[25] R. Kronland-Martinet, Ph. Guillemain, and S. Ystad, "Modelling of natural sounds using time-frequency and wavelet representations," Organised Sound **2**, 179–191 (1997).

[26] M. Aramaki, J. Bensa, L. Daudet, P. Guillemain, and R. Kronland-Martinet, "Resynthesis of coupled piano string vibrations based on physical modeling," J. New Music Res. **30**, 213–226 (2002).

[27] J. D. Markel and A. H. Gray, *Linear Prediction of Speech* (Springer-Verlag, New York, 1976).

[28] G. Weinreich, "Coupled piano strings," J. Acoust. Soc. Am. **62**, 1474–1484 (1977). Also contained in Ref. 36. See also Sci. Am. **240**, 94 (1979).

[29] Ph. Guillemain and R. Kronland-Martinet, "Characterization of acoustic signals through continuous linear time-frequency representations," IEEE Special Issue on Time-Frequency and Time-Scale Analysis, **84**, 561–585 (1996).

[30] H. A. Conklin, Jr., "Design and tone in the mechanoacoustic piano. III. Piano strings and scale design," J. Acoust. Soc. Am. **100**, 1286–1298 (1996).

[31] J. Bensa, F. Gibaudan, K. Jensen, and R. Kronland-Martinet, "Note and hammer velocity dependence of a piano string model based on coupled digital waveguides," *Proceedings of the International Computer Music Conference, La Havana, Cuba* (Computer Music Association, 2001), pp. 95–98.

[32] N. Giordano, "Mechanical impedance of a piano soundboard," J. Acoust. Soc. Am. **103**, 2128–2133 (1998).

[33] N. Giordano, "Sound production by a vibrating piano soundboard: Experiment," J. Acoust. Soc. Am. **104**, 1648–1653 (1998).

[34] M. Van Valkenburg, *Network Analysis* (Dover, New York, 1975).

[35] J. O. Smith III, Music applications of digital waveguides. Technical Report STAN-M-39, CCRMA, Music Dept., Stanford University, 1987. A compendium containing four related papers and presentation overheads on digital waveguide reverberation, synthesis, and filtering. CCRMA technical reports can be ordered by calling (650)723-4971 or by sending electronic mail requests to info@ccrma.stanford.edu

[36] *Five Lectures on the Acoustics of the Piano*, edited by A. Askenfelt (Royal Swedish Academy of Music, Stockholm, 1990). Lectures by H. A. Conklin, A. Askenfelt, E. Jansson, D. E. Hall, G. Weinreich, and K. Wogram. Sound example CD included. Publication number 64. Available online at http://www.speech.kth.se/music/5_lectures/

Blue whale (*Balaenoptera musculus*) sounds from the North Atlantic

David K. Mellinger[a)]
Bioacoustics Research Program, Cornell University, 159 Sapsucker Woods Road, Ithaca, New York 14850-1999 and Cooperative Institute for Marine Resources Studies, Oregon State University, 2030 SE Marine Science Drive, Newport, Oregon 97365[b)]

Christopher W. Clark[c)]
Bioacoustics Research Program, Cornell University, 159 Sapsucker Woods Road, Ithaca, New York 14850-1999

(Received 30 August 2002; revised 18 May 2003; accepted 30 May 2003)

Sounds of blue whales were recorded from U.S. Navy hydrophone arrays in the North Atlantic. The most common signals were long, patterned sequences of very-low-frequency sounds in the 15–20 Hz band. Sounds within a sequence were hierarchically organized into phrases consisting of one or two different sound types. Sequences were typically composed of two-part phrases repeated every 73 s: a constant-frequency tonal ''A'' part lasting approximately 8 s, followed 5 s later by a frequency-modulated ''B'' part lasting approximately 11 s. A common sequence variant consisted only of repetitions of part A. Sequences were separated by silent periods averaging just over four minutes. Two other sound types are described: a 2-5 s tone at 9 Hz, and a 5–7s inflected tone that swept up in frequency to ca. 70 Hz and then rapidly down to 25 Hz. The general characteristics of repeated sequences of simple combinations of long-duration, very-low-frequency sound units repeated every 1–2 min are typical of blue whale sounds recorded in other parts of the world. However, the specific frequency, duration, and repetition interval features of these North Atlantic sounds are different than those reported from other regions, lending further support to the notion that geographically separate blue whale populations have distinctive acoustic displays. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1593066]

PACS numbers: 43.80.Ka [WA]

I. INTRODUCTION

Several balaenopterid species of mysticete whales are known to produce intense, very-low-frequency sounds (<100 Hz). The biological functions of these sounds are not known with certainty, but several suggestions have been made. Patterson and Hamilton (1964) suggested that fin whale (*Balaenoptera physalus*) 20 Hz pulses could be used for echolocation, with a whale hearing the echo of its signals off large ocean features such as continental shelves and seamounts. Evans (1967) suggested that the sounds might specify the location and breeding condition of an individual. Payne and Webb (1971) theorized that low-frequency sounds from blue and fin whales might be used to communicate with other members of an extended herd scattered over thousands of kilometers. Watkins *et al.* (1987) suggested that fin whale sounds were produced by males and served a function similar to that for humpback song. Clark and Ellison (2003) emphasize the influence of environmental factors on the evolution of signal features optimized for long-range detection, and conclude that the long, patterned sounds of blue (*Balaenoptera musculus*) and fin whales function primarily as male advertisement displays with a possible secondary function of echo-ranging. Croll *et al.* (2002) recently provided the first conclusive evidence in support of the male display hypothesis by showing that in fin whales, only males produce long patterned sequences.

Long, loud, very-low-frequency underwater sounds with durations of 10 s or more and fundamental frequencies below 30 Hz were reported by several researchers as early as 1965, and were thought to come from biological sources (Weston and Black, 1965; Kibblewhite *et al.*, 1967). Six years later, Cummings and Thompson (1971a, b) expressed confidence in identifying the blue whale as the source of some of these sounds. This identification was based on visual sightings of blue whales, the absence of any other visible large marine mammal in the immediate vicinity, and the associated changes in the sounds' received intensities with the movements of a few whales toward and away from the recording hydrophone.

Most early descriptions of blue whale sounds were from the Pacific Ocean (Northrop *et al.*, 1971; Cummings and Thompson, 1977; Thompson and Freidl, 1982). In addition to providing basic descriptions of sounds (Thompson *et al.*, 1996; Rivers, 1997), researchers have used blue whale sounds to locate and track whales (Stafford *et al.*, 1998), to describe local movements (McDonald *et al.*, 1995) and larger-scale seasonal distributions (Stafford *et al.*, 1998, 1999a) of individuals, and to investigate the possibility of using sound features to distinguish between different populations (Ljungblad *et al.*, 1998; Stafford *et al.*, 1999a, b). Alling *et al.* (1991) recorded blue whales, which may have

[a)]Electronic mail: david.mellinger@oregonstate.edu
[b)]Address for correspondence.
[c)]Electronic mail: cwc2@cornell.edu

 0001-4966/2003/114(2)/1108/12/$19.00

been the pygmy subspecies (*B. m. brevicauda*), off Sri Lanka. Clark and Fristrup (1997) provided basic descriptions and illustrations of common blue whale sounds from different oceans. Sounds from a North Atlantic blue whale have been described only once based on a single recording of seven sounds in the presence of a single individual in the St. Lawrence River (Edds, 1982).

With the end of the cold war, we were granted limited access to the U.S. Navy Sound Surveillance System (SOSUS), an ocean-scale network of fixed hydrophone arrays originally deployed for listening for Soviet submarines (Costa, 1993; Gagnon and Clark, 1993; Clark and Mellinger, 1994; Clark, 1994; Nishimura and Conlon, 1994). By monitoring the Atlantic arrays, it was possible to detect whales over large areas of the North Atlantic.

Here we describe a relatively large sample of blue whale sounds collected via Navy SOSUS arrays. This acoustic recording effort was not systematic or uniform throughout the North Atlantic. Rather it was an attempt to collect representative sounds samples from blue, fin, humpback (*Megaptera novaeangliae*) and minke (*B. acutorostrata*) whales as part of a larger ''dual-use'' program aimed at demonstrating the value of SOSUS for basic research. For reasons related to the process of declassifying acoustic data, the vast majority of sound recordings reported here were collected from the western North Atlantic arrays, with only a few recordings from the northeastern North Atlantic. One of us (CWC) did routinely have access to data from other parts of the SOSUS network, both Atlantic and Pacific oceans, and occasionally from ships with towed arrays. These provided numerous opportunities to compare blue whale sounds from different regions of the North Atlantic. The acoustic recording collection was also augmented by the systematic documentation of the occurrence of blue, fin, humpback and minke whale sounds throughout the western North Atlantic (Mellinger *et al.*, 2000; Clark and Gagnon, to appear) and northeastern North Atlantic (Charif *et al.*, 2001).

Almost all the blue whale sounds reported here were in the infrasonic frequency range and occurred in patterned sequences. The exceptions include several unusual sound types that are somewhat similar to sounds reported for blue whales in other parts of the world (Thompson *et al.*, 1996; Ljungblad *et al.*, 1997; Bass and Clark, 2002). All recorded sounds were easily distinguishable from the infrasonic, patterned sounds of fin whales (Watkins *et al.*, 1987).

All sounds described here are characterized by their frequency, duration, and timing patterns. We also provide data on the percentage of time that an animal produced sounds, a value often referred to as duty cycle. Duty cycle is an important parameter in models estimating the probability of an acoustic detection because it quantifies how often an animal is acoustically available for detection. All of these acoustic descriptions bear on the possible use of acoustic cues to detect blue whales during marine mammal surveys, and provide information that may be useful for automatically locating and tracking blue whales. In addition, North Atlantic blue whale vocalizations are briefly compared to those from other oceans to reveal some basic acoustic similarities and differences between populations.

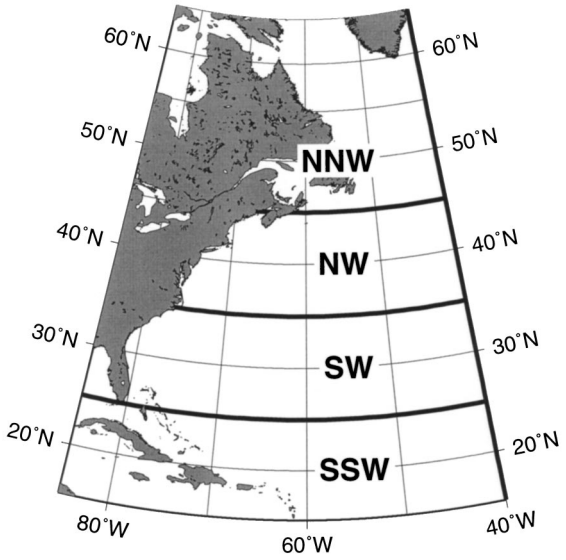


FIG. 1. Geographic locations of the recording areas shown in Table I.

II. METHODS

Recordings in the western North Atlantic were collected from the U.S. Navy's SOSUS system between December 1992 and December 1994. Geographic areas from which most of these sounds were collected are indicated in Fig. 1. One additional recording was from the eastern North Atlantic, east of the ''NNW'' region of Fig. 1. Many of the details of the SOSUS system remain classified, but the lack of those details here does not influence the validity or resolution of the data presented. The fixed arrays are installed to receive sounds that have traveled though the deep sound channel (SOFAR channel) (Urick, 1983), making it possible to detect blue whale sounds over enormous distances.

Sounds received on the hydrophones of each array were input to the SOSUS beamformer system, and selected beamformed data from different arrays were recorded and declassified for later analysis. The low-frequency rolloff point of the SOSUS system was below the lowest frequency of blue whale sounds. Sounds were digitally recorded with 15-bit resolution on a TEAC RD-135 data recorder at a sampling rate of 12 kHz on 2 h DATs. All digital recordings were later transferred to Macintosh and Sun computers. Sounds were digitally filtered and decimated to sampling rates of 400, 200, or 100 Hz, depending on the frequency of the whale vocalization analyzed. The 3-dB roll-off points of the low-pass filters used in downsampling were at 194, 93, and 43 Hz, respectively.

Sound recordings were displayed as continuous spectrograms using the bioacoustic sound software packages Canary (Charif *et al.*, 1995) and Osprey (Mellinger, 1994). Sounds were identified as produced by blue whales based on their similarity in frequency contour and duration to the Atlantic blue whale sounds reported by Edds (1982), and by their general similarity in frequency, duration, and timing to other sounds recorded elsewhere in the world when blue whales were seen (Cummings and Thompson, 1971a; Thompson *et al.*, 1996; Rivers, 1997; Clark and Fristrup, 1997; Stafford

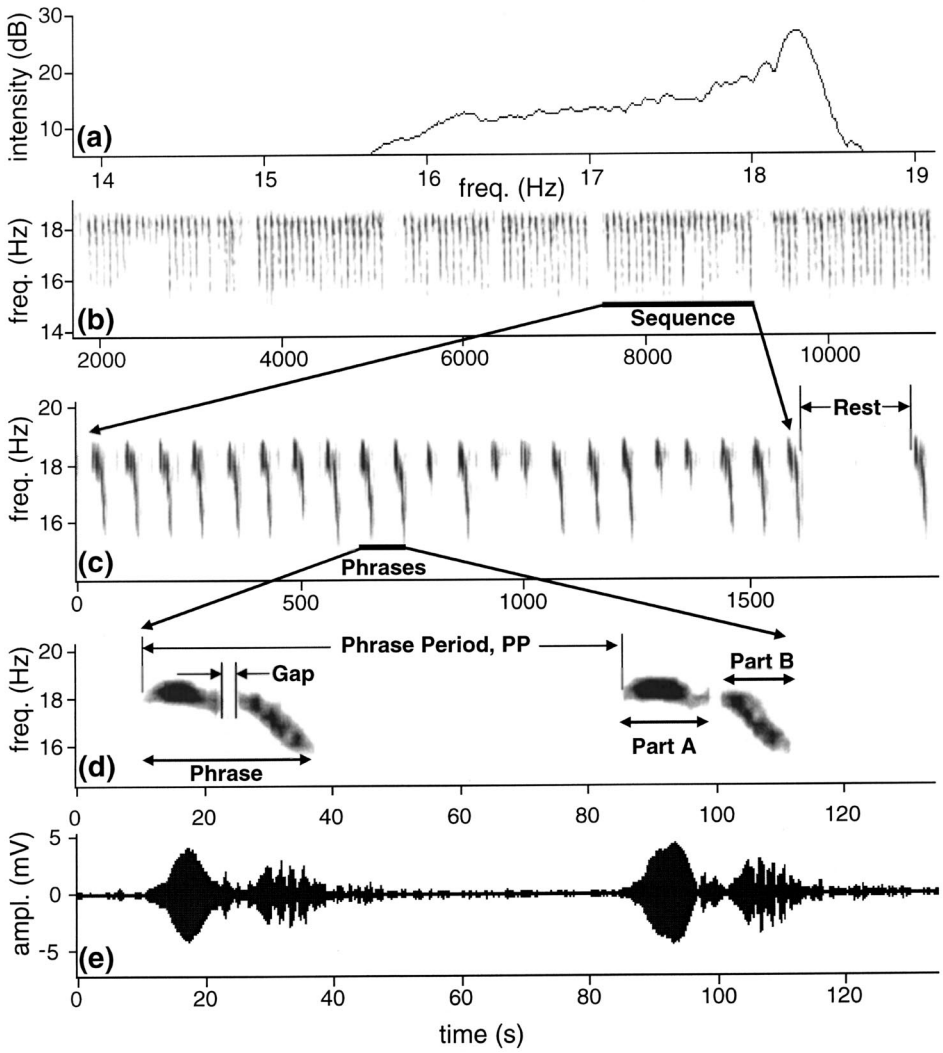


FIG. 2. Blue whale sounds illustrating the hierarchical nature of these vocalizations and terminology. (a) Spectrum for a sequence of phrases. This recording's intensity was not calibrated, so only relative measures of intensity are meaningful. (b) A portion of a bout containing several successive sequences. (c) A single sequence containing 22 phrases, followed by a rest. (d) Two phrases from that sequence showing part A, part B, the gap between them, and the phrase period. (e) The time series representation of the sound in (d) showing the relative intensity of the various parts.

et al., 1998; Ljungblad *et al.*, 1998; Thode *et al.*, 2000; McCauley *et al.*, 2001; McDonald *et al.*, 2001).

A stereotyped, regularly repeating series of sounds is referred to as a *sequence* (i.e., a song in humpback whale terms). A series of repeated sequences was identified as coming from the same animal using several cues based on visual inspection of a continuous spectrogram display. Typically, sequences were sufficiently uncommon that it was rare to have two sequences appearing simultaneously on the same display. Within a sequence, successive sounds typically appeared similar in intensity, frequency contour, and reverberation characteristics. In the rare cases when sound characteristics in a display appeared to change quickly, the sounds on either side of the change were treated as coming from two different whales. In the rare case when two (or very rarely, three) sequences were present at the same time, differences in intensity, frequency contour, or multipath characteristics were usually obvious, making it simple to distinguish sounds from the different animals. Usually in these cases it was possible to analyze only sounds from the more intense sequences, but sometimes sounds from two whales could be analyzed. In cases where display features suggested that two or more sequences were present, but they could not be easily distinguished, the recording was discarded from the analysis set.

Based on these selection criteria, a series of sequences was determined as coming from the same animal. From observations and analysis of SOSUS displays, some blue whales will produce long series of sequences for many days (see Tyack and Clark, 2000, Fig. 4.17), referred to as a *bout*. Because recording duration was limited by the times at which we were able to make recordings at a Navy facility, we could not know if the first or last recorded sequence was the start or end of a bout, respectively. Therefore, in all further discussion, a series of sequences recorded from the same animal during the same recording session is referred to as a *session* rather than a bout, and a session represents the vocal behavior of a single whale. The total number of individuals contributing to the sample of sessions is unknown.

There is a clear hierarchical organization in blue whale sequences, as illustrated in Fig. 2. The basic elements are individual sound units (Payne and McVay, 1971), here called *parts*, consisting of a frequency contour with a steady frequency or a constant rate of frequency change. Combinations of parts are organized into *phrases*. One or more phrases are repeated in a regular cadence to form a *sequence*, and se-

TABLE I. Recording locations (see Fig. 1) and dates. Each number in the table is the number of sessions recorded at the date and in the region specified. Recording locations for the seven sessions in the ''Unknown'' column have been lost.

	Region				
	SSW	SW	NW	NNW	Unknown
Dec. 1992					2
Jan. 1993	1			2	
Mar. 1993	1	2			
Dec. 1993	4	5	3		
Jan. 1994	8	7	7		
Feb. 1994	3	13	1		
Mar. 1994		3			
Apr. 1994				2	
Sep. 1994	4	8	1		
Oct. 1994	1	1	1		5
Dec. 1994	3		2		

quences are repeated to form a *session*. Sequences are separated by a period of silence referred to as a *gap*. The time from one phrase onset to the next is referred to as the *phrase period (PP)*. PPs were categorized into two types. A *regular PP* is the silent period between two phrases within the same sequence. A *rest* is the silent period between the last phrase of one sequence and the first phrase of the next sequence. In blue, fin, and humpback whales, rests are typically several minutes long and often coincide with the whale coming to the surface to breathe (Cummings and Thompson, 1971a; Payne *et al.*, 1983; Watkins *et al.*, 1987). Sequence durations were measured from the start of the first sound in a sequence to the end of the last sound in the sequence. Rests were measured as the duration between the end of the last sound before a rest and the start of the first sound after the rest.

The duty cycle of each session was calculated by examining the time from the start of the first sequence of the session to the start of the last sequence. The durations of all sounds recorded within this time period were summed and divided by the duration of the time period. A sequence was included in the duty-cycle analysis only if the entire sequence and its succeeding gap were recorded. By this procedure, the duty cycle measure is biased high since the total time period does not include silent gaps between bouts.

Each part of each phrase in a session was measured for a set of relatively simple acoustic features: start time, end time, start frequency, and end frequency. These measurements were used to calculate values for part duration, gap duration, phrase duration, regular PP, rest duration, sequence duration, number of phrases/sequence, and duty cycle. All measurements were made from spectrograms. Sounds were converted into spectrograms with frame length of 2.56 s, 75% overlap, Hamming window, and FFT size of 5.12 s (i.e., with zero-padding to double the length of the frame). The frame length as measured in samples was 256, 512, or 1024 samples for sampling rates of 100, 200, or 400 Hz, respectively. These parameters resulted in an effective filter bandwidth of 1.6 Hz, a frequency grid spacing of 0.2 Hz, and a temporal resolution of 0.6 s.

Normally the start or end of a sound merged into the background noise, making it difficult to precisely measure a sound's initial and final frequency and time values. To partially control for variation in signal quality and ambient noise, recordings were subjectively graded from 1 to 4 according to visual clarity in the spectrogram. The very worst, quality-4 vocalizations were considered unusable and were not even recorded. The quality-3 recordings were good enough only for measuring the time at which each phrase occurred, so that phrase periods could be calculated; neither frequency measurements nor phrase duration measurements were possible. Because data were recorded principally when blue whale sounds were clearly identifiable, there were relatively few quality-3 recordings. To measure the time at which phrases occurred, a single point of relatively high intensity in the frequency contour of each phrase was selected, and the time of this point was recorded. These times were then used for further analysis of phrase period as described below. It was often the case for quality-3 recordings that only this most intense point could be readily discerned, with other parts of each phrase appearing indistinct. For all quality-2 and quality-1 sounds, spectrogram intensity was standardized by setting it such that background noise was just visible in the spectrogram image. Start and end time measurements of contours were made at the points where the frequency contour appeared out of, or disappeared back into, background noise. Measurements of a sound's frequency modulation (FM) contour were made at the start and end of the contour, with the start and end frequency values selected as the highest-amplitude value of the frequency spectrum at the start or end of the contour, respectively.

Another form of distortion was multiple arrivals of the same vocalization, caused by multipath propagation. In spectrograms, these multiple arrivals usually appeared as two or more repetitions of the FM portion of a sound, separated in

TABLE II. Acoustic characteristics of blue whale A-B sequences. ''*n* sessions'' is the total number of sessions measured for sequence measurements. ''*n* sequences'' is the total number of sequences for all sessions measured. These *n*'s vary because different characteristics were measurable in different sessions. ''Mean ± s.d.'' is the mean of averages for sessions and the standard deviations of averages for sessions, respectively. ''Min.–Max. range'' is the minimum and maximum range of sequence per session.

Feature	Feature measure	*n* sessions	*n* sequences	Mean ± s.d.	Min.–Max. range
Sequence	No. of phrases/session	69	298	10.1 ± 3.2	1–25
	Duration	62	251	678.9 ± 234.4 s	4.8–1819.3 s
Rest	Duration for all rests	74	346	237.7 ± 75.5 s	81.3–1634.6 s
	Duration for rests < 500 s	72	325	193.6 ± 46.0 s	81.3–492.2 s

TABLE III. Numbers and percentages of the three different phrase types, out of 3839 total phrases in 80 sessions for which phrase type could be determined.

Phrase type	No. of phrases	No. of sessions with at least one phrase
A-B	2522 (65.7%)	71 (89%)
A-only	1294 (33.7%)	67 (84%)
B-only	23 (0.6%)	9 (11%)

time by 1–3 s. When multipath effects were present, the clearest arrival of the vocalization was chosen as the one to measure. In order to get accurate intersound timing measurements, the corresponding multipath arrival was measured on successive vocalizations as well. Choosing which multipath arrival to measure required looking at an entire sequence of vocalizations to see which arrival remained the clearest over time.

III. RESULTS

A. Vocalization types and characteristics

A total of 236.1 h of recordings from the western North Atlantic from December 1992 through December 1994 (Table I) were analyzed for blue whale sounds. Tables II–IV list sample sizes and measurement statistics for sequences and phrases. Overall, analysis resulted in the identification of 90 sessions, which contained a total of 298 complete sequences and 4635 phrases. Of these 90 sessions, 80 were of a quality high enough to allow measurements of sound frequencies and durations for at least some of the sequences (20 sessions contained some quality-1 sequences and 67 sessions contained some quality-2 sequences; 7 of these sessions contained both quality-1 and quality-2 sounds). For the remaining 10 quality-3 sessions, only phrase timing information could be extracted. Six of the 90 sessions included overlapping sequences, in which two whales were present on the same recording.

1. Sequences

At least one complete sequence was recorded in 69 sessions, resulting in 298 complete sequences containing a total of 3006 phrases (see Table II). Across sessions there was a high level of variability in the number of phrases per sequence, ranging from 1 (a single isolated phrase) to 25. If p_i is the average number of phrases per sequence within session i, then the mean number of phrases per sequence averaged over all sessions, $\overline{p_i}$, was 10.1 ± 3.2 (mean $\pm$ s.d., $n = 69$), with a range of 2.2–25.

2. Phrases

The most common phrase type consisted of two parts, as illustrated in Figs. 3(a) and (b). The first part of each phrase was a tonal sound, either of constant frequency or with a very modest change in frequency. The second part was a tonal sound descending in frequency. Following the naming conventions used for North Pacific blue whale sequences, the two components of this phrase are designated *part A* and *part B*, and a phrase having both part A and part B is referred to as an *A-B phrase*. Part A was usually separated from part B by a short silent gap, but sometimes was contiguous with part B to form a continuous frequency contour. There were two other, less common phrase variations, one in which part B was absent, the *A-only phrase* [Figs. 3(c) and (d)], and one in which part A was absent, the *B-only phrase*.

Table III lists the numbers and proportions of different phrase types found in the 80 sessions for which phrases could be measured. All 80 sessions included at least one A-B or one A-only phrase, but the patterns varied. Most sequences included some mixture of A-B (89%) and A-only (84%) phrases—typically in a seemingly random order—but some included just A-B phrases and some just A-only phrases (see Fig. 4). Successive sequences from the same whale showed similar patterns, so that, for instance, a whale making A-only phrases in one sequence would continue to make A-only phrases in later sequences. The B-only phrase type was rare. It was made in only 9 (11.2%) of the 80 sessions and accounted for only 23 (0.6%) of the 3839 measured phrases.

Very rarely, phrases had harmonics, as shown in Fig. 5. On average, the second harmonic was less intense than the first harmonic by 19.8 ± 1.6 dB (mean $\pm$ s.d.). It is possible for harmonic distortion to occur in a sound signal because of

TABLE IV. Acoustic characteristics of blue whale A-B phrases. "*n* sessions" is the total number of sessions measured for sequence measurements. "*n* total" is the total number of phrases for all sessions measured. "Mean $\pm$ s.d." is the mean of averages for sessions and the standard deviations of averages for sessions, respectively. "Min.–Max. range" is the minimum and maximum range over all individual phrases, not the minimum–maximum range for session averages. Sample sizes (n) vary because different sessions contained different types of phrases, and because for some sessions, only measurements of phrase period were possible.

Feature	Feature measure	*n* sessions	*n* total	Mean $\pm$ s.d.	Min.–Max. range
Unit-Part A	start frequency	80	3816	18.5 ± 0.11 Hz	17.8–19.4 Hz
	end frequency	80	3816	18.3 ± 0.16 Hz	17.1–19.0 Hz
	duration	80	3816	8.2 ± 2.22 s	1.7–28.4 s
Unit-Part B	start frequency	72	2545	18.5 ± 0.13 Hz	17.5–19.1 Hz
	end frequency	72	2545	15.7 ± 0.27 Hz	14.3–17.6 Hz
	duration	72	2545	11.4 ± 1.84 s	4.1–19.7 s
Interunit gap (A-B gap)	duration	72	2522	4.6 ± 1.79 s	0.0–14.18 s
Phrase	duration	80	3839	17.7 ± 5.39 s	2.3–43.8 s
Regular phrase period	duration	90	4025	73.3 ± 5.0 s	34.8–98.9 s

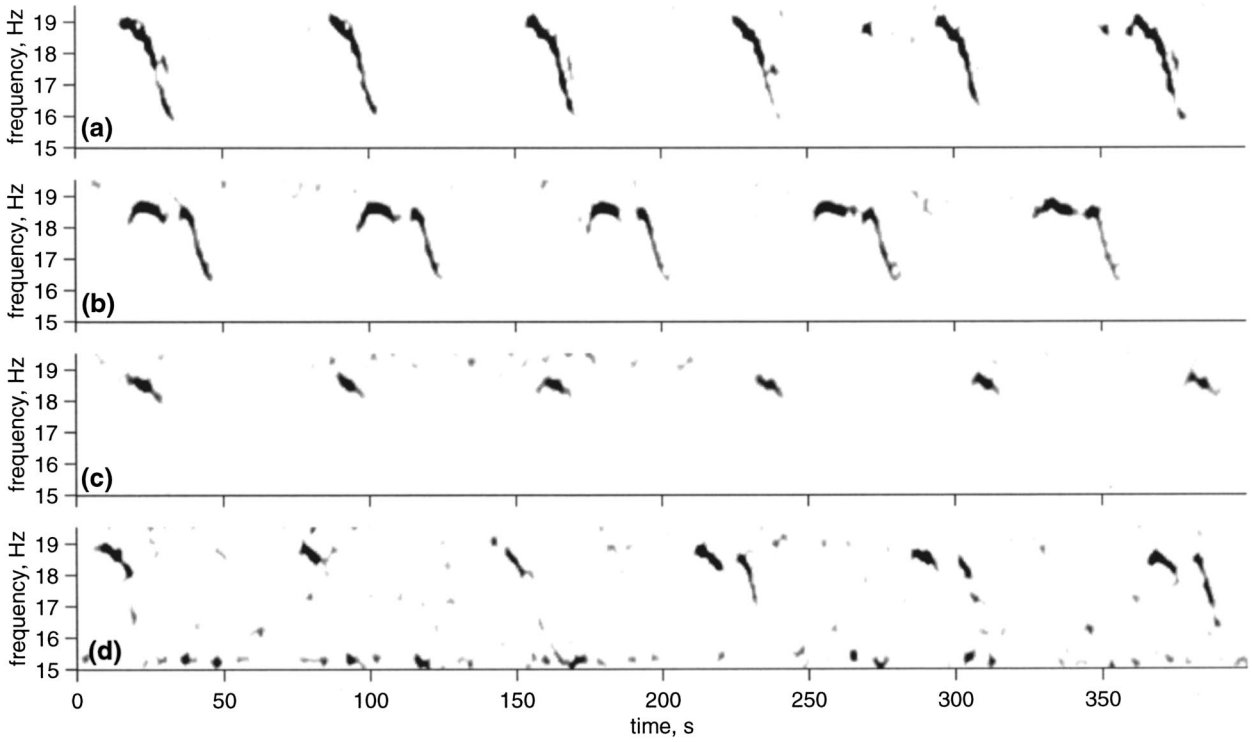


FIG. 3. Examples of blue whale partial sequences illustrating some of the typical types of phrases. (a) Series of the most common blue whale phrase type in which part B follows part A without a gap and forms a continuous frequency contour (Grand Banks, Newfoundland, 18 December 1992). A slight echo is evident just after some part B sounds. (b) Another common phrase type in which part A is followed by a short silent gap before part B, and the start of part A sweeps up slightly in frequency (near Bermuda, 11 January 1994). (c) A sequence consisting of only part A sounds (East of Bermuda, 21 December 1994). (d) A sequence consisting of a mixture of A-only phrases and A-B phrases (East Coast of the United States, 20 December 1993). In this case, part B starts at a higher frequency than the end of part A. (Spectrogram parameters: frame size 5.12 s, 87.5% overlap, FFT size 20.48 s, Hamming window. These parameters are used in all succeeding spectrograms unless noted otherwise.)

clipping at some stage in the recording process. However, this does not appear to be the case in any of these recordings since the time waveform before resampling did not appear to be clipped. Also, the fin whale 20 Hz waveforms present in these recordings, which were as great as 7 dB higher in amplitude than the blue whale sounds, did not have harmonics and were not clipped.

Table IV shows statistics for phrase measurements. From these data, a typical sequence was composed of a variable number of two-part phrases. A typical phrase lasted *ca.* 18 s and was repeated *ca.* every 73 s. The first part of a phrase was a constant 18–19 Hz tone lasting *ca.* 8 s. The second part was an 18–15 Hz downsweep lasting *ca.* 11 s and separated from the first part by 5 s of silence. Phrase measures had low coefficients of variation for frequency characteristics and moderate to high levels for duration characteristics, a reflection of the stereotypy in frequency content and the influence of propagation on received amplitude.

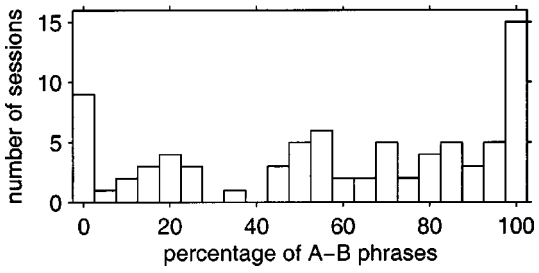


FIG. 4. Percentage of A-B phrases made in the sessions ($n=80$). *X*-axis: percentage of A-B phrases out of all phrases in a session; thus, the vertical bar at 0 represents those whales that made entirely A-only sounds, while the bar at 100 represents whales that made entirely A-B phrases, and intermediate bars represent whales that produced both types of phrases in various proportions. *Y*-axis: Number of sessions in which the given percentage of A-B phrases was found.

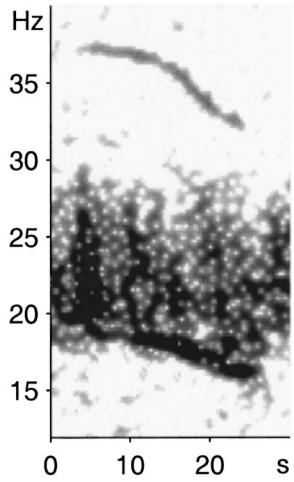


FIG. 5. A-B phrase with harmonic. The harmonic at 33–36 Hz was not caused by clipping during the recording process (see text). This spectrogram also contains fin whale sounds in the 20–28 Hz band. Recorded in the eastern North Atlantic, 11 April 1994.

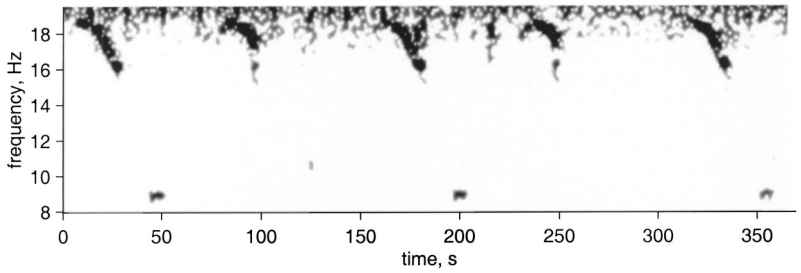


FIG. 6. The 9 Hz sound. This sound occurs only in conjunction with A-B phrases, and only after part B. Recorded off Newfoundland, 18 December 1992.

3. The 9 Hz sound

On a few recordings from several sites in the western North Atlantic, a brief sound at 9 Hz was observed after the end of a part B. Examples of this sound, named the *9 Hz sound*, are shown in Fig. 6. The duration of this sound was difficult to measure, since multipath effects were evident in every recording in which it was observed. However, duration was estimated to be in the range of 2–5 s. This sound was found only in association with A-B phrases, and always after the end of part B. It was observed only on the very clearest recordings, and its intensity was typically 14 dB lower than part A. Only a few recordings had a high enough signal-to-noise ratio for this 9 Hz sound to be detectable, so the fraction of sessions that included it could not be determined. On some recordings, the 9 Hz sound appeared after every part B sound, while for other sessions, it appears after only some of the part B sounds.

4. The arch sound

Another type of vocalization, the *arch sound*, was recorded over a period of several days in December 1994 in the eastern North Atlantic. Some examples of this sound are shown in Fig. 7. Navy personnel recognized this as an unusual whale sound and specifically noted that it occurred in sequences coincidentally with a long series of A-B phrase sequences. They archived the acoustic data from multiple arrays while collecting some locations for both the arch sounds and sequences. Several weeks later one of us (CWC) visited the Navy facility, converted the multi-array data onto DATs, and reprocessed these data for locations.

The sequences of locations for the A-B phrases and the arch sounds throughout an 8-hour period produced two nearly identical tracks with the locations of A-B phrases interleaved with arch sound locations. Arch sound sequences alternated with sequences of A-B phrases in a very obvious and predictable manner. Repeatedly, within tens of seconds after termination of an A-B sequence an arch sequence began. Within tens of seconds after termination of the arch sequence an A-B sequence began. The time gaps between these antiphonal events were typically less than the times between adjacent A-B phrases. There are only a few reasonable explanations for this coincidence in A-B phrase and arch phrase locations and their antiphonal occurrence. Either one animal produced both the A-B phrase sequences, typical of blue whales in the North Atlantic, and the arch sound sequences, or two animals in very close proximity alternated in their production of A-B phrase sequences and arch sound sequences. Although several fin whale sequences were evident on these same recordings, all were faint and highly distorted with multi-path. Based on these analysis and considerations, the arch sound is believed to have come from a blue whale.

Arch sound frequency contours started with an upsweep, reached a peak frequency, and then descended in frequency. Some arch sounds remained nearly constant in frequency for several seconds at the highest point in the arch before de-

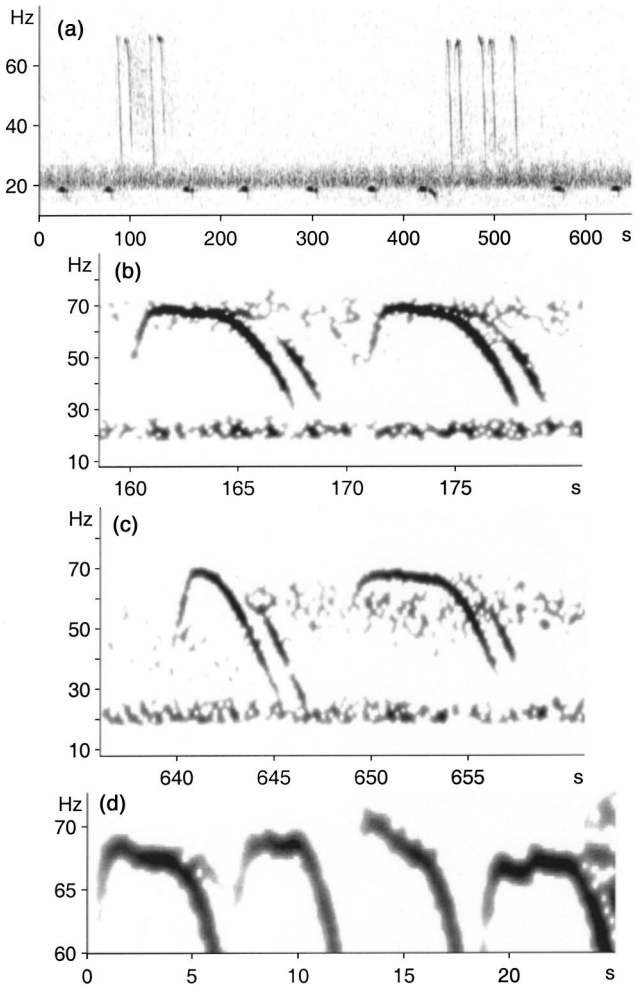


FIG. 7. Arch sound examples. (a) Two groups of arch sounds within a sequence of blue whale phrases. Grouping of arches, and pairing within groups, are evident. (b) Expanded view of two typical arches, each with a multipath echo. (c) Two forms of an arch, one that descends soon after reaching the peak frequency (left) and one that remains at peak frequency for several seconds before descending. (d) Close-up of four arch sounds to show the variation in frequency contour. All recordings are from the eastern North Atlantic, 18–19 December 1994. Spectrogram parameters for (a)–(c): frame size 1.28 s, 75% overlap 75%, FFT size 2.56 s, Hamming window; for (d), same parameters except 87.5% overlap and FFT size 5.12 s.

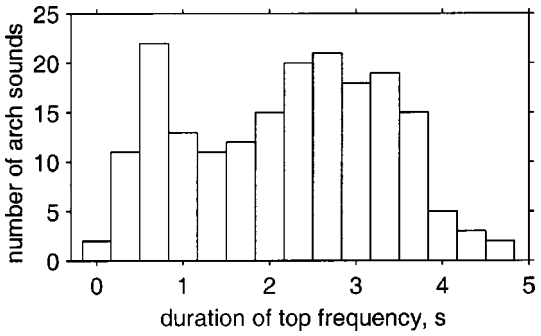


FIG. 8. Histogram showing the variation in duration of peak frequency in arch sounds ($n=189$).

scending, while others descended immediately (Fig. 8). Those that did remain nearly constant near the peak frequency exhibited variable frequency modulation during the peak frequency of the sound [Fig. 7(d)]. Arch sounds usually occurred in groups of 4–8, although there was obvious variability in the number and form of these sounds within and between groups. Within these groups, the arches often occurred in pairs [Fig. 7(a)], with one arch following another within 2–3 s and the next arch or arch pair occurring 9–18 s later. Pairing was the most common subgrouping, but sometimes larger subgroupings were seen, with 3–5 arches occurring in rapid succession.

The 189 measurable arch sounds started at a frequency of 56 ± 7.3 Hz (mean$\pm$s.d.), ascended to a peak frequency of 69 ± 1.1 Hz over 1.0 ± 0.3 s, stayed close to that peak for 2.2 ± 1.2 s (range 0.0–4.6 s) while changing frequency only slightly to 66.7 ± 1.2 Hz, and then descended to 35 ± 3.4 Hz over 3.1 ± 0.3 s. Overall arch duration was 6.3 ± 1.2 s. Peak intensity of the arches averaged 2–3 dB lower than the peak intensity of adjacent A-B phrases.

B. Timing patterns

Phrase periods (PPs) in all 90 sessions were measurable, even though individual phrase structures (A-only, A-B, or B-only) sometimes could not be determined. Figure 9 illustrates an example of a typical sound timing pattern. A-B phrases of this session occurred an average of every 75.5 ± 4.3 s ($n=72$), except for occasional rests.

To examine the data for evidence of surfacing behavior, each PP was categorized as being either a regular PP or a rest. Figure 10 shows a histogram of all PPs. From this distribution, 100 s was used as the criterion for distinguishing a regular PP (<100 s) from a rest ($\geqslant100$ s). Tables II and IV show the characteristics of regular PPs and rests for all sessions. The distribution of rest times is shown in the inset of

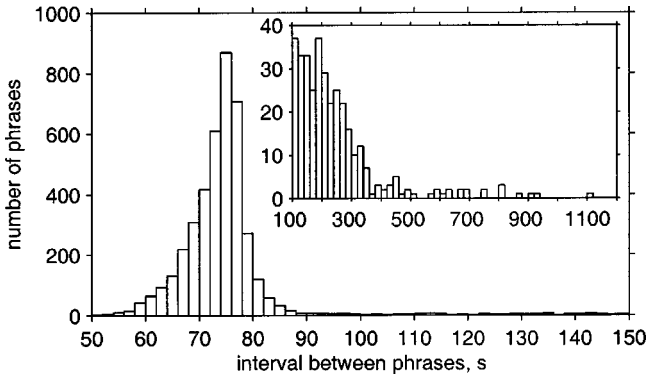


FIG. 10. Frequency distribution of phrase period (PP) values ($n=4371$) for all measured sessions ($n=90$). Inset: Frequency distribution for PP values between 100 and 1200 s ($n=346$). Absolute numbers are larger in the inset than in the main histogram because the histogram bin width is 20 s instead of 2 s.

Fig. 10. The PPs for rests of less than about 500 s appear to fit the upper half of a normal distribution quite well, while rests longer than 500 s do not, and so may represent a different behavior. In light of this fact, Table IV also shows statistics for only those rests less than 500 s.

A total of 63 sessions were clear enough to measure sequence duration. Statistics of all sequence durations are listed in Table II and their distribution shown in Fig. 11. Sequences lasted 0.1–30.3 min, with the shortest sequences consisting of a single phrase.

A total of 57 sessions were recorded with high enough quality to allow measurement of the duty cycle–i.e., of both sequence duration and duration of the succeeding rest period. The distribution of duty cycles is shown in Fig. 12. The mean of this distribution is $19.0\%\pm6.3\%$ (mean$\pm$s.d.), with a range of 2.9%–28.7%.

IV. DISCUSSION

These analyses provide the first quantitative description of basic characteristics of sounds from North Atlantic blue whales. Although these measurements for A-B phrases are based only on recordings from the western North Atlantic, spectrographic examples from the eastern North Atlantic SOSUS arrays throughout this same period revealed similarly long, patterned sequences composed from similar A-B phrases.

Along with this descriptive documentation, these analyses provide data possibly useful for distinguishing between different blue whale populations, deducing mechanisms of vocal production, or estimating surface and dive times. Recent efforts to identify the sex of vocally active blue whales

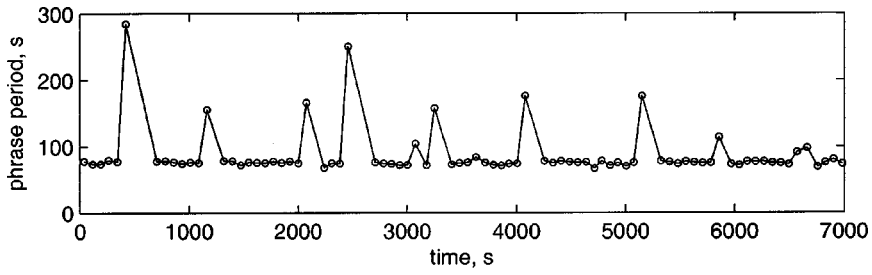


FIG. 9. Phrase period (PP) values for a typical sequence of A-B phrases.

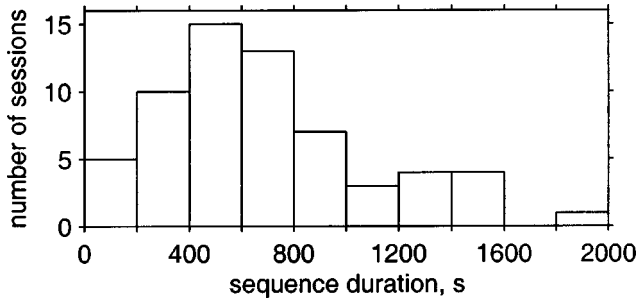


FIG. 11. Frequency distribution of the mean sequence duration for all sessions with at least one measurable sequence ($n=62$).

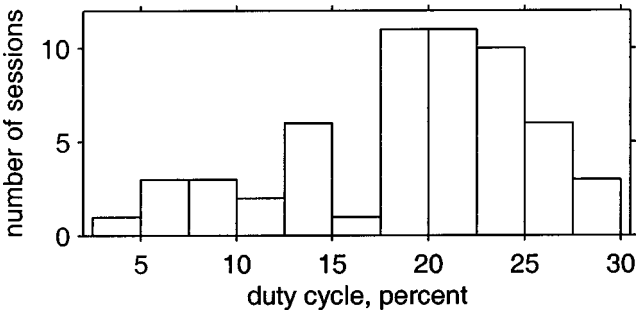


FIG. 12. Frequency distribution of duty cycle for all measurable sessions ($n=57$).

do not contradict the hypothesis that males produce these long patterned sequences (McDonald *et al.*, 2001), while recent work on fin whales (Croll *et al.*, 2002) empirically validates Watkins's hypothesis that fin whale 20 Hz sequences are produced by males (Watkins, 1981; Watkins *et al.*, 1987).

A. Interregion comparisons

There are some obvious similarities between the patterned sounds of blue whales in different regions of the world (Table V). Blue whales recorded in the North Atlantic (these data), northeastern Pacific (McDonald *et al.*, 1995; Rivers, 1997), northwestern Pacific (Stafford *et al.*, 2001), off Chile (Cummings and Thompson, 1971a; Shimada, 1999), on the Madagascar Plateau (Ljungblad *et al.*, 1998), off western Australia (McCauley *et al.*, 2001), and at latitudes greater than 60 °S off Antarctica (Ljungblad *et al.*, 1998) all emit long, multi-part, 15–40 Hz phrases (see also Clark and Fristrup, 1997, Fig. 6). All produce tonal, constant-frequency or FM sounds, and many also produce amplitude-modulated tones, emitted in long, stereotyped sequences. Phrase durations vary considerably, from 18 to 60 s, but are long compared to those for other large baleen whales. Phrase periods also vary considerably, from *ca.* 70 to 120 s, but are invariably longer than the phrase periods of other large baleen whales (Clark and Ellison, 2003).

There are obvious differences between sounds from the North Atlantic (these data) and blue whale sounds recorded elsewhere. As shown in Table V, the only other region with two-part phrases is the northeast Pacific. North Atlantic A-B phrases are easily distinguished from northeast Pacific A-B phrases. North Atlantic A-B phrase duration (from start to end) is about one-third as long and both parts A and part B are about half as long as those for the northeast Pacific, and part A in the North Atlantic is not amplitude modulated.

North Atlantic (these data) and northeastern Pacific (Clark and Fristrup, 1997; Stafford *et al.*, 1999a, b, 2001; McDonald *et al.*, 2001) blue whales also emit a low-intensity, very-low-frequency sound following a tonal, FM downsweep (the 9 Hz sound in the North Atlantic, and a sound at 11 Hz in the northeastern Pacific). A similar sound may exist in other regions, but detecting it probably requires high signal-to-noise recordings and careful scrutiny of those recordings.

There are also some obvious differences between the reported sound sequences of blue whales in different regions of the world. Most obviously, the frequency contours of phrases from the different regions are different; a glance at the spectrograms in the references cited above (see Table V) is enough to allow one to distinguish blue whale sounds from the various regions. A similarly striking difference is that phrases from some regions have two parts and others have three or four, with region-specific differences in phrase gaps. For northeastern Pacific and Madagascar blue whales, phrase parts can be separated by long silences of up to 25 s, while in other regions, phrase parts are separated by gaps of only a few seconds. The rates of frequency modulation for tonal phrase parts also varies. Antarctic and Madagascar blue whales produce FM sounds that descend rapidly (13 Hz in 2 s or less), while whales recorded in other regions produce FM sounds that change frequency much more slowly (e.g., 3 Hz in 11 s for the North Atlantic). Harmonic structure in blue whale sounds is known to be present in the northeastern Pacific (Thompson *et al.*, 1996) and off western Australia (McCauley *et al.*, 2001), and possibly occurs in other regions (e.g., Cummings and Thompson, 1971a; Ljungblad *et al.*, 1997). [Sidebands in spectrograms due to amplitude-modulated vocalizations and long spectrogram frame sizes are sometimes mistaken for harmonics; Watkins (1967) and Bradbury and Vehrencamp (1998) discuss these differences.] The fundamental frequency band varies as well. For example, in the Southern Hemisphere (from Chile and the Antarctic), the fundamental frequency is 18–28 Hz, compared to 17–20 Hz in the North Pacific or 15.5–18.5 Hz in the North Atlantic. Variability in simple acoustic characteristics provides a potential mechanism for distinguishing between sounds from these different regions and suggest that acoustical differences between different regions could reflect seasonal distributions, and potentially the movements of animals from these different populations (Stafford *et al.*, 1999a, 2001; Clark *et al.*, 2002). Further studies spanning large geographic areas are needed to determine the extent to which acoustic differences are associated with management units such as stocks. Similarly, genetic studies are needed to determine the extent to which acoustic differences correspond to different populations or subpopulations.

B. Sound types

Prior to this report, the 9 Hz sound had not been described for North Atlantic blue whales. When the 9 Hz sound was detected within a sequence, it was often present after every A-B phrase. Also, it often occurred when A-B phrases

TABLE V. Approximate mean characteristics of published blue whale sounds produced from various regions of the world. CF is a constant-frequency or nearly constant-frequency tone; FM is a frequency-modulated tone; and AM is an amplitude-modulated (pulsive) sound at a near-constant frequency.

Location	Phrase description	Fundamental frequency (Hz)	Duration (s)	Harmonics?	Phrase duration (s)	Phrase period (s)
North Atlantic (these data)	Part A: CF	18.5	8	rarely	18	73
	Part B: FM downsweep	18.5 to 15.7	11	rarely		
Northeastern Pacific (McDonald *et al.*, 1995)	Part A: AM	17.5	19	yes	62	122
	Part B: FM downsweep	19 to 17	19	yes		
Northwestern and north-central Pacific (Stafford *et al.*, 2001)	Part A: CF	20.2	6	no	23	81
	Part B: FM downsweep	20.2 to 18.4	2.5	no		
	Part C: FM up-and-downsweep	18.2 to 19.3 to 18.4	10	no		
Chile (Cummings and Thompson, 1971a)	Part A: AM	23	14	no	37	102
	Part B: AM	23	10	no		
	Part C: AM	23	8.5	no		
Madagascar pygmy blue (Ljungblad *et al.*, 1998)	Unit A_1: AM	38	4	no	60	95
	Unit A_2: CF	38	11	no		
	Unit B_1: FM downsweep	38 to 25	1.5	no		
	Unit B_2: CF	25	17.5	no		
Antarctic, 17 W longitude (Ljungblad *et al.*, 1998)	Part A: CF	28	10	no	30	75
	Part B: FM downsweep	28 to 20	2	no		
	Part C: FM downsweep	20 to 18	10	no		
Northern Indian (Alling *et al.*, 1991; Alling, 2003)	Unit 1: AM	27	12	no	123	199
	Unit 2: AM and FM	50 to 70	11	no		
	Unit 3: CF	122	29	no		
	Unit 4: AM	35	9	no		
Eastern Indian (McCauley *et al.*, 2001)	Component Ia: AM	19	21	yes?	102	180
	Component Ib: CF	21	22	yes		
	Component II: FM upsweep	20 to 26	23	yes		
	Component III: CF	19	22	yes		

were alternating with A-only phrases. In some cases it occurred irregularly after some, but not all, A-B phrases in a series, and sometimes after two consecutive A-B phrases. The organized occurrences of the 9 Hz sound in association with typical phrases indicates that our detection of the 9 Hz sound was independent of received level or ambient noise conditions. This further implies that production of the 9 Hz sound is not an obligate, involuntary component of sequence sound production, and is therefore controlled by the whale on a phrase-by-phrase basis. Interestingly, this sound always appeared at a time/frequency point that was an extension of the descending frequency sweep of part B, as if the whale's frequency change had continued and it had merely gone silent until reaching 9 Hz. We hypothesize that there is a functional relationship between the occurrence of the 9 Hz sound, and similarly associated very-low-frequency sounds, and the sound production mechanism.

A similar type of arch sound has been recorded from blue whales off southern California (Bass and Clark, 2002), although until now, it had not been reported for blue whales in the North Atlantic. Ljungblad *et al.* (1997) attributed shorter duration (1–2 s) arch sounds in the 190–280 Hz frequency band to blue whales off southern Australia. For arch sounds reported here, the downswept ending resembles a sound from the northeastern North Pacific described as a ''short pulse'' by Thompson *et al.* (1996) and subsequently labeled the ''Type D'' call (McDonald *et al.*, 2001), and recorded in the presence of socially active blue whales off southern California (Bass and Clark, 2002). It also resembles the ''Type IV'' sound from the northwestern North Pacific described by Stafford *et al.* (2001). The sounds described by Thompson *et al.* (1996) occurred sporadically in clusters, as did the arch sounds described here. These sounds are somewhat similar to sounds recorded from fin whales (Watkins,

1981; Edds-Walton, 1997). For both species this sound type occurs in the 30–80 Hz frequency band and descends in frequency. However, the blue whale FM sounds in this band are longer than those from fin whales and rise in frequency before descending in frequency.

C. Vocalization timing

The data presented here on sequences of consecutive sounds, and the periods of silence between them, possibly give information about dive and surface respiration behavior. In this regard, the data must be interpreted carefully: some observers (e.g., Cummings and Thompson, 1971a) report that silences between blue whale sound sequences corresponded with respiratory cycles, while others (e.g., Edds, 1982) report possible sound production during a respiratory sequence. Humpback singers are known to come to the surface to breathe while continuing to sing and do not always breathe during the silent periods between songs (Cerchio *et al.*, 2003). More data on blue whale surfacing and vocalization behavior are needed to clarify this issue.

The distribution of rests less than about 500 s appears to fit the upper half of a normal distribution quite well, while the distribution of all rests (including those >500 s) does not: the distribution becomes right-skewed because the 21 rests longer than 500 s are far more than the number expected from a normal distribution, which is less than two. This leads to the hypothesis that the behavior associated with rests less than 500 s may constitute a different behavior than for rests greater than 500 s. Perhaps the shorter rests happen during normal blow cycles and the longer rests occur because of some other type of resting behavior, or some type of interruption. Testing this hypothesis will in all likelihood require a combination of acoustic and nonacoustic methods such as visual observation or tagging.

The timing information for singing behavior as presented here could be useful for estimating relative abundance in an acoustic or visual-acoustic survey. An animal's vocalization pattern has important implications for at least two types of acoustic surveys. In one type (e.g., van Parijs *et al.*, 2002), the number of individual vocalizing animals in a given region is counted over some time period. Statistics for duty cycle can be used to estimate how many animals are available to be detected per unit time, and phrase duration and phrase period can be used for estimating the probability that one individual's vocalizations will overlap those from another. Another type of acoustic survey involves measuring the total acoustic energy in the species' frequency band at one or more sensors, then applying inversion methods to estimate the number of individuals vocalizing. For this method, the estimated number of individuals contributing to the total energy depends on a suite of sound production measures, including phrase period and duty cycle. The influence of other important factors such as season, location, time of day, and source level could not be determined from this data set. Our samples were biased toward times and sensors with blue whale sounds, and source level estimates are not available for reasons of Navy security.

It is hoped that the information presented here will be useful in future blue whale research. This includes investigations to estimate abundance, study communication behavior, determine seasonal distributions and migration routes, and arrive at a more complete understanding of this poorly known, cosmopolitan species.

ACKNOWLEDGMENTS

Special thanks to Commodore John Parrish, Captain Kirk Evans, Dennis Conlon, Hank Fleming, LCDR George Gagnon, Clyde Nishimura, CDR Lysa Olsen and the many personnel of the various Naval Ocean Processing Facilities throughout the North Atlantic who contributed to the data collection. Without their vision, leadership and dedicated efforts none of these data or insights would have become available. Thanks also to Julie A. Rivers for assistance measuring data the first time. This work was funded by the Office of Naval Research Grants Nos. N00014-00-F-0395 and N00014-03-1-0099 to DKM, and ONR Grant No. N00014-93-1-0431 and Naval Research Laboratory Contract No. N00014-94-C-6016 to CWC.

Alling, A. (**2003**). Personal communication.

Alling, A., Dorsey, E., and Gordon, J. (**1991**). "Blue whales (*Balaenoptera musculus*) off the northeast coast of Sri Lanka: Distribution, feeding and individual identification," in *Cetaceans and Cetacean Research in the Indian Ocean Sanctuary*, edited by S. Leatherwood and G. Donovan (United Nations Environment Programme, Nairobi), pp. 247–258.

Bass, A. H., and Clark, C. W. (**2002**). "The physical acoustics of underwater sound communication," in *Acoustic Communication*, edited by A. M. Simmons, R. R., Fay, and A. N. Popper (Springer, New York), pp. 15–64.

Bradbury, J. W., and Vehrencamp, S. L. (**1998**). *Principles of Animal Communication* (Sinauer, Sunderland, MA).

Cerchio, S., Frankel, A., and Jacobsen, J. (**2003**). Personal communication.

Charif, R. A., Clapham, P. J., and Clark, C. W. (**2001**). "Acoustic detections of singing humpback whales in deep waters off the British Isles," Marine Mammal Sci. **17**, 751–768.

Charif, R. A., Mitchell, S. G., and Clark, C. W. (**1995**). Canary 1.2 User's Manual. Tech. Rept., Cornell Lab. Ornithol., Ithaca, NY.

Clark, C. W. (**1994**). "Application of U.S. Navy underwater hydrophone arrays for scientific research on whales," Rep. Intl. Whal. Commn. **44**, 1–12.

Clark, C. W., and Ellison, W. T. (**2003**). "Potential use of low-frequency sounds by baleen whales for probing the environment: Evidence from models and empirical measurements," in *Advances in the Study of Echolocation in Bats and Dolphins*, edited by J. A. Thomas and R. A. Kastelein (Plenum, New York).

Clark, C. W., and Fristrup, K. M. (**1997**). "Whales '95: A combined visual and acoustic survey of blue and fin whales off Southern California," Rep. Intl. Whal. Commn. **47**, 583–600.

Clark, C. W., and Gagnon, G. C. (**to appear**). "Low-frequency vocal behaviors of baleen whales in the North Atlantic: Insights from IUSS detections, locations, and tracking from 1992 to 1996," to appear, J. Underwater Acoust.

Clark, C. W., and Mellinger, D. K. (**1994**). "Application of Navy IUSS for whale research," J. Acoust. Soc. Am. **96**, 3315(A).

Clark, C. W., Borsani, F., and Notarbartolo-di-Sciara, G. (**2002**). "Vocal activity of fin whales, *Balaenoptera physalus*, in the Ligurian Sea," Marine Mammal Sci. **18**, 286–295.

Costa, D. P. (**1993**). "The secret life of marine mammals: novel tools for studying their behavior and biology at sea," Oceanography **6**(3), 120–128.

Croll, D. A., Clark, C. W., Acevedo, A., Tershy, B., Flores, S., Gedamke, J., and Urban, J. (**2002**). "Only male fin whales sing loud songs," Nature (London) **417**, 809.

Cummings, W. C., and Thompson, P. O. (**1971a**). "Underwater sounds from the blue whale, *Balaenoptera musculus*," J. Acoust. Soc. Am. **50**, 1193–1198.

Cummings, W. C., and Thompson, P. O. (**1971b**). "Bioacoustics of marine mammals; R/V HERO cruise 70-3," Antarctic J. United States **6**(5), 158–160.

Cummings, W. C., and Thompson, P. O. (**1977**). ''Long 20-Hz sounds from blue whales in the Northeast Pacific,'' Proc. Second Conf. Biol. Mar. Mamm., p. 73.
Edds, P. L. (**1982**). ''Vocalizations of the blue whale, *Balaenoptera musculus*, in the St. Lawrence River,'' J. Mammal. **63**, 345–347.
Edds-Walton, P. L. (**1997**). ''Acoustic communication signals of mysticete whales,'' Bioacoustics **8**, 47–60.
Evans, W. E. (**1967**). ''Vocalizations among marine mammals,'' in *Marine Bio-acoustics*, edited by W. N. Tavolga (Pergamon, New York), Vol. 2, pp. 159–186.
Gagnon, G. J., and Clark, C. W. (**1993**). ''The use of U.S. Navy IUSS passive sonar to monitor the movements of blue whales,'' Proc. Tenth Biennial Conf. Biol. Mar. Mamm., p. 50.
Kibblewhite, A. C., Denham, R. N., and Barnes, D. J. (**1967**). ''Unusual low-frequency signals observed in New Zealand waters,'' J. Acoust. Soc. Am. **41**, 644–655.
Ljungblad, D. K., Clark, C. W., and Shimada, H. (**1998**). ''A comparison of sounds attributed to pygmy blue whales (*Balaenoptera musculus brevicauda*) recorded south of the Madagascar Plateau and those attributed to 'true' blue whales (*Balaenoptera musculus*) recorded off Antarctica,'' Rep. Intl. Whaling Commn. **48**, 439–442.
Ljungblad, D. K., Stafford, K. M., Shimada, H., and Matsuoka, K. (**1997**). ''Sounds attributed to blue whales recorded off the southwest coast of Australia in December 1995,'' Rep. Intl. Whal. Commn. **47**, 435–439.
McCauley, R. D., Jenner, C., Bannister, J. L., Burton, C. L. K., Cato, D. H., and Duncan, A. (**2001**). ''Blue whale calling in the Rottnest trench-2000,'' Western Australia Centre for Marine Science and Technology, Curtin University, No. R2001-6.
McDonald, M. A., Hildebrand, J. A., and Webb, S. C. (**1995**). ''Blue and fin whales observed on a seafloor array in the northeast Pacific,'' J. Acoust. Soc. Am. **98**, 712–721.
McDonald, M. A., Calambokidis, J., Teranishi, A. M., and Hildebrand, J. A. (**2001**). ''The acoustic calls of blue whales off California with gender data,'' J. Acoust. Soc. Am. **109**, 1728–1735.
Mellinger, D. K. (**1994**). Osprey 1.2 Guide. Tech. Rept., Cornell Lab. Ornithol., Ithaca, NY.
Mellinger, D. K., Carson, C. D., and Clark, C. W. (**2000**). ''Characteristics of minke whale (*Balaenoptera acutorostrata*) pulse trains recorded near Puerto Rico,'' Marine Mammal Sci. **16**, 739–756.
Nishimura, C. E., and Conlon, D. M. (**1994**). ''IUSS Dual Use: Monitoring whales and earthquakes using SOSUS,'' Marine Technol. Soc. J. **27**(4), 13–21.
Northrop, J., Cummings, W. C., and Morrison, M. F. (**1971**). ''Underwater 20-Hz signals recorded near Midway Island,'' J. Acoust. Soc. Am. **49**, 1909–1910.
Patterson, B., and Hamilton, G. R. (**1964**). ''Repetitive 20 cycle per second biological hydroacoustic signals at Bermuda,'' in *Marine Bio-acoustics*, edited by W. N. Tavolga (Pergamon, New York), pp. 125–145.
Payne, K. B., Tyack, P. L., and Payne, R. S. (**1983**). ''Progressive changes in the songs of humpback whales (*Megaptera novaeangliae*): a detailed analysis of two seasons in Hawaii,'' in *Communication and Behavior of Whales*, edited by R. S. Payne (Westview, Boulder), pp. 9–57.
Payne, R. S., and McVay, S. (**1971**). ''Songs of humpback whales,'' Science **173**, 587–597.
Payne, R. S., and Webb, D. (**1971**). ''Orientation by means of long-range acoustic signaling in baleen whales,'' Ann. N.Y. Acad. Sci. **188**, 110–142.
Rivers, J. A. (**1997**). ''Blue whale, *Balaenoptera musculus*, vocalizations from the waters off central California,'' Marine Mammal Sci. **13**, 186–195.
Shimada, H. (**1999**). Personal communication.
Stafford, K. M., Fox, C. G., and Clark, D. S. (**1998**). ''Long-range acoustic detection and localization of blue whale calls in the northeast Pacific Ocean,'' J. Acoust. Soc. Am. **104**, 3616–3625.
Stafford, K. M., Nieukirk, S. L., and Fox, C. G. (**1999a**). ''An acoustic link between blue whales in the eastern tropical Pacific and the northeast Pacific,'' Marine Mammal Sci. **15**, 1258–1268.
Stafford, K. M., Nieukirk, S. L., and Fox, C. G. (**1999b**). ''Low-frequency whale sounds recorded on hydrophones moored in the eastern tropical Pacific,'' J. Acoust. Soc. Am. **106**, 3687–3698.
Stafford, K. M., Nieukirk, S. L., and Fox, C. G. (**2001**). ''Geographic and seasonal variation of blue whale calls in the North Pacific,'' J. Cetacean Res. Mgmt. **3**, 65–76.
Thode, A. M., D'Spain, G. L., and Kuperman, W. A. (**2000**). ''Matched-field processing, geoacoustic inversion, and source signature recovery of blue whale vocalizations,'' J. Acoust. Soc. Am. **107**, 1286–1300.
Thompson, P. O., and Friedl, W. A. (**1982**). ''A long-term study of low frequency sounds from several species of whales off Oahu, Hawaii,'' Cetology **45**, 1–19.
Thompson, P. O., Findley, L. T., Vidal, O., and Cummings, W. C. (**1996**). ''Underwater sounds of blue whales, *Balaenoptera musculus*, in the Gulf of California, Mexico,'' Marine Mammal Sci. **12**, 288–293.
Tyack, P. L., and Clark, C. W. (**2000**). ''Communication and acoustical behavior in dolphins and whales,'' in *Hearing by Whales and Dolphins*, edited by W. W. Au, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 156–224.
Urick, R. J. (**1983**). *Principles of Underwater Sound*, 3rd ed. (McGraw–Hill, New York).
Van Parijs, S., Smith, J., and Corkeron, P. (**2002**). ''Using calls to estimate the abundance of inshore dolphins: a case study with Pacific humpback dolphins, *Sousa chinensis*,'' J. Appl. Ecol. **39**, 853–864.
Watkins, W. A. (**1967**). ''The harmonic interval: fact or artifact in spectral analysis of pulse trains,'' in *Marine Bio-Acoustics*, edited by W. N. Tavolga (Pergamon, New York), pp. 15–43.
Watkins, W. A. (**1981**). ''Activities and underwater sounds of fin whales,'' Technical Report 33, Whaling Res. Inst., Tokyo.
Watkins, W. A., Tyack, P., Moore, K. E., and Bird, J. E. (**1987**). ''The 20-Hz signals of finback whales (*Balaenoptera physalus*),'' J. Acoust. Soc. Am. **82**, 1901–1912.
Weston, D. E., and Black, R. I. (**1965**). ''Some unusual low-frequency biological noises underwater,'' Deep-Sea Res. Oceanogr. Abstr. **12**, 295–298.

The sonar beam pattern of a flying bat as it tracks tethered insects

Kaushik Ghose and Cynthia F. Moss
Neuroscience and Cognitive Science Program, Department of Psychology, The University of Maryland, College Park, Maryland 20742

(Received 31 July 2002; revised 25 April 2003; accepted 12 May 2003)

This paper describes measurements of the sonar beam pattern of flying echolocating bats, *Eptesicus fuscus*, performing various insect capture tasks in a large laboratory flight room. The beam pattern is deduced using the signal intensity across a linear array of microphones. The positions of the bat and insect prey are obtained by stereoscopic reconstruction from two camera views. Results are reported in the form of beam-pattern plots and estimated direction of the beam axis. The bat centers its beam axis on the selected target with a standard deviation (σ) of 3°. The experimental error is $\pm 1.4°$. Trials conducted with two targets show that the bat consistently tracks one of the targets with its beam. These findings suggest that the axis of the bat sonar beam is a good index of selective tracking of targets, and in this respect is analogous to gaze in predominantly visual animals. [DOI: 10.1121/1.1589754]

PACS numbers: 43.80.Ka, 43.66.Gf, 43.66.Qp [WA]

I. INTRODUCTION

Echolocating bats can orient, forage, and perform other perceptually guided tasks in complete darkness by emitting ultrasonic vocal signals and analyzing the echoes returning from objects in their environment.[1] In this respect, bats provide an opportunity to study the use of audition in spatial tasks, which may be accomplished in other animals by using vision.

We studied *Eptesicus fuscus*, a bat species that echolocates with frequency modulated (FM) signals. Each sonar signal consists of several harmonically related frequency sweeps. The *E. fuscus* echolocation call time–frequency structure changes as the bat searches for, approaches, and captures insect prey (Fig. 1). This species forages mainly in open spaces, but has been reported to pursue prey near vegetation.[2]

The timing, duration, and spectral characteristics of each sonar pulse influence the echo information available to the bat's acoustic imaging system. While searching for prey, *E. fuscus* uses long (15–20 ms) pulses with a shallow frequency sweep. The fundamental frequency sweeps from approximately 28 to 22 kHz. The rate of production may be as low as 5–10 Hz. Upon detecting a prey item, the bat approaches it, shortening the pulses to 2–5 ms and increasing bandwidth (fundamental sweeping from 60 to 22 kHz). During the terminal phase the pulses may be as short as 0.5–1 ms, with the fundamental sweeping from about 40 to 12 kHz and produced at rates of up to 150–200 Hz in the terminal (or feeding) buzz[3,4] (see Fig. 1). Vocalizations cease when the bat is about 10–15 cm from the prey (which is approximately 30–50 ms prior to contact with the prey). The sequence is completed with a capture attempt using the tail membrane (arranged like a scoop), the wing tips (to push the prey towards the mouth), or in rare instances, directly with the mouth. The longer duration search signals have only been recorded from bats foraging in wide-open spaces and not in the lab.[4]

The spatial characteristics of the sonar beam also influence the echoes received by the bat. Hartley and Suthers[5] measured the beam pattern of a stationary, anesthetized *E. fuscus* resting on a platform and stimulated to vocalize by applying electrical pulses to a vocal-motor area of the brain. The results of this study showed that the sonar beam of *E. fuscus* is broad, but not omnidirectional. The sonar beam has a main lobe directed along the midline and slightly downwards, its vertical position rising slightly at higher frequencies. The main lobe intensity drops by 3 dB at 35° off midline. There is a ventral lobe below the main lobe, and weaker by about 6 dB compared to peak intensity.

The directionality suggests that objects closer to the beam axis (the direction of the peak of the main lobe of the beam) will return stronger echoes than objects located more laterally. We propose that the bat maximizes the signal-to-noise ratio of returning echoes by directing its vocalization beam at the location of a prey item. Therefore, we hypothesize that the bat's aim of its sonar beam in the direction of a target is a natural motor action associated with target selection and tracking. We test this hypothesis by recording the sonar beam patterns produced by bats catching tethered insects in a flight room. We use these data to calculate the direction of the beam axis with respect to the target.

II. METHODS

A. Behavioral tasks

Four echolocating bats of the species *E. fuscus* were used for the study. The bats were allowed to fly in a large room (7×6 m) whose walls were covered with sound-absorbent foam (Sonex-1) to dampen reverberation and enable recordings of bat vocalizations. The bats were trained to take a mealworm (target) from a tether while in flight. The target could be moved in a circular path by a motor-operated boom positioned just below the ceiling. It could also be dropped into the flight space by a trap-door mechanism

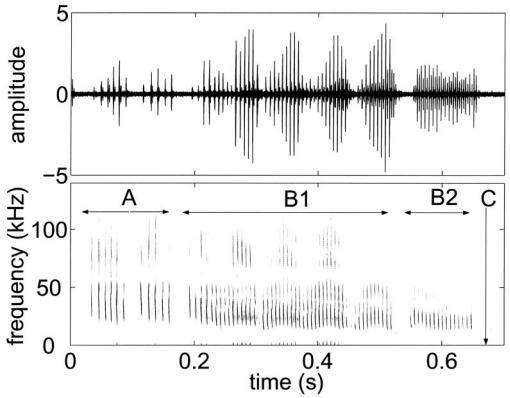


FIG. 1. The top panel shows the time waveform of a series of *E. fuscus* vocalizations recorded in the laboratory. The bottom panel shows the spectrogram of this signal. Different stages of foraging are marked out. **A** is the approach phase, **B1** is buzzl, **B2** is buzz2, while **C** refers to the time of contact of the bat with the prey.

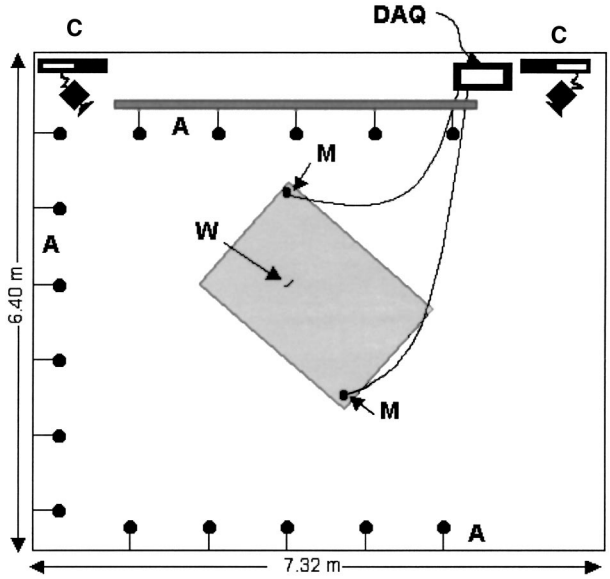


FIG. 2. Plan view of flight room and array layout. **A**: Microphone array; **M**: Ultrasound advice microphones; **C**: High-speed digital video cameras running at 240 frames per second; **DAQ**: Data acquisition systems; IoTech WaveBook, 2 channels at 250 kHz each and National Instruments AT-MIO-16-E-1 board, 16 channels at 20 kHz each; **W**: Tethered worm. Shaded area represents the calibrated space (within which the path of the bat may be accurately reconstructed from the camera views).

mounted just under the ceiling. The trap door was padded to minimize noise as it opened. Microphones placed on the floor of the room did not pick up any sound when the trap door opened. We cannot, however, rule out there being some sound associated with the trap-door opening that the bat could hear.

One behavioral task consisted of releasing the tethered target from the trap door at a random point in time as the bat flew by. In this manner the bat was presented with a target whose location (over an area of approximately 2 m^2) was unknown until the trap door was opened. Analysis of the beam direction before and after the target presentation enabled us to study one aspect of the orienting behavior of the bat as it detects and then attacks prey. The four bats had previously been trained to take targets from a tether and had been used the previous year for studying their vocalization behavior as they caught tethered insects in the laboratory. There was no training time required for the bats during the current set of experiments besides 1 week of ''warm-up'' flying at the start of the season, after which the bats, vocal behavior was recorded as they caught tethered insects. Data was collected in the form of insect capture trials set up by the experimenter; each trial consisted of a segment of data that contained one and sometimes more attempts by the bat to capture the target.

B. Array recordings

The array consisted of 16 Knowles FG3329 microphones arranged in a planar U-shape along three walls of the flight room (see Fig. 2). The linear spacing between the microphones was 1 m, and the height of the microphones was 0.9 m above the floor. Each microphone was extended from the wall mounting by a thin (3-mm-diameter) steel rod 0.3 m long. This served to reduce the overlap between the original sound and any residual echoes from the sound-proofing panels or mounting base. In order to compute the beam pattern for a given frequency band, the information required is the intensity of the signal in that band. This information can be obtained from both the bandpass signal as well as the envelope of that signal, provided the signal is narrow band, or it can be broken up into segments that are narrow band, as shown here.

Let $f(t)$ be the measured signal, let $f_a(t)$ be the analytical signal for $f(t)$, and let $\hat{f}(t)$ be the Hilbert transform of $f(t)$, such that

$$f_a(t) = f(t) + j \cdot \hat{f}(t). \tag{1}$$

We know that the envelope of $f(t)$ is

$$|f_a(t)| = \sqrt{f(t)^2 + \hat{f}(t)^2}. \tag{2}$$

Therefore, the integral of the square of the envelope from time t_1 to t_2 reduces to

$$\int_{t_1}^{t_2} |f_a(t)|^2 dt = \int_{t_1}^{t_2} f(t)^2 dt + \int_{t_1}^{t_2} \hat{f}(t)^2 dt. \tag{3}$$

We recognize the first term to be the energy of the signal over the time t_1 to t_2. If we assume that the signal over this time period has primarily one frequency component, then $\hat{f}(t)$ is merely a phase-shifted version of $f(t)$. If we further assume that the time interval $t_1 - t_2$ is much larger than the period of the signal $f(t)$, then $\int_{t_1}^{t_2} \hat{f}(t)^2 dt \simeq \int_{t_1}^{t_2} f(t)^2 dt$, which gives us

$$\int_{t_1}^{t_2} |f_a(t)|^2 dt \simeq 2 \int_{t_1}^{t_2} f(t)^2 dt. \tag{4}$$

This result [Eq. (4)] shows that integrating the square of the envelope of a bandpassed version of a bat call will give us the signal intensity in that band. Simulations using recorded bat vocalizations confirm this result. As described above, the sonar vocalizations of *E. fuscus* are frequency sweeps com-

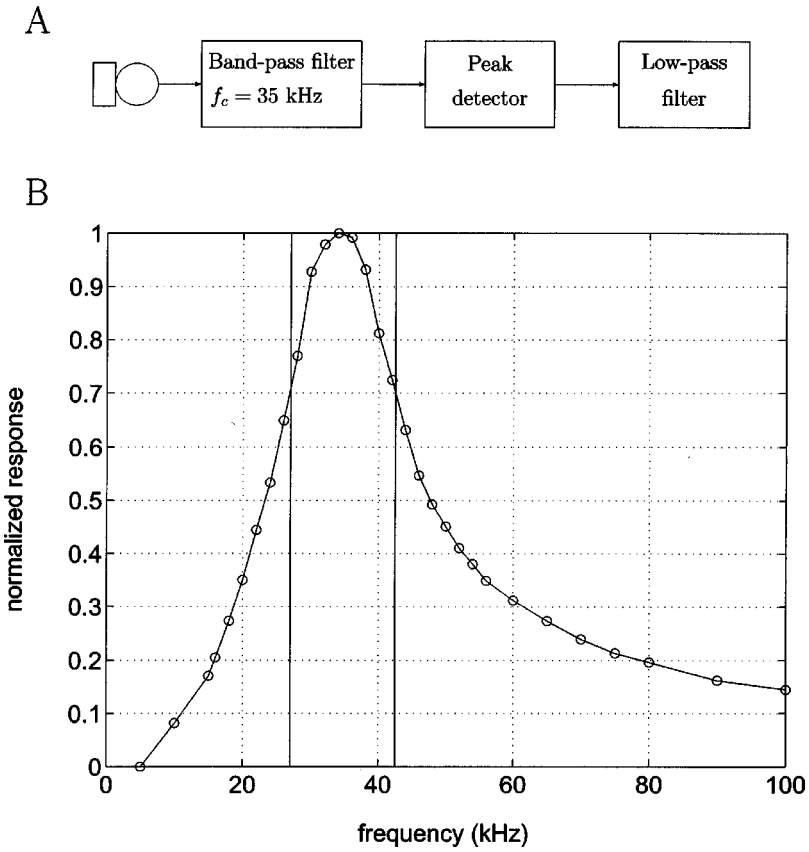


FIG. 3. (A) Schematic of signal-processing hardware. (B) Filter characteristics of the bandpass filter used. The *x* axis shows the frequency, while the *y* axis shows the normalized response. The vertical bars correspond to the 3-dB (half-power) points, i.e., the start and stop frequency. Examples of bandpass signal and envelope extraction may be seen in Figs. 4(A) and (B).

posed of a fundamental and several harmonics. By bandpass filtering this signal we can meet the required criteria.

The frequency content of the envelope for the echolocation signals is related to the duration of the signals. The shortest signal durations occur during the terminal buzz phase of insect capture and are on the order of 0.5–1 ms, which implies that the upper limit frequency content of the *envelope of the whole signal* is around 2 kHz. Assuming conservatively that the envelope of a bandpass of this signal has a duration of 0.25 ms; this places the frequency content of the envelope at around 4 kHz. Therefore, a sampling rate of 20 kHz captures the envelope with good fidelity. This reduces the data acquisition requirements for sonar signal recordings from an array of microphones by a factor of 12.5 (assuming a sampling rate of 250 kHz is sufficient to record the broadband signal).

The frequency content of the sonar signals of *E. fuscus* hunting insect prey in the lab varies widely, with higher frequency content during the early approach phase of insect pursuit and lower frequency content during the terminal buzz phase. By choosing a frequency band centered at 35 kHz we discovered that we could record signals during all foraging stages [the typical signal-to-noise ratio (SNR) using this method was estimated at 20 dB for the bat vocalizations].

In order to determine the beam pattern from a flying bat, the distance-dependent attenuation of the sonar signals must be corrected.[6] This correction (detailed in the Data Processing section) has two components. One is the spherical attenuation loss and depends only on the distance between the bat and a given microphone. The other is the absorption of energy as the sound is propagated through the air. This is dependent on both distance and frequency.

Keeping these factors in mind, we developed the scheme outlined in Fig. 3(A). The signal from each microphone was fed to an amplifying bandpass filter which extracts signal components centered around 35 kHz. All circuits were constructed with off-the-shelf components soldered onto custom-printed circuit boards. The frequency characteristics of the filter used are shown in Fig. 3(B).

This signal was then fed to a peak detector circuit which extracted the envelope of this bandpassed signal. The envelope was smoothed by a low-pass filter and then digitized. Examples of synthetic and bat sonar signals received at a microphone and their bandpass filtered, smoothened envelopes may be seen in Fig. 4.

Signal digitization was done by a National Instruments Data Acquisition Board (AT-MIO-16-E-1, 12 bit, 50-ns clock, 8-s rolling buffer) controlled by a PC running a C program.

C. Broadband microphone recordings

In addition to the array microphones, we used two Ultrasound Advice SM2 microphones and SP2 amplifiers [flat response up to ($\pm$2 dB) 40 kHz, 5-dB drop from 40 to 100 kHz]. The microphone signal was further amplified and filtered by active filters (Stanford Research Systems model SR 650 digital filter, bandpass set at 10–99 kHz). These microphones recorded full bandwidth vocalization waveforms. The

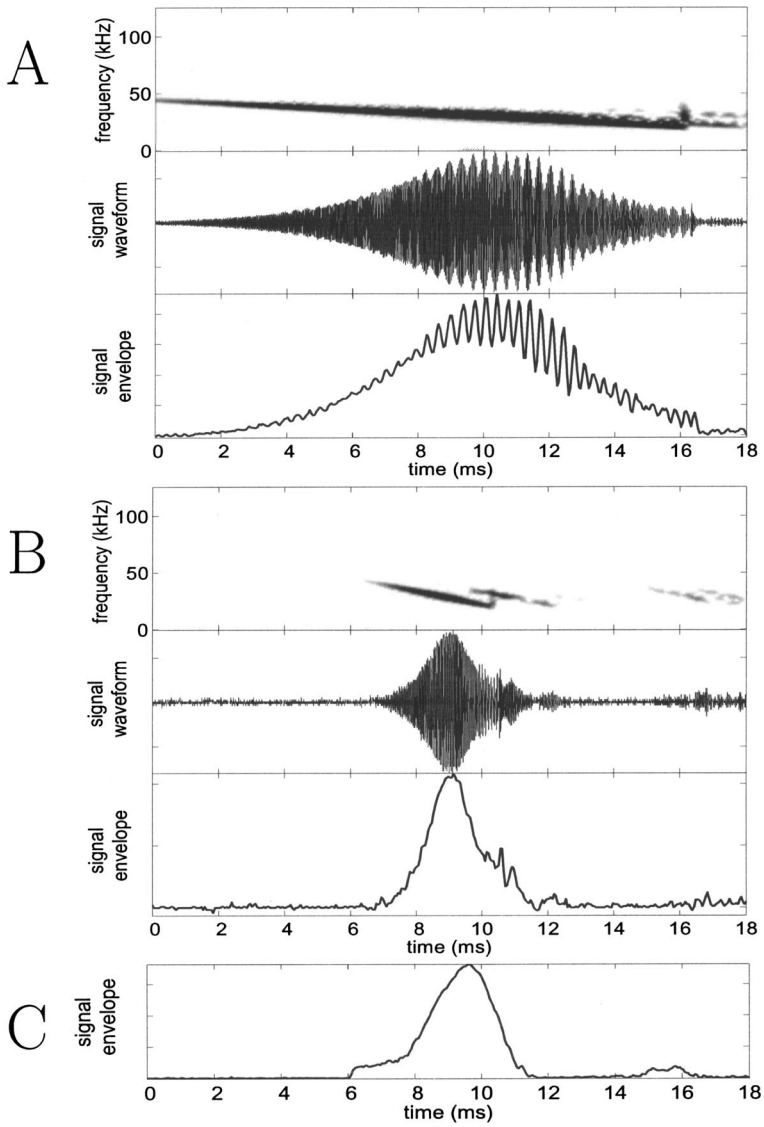


FIG. 4. Panels (A) and (B) show recordings taken by pinging the array with frequency sweeps from an emitter. (A) shows that for a long (shallow) sweep there is more overlap, between the incident sound and the returning echo, and the beats are more prominent. (B) shows that for short sweeps there is less overlap. The top panel in each is the spectrogram of the bandpassed signal received at one of the array microphones, the middle panel shows the time waveform of that signal, while the bottom panel shows the envelope extracted by the array hardware. The interaction between the incident sound and an overlapping echo shows up as a beat. In both (A) and (B), the emitter was placed in the plane of the array so as to maximize the echo returning to the microphone from the array backend. Due to limitations of the signal generator used to produce the emitted sounds, each frequency sweep has a brief glitch as it resets to the start frequency and this is visible as a vertical streak in the spectrogram. This does not change any results. (C) shows the envelope signal taken from an array circuit during a trial with a flying bat. In general, the bat sounds recorded at the array do not show apparent effects of overlapping echoes. A detailed explanation is given in the text.

signals were digitized using an IoTech Wavebook 512 at 250 kHz per channel (12 bit, 8.19-s rolling buffer) run by a Dell laptop computer.

D. Cameras and calibration

Two Kodak MotionCorder digital video cameras running at 240 Hz were used to record the flight paths of the bats and the locations of insect targets and microphones. The cameras were operated under long wavelength lighting (>650 nm, red filters, Reed Plastics, Rockville, MD), to ensure that the bats were not using vision in the insect capture task.[7] The digital frames stored on the camera buffers were downloaded onto analog tape. Relevant sections of the video record were then redigitized using a MiroVideo DC30 capture board. Motion analysis software from Peak Performance Technologies (Motus) was used to convert the images of the bat and other objects from the two camera recordings into three-dimensional coordinates. A calibration frame supplied by Peak Performance was used for this transformation. Since the array was outside the space covered by the calibration frame, manual measurements were made that enabled us to compute the array coordinates in the camera reference frame.

E. Triggering and synchronization

Data acquired by the three digitizing systems were continuously stored on rolling buffers. When the trial was judged to be complete (usually after a capture or capture attempt) the same end trigger was fed to all three systems to capture the last 8 s of data.

III. DATA ANALYSIS

A. Beam-pattern computation

The signals from each microphone were segmented to select out the vocalizations and exclude the echoes. The received intensity I_r was computed from the envelope. This intensity value was corrected for spherical loss and atmospheric attenuation to give I_c as shown in Eq. (5). Values for the attenuation coefficient were obtained from standard

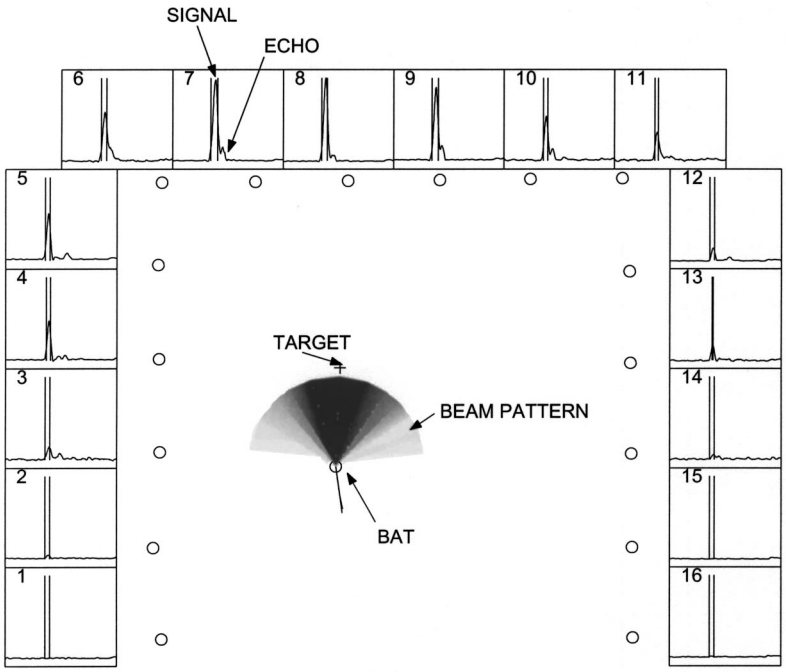


FIG. 5. Beam-pattern reconstruction. Central panel shows the reconstructed beam pattern. The 16 circles along the edges of the panel are the positions of the microphones in the array. The pattern is normalized such that the peak intensity has a value of 1.0 and is colored black. Lighter colors denote progressively lower intensities. The circle at the center of the beam pattern represents the position of the bat. The + symbol represents the position of the worm. The thin curved line terminating at the bat's position is the trajectory of the bat up to that frame. The straight line drawn from the bat represents the direction of motion of the bat (in this frame the two overlap). Surrounding panels (numbered 1 through 16) show the envelope signals digitized from each microphone. All the side panels have the same scales. Twenty ms of data are shown. The signal on each panel is time shifted to compensate for the time of travel of the sound from the bat to the corresponding microphone. As a result the direct signal from the bat (first sound) lines up on all the panels. A fairly loud echo (second smaller bump) does not as its source is at a different position. This makes it easier to segment the signals and discard the echoes. The segmentation for the vocalization shown is depicted as two vertical bars bracketing the relevant portion of the envelope trace.

tables (ISO 9613—1, acoustics, and cross checked against an ASA Acoustics Handbook[8]). The corrected intensity was calculated as

$$I_c = I_r r^2 \cdot 10^{1/10 \cdot r\alpha}, \tag{5}$$

where r is the distance between the microphone and the bat. Software for this calculation was written in MATLAB.

The overall beam pattern was then reconstructed, as shown in Fig. 5.

B. Beam-axis computation

According to our hypothesis, the bat aims its sonar beam at a target of interest. Assuming the beam to be symmetrical, adding up direction vectors from the bat to each microphone, weighted by the corrected intensity at that microphone, results in a vector whose direction is an objective estimate of the beam axis, regardless of the actual profile of the beam. This is given by Eq. (6)

$$\mathbf{H} = \sum_n \mathbf{I}_i, \tag{6}$$

where $\mathbf{I}_i$ is the vector drawn from the bat to microphone i with magnitude proportional to the corrected intensity. $\mathbf{H}$ is the resultant, whose direction is the estimate of the beam axis.

C. Errors due to array geometry

Figure 6 shows simulation results for beam-axis computations for six different head orientations. The simulated beam pattern is shown at the center of the array. This beam is then ''emitted'' at different positions in the space enclosed by the array, and the estimated beam directions are computed from the signals received by the array elements. The results are shown as black arrows. As can be seen from Figs. 6(A)–(E), only if the source of the signal is close to the array (around 1 m) do we see edge effects which warp the estimate. During experiments we only use the data collected within the calibrated space, which is more than 1 m from the array boundary. In addition, as expected, if the beam points out of the space enclosed by the array we get a biased estimate of beam direction. This is illustrated in Fig. 6(F). (If the array were constructed to be a ring, this error would not be present. More microphones were not added because of limitations in the data acquisition hardware and due to difficulties in placing an array segment on the fourth wall of the flight room.)

D. Calibrations with frequency sweeps

The array was tested using an emitter mounted on a tripod at the center of the array and oriented in different directions. The emitter produced frequency sweeps starting from 50 kHz and sweeping down to 20 kHz. The signals were recorded using the array, and the emitter itself was filmed using the video cameras. Two markers were attached to the emitter, and these were used to reconstruct the direction the emitter was pointing. The signals recorded at the array were analyzed in the same manner as real bat signals and the direction of the beam was computed as described previously. This was compared against the reference direction computed from two markers attached to the emitter.

Three calibrations were done from two positions in the calibrated space, and the results of the calibration are illustrated in Fig. 7. If the array computation did not need any correction, then the traces would be a horizontal line running along zero. The average of these traces between the two vertical dotted lines at $-50°$ and $+120°$ was used to create a calibration curve to map the measured beam axis to a corrected beam axis. The final beam-axis compututation results in an error of $\pm 1.4°$.

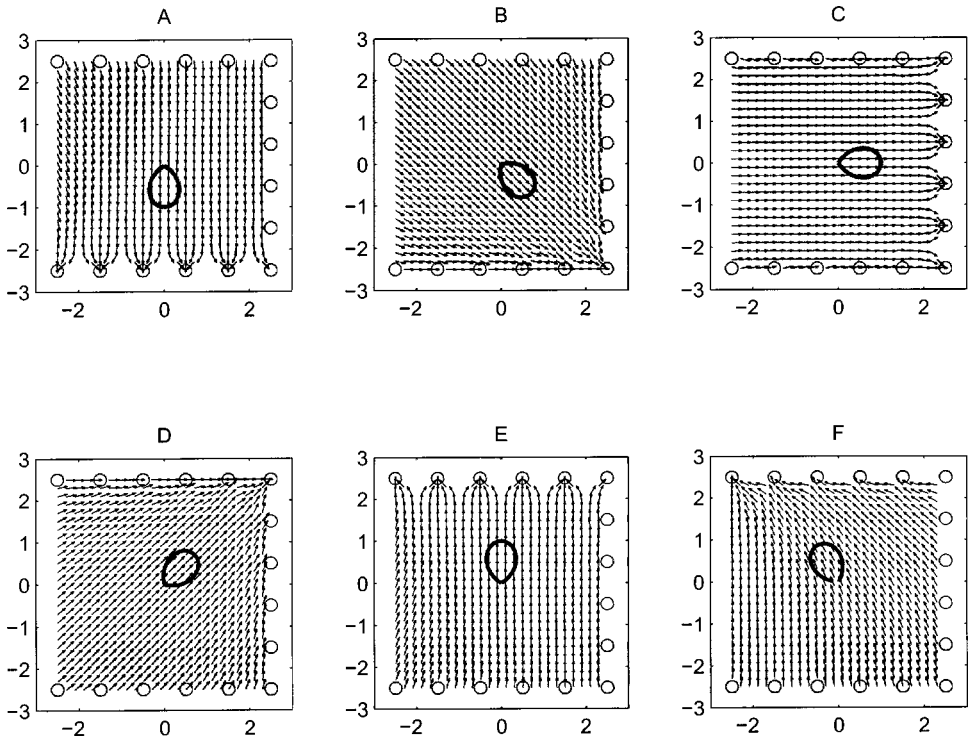


FIG. 6. Plots of estimated head aim at different points within the space enclosed by the array (small black arrows) with a polar plot of the beam intensity profile (bold pattern) overlaid at the center. The direction the beam is pointing in corresponds to the peak of the profile. Plots (A) to (E) demonstrate that errors in computing head aim grow large only when the source is close to the edge of the array (around 1 m). (F) demonstrates that if the beam is directed out of the space enclosed by the array estimates become biased even near the center of the array. The *x*- and *y* axes tick marks are in meters.

E. Effect of echoes on the estimate

Echoes that overlap with the original bat vocalization at the array microphone change the envelope of the received signal. The bat vocalizations are frequency sweeps, and the interaction between incident sound and overlapping echo takes the form of ''beats'' in the envelope. This is illustrated in Figs. 4(A) and (B), which show the results of ensonifying the array with an emitter placed level with the array and producing frequency sweeps. Steeper sweeps, shown in Fig. 4(B), result in less of an overlap zone and fewer beats than shown in (A), since the interacting frequencies are further apart. The modulation depth of the beats depends on how strong the echo is relative to the direct emission. Figures 4(A) and (B) illustrate the largest echo effects, since the emitter is placed in the plane of the array (0.9 m above the ground), and the array microphones received a relatively large echo from the base of the microphone support. In general, the bats do not fly so low in the room (the average altitude of the bats is about 1.5 m off the ground and this is probably influenced by the height at which prey items are usually presented). Thus, the echoes that interact at the array microphones are typically from the walls or floor. These echoes are greatly attenuated (due to the sound-absorbent foam used). In addition, the path the echo travels is larger and the overlap with the incident sound is less. This is illustrated in Fig. 4(C). In practice, modulation of the sound at the microphone array due to loud echoes overlapping with the incident sound is rarely observed. In addition, runs were done with the emitter placed in the plane of the array and producing shallow frequency sweeps so as to intentionally corrupt the readings with echoes. Analysis of these runs show that the error introduced by echoes remains within the tolerance ($\pm 1.4°$) of the method.

F. Limitations of a linear array

The sonar beam of the bat extends in both azimuth and elevation. A linear array takes only a slice through the three-dimensional structure of the beam. Therefore, the exact shape and amplitude of the beam pattern recorded by a linear array depends on the vertical orientation of the beam. This means that absolute measurements of the beamwidth and intensity cannot be taken from our data. The conclusions about beam axis remain valid for a bat with its head held roughly level with the horizon. The bat's beam is not of circular cross

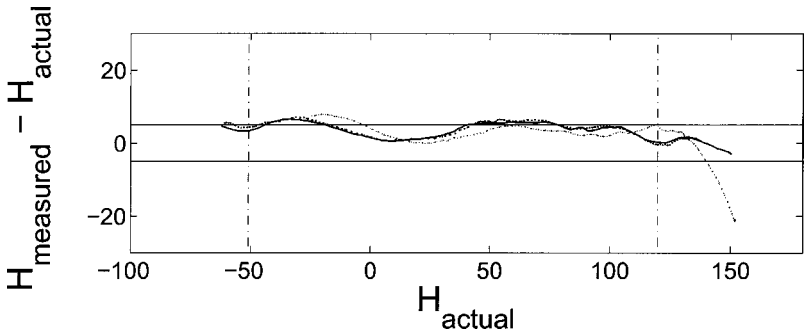


FIG. 7. This graph summarizes the calibration runs. The *y* axis shows the angular difference between the emitter direction observed from the video ($\mathbf{H}_{\text{actual}}$) and the beam center estimated from the array data ($\mathbf{H}_{\text{computed}}$) plotted against $\mathbf{H}_{\text{actual}}$. The two horizontal lines mark $\pm 5°$. This graph illustrates the edge effect predicted by the simulations (see Fig. 6). The edge effect is seen as an increase in bias of the error towards one direction as $\mathbf{H}_{\text{actual}}$ begins to approach the edge of the array. The average of these traces between the two vertical dotted lines at $-50°$ and $+120°$ was used to create a calibration curve to map the measured beam axis to a corrected beam axis. The final beam axis compututation results in an error of $\pm 1.4°$.

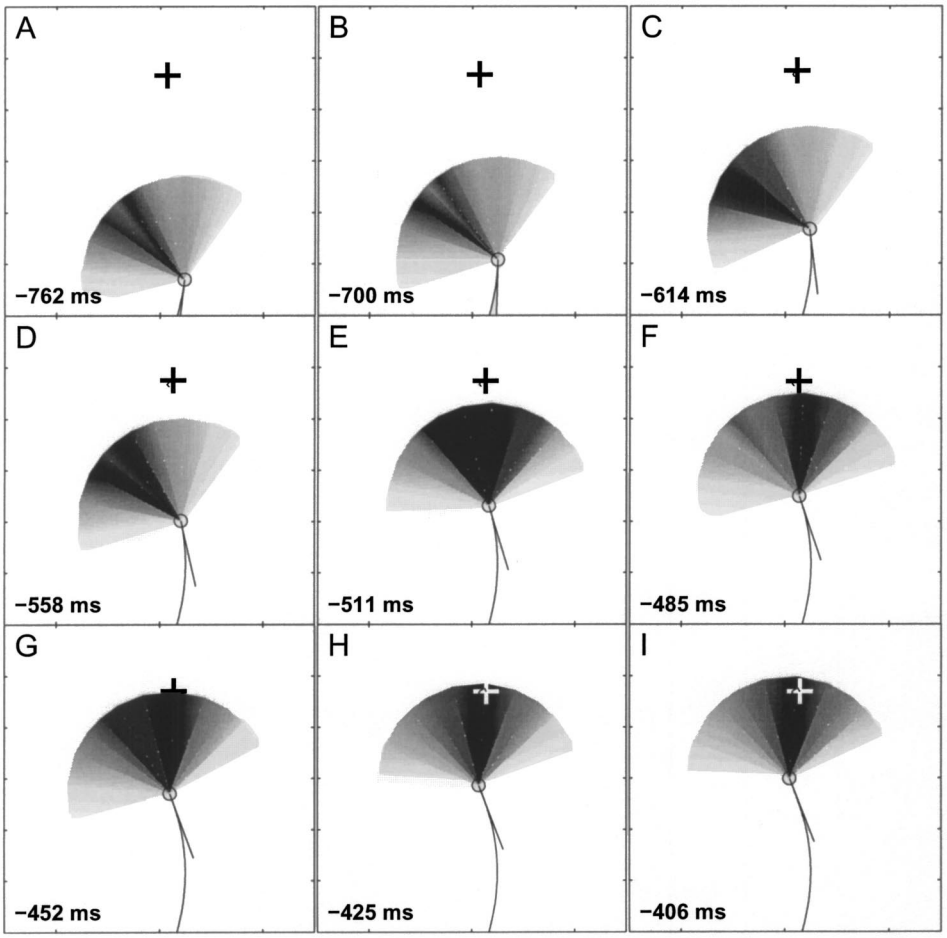


FIG. 8. Beam patterns of several vocalizations from a bat intercepting a tethered meal worm. The meal worm is denoted by +; the bat is denoted by a circle with a line extending to show the velocity vector of flight, which is assumed to be approximately aligned to the body. The times indicate milliseconds before contact. The circles at the borders of the panels denote the positions of the microphones. Note how a scanning motion (A,B,C,D) narrows down (E,F,G) and then changes to a ''lock-on'' motion (H,I) as the bat searches for then selects the target. Also note the split that appears in vocalization patterns A and B. This is discussed in the text. Microphone positions are not shown, but the orientation of the plot is identical to that in Fig. 5.

section (and indeed may have a prominent ventral lobe;[5] also see the ''horns of the bat'' subsection later), and so the beam pattern recorded by a linear array will be distorted if the bat rotates its head relative to the horizontal.

IV. RESULTS

A. Beam patterns

The basic data from the experiments are the beam patterns measured as the bat selects, tracks, and then captures a target. A sequence showing beam patterns from successive vocalizations is shown in Fig. 8. These show clearly how the bat first scans the space around it with the beam [Figs. 8(A), (B), and (C)] and then aligns its beam with the target [Figs. 8(D), (E), and (F)]. Also note the ''notch'' in the beam patterns in (A), (B), and (D). The notch may be due to the orientation of the bat's head with respect to the horizontal microphone array and the ventral lobe of the beam. This is discussed in a later subsection, ''horns of the bat.'' We also made animations of the beam patterns recorded from several trials, and these are available as .avi files on our website http://www.bsos.umd.edu/psyc/batlab/jasa03/. A brief description of the animations on the website is given in Table I.

B. Beamwidth

The measurements were used to find the half-power points of the beam (where the intensity is 3 dB below the peak). Figure 9(A) is a frequency histogram of the full -3 dB beamwidths obtained by this method. The four traces correspond to data from the four bats. Figure 9(B) shows a scatter plot of the beamwidths against the range from the target when they were obtained. There is no significant correlation between beam width and range to target ($r=-0.0252$, $p>0.1$). Most of the data points are obtained with the bat within 1 m of the target. The mean value of -3-dB beamwidth from all the bats is 70°.

C. Tracking accuracy

Using Eq. (6) the axis of the beam can be obtained. The angular deviation between the beam axis and the target (the tracking angle) for 13 trials was analyzed and the results are summarized in Fig. 10. Figure 10(A) shows the tracking angle plotted against time to contact with the target. During the last 300 ms of attack the bat locks it beam with a standard deviation (σ) of 3° onto the target. Figure 10(B) shows the tracking angle plotted against range to target. This shows that within 0.5 m of capture the bat has locked its beam onto the target with a σ of 3°. Figure 10(C) shows the interpulse interval plotted against time. Figures 10(D) and (E) show the distribution of tracking angles at different stages. (D) shows data taken when the bat was more than 300 ms from target contact, while (E) shows data taken when the bat was within 300 ms of contact.

TABLE I. Beam-pattern animation descriptions. All files are found at http://www.bsos.umd.edu/psyc/batlab/jasa03/

File name (.avi)	Description
2001.09.18.2.01 2001.09.18.2.01 split 2001.09.18.2.01 splitpoplar	The bat flies in from the far end of the array. The black persistent lines represent the computed beam axis for each vocalization. The worm is dropped into the flight space at frame 78. The bat directs its beam initially to the left of its flight path up to frame 132, then starts to ping in the direction of the target (ahead of it) from frame 143 onwards. It increases its repetition rate noticeably from frame 169 onwards. The 2001.09.18.2.01split animation shows the view from one of the infrared cameras. The 2001.09.18.2.01splitpolar animation shows polar plots of the beam pattern.
2001.10.02.1.01 2001.10.02.1.01 split	The bat takes a sharp turn to its right, flying towards the room center. The target is dropped from the trap door in frame 100. The bat first directs its beam towards the target at frame 199, and makes a sharp turn left to try and intercept it. The bat hits the target but fails to capture it. The target remains swinging on the tether. The bat flies past, then makes a sharp 180° turn starting at frame 406 and directs its beam in the direction of the target. It picks up pursuit of the target at frame 535, noticeably increasing its repetition rate at frame 545. This attempt ends in a successful capture.
2002.08.20.3.02	The bat flies towards the center of the room. The black square represents an inedible block of foam. The bat vocalizes ahead of its flight path. The target is dropped at frame 25. The bat initially ''inspects'' the inedible foam block (frames 119 to 181) then directs its beam to the target from frame 184 onwards.
2001.06.12.1.03	The bat attempts to capture a tethered meal worm being moved in a circle about 0.5 m in diameter. The bat keeps its beam centered on the target throughout, even though it gives up pursuit after making a complete circuit. Beam-pattern data are not available for part of the pursuit (during which the beam was directed where there were no microphones).

D. The "horns of the bat"

Referring to vocalizations shown in panels (A), (B), and (D) of Fig. 8, the beam seems to be split in two, i.e., displaying two spatially separate energy peaks. The remaining vocalizations seem to have one large lobe. Polar plots of normalized intensity for a single beam and a ''notched'' beam are shown in the right panels of Figs. 11(A) and (B), respectively. The left panels show the image from one of the cameras at the instants these beam patterns were measured. The image of the bat is circled. We confirmed that this notch was not due to measurement error (e.g., malfunction in some of the array elements). As shown in Fig. 11(B), we discovered that in some trials the notch occurred when the bat was clearly banking during a turn. We do not know if the head is tilted during the bank.

V. DISCUSSION

A. Tracking accuracy

From our experiments we conclude that the big brown bat, *E. fuscus*, tracking tethered insects, centers its beam axis on the target with a standard deviation (σ) of 3° during the terminal phase of insect capture. The method used here introduced an error of $\pm 1.4°$. The value of target accuracy we obtain is lower than the accuracy reported by Masters with measurements taken from a stationary bat tracking a smoothly moving target from a platform[9] which was given as 1°. However, in the Masters, Moffat, and Simmons study the authors applied a lag and gain correction to the bat's actual head motion to arrive at the value. The actual head motion, as reported in that paper, appeared to follow the target motion with errors of up to 10°. The bat seemed to follow the target accurately when it was sweeping past the front of its observing platform, but as the target rotated to more extreme angles the bat did not orient to follow it completely. Webster and Brazier,[10] using photographs of bats attacking free-flying insect prey, arrived at the slightly looser value of 5°, but the accuracy of the method used was not mentioned.

Given that the 3 dB width of the beam is around 70°, a standard deviation of 3° in directing the beam onto the target is unlikely to be due to the bat's need to maintain a good echo return from the target. We cannot say from these experiments what other advantage there may be to centering the tracked target. One review[11] suggests that the bat's azi-

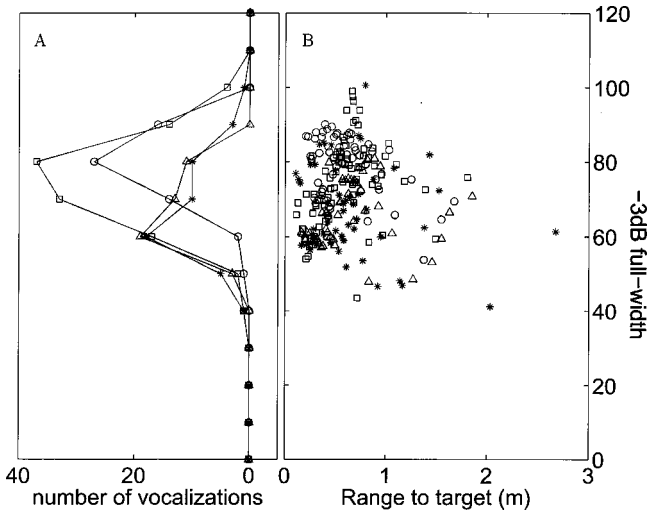


FIG. 9. (A) shows a frequency histogram of the computed beamwidths over 13 trials and 4 bats. (B) shows the data from which this histogram was made plotted against the range to target at which the measurements were taken. The data from different bats are shown as different symbols.

muthal localization acuity is greatest in a narrow (~10°) zone directly in front of it. If this is correct, then the bat may be centering the target while tracking in order to keep it in this high localization acuity zone. Neural recordings from the inferior colliculus of the mustached bat show that the thresholds of all binaural neurons are lowest at the horizontal midline independent of the neuron's frequency selectivity,[12] suggesting that for mustached bats, at least, there is a preference for processing echoes from directly ahead. Studies on the localization ability of the bottle-nosed dolphin indicate that the minimum audible angle (MAA) directly in front of the animal for broadband clicks is around 0.9° in azimuth.[13] The MAA in more lateral positions has not been studied.

Assuming that the axis of the beam bears a constant relation to the bat's head, another hypothesis may be that a type of beamforming operates in the bat's acoustic system. In this beamforming operation, signals that arrive simultaneously in both ears (i.e., from the center line) are enhanced compared to signals from more off-axis targets.

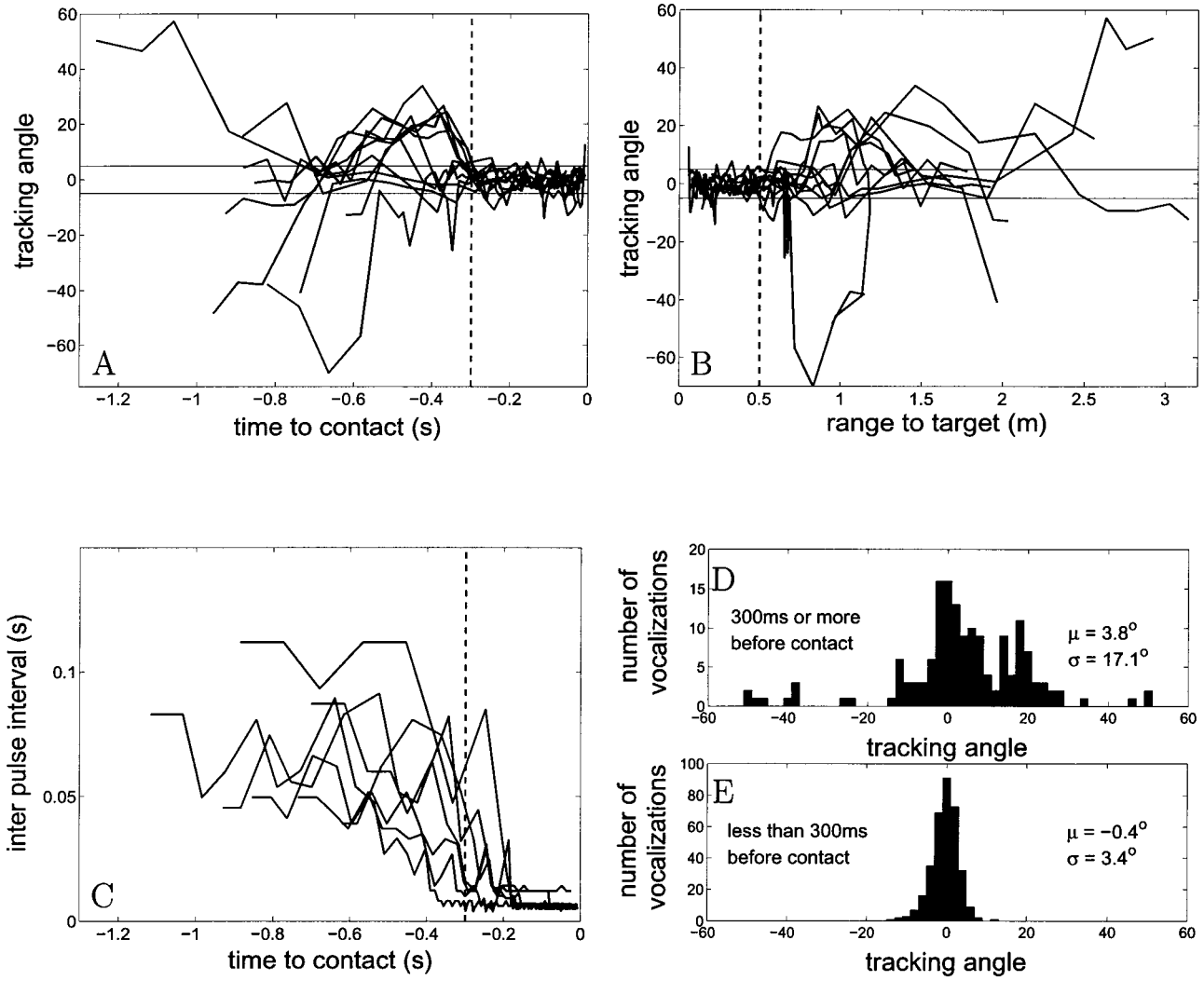


FIG. 10. This plot summarizes the results of analyzing the angular deviation between the beam axis and the target (the tracking angle) for 13 trials. (A) shows the tracking angle for each trial plotted against time to contact with the target (zero being time of contact). The vertical dotted line marks 300 ms before contact. (B) is a plot of tracking angle against range to target. The vertical dotted line marks 0.5 m to target. In plots (A) and (B) the solid horizontal lines mark ±5°. (C) shows the interpulse interval plotted against time. (D) and (E) show the distribution of tracking angles at different time periods before target contact. (D) shows data when there is more than 300 ms to contact, while (E) shows data when the bat is within 300 ms of contact.

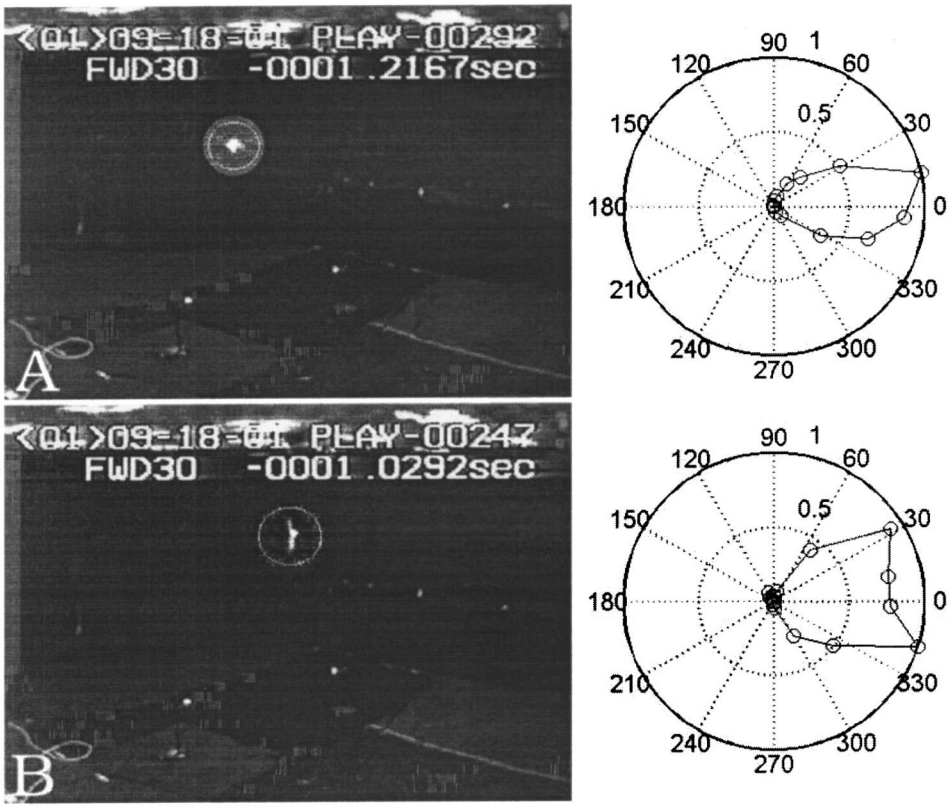


FIG. 11. The images in the left column are taken from one of the video camera records of a trial. The location of the bat is circled. The images are roughly 190 ms apart in time. The right column shows the polar plot of intensity (maximum intensity normalized to 1.0, and represented by the outermost ring of the polar plot) from the vocalizations made during the respective frames. The circles denote actual intensity data points. In (A) the bat is in level flight, heading parallel to the plane of the camera. Note that the beam pattern has a single large lobe. In (B) the bat is banking sharply, as can be deduced from the relative positions of the wings. It is moving into the plane of the camera. Note that the beam pattern now has a prominent notch. Animations of this trial may be seen at http://www.bsos.umd.edu/psyc/batlab/jasa03/. The original gray levels of the camera images have been remapped in a nonlinear fashion to enhance the images.

B. Beam axis as an index of selection and tracking

Our data suggest that the sonar beam direction of an echolocating bat is a useful index of its selection and tracking behavior during prey capture. The bat points its beam around the flight space using a gradual scanning motion while searching for prey. When prey is presented to the bat (e.g., by dropping it into the flight space using a trap door) the scanning pattern shifts towards the position of the target. Finally, the bat ''locks'' its sonar beam onto the target and tracks it closely. The lock-on behavior precedes the high vocalization repetition rates characteristic of the terminal phase by 50–100 ms [see Figs. 10(A) and (C)]. This may reflect a sequential process of first localizing an object, directing the beam towards it, and then identifying it as a prey item to capture. It may also indicate different latencies for motor pathways mediating head orientation and vocalization control.

The following animation illustrates that the bat may direct its beam sequentially at different objects before deciding to attack one [http://www.bsos.umd.edu/psyc/batlab/jasa03/2002.08.20.3.02.avi] .

The lock-on behavior is observed even when the prey and bat are moving in a tight circle, and the bat is not within catching distance of the prey, as illustrated in animation [http://www.bsos.umd.edu/psyc/batlab/jasa03/2001.06.12.1.03.avi]. This animation also demonstrates that the bat may orient its beam up to 90° off its flight path (''looking over its shoulder'') in order to maintain lock-on to the target. It appears that pointing its beam at a target of interest is a deliberate strategy adopted by the bat.

It is important, at this point, to note that the relationship between the beamwidth and the spatial limits of target perception by the bat are unknown. The limits will possibly depend on a combination of the size of the target, orientation of the pinnae, and intensity of the vocalization in addition to the direction and width of the beam.

C. The horns of the bat

We consider here why we observe a notch in some beam patterns. We noted that (a) the notch could ''travel'' from one microphone (or two adjacent microphones) to the other and (b) during the same trial we could get a combination of ''normal'' and ''notched'' beam patterns, implying that it was not an artifact due to a bad microphone. We hypothesize that the notch is due to a strong ventral lobe, perhaps more prominent than that measured by Hartley, which was 6 dB below the main lobe intensity.[5] Whenever the bat's head is sufficiently tilted with respect to the horizontal, the cross section of the sonar beam taken by the linear array would pick up the two lobes. In other cases, when the head is level with respect to the array, the cross section consists of one lobe. In support of this hypothesis, we noted that in some trials the notch appears during sharp banking turns by the bat (as estimated from the positions of the wings), e.g., see Fig. 11(B). During a banking turn, it is likely the head is also tilted with respect to the horizon. The notch is probably also not due to a shadowing effect of the beam by the target since it is sometimes observed when the beam is directed away from the target, or when there is no target in the room (e.g., Fig. 8).

D. Comparison with related work

Previous work using microphone arrays to record bat vocalizations have been conducted in the field, and the main

aims of these studies have been to estimate bat position and vocalization source levels. Jensen and Miller[14] used a vertical array of three microphones to study the variation of bat vocalization intensity with altitude. The array data were also used to localize the bat's position with respect to the microphones in the array. Holderied[15] used two microphone clusters to track bats up to a range of 35 m in the field and study source levels. These studies were not designed with the intent of studying the beam pattern directly, but have revealed indirect effects of the beam, such as periodic variations in received intensity, which may be attributed to the bat pointing its beam in different directions (i.e., scanning) while in flight.

Møhl *et al.*[16] recorded sperm whale vocalizations using an array of hydrophones. They used these data to localize the animals and deduce the directionality of their emissions. More controlled measurements of the beam patterns of stationary dolphins have been taken.[17] In comparison to bats, dolphins have a much narrower half-power beamwidth (10° compared to 70°). The peak of the main lobe seems to be directed upward of the snout axis by 5°, in contrast to *E. fuscus*, where the main lobe seems to be directed 10° below the snout. The differences in the width of the sonar beams of bats and marine mammals may be related to differences in the physical structure of the head as well as differences in signal generation and acoustics in air and underwater. Interaural time and intensity cues for localizing sound underwater are less salient than in air. By producing a narrow emission beam, dolphins could conceivably improve their localization ability.

E. Limitations of a linear array

The apparatus used here, a linear array of microphones, is limited in that it takes only a planar cross section of the bats' three-dimensional sonar beam. By using an array that extends in both the vertical and horizontal planes, these results may be extended to observe the vertical tracking behavior of the bat and the position of the notch (the region between the ventral and axial lobes of the beam) when the bat tracks prey.

VI. CONCLUSION

These experiments are the first measurements of the bat's sonar beam pattern as it tracks and intercepts prey in flight. There has been work on the sonar beam of a stationary anesthetized bat[5] where the sonar beam was described in great detail, but for a nonbehaving animal. There has been more extensively reported work on the sonar beam of dolphins and other odontocetes.[17] In these studies, too, the subjects were stationary and not using sonar for a target interception task.

The data presented here suggest that echolocating bats of the species *E. fuscus* direct their beam at a target of interest with an accuracy of about 3°. There may be some analogy between the orienting of the sonar beam by echolocating bats and the orienting of gaze by visual animals like primates. Early experiments by Yarbus on humans have revealed that when viewing the same scene the pattern of eye movements used is influenced by what information the subject is trying to acquire from the viewing.[18] Some experiments have also suggested that covert shifts of visual attention are linked to the preparation to make saccades.[19] Orienting the eyes to a visual stimulus is an important natural action, even though primates can, if needed, covertly attend to a stimulus without repositioning the eyes (For a review see McFadden and Wallman[20]).

We propose that the orientation of the beam may be used as an index that reveals some aspects of the bat's internal state during different behavioral tasks. Specifically, we think that the orienting of the beam may be used to probe what objects in a complex environment the bat is interested in. We also propose that the orienting behavior may be used to measure latencies in various target detection tasks in echolocation, much like eye movements are used in visual paradigms.

ACKNOWLEDGMENTS

We wish to thank Dr. Timothy Horiuchi for help with the design of the envelope detector circuit, for encouraging us when the going got tough and for making valuable suggestions for the improvement of this manuscript. We thank Dr. Jonathan Simon for advice and encouragement. Thanks go to the two anonymous reviewers for their comments on the first draft of this paper, which helped us to refine its presentation. We also thank Dr. Manjit Sahota for help with building the microphone array, and Amaya Perez for technical assistance. Hannah Gilkenson and Kari Bohn helped with the training of the bats used in the experiments. We would like to thank the Knowles company for donating several samples of their FG3329 electret microphone, which were used to build the array used in the experiments. Thanks are also due to Dr. Richard Berg of the UMD-CP Physics Dept. for his help in selecting an appropriate emitter to calibrate the array. This work was supported by NSF Grant IBN0111973 to C.F.M. and the UMD-CP Psychology Dept. Jack Bartlett fellowship to K.G.

[1] D. R. Griffin, *Listening in the Dark* (Yale University Press, New Haven, 1958).
[2] J. A. Simmons, K. M. Eastman, S. S. Horowitz, M. J. O'Farrell, and D. N. Lee, "Versatility of biosonar in the big brown bat, *Eptesicus fuscus*," ARLO **2**(1), 43–48 (2001).
[3] D. R. Griffin, F. A. Webster, and C. R. Michael, "The echolocation of flying insects by bats," Anim. Behav. **8**, 151–154 (1960).
[4] A. Surlykke and C. F. Moss, "Echolocation behavior of big brown bats, *Eptesicus fuscus*, in the field and the laboratory," J. Acoust. Soc. Am. **108**(5), 2419–2429 (2000).
[5] D. J. Hartley and R. A. Suthers, "The sound emission pattern of the echolocating bat, *Eptesicus fuscus*," J. Acoust. Soc. Am. **85**(3), 1348–1351 (1989).
[6] B. Lawrence and J. Simmons, "Measurements of atmospheric attenuation at ultrasonic frequencies and the significance for echolocation by bats," J. Acoust. Soc. Am. **71**(3), 585–590 (1982).
[7] G. M. Hope and K. P. Bhatnagar, "Electrical response of bat retina to spectral stimulation: Comparison of four micropteran species," Experientia **35**, 1189–1191 (1979).
[8] L. L. Beranek, *Acoustics* (Acoustical Society of America, New York, 1986).
[9] W. M. Masters, A. J. M. Moffat, and J. A. Simmons, "Sonar tracking of horizontally moving targets by the big brown bat, *Eptesicus fuscus*," Science **228**, 1331–1333 (1985).
[10] F. Webster and O. Brazier, "Expt. studies on target detection, evaluation

and interception by echolocating bats,'' Aero. Med. Res. Lab, Wright Patterson AFB Tech Rep., 65–172 (1965).
[11] J. A. Simmons, ''Directional Hearing in Echolocating Animals,'' in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer-Verlag, New York, 1987), Chap. 8, p. 217.
[12] E. Covey and J. H. Casseday, ''The lower brainstem auditory pathways,'' in *Hearing by Bats, Springer Handbook of Auditory Research*, edited by A. N. Popper and R. R. Fay (Springer-Verlag, New York, 1995), Chap. 11, p. 262.
[13] D. L. Renaud and A. N. Popper, ''Sound localization by the bottle-nosed porpoise, *Tursiops truncatus*,'' J. Exp. Biol. **63**, 569–585 (1975).
[14] M. E. Jensen and L. A. Miller, ''Echolocation signals of the bat *Eptesicus serotinus* recorded using a vertical microphone array: Effect of flight altitude on searching signals,'' Behav. Ecol. Sociobiol. **47**, 60–69 (1999).
[15] M. Holderied, ''Faster, higher, further,'' talk at the University of Maryland (2002).
[16] B. Møhl, M. Wahlberg, P. T. Madsen, L. A. Miller, and A. Surlykke, ''Sperm whale clicks: Directionality and source level revisited,'' J. Acoust. Soc. Am. **107**(1), 638–648 (2000).
[17] W. W. L. Au, ''Characteristics of the transmission system,'' in *The Sonar of Dolphins* (Springer-Verlag, New York, 1993), Chap. 6, pp. 98–114.
[18] A. L. Yarbus, *Eye Movements and Vision* (Plenum, New York, 1967).
[19] G. Rizzolatti, L. Riggio, I. Dascola, and C. Umilt, ''Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention,'' Neuropsychologia **25**, 31–46 (1987).
[20] S. McFadden and J. Wallman, ''Shifts of Attention,'' in *Vision and Attention*, edited by M. Jenkin and L. Harris (Springer, New York, 2001), Chap. 2, p. 22.

The relationship between acoustic structure and semantic information in Diana monkey alarm vocalization[a)]

Tobias Riede[b)]
Institut für Theoretische Biologie, Humboldt-Universität zu Berlin, Invalidenstrasse 43, 10115 Berlin, Germany

Klaus Zuberbühler[c)]
School of Psychology, University of St. Andrews, St. Andrews, Fife KY16 9JU, Scotland, United Kingdom

(Received 2 July 2002; revised 5 April 2003; accepted 15 April 2003)

Mammalian vocal production mechanisms are still poorly understood despite their significance for theories of human speech evolution. Particularly, it is still unclear to what degree mammals are capable of actively controlling vocal-tract filtering, a defining feature of human speech production. To address this issue, a detailed acoustic analysis on the alarm vocalization of free-ranging Diana monkeys was conducted. These vocalizations are especially interesting because they convey semantic information about two of the monkeys' natural predators, the leopard and the crowned eagle. Here, vocal tract and sound source parameter in Diana monkey alarm vocalizations are described. It is found that a vocalization-initial formant downward transition distinguishes most reliably between eagle and leopard alarm vocalization. This finding is discussed as an indication of articulation and alternatively as the result of a strong nasalization effect. It is suggested that the formant modulation is the result of active vocal filtering used by the monkeys to encode semantic information, an ability previously thought to be restricted to human speech. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1580812]

PACS numbers: 43.80.Ka [WA]

I. INTRODUCTION

Male Diana monkeys produce acoustically distinct alarm vocalization to two of their natural predators, the crowned eagles and the leopards (Zuberbühler *et al.*, 1999; Zuberbühler, 2000a). Field playback experiments have shown that nearby listeners respond to these alarm vocalization by producing their own corresponding alarm vocalizations and by showing characteristic locomotor responses. For instance, when hearing leopard alarm vocalization from a concealed speaker, nearby monkeys tend to approach the site of the suspected predator while continuously producing alarm vocalizations, presumably to signal detection to the predator and futility of further hunting (Zuberbühler, Jenny, and Bshary, 1999). Playback experiments have shown that the monkeys' response is driven by the associated meaning, rather than the vocalizations' mere acoustic features (Zuberbühler *et al.*, 1999). This has been taken to suggest that Diana monkey alarm vocalization are another example of natural semantic communication in animals (Seyfarth and Cheney, 2003).

Comparably little is known about the mechanisms of sound production that underlie these behavioral patterns. There is an increasing consensus among researchers in the field that the source-filter-theory, originally put forward to explain speech production (Fant, 1960), serves as a useful model for mammalian sound production (Andrew, 1976; Owren and Bernacki, 1988; Riede and Fitch, 1999). The theory posits that a vocal signal is produced by the vocal folds (the source), and is subsequently shaped by the resonance properties of the vocal tract (the filter). A number of recent studies suggested that some nonhuman species are capable of vocal-tract filtering by controlling the resonance properties independently of the glottal source (Hauser *et al.*, 1993; Hauser and Schön-Ybarra, 1994; Fitch and Reby, 2001). Owren (1990a,b) showed that eagle, and snake alarm vocalizations of vervet monkeys could be distinguished by measures associated with the source, the filter, and timing. They used synthetic versions of these vocalizations to show that individual subjects based their discrimination on acoustic cues associated with the filter, independent of those associated with the source or timing.

The sound production systems of all mammals exhibit a number of fundamental anatomical and acoustical similarities. The primary acoustic signal is generated at a source, typically the vocal folds of the larynx (the glottal source), which are driven into rapid mechanical oscillations by an expiratory airflow from the lungs. The oscillating vocal folds modulate the airflow through the glottal opening, i.e., the airspace between the vocal folds, producing a time-varying acoustic signal: the glottal source signal. The vocal folds are set into vibrations by the combined effect of subglottal pressure, the viscoelastic properties of the folds, and the Bernoulli effect. The aerodynamic energy is sustained by the subglottal pressure, which is maintained by the muscles of expiration. Recently, it has been shown that the vocal folds constitute a highly nonlinear self-oscillating system best modeled as coupled oscillators (Herzel *et al.*, 1995), resulting in the occurrence of nonlinear phenomena in the vocal repertoire.

[a)]Dedicated to Prof. Günter Tembrock on the occassion of his 85th birthday.
[b)]Present address: Medical Science Program, 315 Jordan Hall, Indiana University, IN 47405. Electronic mail: tobiasriede@web.de
[c)]Electronic mail: kz3@st-and.ac.uk

 0001-4966/2003/114(2)/1132/11/$19.00

Nonlinear phenomena have been demonstrated earlier in a number of non-human mammals (Wilden *et al.*, 1998), but they have not been found in male Diana monkeys alarm vocalization (Riede and Zuberbühler, 2003).

All mammals have a supralaryngeal vocal tract (hereafter referred to as ''vocal tract'') through which the sound generated at the glottal source must pass. Like any tube of air, the air column contained in the vocal tract has resonant modes, which selectively allow certain frequencies in the glottal source to pass and radiate through the mouth or nostrils into the environment. The vocal tract hence acts as a bank of bandpass filters, each of which allows a narrow range of frequencies to pass. The vocal-tract resonances, along with the spectral peaks they produce in the vocal signal, have been termed ''formants.'' Originally, the term was used to describe speech signals (e.g., Fant, 1960; Titze, 1994), but various researchers have used it to describe animal sounds (Lieberman *et al.*, 1969; Nowicki, 1987; Owren and Bernacki, 1988, 1998; Fitch, 1997; Riede and Fitch, 1999). However, many studies have used the term simply to describe spectral concentrations of acoustic energy that appear to be harmonically unrelated to the fundamental frequency (e.g., Hauser *et al.*, 1993; Zuberbühler, 2000a). Unequivocal identification of formants requires an analysis technique that separates the effect of the glottal source from the effect of the vocal tract (e.g., Owren and Bernacki, 1998).

In this study, we investigate the impact of source and vocal-tract parameters on the acoustic structure of Diana alarm vocalizations. We were particularly interested in the question of whether the source-filter theory is suitable to explain the acoustic patterns produced by the Diana monkeys in response to the two predators. We conducted an acoustic analysis based on a linear predictive coding algorithm to determine (i) whether source and tract parameters are independent of each other and (ii) which of the two the monkeys use to convey predator information.

II. METHODS

A. Study site, subjects, and vocalization sample

Data were collected in the Taï National Park, Côte d'Ivoire, between June 1994 and June 1997 in an approximately 40-km^2 study area of primary rain forest surrounding the Centre de Recherche en Ecologie (Université de Abobo-Adjame, Abidjan) research station (5° 50′ N, 7° 21′ W), in the Taï National Park, Côte d'Ivoire. Seven monkey species were regularly observed in the area: the western red colobus (*Colobus badius*), the western black-and-white colobus (*Colobus polykomos*), the olive colobus (*Colobus verus*), the Diana monkey (*Cercopithecus diana*), the lesser white-nosed monkey (*Cercopithecus petaurista*), the Campbell's monkey (*Cercopithecus campbelli*), and the sooty mangabey (*Cercocebus atys*). Diana monkey groups typically consisted of about 20–25 individuals with one adult male and several adult females with their offspring, each occupying and defending a stable home range of less than 1 square kilometer against neighboring groups.

We analyzed the vocalizations of ten different adult males from ten different Diana monkey groups, five responding to an eagle and five to a leopard. We analyzed the first five vocalizations given by each individual, resulting in a sample of 25 eagle and 25 leopard alarm vocalizations. Recordings were made with a Sony WMD6C tape recorder and a Sennheiser microphone (ME88 head with K3U power module) on 90-min type IV metal tapes. The frequency response of the microphone (40 Hz–20 kHz; ±2.5 dB) and the tape recorder (40 Hz–14 kHz, ±3 dB; distortion of 0.1%; signal-to-noise-ratio of 57 dB) are flat and within the frequency range of analysis. Recordings were made at distances ranging from 20 to 50 meters.

B. Acoustic analysis

We digitized all recordings at a 16-bit quantization and a 44-kHz sampling rate using SIGNALIZE software. We performed signal analysis on a PC using the signal-processing software HYPERSIGNAL-MACRO™ and a DSP32C PC system board. We completed the spectrographic analysis by using a 512-point fast Fourier transformation, with 75% frame overlapping, a 44-kHz sampling frequency, and a Hanning window. To avoid aliasing effects we low-passed filtered all vocalization at 22 kHz.

Linear predictive coding (''LPC'') is a spectral modeling technique used to estimate formant frequencies in human speech. LPC uses least-squares curve fitting to estimate the value of a point in a time-domain waveform based on the past N points, where N is the order of the LPC analysis. LPC algorithms then construct the best-fitting all-pole model to account for the waveform. ''All-pole'' means that only vocal-tract resonances (''poles'') are estimated, and not antiresonances (''zeros''). Such a spectral model appears to be a valid first approximation for most human speech signals and the monkey vocalization analyzed here (Markel and Gray, 1976). The specifics of the algorithms used in linear predictive coding are described elsewhere (see Markel and Gray, 1976 for the mathematical details, and Owren and Bernacki, 1998 for application in bioacoustics). In the current analysis we used the autocorrelation technique provided by HYPERSIGNAL™. The technique outputs the coefficients of an Nth-order all-pole digital filter whose frequency response best approximates in a least-squares sense the spectrum of the input signal. Given a broadband source signal and an appropriate model order [typically estimated as $2+2*N$(formants)], LPC analysis can provide an extremely accurate estimate of formant center frequencies in both human speech and animal sounds. Signal analysis was conducted with HYPERSIGNAL-MACRO™ with 12 coefficients and pre-emphasis settings of 0.8 to 0.99. All LPC measurements were visually verified by superimposing the LPC-derived frequency response over a 512-point fast Fourier transform (FFT) of the same time slice, allowing the user to select the optimum number of coefficients for each vocalization by trial and error.

In order to make formant extraction more reliable, acoustic parameters were interpreted using anatomical estimates. Since the recorded signal is a combination of the primary signal spectrum and the transfer function of the vocal tract, peaks in the spectrum can be the result of the source

TABLE I. Pulse duration. Minimum (P-dur min), maximum (p-dur max), mean (p-dur mean) of pulse duration in eagle (individuals 1 to 5) and leopard (individuals 6 to 10) alarm vocalizations. For each individual 10 calls (*N*) (2 calls per bout, 5 bouts per individual) were measured. In individuals 6 and 7, 1 out of 5 bouts contained only 1 call; therefore, 1 other bout was chosen to deliver 3 calls. In individual 9 only 4 bouts were available, and 1 of those delivered only 1 call.

Individual	*N*	p-durat min (ms)	p-durat max (ms)	p-durat mean (ms)	s.d. (ms)	Jitter (%)
			Eagle alarm			
1	10	13.9	19.9	16.4	1.1	6.4
2	10	8.3	21.9	15.9	0.6	8.7
3	10	13.3	22.2	17.4	1.5	6.5
4	10	8.3	20.3	14.1	1.7	9.2
5	10	10.1	24.0	16.9	1.1	8.3
mean	50			16.1	1.5	7.8
			Leopard alarm			
6	10	13.3	29.9	16.7	2.3	8.9
7	10	14.6	25.6	18.3	2.1	6.9
8	7	12.1	22.7	15.6	2.5	6.8
9	10	15.1	26.7	17.4	1.1	8.1
10	10	14.8	28.3	17.4	1.4	7.4
Mean	47	13.3	29.9	17.1	2.1	7.6

(e.g., regular/harmonic or irregular patterns) as well the vocal tract (i.e., formants). That makes it important to sort out if spectral peaks can be considered as formants. The relationship between vocal-tract length and formant position follows Eq. (1):

$$Fn=[(2n-1)*c]/4\mathrm{VTL}. \quad (1)$$

In this equation, *Fn* is the *n*th formant frequency in Hz; *c* is speed of sound (35 000 cm/s in warm humidified air) and VTL is vocal tract length in cm. This relationship is true for a uniform, hard-walled tube closed at one end, but it has also shown to be a good approximation for several nonhuman mammalian species' vocal tracts; e.g., domestic cats (Carterette *et al.*, 1979), rhesus macaques (Fitch, 1997; Rendall *et al.*, 1998), and domestic dogs (Riede and Fitch, 1999). Broadband utterances are particularly well suited for formant extractions (Owren and Linker, 1995) because they reflect the vocal-tract transfer function best. In addition to the male Diana monkey vocalization, our focus of interest in this paper, we also examined broadband female vocalizations for formant characteristics. Females are approximately 20% smaller than males, suggesting also to have an at least 20% shorter VTL (Fitch, 1997; Riede and Fitch, 1999).

C. Acoustic parameters

The basic acoustic unit in the Diana monkey alarm vocalization is the pulse (Riede and Zuberbühler, 2003). The pulse, which resembles a damped oscillation, is a rapid-amplitude transition of the signal from a baseline value to a higher or lower value, followed by rapid return to baseline. We measured pulse duration (Table I) as the interval between two pulse onsets. Pulse duration corresponds to the fundamental frequency of the oscillating vocal folds. The next higher acoustic units of Diana monkey alarm vocalizations are calls (Fig. 1), whose duration we also measured. Note that this terminology differs from the one used in a previous study (Zuberbühler, 2000a). A call consists of a series of pulses of varying duration. This cycle-to-cycle variability in fundamental frequency was termed jitter (Table I, Titze, 1994), calculated as the ratio of standard deviation to mean of the pulse duration per call. For each individual a mean jitter was calculated based on ten calls (except in one individual, when only seven calls were available). The final temporal parameter was bout duration, calculated as the overall duration of continuous acoustic energy. As spectral parameters we measured formants and formant bandwidth. Formant bandwidth is the size of the formant in the spectral representation. Two points are identified on the slopes of the resonance curve, where the response is 3 dB lower than at the peak. The difference in frequency between the 3-dB points defines the formant bandwidth. The size of the formant bandwidth is determined by the amount of attenuation in the vocal tract. Statistical analyses were based on monovariate Mann-Whitney *U*-tests, nested analysis of variance and factor analysis (SPSS 10.0).

III. RESULTS

A. Source acoustics

Pulse duration varied in eagle alarm vocalization between 8.3 and 24 ms (mean±s.d. 16.1±1.5), suggesting a fundamental frequency between 42 and 120 Hz (mean 62 Hz). It varied in leopard alarm vocalizations between 13.3 and 29.9 ms (mean±s.d. 17.1±2.1), suggesting a fundamental frequency between 33 and 75 Hz (mean 58 Hz). Pulse duration differed significantly between eagle and leopard alarm vocalizations ($N_1=50$, $N_2=42$, $U=649$, $P<0.001$, Mann-Whitney *U*-test, two-tailed), although the individual values overlapped broadly (Table 1). In both vocalization types, the pulses occurred very regularly, suggesting a rigid vibration pattern of the oscillating glottal source. Jitter did not differ significantly between eagle and leopard alarm vocalization (eagle alarm vocalization: mean 7.8%, leopard alarm vocalization: mean 7.6%; $N_1=5$, $N_2=5$, $U=13$, $P>0.2$, Mann-Whitney *U*-test, two-tailed). Call duration did

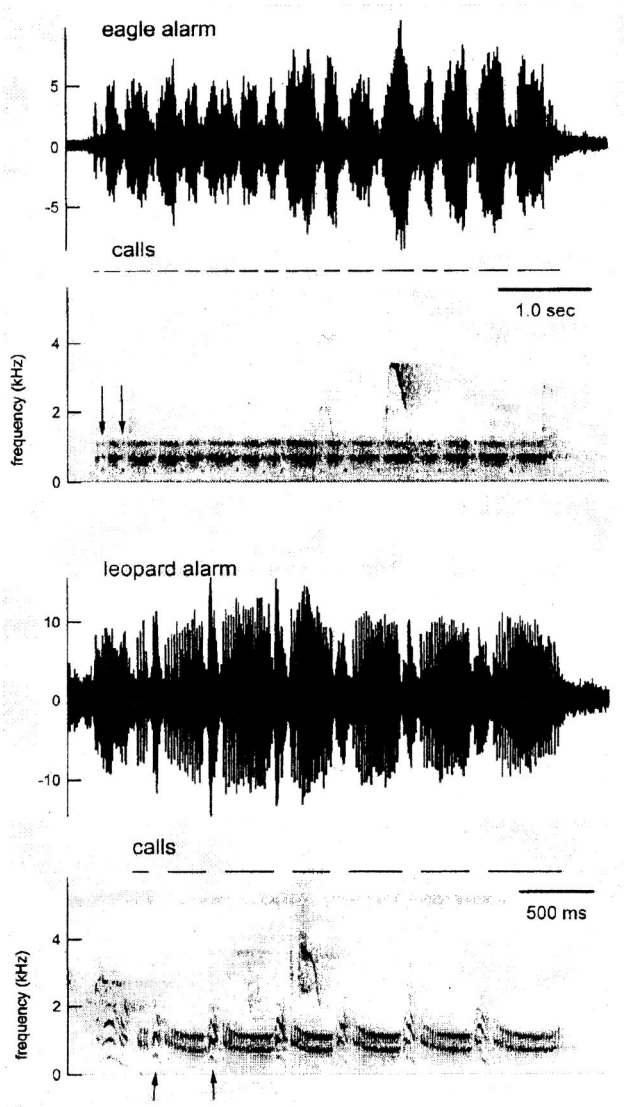


FIG. 1. Example of a male eagle alarm vocalization and a male leopard alarm vocalization. Top: time series, bottom: spectrogram. Calls are indicated by horizontial lines. The first two intercall elements in both bouts are indicated by arrows. In the range between 2 and 5 kHz several structures are visible (arch-like frequency bands), caused by alert calls given by other Diana monkey group members.

not differ significantly between eagle and leopard alarm vocalizations (leopard alarm vocalization: range: 56 and 566 ms; eagle alarm vocalizations range: 23 and 717 ms; $N_1=25$, $N_2=25$, $U=287$, $P>0.3$, Mann-Whitney U-test, two-tailed, Table II). Bout duration differed significantly between eagle (mean$\pm$s.d., 3615 ± 1392 ms) and leopard alarm vocalizations (mean$\pm$s.d., 2206 ± 1321 ms) ($N_1=25$, $N_2=25$, $U=146$, $P<0.001$), although the individual values overlapped broadly (Table II).

B. Tract acoustics

1. Identifying formants

The average power spectrum of five subsequent pulses of a male leopard alarm vocalizations is depicted in Fig. 2: Two prominent peaks are visible between 600 and 1500 Hz, both identified by LPC analysis. The first peak ranged between 690 and 1000 Hz (Table III). Assuming that this peak has formant properties, then Eq. (1) suggests a vocal-tract length between 8.8 and 12.7 cm, which is anatomically reasonable. There was a significant difference between the $F1$ values at the beginning of the calls in eagle and leopard alarm vocalizations ($N_1=50$, $N_2=51$, $U=40$, $P<0.001$, Mann-Whitney U-test, two-tailed, Table III). LPC analysis depicted a second peak at around 1300 Hz. To consider this peak as a second formant is difficult to reconcile with Eq. (1) since it presupposes a substantially longer vocal tract of approximately 20 cm. The second peak was close to the first one, even during modulation, and it exhibited a broader bandwidth. In some high-quality recordings with low background sound-pressure level, a third peak near 2800 Hz was visible. Assuming that this peak has formant properties (i.e., second formant), then Eq. (1) suggests a vocal-tract length similar to the one predicted by the first peak. In the female alarm vocalization (Fig. 3) two separate peaks were visible at 1000 and at 2800 Hz, both again suggesting a VTL of roughly 9 cm. An additional peak around 1300 Hz was absent in female vocalization.

2. Formant behavior

Leopard and eagle alarm vocalizations differed most strongly in the downward modulation of the first formant ($\Delta F1$-start; Figs. 1 and 2 and Table IV). Although present in both alarm vocalization types, the modulation was three times stronger in the leopard alarm vocalization and there was little overlap between vocalization types (Table IV; $N_1=50$, $N_2=51$, $U=174$, $P<0.001$, Mann-Whitney U-test, two-tailed), and although the variability between individuals is significant, the nested ANOVA indicates that the variability between ''eagle'' and ''leopard'' calls in the parameter ''formant transition'' is even higher (Table V).

Simultaneously, the second prominent peak at around 1300 Hz modulated downwards, but to a lesser extent. There was a significant difference between the first and the second peak difference at the beginning and in the middle of the call ($T=1064$; $P<0.05$, Wilcoxon test, two-tailed), suggesting that the first peak modulated stronger than the second one. Although $\Delta F1$-start was defined as the difference between the beginning and the middle of the call, the actual modulation reliably occurred during the first four to six pulses, corresponding to less than 20% of all pulses in the entire call (Fig. 4). Finally, the modulation of $F1$ between the middle and the end of the call was not significantly different between two alarm vocalization types, and individual values overlapped strongly ($N_1=50$, $N_2=50$, $U=1166$, $P>0.05$, Mann-Whitney U-test, two-tailed). A factor analysis indicated that 35% of overall acoustic variability was explained by the formant downward modulation.

IV. DISCUSSION

In this study we were interested in the acoustic structure of Diana monkey alarm vocalization and in how the various acoustic parameters segregated eagle from leopard alarm vocalizations. Our analyses suggested that the spectral peaks in the alarm vocalizations have formant properties, i.e., that the spectral concentrations of acoustic energy are harmonically

TABLE II. Acoustic parameters of eagle and leopard alarm vocalizations. Minimum, maximum, and mean number of calls (no call min; no call max; no call mean); minimum, maximum, and mean of the call duration (call dura min, call dura max, call dura mean) and bout duration of 5 individuals each represented with 5 Eagle and 5 Leopard alarm bouts.

	No call min	No call max	No call mean (ms)	Call dura min (ms)	Call dura max (ms)	Call dura mean (ms)	Bout duration (ms)
				Eagle Alarm			
Ind.1	4	7	5.8	64	717	294	2107
Ind.2	8	15	10.6	58	515	256	3593
Ind.3	10	16	13.6	23	495	242	4439
Ind.4	6	19	11.0	39	604	232	3490
Ind.5	10	23	15.0	56	472	199	4446
Mean			11.2			236±122	3615±1392
						$N=279$	$N=25$
				Leopard alarm			
Ind.6	1	17	8.6	56	482	264	3308
Ind.7	1	13	5.2	71	424	213	1584
Ind.8	2	6	3.6	73	369	221	1066
Ind.9	3	12	6.4	73	506	197	2038
Ind.10	6	9	7.2	73	566	293	3030
Mean			6.2			243±116	2205±1321
						$N=155$	$N=25$

unrelated to the fundamental frequency. A formant downward modulation at the beginning of the call most reliably distinguished between the two alarm vocalization types because there was little to no overlap between individual vocalizations, monovariate statistical analyses yielded the highest P values, and a factor analysis showed that 35% of overall acoustic variability could be explained by this one parameter.

A. Source acoustics

Although source acoustic parameters and temporal cues (pulse duration, call duration, bout duration, number of calls per bout, jitter) differentiated eagle and leopard alarm vocalization to various degrees, they were unable to distinguish between the two vocalization types: the parameter distribution was strongly overlapping between both types of vocalization.

Our analyses showed that the pulse character of the alarm vocalization was surprisingly consistent. Earlier studies considering the vocal source as a nonlinear system found much higher degrees of irregularities in vocal utterances. For example, up to 15% of human infant cries and animal vocalization contained nonlinear phenomena (Robb and Saxman, 1988; Wilden *et al.*, 1998; Riede *et al.*, 2000). Nonlinear phenomena were virtually absent in male Diana monkey vocalizationa and pulses were not interrupted by any other vibration modes of the vocal folds, except for the short harmonic intercall elements. Specialized system adjustments and the low fundamental frequency could account for the remarkable stability of the oscillating system (but see Mergell *et al.*, 1999).

The spectrum of a pulse at the source is expected to be broadband (Titze, 1994; Au, 1993). This suggests that prominent peaks in the spectrum of the emitted signal are due to filtering effects in the vocal tract. Because of their broad bandwidth, pulses are well suited to picture the resonance characteristics of the vocal tract. This characteristic, as well as the robustness of this source signal and the very low fundamental frequency let Riede and Zuberbühler (2003) to hypothesize that male Diana monkey alarm calls are adapted

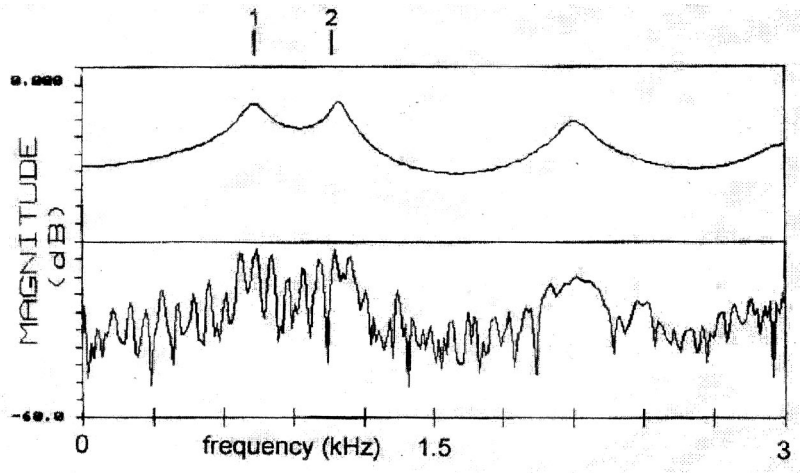


FIG. 2. Spectrum (lower curve) and LPC spectrum (upper curve) of a 100-ms segment of a male eagle alarm call. Two peaks are indicated (“1” and “2”). A third peak around 2100 Hz does not belong to the monkey call.

TABLE III. First formant at the beginning of the call. Mean and standard deviation of the first peak ($F1$) and the second peak ($F2$) at the beginning of the call ($F1$-start±s.d. and $F2$-start±s.d.). Mean and standard deviation of the bandwidth of the first ($B1\pm$s.d.) and second ($B2\pm$s.d.) peak at the beginning of the call.

	N	$F1$-start±s.d. (Hz)	$F2$-start±s.d. (Hz)	$B1\pm$s.d. (Hz)	$B2\pm$s.d. (Hz)
			Eagle alarm		
Individual 1	10	753±52	1298±45	132±53	135±52
Individual 2	10	744±18	1194±20	71±17	117±36
Individual 3	10	692±19	1221±238	69±19	109±41
Individual 4	10	824±26	1227±134	95±25	102±24
Individual 5	10	802±47	1287±82	88±29	120±34
Mean		763±58	1246±130	91±38	117±38
			Leopard alarm		
Individual 6	10	916±67	1371±155	69±16	154±55
Individual 7	10	947±46	1339±68	87±15	90±37
Individual 8	7	907±36	1466±68	86±40	181±35
Individual 9	10	996±71	1473±123	78±32	159±39
Individual 10	10	944±53	1447±142	80±25	167±40
Mean		944±63	1407±125	80±26	150±51

for a more elaborated vocal tract performance, i.e. formant variability. Preliminary video studies suggest a fourth characteristics of this system. There is a distinct possibility that the alarm calls investigated in this study are product of vocal fold vibration caused by air flowing *into* the lungs, rather than the more commonly observed expiratory sound production mechanism in mammals. In this case, the brief but highly harmonic inter-call elements seen in the spectrograms (Fig. 1) would have to be considered the product of vocal fold vibration caused by expiration. Inspiration-caused sound production could also explain the remarkable and otherwise rarely observed absence of nonlinear phenomena during call-

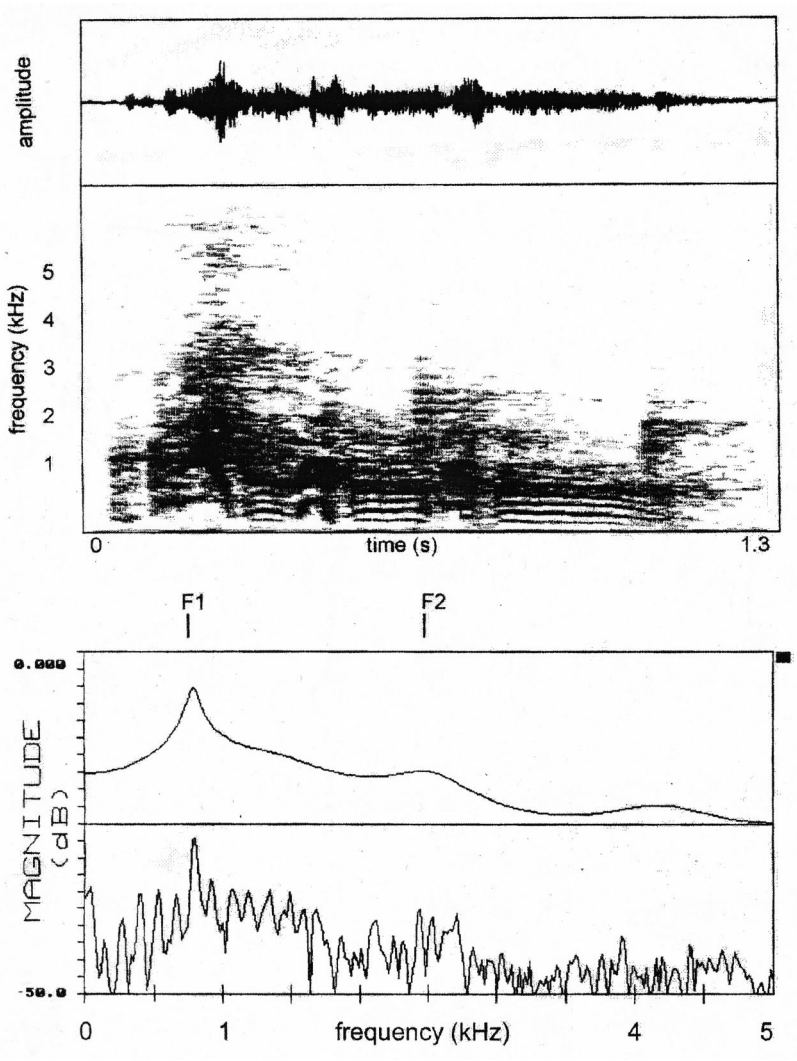


FIG. 3. Example of a female eagle alarm vocalization. Top: time series, middle: spectrogram. Bottom: a 100-ms segment is given in the frequency domain (lower curve) plus a LPC spectrum is given (upper curve). Two peaks are indicated ($F1$,$F2$) representing the first and second formant.

TABLE IV. Formant characteristic of eagle and leopard alarm calls. Mean ($\Delta F1$-start)±standard deviation (s.d.) of the down modulation of the first formant measured between the beginning and the middle of the call. Mean ($\Delta F1$-end)±standard deviation (s.d.) of the modulation of the first formant measured between the middle and the end of the call. For each individual a certain number of calls (N) were considered for measurements.

	N	$\Delta F1$-start±s.d. (Hz)	$\Delta F1$-end±s.d. (Hz)
		Eagle alarm	
Individual 1	10	55.8±60	8.5±33
Individual 2	10	21.6±14	−2.1±23
Individual 3	10	23.5±25	−4.3±22
Individual 4	10	66.5±23	−4.3±28
Individual 5	10	36.5±30	−10.7±18
Mean		50.8±34	−2.5±25
		Leopard alarm	
Individual 6	10	187.3±45	25.9±95
Individual 7	10	187.3±59	19.2±56
Individual 8	7	89.9±46	21.5±32
Individual 9	10	128.0±65	−16.4±23
Individual 10	10	185.0±72	6.6±46
Mean		153.5±68	11±55

ing, perhaps the product of a highly stable oscillating system. High resolution video analyses of the respiratory movements of the thorax during calling will be necessary to resolve this issue. Although we expect that the actual direction of airflow during call production has no direct implications for the explanation of the formant nature of the calls (Robb *et al.* 2001), the evolutionary insinuations would be quite intriguing. What selection pressures could have lead to the invention of an inspiration-based semantic communication system in the evolutionary history of this taxon?

B. Tract acoustics

Equation (1) makes a number of specific predictions regarding the location of the different formant frequencies. A crucial element in the equation is vocal-tract length, a parameter for which no anatomical data are available at the moment. However, previous studies have shown that in mammals there is strong positive correlation between skull length and vocal-tract length, and (in resting position of the larynx) that VTL is maximally as long as skull length but mean VTL is shorter than mean skull length (Fitch, 1997; Riede and Fitch, 1999; Fitch, 2000). Table VI lists skull lengths for Diana monkeys from various museum specimens, measured as the distance between the front of the incisor teeth and the *Protuberantia occipitalis*. Based on the relationship found by Fitch (1997) and Riede and Fitch (1999) the skull data (Table VI) suggest a vocal-tract length of maximally 11.5 cm, which predicts [according to Eq. (1)] a first formant at around 760 Hz, and a second formant at around 2280 Hz. LPC analysis depicts a peak in male and female vocalization near 800 Hz, strongly suggesting that this peak is the equivalent of a first formant. The 11.5-cm VTL also predicts [according to Eq. (1)] a second formant around 2280 Hz. LPC depicts a second peak at 2800 Hz, in males only in high-quality recordings (close distance between microphone and vocalizer, and very low background noise), and more regular in female vocalizations. This suggests that this peak is possibly the equivalent of the second formant. LPC depicts regularly a peak around 1300 Hz in male alarm vocalization, and it will be discussed below how this peak could be explained.

TABLE V. Nested ANOVA of the relationship between alarm call type (eagle versus leopard) and individual specificity.

	DF	Sum of squares	Mean square	F-ratio	Probability level
Call type	1	342 381	342 381	33.45	<0.001
Individual identity	8	81 880	10 235	4.41	<0.001
S	87	202 142	2 323		
Total	96	626 404			

C. The effects of changes in vocal tract diameter

Our calculations are based on the assumption that the formants are created in a uniform tube with no significant changes in tube diameter. It is well known that deviation from such uniformity often results in formant shifts. In humans, nonuniformity of the vocal tract is common, and speech vowels are prominent acoustic products of such nonuniformity (Fant, 1960). The consistent peaks of male Diana monkey vocalization at around 1300 Hz (see Fig. 3), therefore, could be the result of changing tube diameters in the male vocal tract. It is interesting that to the human ear, the calls of an eagle alarm vocalization strongly resemble the human vowel /o/. In contrast, the calls of a leopard alarm vocalization strongly resemble a vowel transition from /ʌ/ to /o/. In human speech, the second formants of both /ʌ/ and /o/ are lowered and they approach the first formant (Story *et al.*, 1996), similar to the monkey vocalizations. Therefore, one interpretation of these data is that, similar to human speech sounds, the second peak in male alarm vocalization around 1300 Hz represents the second formant ($F2$), due to variation in vocal-tract diameter and hence deviating from the predictions made by Eq. (1). Detailed anatomical work will be necessary to resolve this issue.

D. The effects of nasalization

An alternative explanation for the presence of acoustic energy at around 1300 Hz in male alarm vocalizations is provided by nasalization. Nasalization is produced by coupling the oral and nasal cavities via the velopharyngeal opening (*Ostium intrapharyngeum*). During this process the velum is lowered, resulting in coupling between the nasal cavity and the oral vocal tract. Acoustically, nasalization replaces the sharp spectral peak of the first formant ($F1$) by two peaks, the oral and the nasal pole (Fant, 1960), which widens the bandwidth of the first formant (Dang and Honda, 1996). Hence, nasalization shifts the natural frequencies of the oral part of the vocal tract by adding pole-zero pairs to the vocal-tract transfer function. These acoustic effects are strongest at low frequencies, in the vicinity of the first formant (House and Stevens, 1956; Stevens, Fant, and Hawkins, 1987; Maeda, 1993; Dang and Honda, 1996). In the higher frequencies, nasalization may introduce shifts in formants, modification in formant amplitudes, and additional spectral peaks. However, these effects are not as consistent

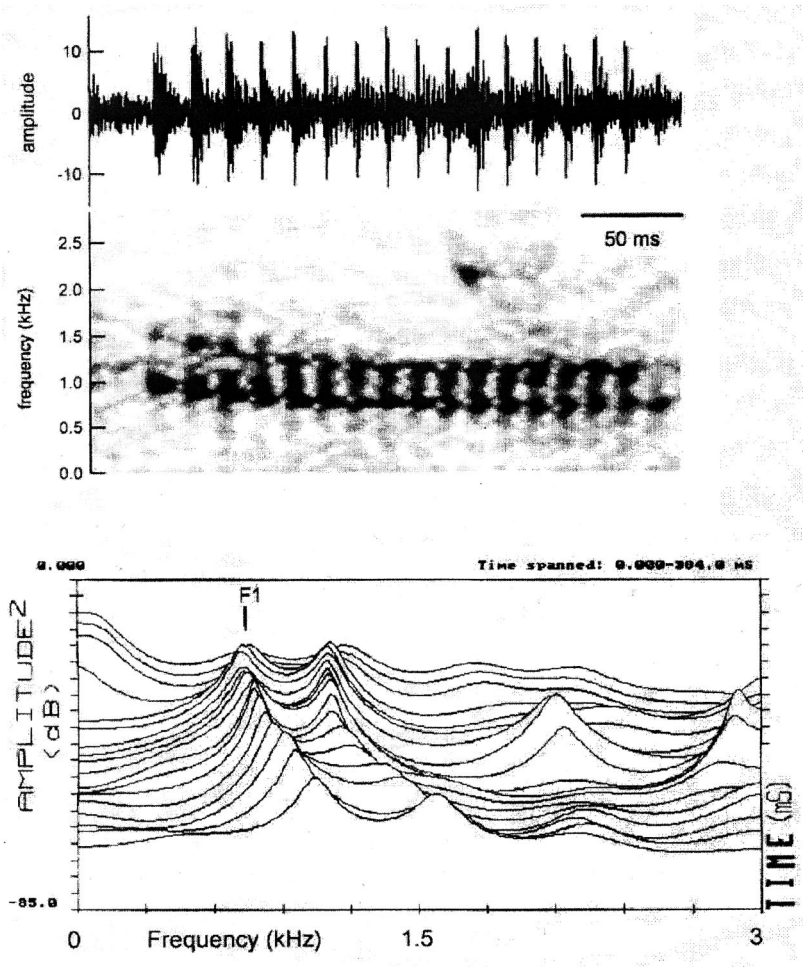


FIG. 4. Example of a male leopard alarm call. From top to bottom: time series, spectrogram, and waterfall representation of the LPC curves. The bottom waterfall representation is a three-dimensional display and shows the Fourier transform spectra of several time slices, altogether 0.384 s. The first formant ($F1$) is indicated in the waterfall representation.

as those in the vicinity of the first formant. Acoustic modeling showed that nasalization can also have secondary spectral effects, for example by reducing the amplitude of higher frequencies, i.e., second and third formants, possibly due to the strong dampening in the nasal tract (House and Stevens, 1956, p. 225).

In acoustic studies on nonhuman mammals, nasalization has been used to explain the appearance of subharmonics in the spectrogram of rhesus monkey (*Macaca mulatta*) vocalizations (Hauser, 1992). However, this interpretation is controversial because other work has shown that subharmonics are considered nonlinear phenomena, caused by vocal-fold vibration (Wilden *et al.*, 1998). Similarly, selective dampening of the fundamental frequency of siamang vocalization has been attributed to nasalization (Haimoff, 1983). However, an alternative explanation for this finding is that the zeros of the oral vocal tract were responsible for filtering the source signal in these animals. In Diana monkey alarm vocalizations, however, the two spectral peaks visible in the range of the first formant might be the result of nasalization. The lower peak around 800 Hz likely represents the first oral tract resonance, while the second peak around 1300 Hz could be the result of the first nasal resonance ($F1_n$). Thus, in Diana monkeys, nasalization can readily explain (i) the lack of acoustic energy above 1500 Hz due to dampening by the nasal tract; (ii) the close proximity of the first two peaks due to its tendency to replace the first formant by two peaks, the oral and the nasal pole.

TABLE VI. Skull length measurments (in cm) from skulls

	Skull length (cm)
1 Adult male (no: 2578)[a]	12.1
4 Adult males[b]	11.2±0.9
1 Adult female (no: 2620)[a]	10.7
8 Adult females[b]	9.9±0.3
2 Adult females[c]	both 10.3
4 Adult males[d]	11.5±0.54
3 Adult females[d]	9.84±1.1

[a]From the Paris Natural History museum, Laboratoire Mammifères et Oiseaux.
[b]http://1kai.dokkyomed.ac.jp/mammal/en/species/cercopithecus diana.html
[c]From the Indiana University at Bloomington Dept. of Anthropology collection.
[d]From the American Museum of Natural History at New York.

E. The effect of extralaryngeal cavities

It has been suggested that additional oral cavities branching off the oral vocal tract could affect formant behavior. For example, various guenons are known to possess pairwise or singular forms of air sacs, which develop directly from the laryngeal or pharyngeal cavity (Gautier, 1971). Perforation of air sacs reduced the amplitude of the vocalization, enriched the spectral pattern because more harmonics were visible in the spectrogram, and introduced irregular noise (Gautier, 1971). Recent modeling work on the effects of add-

ing a side branch (like the air sac) to the oral vocal tract leads to the consistent introduction of zeros into the transfer function and to an overall spectrum that resembled that of the Diana monkey alarm vocalizations (Jackson *et al.*, 2001).

F. Formant frequency modulations

Leopard and eagle alarm vocalizations differed most prominently in the downward modulation of the first formant at the beginning of each call (Fig. 1 and Table IV) with leopard alarm vocalizations exhibiting a threefold stronger downward modulation than eagle alarm vocalizations. The strength of downward modulation was very consistent and differentiated the two alarm vocalizations exclusively, suggesting that one single parameter differentiated between the two alarm vocalization types to a large degree. The formant differences at the beginning of a call indicated different configurations of the vocal tract, for instance in relative length, cross-sectional areas, or coupling with the nasal tract.

To human listeners the downward modulation of the first formant provide the perceptual impression of a transition from a /ʌ/ vowel to an /o/-like vowel. Anatomically the /ʌ/–/o/ transition is mainly correlated to an increase in volume of the pharyngeal chamber with minimal vocal-tract elongation (Story *et al.*, 1996). In this transition both formants modulate downwards, although the first formant does it more strongly than the second one. According to Peterson and Barney (1952) and Story *et al.* (1996), the /ʌ/ to /o/ transition is caused by a first and second formant decrease, similar to the findings in male Diana monkey alarm vocalizations. In the monkey alarm vocalizations both peaks modulated downwards simultaneously, while the first one did so more strongly than the second one. The articulatory maneuver in the frontal oral cavity responsible for the shift from /ʌ/ to /o/ appeared to be similar in the leopard alarm vocalizations, namely a lifting of the mandible combined with protrusion of the lips. Mandible lifting by female rhesus macaques (*Macaca mulatta*) caused decreases in the dominant frequencies of coo vocalization (Hauser *et al.*, 1993), although the study did not distinguish between dominant frequencies and formant frequencies.

A number of additional mechanisms could be causally related with the observed formant downward modulation. Vocal-tract elongation has the effect of decreasing formant frequencies [see Eq. (1)], either by lip protrusion or by larynx lowering, as recently demonstrated for red deer (Fitch and Reby, 2001).

G. Perceptual salience of the observed variation in formant modulation

Although no direct evidence is available, it is reasonable to assume that Diana monkeys are sensitive to variation in formant behavior and able to use this acoustic cue as a base for important behavioral decisions for the following reasons. First, humans classify vowels primarily on the basis of the two lowest formant frequencies (Peterson and Barney, 1952; Bogert and Peterson, 1957; Kent, 1978, 1979), suggesting that Diana monkey alarm vocalization provide sufficient acoustic information for accurate discriminations. Second, various birds and mammals have been tested successfully on their discriminative abilities on human vowels (baboons, Hienz and Brady, 1988; dogs, Baru, 1975; cats, Dewson, 1964; blackbirds and pigeons, Hienz *et al.*, 1981). For example, Japanese macaques (*Macaca fuscata*) were tested on single- and multiformant tone complexes, revealing formant-frequency discrimination abilities at 500 Hz and 1.4 kHz (Sommers *et al.*, 1992), which indicates that the formant downward modulation in Diana monkey alarm vocalization is perceptually salient to primate recipients. In chacma baboons (*Papio cyncephalus ursinus*) (Hienz and Brady, 1988) and Japanese macaques (*Macaca fuscata*) (Sinnott, 1989) it was shown that individuals readily discriminated among typical American-English vowel sounds. Japanese macaques and Sykes monkeys (*Cercopithecus albogularis*) were as good as human subjects in this task (Sinnott and Kreiter, 1993; Sinnott *et al.*, 1997). Sinnott and Kreiter (1993) synthesized a steady-state vowel continuum (formant behavior similar in appearance as in the Diana monkey leopard alarm vocalizations) by varying the first and second formant. Subjects heard a given vowel background and responded to changes towards a second vowel target, and monkeys showed similar sensitivity like humans (Sinnott and Kreiter, 1993). Some studies suggested that, similar to humans, nonhuman primates give the first formant more importance than the second formant (Kojima and Kiritani, 1989; Brown and Sinnott, 1990; Sinnott *et al.*, 1997).

Several studies have suggested that formant frequencies are likely to play an important role in nonhuman primate communication. Owren and Bernacki (1988) and Owren (1990c) used operant techniques to demonstrate that spectral features potentially related to formants are discriminated by vervet monkeys in their own species-specific vocalizations. In baboon vocalizations, Owren *et al.* (1997) found a relationship between changes of formant-related spectral peak patterns and social context. Rendall *et al.* (1999), looking at baboon grunt vocalizations, found that formant characteristics were correlated with social context. Rhesus macaque coo vocalizations and grunts are the effects of vocal-tract filtering, but these vocalizations are more likely to be related to individual differences than external events (Rendall *et al.*, 1998). This study reveals that in certain contexts nonhuman primates are able to engage vocal-tract changes to encode important events in the environment, a defining feature of human speech production.

ACKNOWLEDGMENTS

We thank Michael Owren, Robert Seyfarth, and one anonymous reviewer for comments on an earlier version of the manuscript. We are grateful to Gillian Page for providing us with video footage of male calling behavior. This work was supported by a fellowship within the Postdoc Program of the German Academic Exchange Service (DAAD) to TR.

Au, W. (**1993**). *The Sonar of Dolphins* (Springer, New York).
Andrew, R. J. (**1976**). ''Use of formants in the grunts baboons and other nonhuman primates,'' Ann. (N.Y.) Acad. Sci. **280**, 673–693.
Baru, A. V. (**1975**). ''Discrimination of synthesized vowels [a] and [i] with

varying parameters (fundamental frequency, intensity, duration and number of formants) in dog,'' in *Auditory Analysis and Perception of Speech*, edited by G. Fant and M. A. A. Tatham (Academic, New York).
Bogert, B. P., and Peterson, G. E. (**1957**). ''The acoustics of speech,'' in *Handbook of Speech Pathology*, edited by L. E. Travis (Appleton-Century-Cofts, New York), pp. 109–173.
Brown, C. H., and Sinnott, J. M. (**1990**). ''The perception of complex acoustic patterns in noise by Blue monkey (*Cercopithecus mitis*) and human listeners,'' Int. J. Comp. Psychology **4**, 79–90.
Brown, C. H., and Cannito, M. P. (**1995**). ''Modes of vocal variation in Sykes's monkey (*Cercopithecus albogularis*) squeals,'' J. Comp. Psychol. **109**, 398–415.
Carterette, E. C., Shipley, C., and Buchwald, J. S. (**1979**). ''Linear prediction theory of vocalization in cat and kitten,'' in *Frontier of Speech Communication Research*, edited by B. Lindblom and S. Öhman (Academic, New York), pp. 245–257.
Dang, J., and Honda, K. (**1996**). ''A new method for measuring antiresonance details of the vocal transmission characteristics: An experimental study of acoustic tubes,'' J. Acoust. Soc. Jpn. (E) **17**, 93–99.
Dewson, J. H. (**1964**). ''Speech sound discrimination by cats,'' Science (Washington, DC, U.S.) **141**, 555–556.
Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).
Fitch, W. T. (**1997**). ''Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques,'' J. Acoust. Soc. Am. **102**, 1213–1222.
Fitch, W. T. (**2000**). ''Skull dimensions in relation to body size in nonhuman mammals: The causal bases for acoustic allometry,'' Zoology **103**, 40–58.
Fitch, W. T., and Reby, D. (**2001**). ''The descended larynx is not uniquely human,'' Proc. R. Soc. London, Ser. B **268**, 1669–1675.
Gautier, J. P. (**1971**). ''Etude morphologique et fonctionnelle des annexes extra-laryngées des cercopithecinae; liason avec les cris d'espacement,'' Biol. Gabon. **7**, 229–267.
Haimoff, E. H. (**1983**). ''Occurrences of anti-resonance in the song of the siamang (*Hylobates syndactylus*),'' Am. J. Primatol **5**, 249–256.
Hauser, M. D. (**1992**). ''Articulatory and social factors influence the acoustic structure of rhesus monkey vocalization: A learned mode of production?'' J. Acoust. Soc. Am. **91**, 2175–2179.
Hauser, M. D., Evans, C. S., and Marler, P. (**1993**). ''The role of articulation in the production of rhesus monkey, *Macaca mulatta*, vocalization,'' Anim. Behav. **45**, 423–433.
Hauser, M. D., and Schön-Ybarra, M. (**1994**). ''The role of lip configuration in monkey vocalization: Experiments using xylocaine as a nerve block,'' Brain Lang **46**, 232–244.
Herzel, H., Berry, D., Titze, I. R., and Steinecke, I. (**1995**). ''Nonlinear dynamics of the voice: Signal analysis and biomechanical modelling,'' Chaos **5**, 30–34.
Hienz, R. D., Sachs, M. B., and Sinnott, J. M. (**1981**). ''Discrimination of steady-state vowels by blackbirds and pigeons,'' J. Acoust. Soc. Am. **70**, 699–706.
Hienz, R. D., and Brady, J. V. (**1988**). ''The acquisition of vowel discrimination by nonhuman primates,'' J. Acoust. Soc. Am. **84**, 186–194.
House, A. S., and Stevens, K. N. (**1956**). ''Analog studies of the nasalization of vowels,'' J. Speech Hear Disord. **21**, 218–232.
Jackson, M. T., Espy-Wilson, C., and Boyce, S. E. (**2001**). ''Verifying a vocal tract model with a closed side branch,'' J. Acoust. Soc. Am. **109**, 2983–2987.
Kent, R. D. (**1978**). ''Imitation of synthesized vowels by preschool children,'' J. Acoust. Soc. Am. **63**, 1193–1198.
Kent, R. D. (**1979**). ''Isovowel lines for the evaluation of vowel formant structure in speech disorders,'' J. Speech Hear. Disord. **44**, 513–521.
Kojima, S., and Kiritani, S. (**1989**). ''Vocal-auditory functions in the chimpanzee: Vowel perception,'' Int. J. Primatol. **10**, 199–213.
Lieberman, P. H., Klatt, D. H., and Wilson, W. H. (**1969**). ''Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates,'' Science (Washington, DC, U.S.) **164**, 1185–1187.
Maeda, S. (**1993**). ''Acoustics of nasalization and articulatory shifts in French nasal vowels,'' in *Nasals, Nasalization, and the Velum*, edited by M. K. Huffman and R. A. Krakow, Phonetics and Phonology, Vol. 5 (Academic, San Diego), pp. 147–170.
Markel, J. D., and Gray, A. H. (**1976**). *Linear Prediction of Speech* (Springer, New York).
Mergell, P., Fitch, W. T., and Herzel, H. (**1999**). ''Modeling the role of nonhuman vocal membranes in phonation,'' J. Acoust. Soc. Am. **105**, 2020–2028.
Nowicki, S. (**1987**). ''Vocal tract resonances in oscine bird sound production: Evidence from bird songs in a helium atmosphere,'' Nature (London) **325**, 53–55.
Owren, M. J. (**1990a**). ''Acoustic classification of alarm vocalization by vervet monkeys (*Cercopithecus aethiops*),'' J. Acoust. Soc. Am. **101**, 2951–2963.
Owren, M. J. (**1990b**). ''Acoustic classification of alarm vocalization by vervet monkeys (*Cercopithecus aethiops*) and humans (*Homo sapiens*). I. Natural vocalization,'' J. Comp. Psychol. **104**, 20–28.
Owren, M. J. (**1990c**). ''Acoustic classification of alarm vocalization by vervet monkeys and (*Cercopithecus aethiops*) and humans (*Homo sapiens*). II. Synthetic vocalization,'' J. Comp. Psychol. **104**, 29–40.
Owren, M. J., and Bernacki, R. H. (**1988**). ''The acoustic features of vervet monkey alarm vocalization,'' J. Acoust. Soc. Am. **83**, 1927–1935.
Owren, M. J., and Linker, C. D. (**1995**). ''Some analysis techniques that may be useful to acoustic primatologists,'' in *Current Topics in Primate Vocal Communication*, edited by E. Zimmermann, J. Newman, and U. Jürgens (Plenum, New York).
Owren, M. J., and Bernacki, R. H. (**1998**). ''Applying Linear Predictive Coding (LPC) to frequency-spectrum analysis of animal acoustic signals,'' in *Animals Acoustic Communication*, edited by S. L. Hopp, M. J. Owren, and C. S. Evans (Springer, Berlin), pp. 129–162.
Owren, M. J., Seyfarth, R. M., and Cheney, D. L. (**1997**). ''The acoustic features of vowel-like grunt vocalization in chacma baboons (*Papio cyncephalus ursinus*): Implications for production processes and function,'' J. Acoust. Soc. Am. **101**, 2951–2963.
Peterson, G. E., and Barney, H. L. (**1952**). ''Control methods used in a study of vowels,'' J. Acoust. Soc. Am. **24**, 175–184.
Rendall, D., Owren, M. J., and Rodman, P. S. (**1998**). ''The role of vocal tract filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalization,'' J. Acoust. Soc. Am. **103**, 602–614.
Rendall, D., Seyfarth, R. M., Cheney, D. L., and Owren, M. J. (**1999**). ''The meaning and function of grunt variants in baboons,'' Anim. Behav. **57**, 583–592.
Riede, T., and Fitch, T. (**1999**). ''Vocal tract length and acoustics of vocalization in the domestic dog (*Canis familiaris*),'' J. Exp. Biol. **202**, 2859–2867.
Riede, T., Herzel, H., Mehwald, D., Seidner, W., Trumler, E., Böhme, G., and Tembrock, G. (**2000**). ''Nonlinear phenomena in the natural howling of a dog–wolf mix,'' J. Acoust. Soc. Am. **108**, 1435–1442.
Riede, T., and Zuberbühler, K. (**2003**). ''Pulse register phonation in Diana monkey alarm calls,'' J. Acoust. Soc. Am. **113**, 2919–2926.
Robb, M. P., and Saxman, S. H. (**1988**). ''Acoustic observations in young children's non-cry vocalization,'' J. Acoust. Soc. Am. **83**, 1876–1882.
Robb, M. P., Chen, Y. , Gilbert, H. R., and Lerman, J. W. (**2001**). ''Acoustic comparison of vowel articulation in normal and reverse phonation,'' J. Speech, Lang. Hear. Res. **44**, 118–247.
Seyfarth, R. M., and Cheney, D. L. (**2003**). ''Signalers and receivers in animal communication,'' Annu. Rev. Psychol. **54**, 145–173.
Sinnott, J. M. (**1989**). ''Detection and discrimination of synthetic English vowels by Old World monkeys (*Cercopithecus, Macaca*) and humans,'' J. Acoust. Soc. Am. **86**, 557–565.
Sinnott, J. M., and Kreitler, N. (**1993**). ''Differential sensitivity to vowel continua in Old World monkeys (*Macaca*) and humans,'' J. Acoust. Soc. Am. **89**, 2421–2430.
Sinnott, J. M., Brown, C. H., Malik, W. T., and Kressley, R. A. (**1997**). ''A multidimensional scaling analysis of vowel discrimination in humans and monkeys,'' Percept. Psychophys. **59**, 1214–1224.
Sommers, M. S., Moody, D. B., Prosen, C. A., and Stebbins, W. C. (**1992**). ''Formant frequency discrimination by Japanese macaques (*Macaca fuscata*),'' J. Acoust. Soc. Am. **91**, 3499–3510.
Stevens, K. N., Fant, G., and Hawkins, S. (**1987**). ''Some acoustical and perceptual correlates of nasal vowels,'' in *Festschrift for Ilse Lehiste*, edited by R. Channon and L. Shockey (Foris, Dordrecht), pp. 241–245.
Story, B. H., Titze, I. R., and Hoffman, E. A. (**1996**). ''Vocal tract area functions from magnetic resonance imaging,'' J. Acoust. Soc. Am. **100**, 537–554.
Tembrock, G. (**1996**). *Akustische Kommunikation der Säugetiere* (Wissenschaftl. Buchgesell., Darmstadt).
Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ).

van den Berg, J. (**1958**). ''Myoelastic-aerodynamic theory of voice production,'' J. Speech Hear. Res. **1**, 227–244.
Wilden, I., Herzel, H., Peters, G., and Tembrock, G. (**1998**). ''Subharmonics, biphonation, and deterministic chaos in mammal vocalization,'' Bioacoustics **9**, 171–196.
Zuberbühler, K. (**2000a**). ''Referential labeling in Diana monkeys,'' Anim. Behav. **59**, 917–927.
Zuberbühler, K. (**2000b**). ''Causal cognition in a nonhuman primate: Field playback experiments with Diana monkeys,'' Cognition **76**, 195–207.
Zuberbühler, K., Cheney, D. L., and Seyfarth, R. M. (**1999**). ''Conceptual semantics in a nonhuman primate,'' J. Comp. Psychol. **113**, 33–42.
Zuberbühler, K., Jenny, D., and Bshary, R. (**1999**). ''The predator deterrence function of primate alarm vocalization,'' Ethology **105**, 477–490.

The monopulsed nature of sperm whale clicks

Bertel Møhl,[a] Magnus Wahlberg, and Peter T. Madsen[b]
Department of Zoophysiology, Institute of Biological Sciences, University of Aarhus, DK-8000 Aarhus C, Denmark

Anders Heerfordt
Computer Sciences, Corporation Scandinavia, Retortvej 8, DK-1780 V, Copenhagen, Denmark

Anders Lund
Department of Animal Behaviour, Zoological Institute, University of Copenhagen, Tagensvej 16, DK-2200 N, Denmark

(Received 12 August 2002; revised 14 December 2002; accepted 15 April 2003)

Traditionally, sperm whale clicks have been described as multipulsed, long duration, nondirectional signals of moderate intensity and with a spectrum peaking below 10 kHz. Such properties are counterindicative of a sonar function, and quite different from the properties of dolphin sonar clicks. Here, data are presented suggesting that the traditional view of sperm whale clicks is incomplete and derived from off-axis recordings of a highly directional source. A limited number of assumed on-axis clicks were recorded and found to be essentially monopulsed clicks, with durations of 100 μs, with a composite directionality index of 27 dB, with source levels up to 236 dB *re*: 1 μPa (rms), and with centroid frequencies of 15 kHz. Such clicks meet the requirements for long-range biosonar purposes. Data were obtained with a large-aperture, GPS-synchronized array in July 2000 in the Bleik Canyon off Vesterålen, Norway (69°28′ N, 15°40′ E). A total of 14 h of sound recordings was collected from five to ten independent, simultaneously operating recording units. The sound levels measured make sperm whale clicks by far the loudest of sounds recorded from any biological source. On-axis click properties support previous work proposing the nose of sperm whales to operate as a generator of sound.
[DOI: 10.1121/1.1586258]

PACS numbers: 43.80.Ka, 43.64.Tk, 43.30.Vh [WA]

I. INTRODUCTION

Since the first detailed description of the properties of sperm whale clicks by Backus and Schevill (1966), the unique, multipulsed nature of these clicks has been their ''trademark:'' regularly spaced pulses of sound of a few ms duration and with a decreasing amplitude. The interpulse interval (IPI) is on the order of 5 ms, and three or more pulses may be found in a click. Thus, the duration of a click may reach 20 to 30 ms (see Fig. 1).

Early investigators of sperm whale clicks (Backus and Schevill, 1966; Dunn, 1969; Levenson, 1974; Watkins, 1980) reported source levels to be moderate (170–180 dB *re*: 1 μPa), directionality to be low or absent, and the spectrum of the clicks to peak in the 2- to 8-kHz range. However, studies using large-aperture array techniques found source levels between 202 and 223 dB *re*: 1 μPa, a pronounced directionality and spectral emphasis above 10 kHz, as documented in part or all of the following papers: Whitney, 1968; Madsen and Møhl, 2000; Møhl *et al.*, 2000; Thode *et al.*, 2002. Thus, two views on the properties of sperm whale clicks may be said to exist, in the following referred to as the classical view and the large-aperture view, respectively. The present work, using a specially designed array, quantitatively extends the large-aperture view.

The multipulsed nature of sperm whale clicks inspired the dominating theory of sound production mechanics in this whale by Norris and Harvey (1972), explaining the interpulse interval (IPI) of the click by quantitative properties of the nasal anatomy: the length of the spermaceti organ, the velocity of sound in spermaceti, and the distance between the sound-reflecting air sacs at each end of the spermaceti organ. A single pulse, possibly generated by the monkey lips (a massive valve-like structure in the right nasal passage) at the front of the spermaceti organ, was proposed to initiate each click. The multipulsed click was seen as the result of reverberation of the initial pulse between the two air sacs. For each reverberation, as well as from the initial pulse, a part of the sound energy would exit to the water. This was indeed a bold proposal since the nose of sperm whales, responsible for the characteristic, box-shaped appearance of the head, makes up about 1/3 of the total body length, the latter being on the order of 15 meters in adult males. However, the anatomy and basic mechanisms of the supracranial soft anatomy are considered homologous to that found in other, smaller odontocetes (Cranford *et al.*, 1996). Here, a sound-generating function for these tissues is well established.

The Norris and Harvey scheme has recently been supported by sound-transmission experiments within the spermaceti complex (Møhl, 2001; Møhl *et al.*, 2003; Madsen *et al.*, 2003). The scheme is the theory behind acoustic remote sizing of sperm whales, exploiting that the interpulse interval is a function of the length of the spermaceti organ,

[a]Electronic mail: bertel.moehl@biology.au.dk
[b]Present address: The Ocean Alliance/The Whale Conservation Institute, 191 Weston Road, Lincoln, MA 01775.

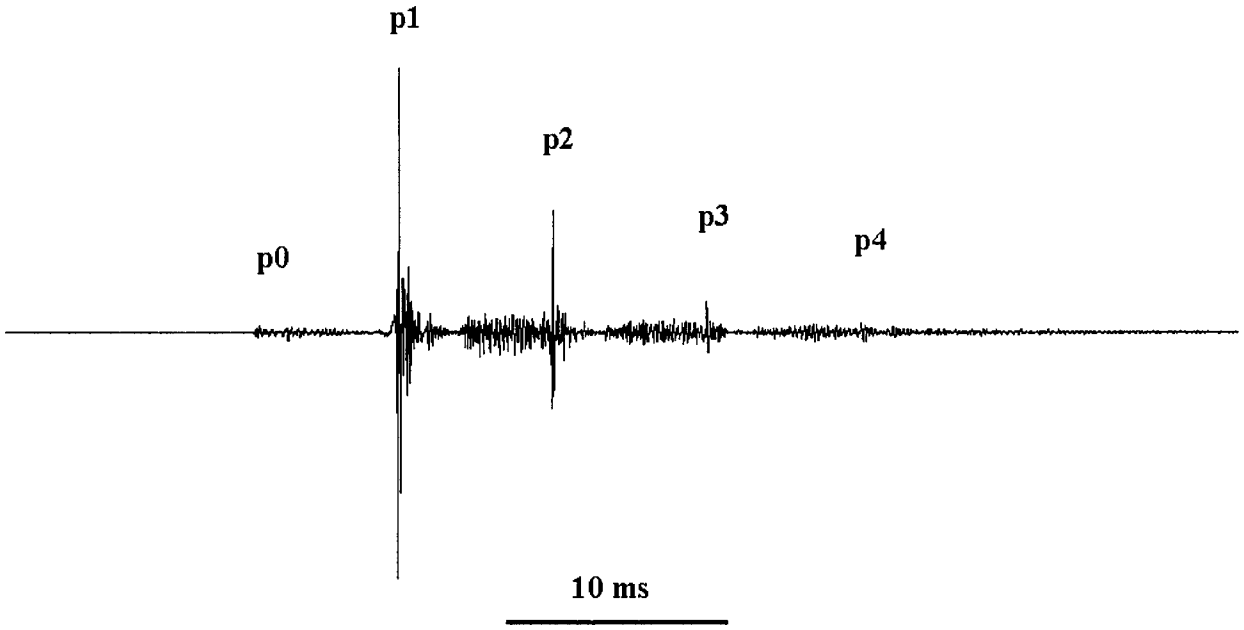


FIG. 1. The classical, multipulse structure of a sperm whale click with pulses labeled from p0 and upwards. The term ''N&H set'' (Norris and Harvey set) is proposed to signify the set of uniformly spaced pulses with decaying amplitude. The p0 pulse has special properties and significance, but is included in the concept of the N&H set. (Surface-reflected pulses of p1 to p3 have been suppressed by editing).

which again is a function of the total body length (Gordon, 1991).

Since the properties of sperm whale clicks, according to the classical view, are quite different from the biosonar clicks known from other odontocetes, it is hardly surprising that little agreement is found in the literature on the function of sperm whale clicks.[1] Norris and Harvey suggested that the multipulsed nature of the clicks would be an advantage for sonar in a cluttered environment, giving the echoes a ''mushy'' appearance. This conjecture seems to be in conflict with accepted theories on sonar (Urick, 1983). Watkins (1980) finds the sonar function less likely, based on the observations of the long duration of clicks, lack of directionality, and his experience that only whales in contact with each other seemed to click. Thus, he favors a communication function of the clicks. In contrast, Goold and Jones (1995) favor the sonar function. They estimate the theoretical detection range of 680 m for a squid target (*Loligo*), based on a specified set of assumptions. Recently, Fristrup and Harbison (2002) pursued the same line of reasoning based on the same (classical) data, but they reached the opposite conclusion, namely that sperm whale clicks are not suited for the detection of cephalopod prey. Finally, Cranford (1999) and Gordon (in Goold and Jones, 1995) see the multipulsed click pattern as a means of signaling size to conspecifics, and conjecture that females prefer mating with males with large IPI.

Our view is that sperm whale usual clicks have a sonar function, and that the multipulsed character of sperm whale clicks is derived from off-axis recordings, which give a distorted representation of the salient properties of the highly directional clicks. Off-axis recordings are not suited for evaluation of the sonar detection range of relevance to the whale. Instead, we hold the trademark of sperm whale clicks (when recorded on axis) to be monopulsed, having an extremely high intensity and directionality. A practical drawback of this trademark is the inherent difficulties to observe on-axis signals. This paper presents evidence for the monopulse click, as well as requirements and methods for obtaining such evidence. We also discuss how this concept relates to the anatomy of the sperm whale nose and to the Norris and Harvey (1972) scheme, as well as the implications of the monopulse click used as a biosonar signal.

II. MATERIAL AND METHODS

A. Environmental conditions

Data were obtained off Andenes, Norway in the period from 12 July to 25 July 2000. Here, an undersea canyon in the continental shelf forms a deep-water gully 18 km from shore. Male, adult sperm whales with a mean length of about 15 m (Wahlberg *et al.*, 1995) abound in this canyon during summer and form the basis of whale safari operations (Ciano and Huerle, 2001), as well as for sound recording operations (Møhl *et al.*, 2000). These waters are part of the Norwegian Coastal Current, running NE at about 1 knot. The velocity of sound decreases gradually with depth from 1477 m/s at the surface to 1468 m/s at 500-m depth, with little further change till the seafloor. The shape of the sound velocity profile agreed with measurements done during previous years (Fig. 2 in Wahlberg *et al.*, 2001).

1. Overview

The recordings were made with an array of up to 10 hydrophones, deployed from 7 vessels (4 yachts, ranging from 12 to 44 tons, and 3 zodiacs). The principles of this array are described in Møhl *et al.* (2001). Basically, each vessel continuously logs its position and time stamps on one track of a DAT recorder, the other track being used for sound recordings. Position and time is obtained from a Garmin 25 GPS receiver, augmented with a dGPS receiver (Magellan

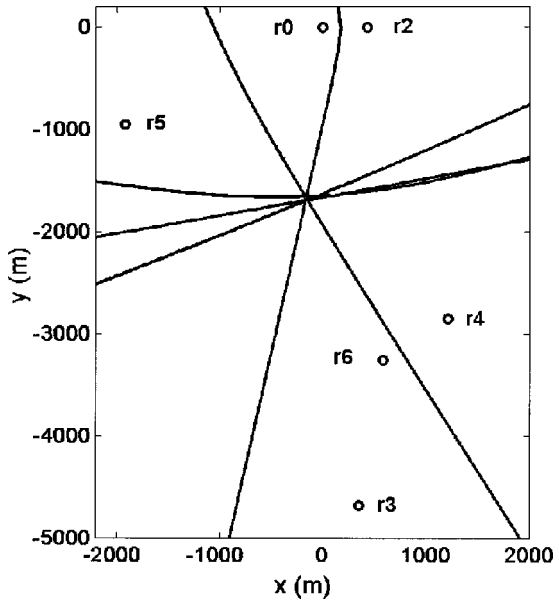


FIG. 2. Example of the recording geometry in 2D format. Relative positions of hydrophones (r0–r6) are marked with circles. Differences in time of arrival of a sperm whale click on pairs of hydrophones are used to generate hyperboloid surfaces, intercepting at the location of the source. Derived sound levels at 1 m from the source of this particular click on hydrophones r3, r6, and r4 were 193, 234, and 185 dB re: 1 μPa peRMS, respectively.

DBR IV beacon receiver), which brings positional uncertainties down to 2.5 m at the 95% level. The position (latitude–longitude format, WGS84 datum), time stamp, and other digital information (see below) are converted to an analog signal by FSK (frequency shift keying) modulation. Each time stamp is in HH:MM:SS UTC format and is recorded as a labeled time marker each second, with a precision limited by the time resolution of the DAT recorder (about 50 μ s). This technique differs from the one used by Møhl *et al.* (2000) by dispensing with synchronization by radio links. Consequently, radio noise is eliminated and the dynamic range is increased. Each node in the array is independent of the other nodes. The other digital information recorded includes tape log information keyed in by the operator, such as positional offset between the hydrophone and the GPS-antenna, sea state, surface activity of whales, attenuator settings, etc., along with the full set of the NMEA sentences generated by the Garmin receiver. The content of the tapes is subsequently transferred to compact disks (CD), preserving the original digitization and linkage between recorded sound, position, time stamps, and tape log information.

With knowledge of sound velocity in water, this array allows for tracking of the sonically active whales in three dimensions, using time-of-arrival information for each click at each hydrophone and the spatial coordinates of the hydrophones (Fig. 2). Distance between source and each hydrophone is computed by methods outlined in Wahlberg *et al.* (2001). The source level at 1 m from the source is calculated for each element of the array, assuming spherical spreading of the sound (see Discussion, Sec. IV) and taking into account frequency-dependent absorption. Since the angular heading of the phonating whale relative to the hydrophone is unknown, the term ASL (apparent source level, see Møhl *et al.*, 2000) is used to emphasize the absence of source heading information.

B. Equipment

Hydrophones were 5 pieces of B&K 8101, 1 piece of Reson 4140, 1 piece of B&K 8105, and 3 pieces of Sonar Products HS150. All hydrophones had a flat (± 2-dB) frequency response in the range from 0.01 to more than 70 kHz. The nominal hydrophone depths were 5, 30, 99, and 327 m. The latter was equipped with a preamplifier, a line driver, and batteries installed in a pressure-resistant canister, connected to the vessel by a steel-armored cable. Calibration was made with B&K 4223 pistonphone calibrators, calibration signals being recorded on every tape at the beginning of each session. The amplifiers were ETEC HA01A, with the high-pass filter set at 0.1 kHz, followed by a two-pole low-pass filter at 11.5 kHz, augmenting the antialias filter of the DAT recorder. The variable level controls of the recorders were clamped in the position of maximum sensitivity, gain being set manually with two-step external attenuators (0 and 40 dB). Recorders were Sony TCD-D3, -D7, -D8, and -D10, sampling at 48 kHz. At one vessel the hydrophone channels were recorded by both DAT recorders and an analog B&K 7606 instrumentation recorder with an upper -6-dB limit at 100 kHz (15 ips), in order to capture possible energy in the sperm whale click above the upper limit (22 kHz) of the DAT recorders.

The positions of the deeper hydrophone were calculated from time-of-arrival measurements by detonators set off at known positions. Sound velocity profiles were obtained with a Star Oddi DST 200 tag, supplemented with hydrographic data from the Institute of Marine Research, Bergen.

C. Analysis

The general approach has been to use the passive sonar equation (Urick, 1983) to derive ASL from received levels, once the position of the whale has been computed from the time-of-arrival differences (TOAD's) *ad modum* Wahlberg *et al.* (2001). The transmission losses (the most influential parameter in the sonar equation) are modeled as spherical spreading plus absorption losses. Sound levels of clicks in the plots below are derived from comparison of the click envelope function with the amplitude of the calibration signal, yielding values in peRMS notation (Møhl *et al.*, 1990). Using this notation, a peRMS measure and a true rms measure of the calibration signal will be identical. With the form factor normally found in sperm whale pulses, a peRMS value of the envelope function is within a few dB of a true rms measure, derived by integration of the waveform over the duration between the -3-dB points of the envelope. True rms values are given for the levels in Table I. The true rms measure is significantly different (yielding lower values) from the peak-to-peak measures, used in most of the literature on odontocete clicks.[2] Energy flux density was calculated using the discrete form of Eq. (11-3) in Au (1993).[3]

For source level estimates, the click had to be identified on more than four platforms, and received level and transmission loss had to be determined. It is a laborious process

TABLE I. Apparent source level, energy flux density, and recording geometry for nine selected clicks of high intensity. Definitions in the text.

Whale ID, time	N[a]	ASL[b] (dB *re*: 1 μPa rms)	Energy (dB *re*: 1 μPa^2 s)	Distance whale–hydrophone (m)	Depth (m)	LEP (m)
A 15_135319[c]	5	236±0.5	196	1779	<600	4
B 15_135431[c]	4	234±0.5	193	1369	<400	40
B 15_135519[c]	5	233±0.6	191	1212	<400	14
C 20_163203	5	231±2	189	1431	1005	228
C 20_163503	5	231+2	190	1009	675	226
C 20_163724	5	226±3	186	866	643	224
D 20_204252	4	229+43/−54[d]	186	254	101	13500
E 21_224509	4+2	229±1	188	1135	639	107
E 21_224829	4+1	228±2	187	787	526	107

[a]Number of receivers. Single number indicates shallow (5–30-m depth) hydrophones, and two numbers indicate shallow and deep (100–400-m) hydrophones.
[b]Root-mean-square intensity over duration restricted by −3 dB *re*: the peak of the click envelope, interpolated with a factor 10 of the sampled data (see Au, 1993). Errors given as 1 s.d.
[c]Array geometry unfavorable for 3D localization. 2D localization is used with depth bounds defined as the seafloor to the surface. These restrictions are incorporated into the calculation of the error in ASL.
[d]Error interval asymmetric due to the logarithmic nature of the transmission loss. The large error for this whale illustrates problems with linear error analysis in certain source-array geometries (cf. Spiesberger and Wahlberg, 2002).

and it is therefore important to select workable click series from the entire material. Accordingly, all tape sequences were scanned manually. Four sets of recordings contained high-level clicks suited for semiautomated click extraction. A property of this approach is that it discards information in the interval between clicks, leaving only some 3% of the original recordings for analysis. Several conditions must be satisfied for automatic click extraction and source location to work, such as good signal-to-noise ratio of the recordings from four or more vessels, absence of other whales and sources of noise, and a suitable geometry between the whale and the set of vessels (Wahlberg *et al.*, 2001). The automated technique was based on a click detector algorithm, which extracted the clicks and aligned them in time. The operator would correct any obvious errors (such as the algorithm detecting the p2 instead of the p1 pulse). Once each click has been timed in the recordings from each vessel, the source position, distance to hydrophones, ASL, etc., can be computed. The purpose of extracting all clicks in a series is that it allows for following signal changes in time and space. The patterns thus obtained are indicative of which parameters are caused by intrinsic properties of the click generator, such as interpulse intervals, and which are consequences of the combination of generator properties with those of the recording geometry, such as directionality and transmission effects. Finally, comparing the development of complete click series at all receivers provides a means for evaluating the consistency of the recordings (e.g., Fig. 3).

III. RESULTS

The main through-going theme of this paper is that the properties of sperm whale clicks differ significantly with as-

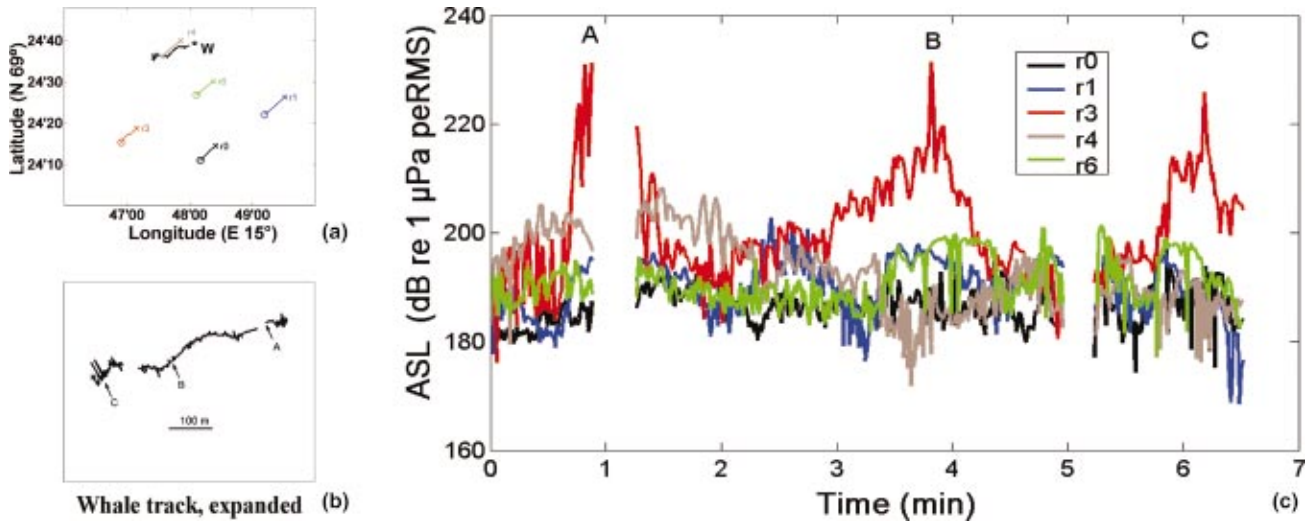


FIG. 3. Acoustic tracking of a sperm whale. (a) Overview of track of whale (westbound surface projected) and array (northeast bound). Hydrophone depths 5–30 m. (b) Whale track expanded. Letters A to C mark whale position during corresponding high-level acoustic events in (c). The small-scale jitter in whale locations is probably caused by uncertainties in the acoustic localization process. (c) ASL for each click at each receiver in the array, using peRMS-notation (see the text).

pect angle. Since direct measures of the heading of the whale are not available, the hydrophone orientation relative to the axis of the animal is in principle unknown. It is assumed that on- and near-on-axis recordings can be distinguished from off-axis recordings by clicks having source levels equal to or above 229 dB *re*: 1 μPa rms. Additionally, on-axis recordings are characterized by p1 pulses being about 40 dB more intense than the remaining pulses of the N&H set, as well as by spectral properties (Sec. III C). This assumption is implicit in the following.

A. Click series

Sperm whale usual clicks are typically emitted in series of tens to hundreds of clicks with a regular or regularly changing repetition rate. A series is defined as limited by silent periods, exceeding the duration of five click intervals in adjacent series (Wahlberg, 2002). Several series may add to form a track. A graphic example of the data from one track with three click series is shown in Fig. 3.

The geometry of recording vessels and whale is given in Fig. 3(a). The vessels are drifting NE with the current, while the whale is moving against the current at a speed through the water of 2 to 3 knots. During the 6 minutes of the track, the whale steadily ascends from a depth of 358 m to 50 m. The detailed, surface-projected 2D track of the whale is in Fig. 3(b), each point of the track representing a position derived from the set of TOADs of a single click at the nodes of the array. The gaps are periods where the whale was silent. The ASL of the click series as seen from the five vessels are plotted in Fig. 3(c), the ordinate given in absolute units, based on calibration, received level, and computed transmission losses. High-level events occur rarely and only at the hydrophone (r3, red) closest to the projected track of the whale. While it would appear that r4 (gray line) in the beginning of the track is on-axis, the r4 levels never get above 210 dB *re*: 1 μPa peRMS. The whale passes this vessel at a depth of some 200 m, and the hydrophone, which is at a depth of 5 m, is likely to be off-axis. The trend of the click rate (not shown) is a small decrease from 1.4 to 1 click/s. The three high-level events in Fig. 3(c) are not accompanied by changes in click rate.

B. Effects of aspect and scan

Directionality of clicks is indicated throughout the recordings by the different appearance of an individual click on the various nodes of the array, as well as in the development over time of a series of clicks, recorded at a single platform. The assumed mechanism for the latter effect is that the whale changes the direction of its sound beam (scanning effect). The difference in waveform of clicks seen from different aspects is illustrated in Fig. 4. The p1 pulse dominates the on-axis signal, and consists of a few cycles (see Fig. 8). The multipulsed pattern is present but not obvious within the dynamic range of a linear plot (Fig. 4, top). In the same click, seen about 20 deg off-axis and amplified (Fig. 4, bottom), the multipulsed pattern is obvious. Notably, the initial event (p0) clearly stands out. The waveform of the off-axis recorded pulses is complicated, consisting of a rather large

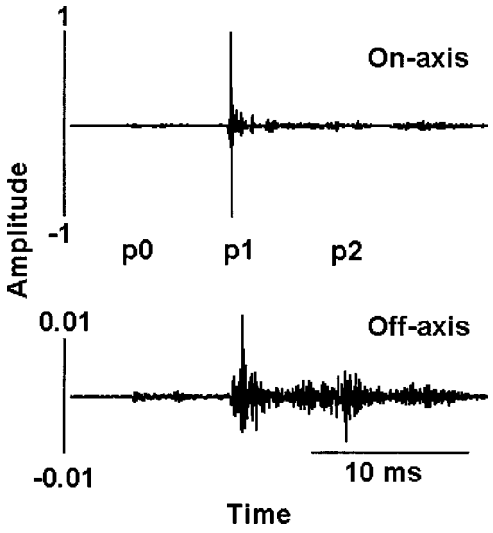


FIG. 4. Same click, recorded at hydrophones r3 (presumably on axis) and r4 (about 20° off axis), obtained 40 s from start of sequence plotted in Fig. 3(c). Gain in lower trace increased by about 40 dB relative to gain in upper trace. The elements of the N&H set are identified.

number of cycles. In between the regularly spaced N&H set considerable click-derived energy occurs, part of which may be surface reflections.

As an example of the scanning effect, the stacked envelopes for a continuous series of 108 clicks, leading up to a presumed on-axis situation, are shown in Fig. 5(a). Note the disproportional increase in p1 intensity towards the end. Also illustrated is the constancy in the timing of the N&H set. In addition to the N&H set, other click-derived pulses occasionally pop up between p1 and p2. In Fig. 5(b), the amplitude of the individual components of the N&H set is given for the same series, relative to a fixed level. It is seen how p1 increases from 12 to about 40 dB above the other pulses, with p2 having a tendency to follow the direction of the changes in p1 amplitude. The p0 amplitude is changing rather little.

Looking at the evolution of the spectra of the p1 pulse of this click series (Fig. 6), it is seen how the spectra of the off-axis clicks before the end of the series have many peaks and notches. The spectra of near-on-axis clicks at the end of the series are smooth, with a broadband appearance and with the frequency of maximum energy shifted upwards.

C. On-axis p1 pulse properties

In on-axis clicks as defined above, only p1 of the N&H set is seen in the waveform displays (Figs. 4, top 7a). The large dynamic range of the envelope function in decibel format reveals the rest of the set, at levels about 40 dB below the p1 pulse (Fig. 7b).

When energy flux density is calculated over the 25-ms time function of the click in Fig. 7(a), 99.6% of the energy is found in the p1 pulse. Energy flux densities for a collection of clicks are given in Table I. For comparison, energy flux densities of clicks from four species of dolphins are lower by 30 to 80 dB (Au, 1993, Table 7.3).

The p1 pulse of an on-axis signal (Fig. 8) has a simple waveform, dominated by a single cycle. The duration of the

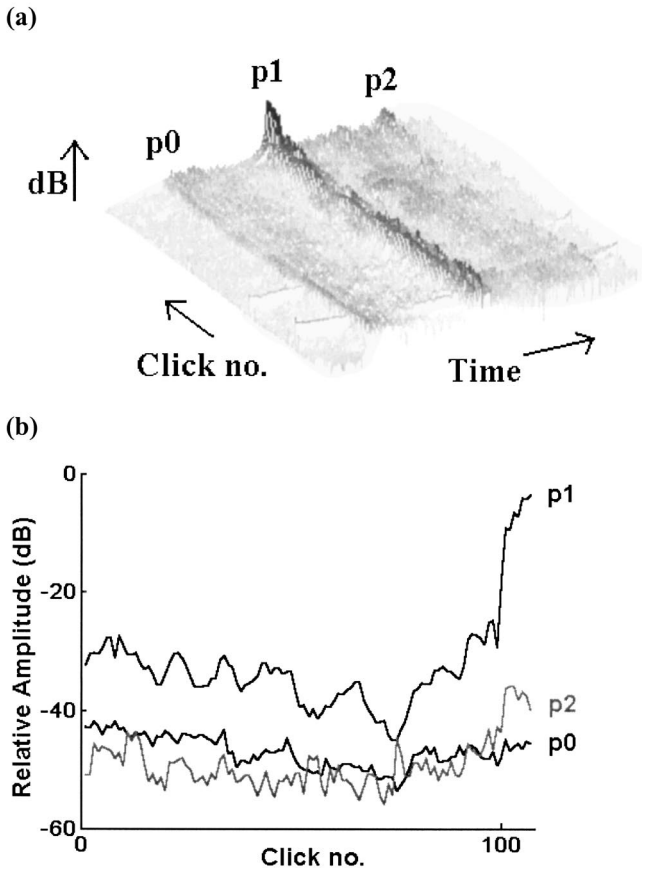


FIG. 5. A series of 108 consecutive sperm whale clicks from a series lasting $1\frac{1}{2}$ min and progressing from an off- to an on-axis condition (last 6 clicks). The whale is approaching from a distance of 1 km and a depth of 0.8 km. Hydrophone depth: 30 m. (a) (waterfall format) shows the log of the envelope functions in 20 ms around each click, aligned by the p1 peak. In (b), the elements of the N&H set for each click have been extracted and the relative amplitude plotted.

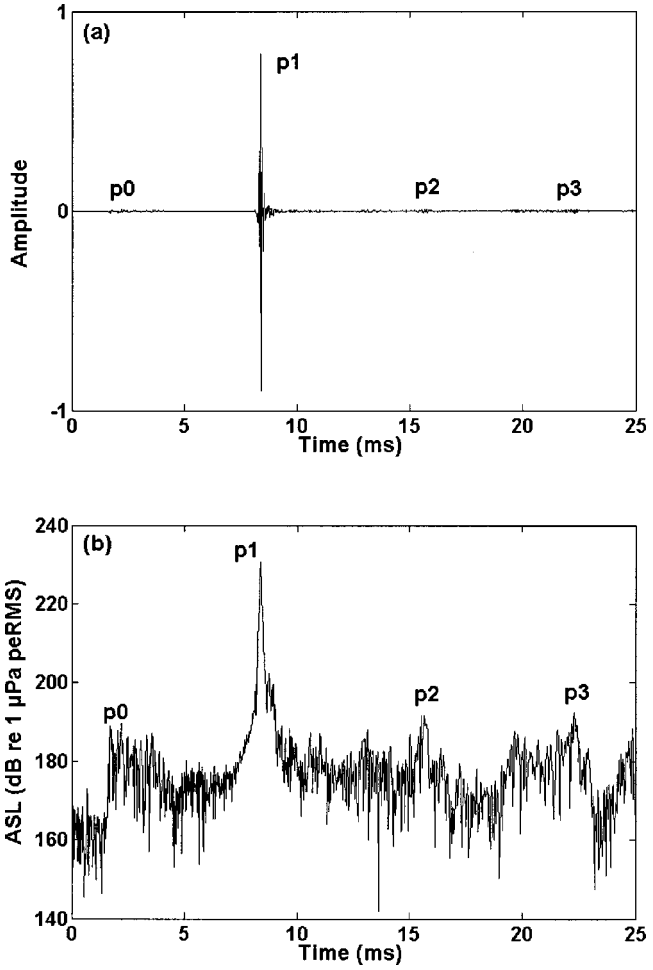


FIG. 7. Time series of an on-axis click (a) in oscillogram format; (b) in *log(envelope)* format. Distance between the whale and the hydrophone is about 1 km.

p1 pulse at the −3-dB limits on the envelope function is 52 μs, 114 μs at the −10-dB limit (Fig. 8).

The Woodward time resolution constant (Au, 1993), defining the theoretical range resolution of a signal for sonar purposes, is 71 μs for the p1 pulse in Fig. 8, corresponding to a range uncertainty of about 11 cm. The product of the time resolution constant and the centralized rms bandwidth (cBWrms; Au, 1993) is 0.29, numerically similar to what is found in *Tursiops* (data from Au, 1993, Table 10.1).

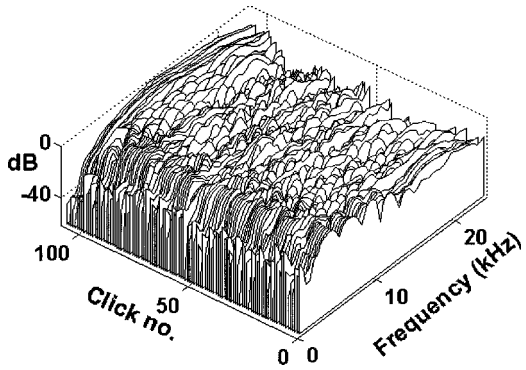


FIG. 6. Spectra of the p1 pulses in the series shown in Fig. 5. Bin width of FFT: 375 Hz; Hanning window.

The spectra of the set of N&H-pulses of this on-axis click are presented in Fig. 9. The p1 spectrum is smooth, peaking around 12 kHz, while the spectra of p2 and p0 have many peaks and notches, possibly indicating a multipath history. As numerical measures of the spectral properties, the centroid frequency (the frequency splitting the spectrum den-

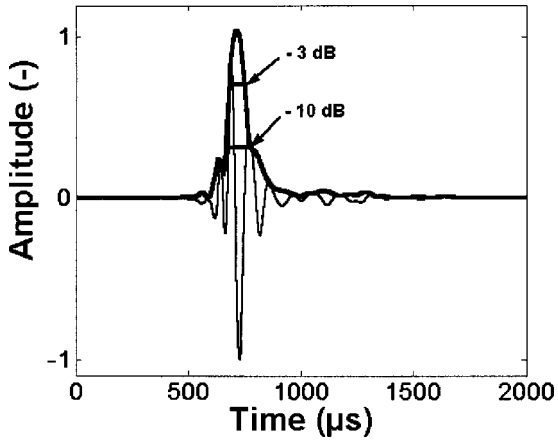


FIG. 8. Waveform and envelope of an on-axis p1 pulse. The −3- and −10-dB limits are added.

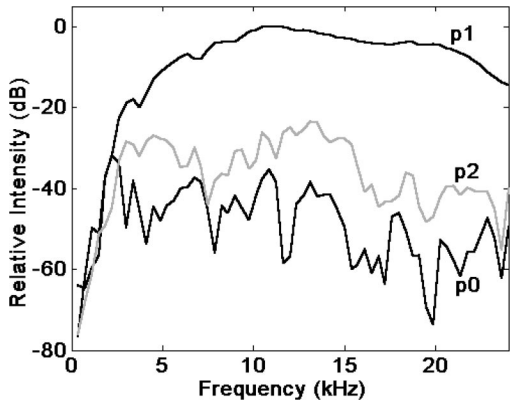


FIG. 9. Spectra of the N&H set of the on-axis click in Fig. 7(a). Bin width of FFT: 375 Hz; Hanning weighting employed.

sity into two equally sized halves; Au, 1993) and the centralized rms bandwidth are given in Table II.

Analysis of on-axis data recorded with the instrumentation recorder shows that the p1 spectrum extends with decreasing magnitude to 40 kHz, above which the signal is masked from noise in the analog recording process.

D. Intensity of on-axis p1

As evidenced by Fig. 3, on-axis events, as defined above, are rare. The maximum SL for any click in our recordings is 236 dB *re*: 1 μPa rms (Table I), with eight other independent events in the range from 226 to 234 dB. In Table I, only the most intense click of a given series is listed. In seven of these events, the whale is on the ascending leg of the dive, with the hydrophones closer to the surface. The general heading of the whales during such events is towards the one hydrophone registering the presumed on-axis event, except in one case, where the hydrophone is rather abeam of the whale's heading. It is noted that the sampling frequency of the position of the whale is dictated by the click rate and probably not coupled to the movements of the whale in a way satisfying the sampling theorem. At the moment of click transmission, the whale may thus point in a direction that is different from the one given by the line between neighboring fixes. For the discussion below, 235 dB *re*: 1 μPa rms is chosen to be representative for the on-axis SL.

In summary, an on-axis click is characterized by its high source level and additionally by having almost all energy contained in the p1 pulse. The spectrum of the p1 pulse has more energy at high frequencies than have off-axis p1 pulses. The waveform and time–bandwidth product of the p1 pulse is similar to that of on-axis clicks from bottlenose dolphins, but the centroid frequency is lower by an order of magnitude.

TABLE II. Spectral properties of N&H pulses of the on-axis click in Fig. 7(a). Definitions in text.

Pulse no.	Centroid frequency (kHz)	cBWrms (kHz)
p0	7.2	5.0
p1	13.4	4.1
p2	10.7	4.2
p3	12.1	3.5

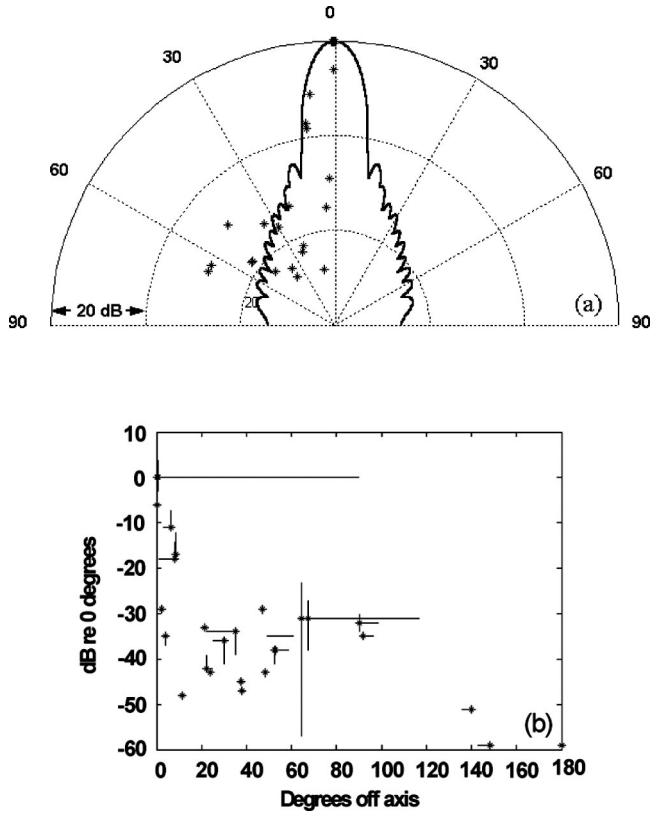


FIG. 10. Composite directionality pattern, based on 6 clicks with on-axis properties recorded by one hydrophone of the array. The radiation pattern is assumed to be rotationally symmetric around the axis of the animal. The dots are plotted as the difference in level and angle from the on-axis recording to recordings from other hydrophones for each click (see Fig. 2 for an example of the recording situation). In (a), the thick line is the theoretical radiation pattern of a circular, 80 cm diameter piston in a baffle, transmitting a p1 pulse as in Fig. 8. (b). 95% error bars, obtained from linear error propagation analysis (see text).

E. Directionality

Since specific information about whereto the whale is directing its beam of sound is unobtainable with the methods used, a radiation pattern in the conventional sense (Au, 1993) cannot be constructed. However, if we assume that a click seen from a hydrophone with ASL$\geqslant$229 dB is on-axis relative to that hydrophone, the aspect angle for each of the other hydrophones can be calculated. If we furthermore assume that the radiation pattern is rotationally symmetric around the acoustic axis, a radiation pattern can be obtained (Fig. 10).

The points plotted in Fig. 10(a) are the data from six clicks in Table I with an ASL of 229 dB or larger (whale D in Table I is excluded due to the large location error). The directional pattern predicted from a plane piston with a diameter of 80 cm (the estimated size of the flat, frontal surface of the junk) emitting the on-axis pulse of Fig. 8, is shown bilaterally (fat line). The half power, half-angle of this function is 4°, and the directional index (DI) is 27 dB.

To illustrate the uncertainties caused by the localization process as determined by linear error propagation analysis (LEP, Wahlberg *et al.*, 2001), the data have been replotted in

Fig. 10(b) with error bars for source level and angle added for each point. The large angular error on axis is caused by a single click (whale B in Table I), where the geometry was unfavorable for localization.

The DI may also be calculated directly from the data set, using the discrete version of Eq. (3.10) in Au (1993)

$$\mathrm{DI} = 10 \log_{10} \frac{2}{\sum_{i=1}^{N} b_i \sin \nu_i \Delta \nu}.$$

In this equation, b_i is the ith sample of an interpolation of the discrete beam pattern; ν is the angle to the acoustic axis; and $\Delta\nu$ is the angular increment between the ith and $i+1$th sample of ν. N is the number of samples, and ν is running from 0 to π radians. This formula assumes that the beam pattern b is rotationally symmetric around the acoustic axis. With the data in Fig. 10(a), the directional index of sperm whale clicks is calculated to be 27 dB, identical to the value derived from the piston model above. However, the precision of this estimate is unknown, due to the variation in the precision of the primary data illustrated in Fig. 10(b).

IV. DISCUSSION

A. Sources of error

The results presented above differ markedly from the classical descriptions of sperm whale clicks (Backus and Schevill, 1966; Levenson, 1974; Watkins, 1980). This is true for the click structure and directionality, as well as for the maximum SL. Differences of 50 to 60 dB in the maximum SL are noteworthy. The results are qualitatively in line with those of three other, large-aperture array recordings (Whitney, 1968; Møhl *et al.*, 2000; Thode *et al.*, 2002). The quantitative differences from the previous large-aperture results are likely a consequence of the low probability of having a single hydrophone in the narrow beam of the whale, combined with differences between the arrays in number of hydrophones, virtual as well as real [surface and bottom reflected signals as used by Thode *et al.* (2002) can be treated as records from virtual hydrophones, mirrored by the surface or the bottom]. The absolute SLs found are significantly higher than the levels reported from any sound-producing species. Since the findings presented are obtained with a novel technique, it will first be discussed if some kind of error could account for such findings.

1. Trivial errors

Trivial errors are for example calibration errors. Equipment and operator malfunctions in the running of a complex setup cannot be totally eliminated, but they can be minimized. The procedure of recording a pistonphone calibrator signal on the tapes of each session, and keeping the digital tape log information about recording conditions inseparable from the sound track, helped to keep trivial errors from proliferating. Redundancy in vessels and hydrophones allowed for omission of the occasional recording that for one reason or another was not acceptable, without falling short of the requirements for acoustic localization. The consistency of the data in Fig. 3, and indeed throughout the data set, is evidence that trivial errors were not common. No recording from any single vessel exhibits a constant bias in received sound level.

2. Localization errors

Errors of a different kind are found in the limited precision of localization. An error in the estimated range between the whale and the hydrophone translates into an error in the estimated ASL. As seen from Table I, column 3, such effects can be quite large, but not large enough to distort the general picture of a number of clicks with ASLs above 230 dB *re*: 1 μPa rms. The localization uncertainty depends on the geometry of the source and the receivers. It is not constant, and therefore it is difficult to apply uncertainty to the DI estimates. The LEP values in Table I and Fig. 10(b) give the composite localization uncertainty for particular clicks. It varies from a few whale lengths to 13.5 km. The LEP values are dominated by uncertainty in the determination of whale depth, a consequence of the majority of the hydrophones being only 5 to 30 m below the surface. Figure 3(b) shows that the click-to-click positional scatter is smaller than the LEP derived uncertainty by about 1 order of magnitude. This is largely explained by the omission of depth information in the 2D plot of Fig. 3(b), as well as by methodological problems with the LEP technique in some source-array geometries (Spiesberger and Wahlberg, 2002).

3. Transmission anomalies

Transmission anomalies have a potential for invalidating the model of spherical spreading. Transmission loss compensation is by far the largest parameter entering the computation of source levels. The sound velocity profile is a fairly simple one (see Sec. II). Ray tracings show some ray bending to be present, resulting in shadow zones for distant whales near the surface, and also creating conditions for caustics (Medwin and Clay, 1998). Such ray bending can change the levels of the received clicks considerably, and may raise questions about the validity of the extreme source levels listed in Table I.

However, the data on relative amplitude changes of the N&H set leading up to a presumed on-axis event should be noted. The rate of increase in levels associated with p1 is neither seen in the other pulses (Fig. 5), nor in off-axis recordings of p1 (Fig. 3c). Any transmission anomaly, such as caustics, would operate equally on all pulses of a set. Another observation is the flattening and expansion of the spectrum towards higher frequencies near on-axis (Fig. 6; also indicated in Møhl *et al.*, 2000, Fig. 7). Caustics will not change the spectral properties of a pulse, only amplitude and phase (Medwin and Clay, 1998). We consequently dismiss caustics as a likely mechanism behind the pattern in Fig. 3(c) and the high levels in Table I. Some other possible effects, such as constructive interference caused by multipath interactions, can be dismissed from the same line of reasoning.

B. Directionality

The combined observations of 1, scarcity of on-axis clicks, 2, their extreme intensity, 3, that high-level clicks are prevailingly recorded in the general course of the whale and

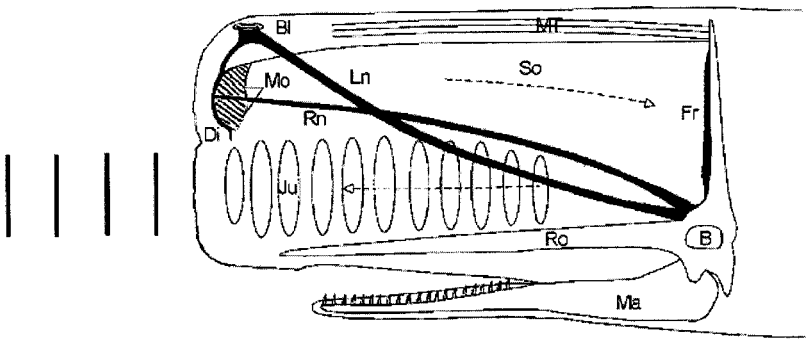


FIG. 11. Diagram of anatomical structures in the sperm whale nose. **B**, brain; **Bl**, blow hole; **Di**, distal air sac; **Fr**, frontal air sac; **Ju**, junk; **Ln**, left naris; **Ma**, mandible; **Mo**, monkey lips; **MT**, muscle-tendon layer; **Ro**, rostrum; **Rn**, right naris; **So**, spermaceti organ. Spermaceti oil is contained in the spermaceti organ and in the spermaceti bodies of the junk. Arrows indicate the assumed sound path from the generating site (**Mo**) back to the reflecting frontal sac (**Fr**) and forward and out through the junk (**Ju**). Sound waves of low divergence are symbolically indicated in front of the whale. (Modified from Madsen *et al.*, 2002a.)

during its ascent (most hydrophones were shallow ones), and 4, that high-level clicks are only recorded by one hydrophone at a time in the set of five or more hydrophones, may all be explained by the high directionality of the clicks. It is emphasized that while Fig. 10 is believed to be the first directionality pattern obtained from any odontocete species in an open ocean environment, it is obtained by a nonstandard method and based on a number of assumptions. The fact that it has been achieved at all was unexpected, since the array was designed to obtain data on source levels, not on directionality.

As the angles between the whale, the ''highlighted'' hydrophone, and the other hydrophones are not controllable, favorable geometries are rare in the material. Particularly, several hydrophones at small angles relative to the highlight direction tend to generate hyperboloid surfaces in the localization calculations that are almost parallel to each other, thereby increasing the positional uncertainty. Another reservation concerns the composite origin of the data, derived from clicks from different whales and recording geometries, and the classification of clicks above or equal to 229 dB *re*: 1 μPa rms as being on-axis clicks. It is consequently not possible to compute a directional index with a meaningful measure of accuracy. Still, the results in Fig. 10 are internally consistent, indicating a half-angle, half-power beam width of about 4°, comparable to that of *Tursiops* (Au, 1993, Table 6.1). Thus, the use by the sperm whale of wavelengths one order of magnitude larger than those of the dolphin is compensated for by other means, of which size is an important factor.

Au *et al.* (1986) show that the radiation pattern of dolphins can reasonably be modeled by radiation from a flat, circular piston with a diameter of 8 cm. Piston modeling of the sperm whale generator is particularly appealing, since the frontal surface of the junk can indeed be flat (Møhl, 2001, Fig. 3) and about 80 cm in diameter in adult males as found off Andenes. Accepting that the supracranial structures of the sperm whale are the generator of sound, and that sound exits the system at the front of the junk (Møhl and Thiele, 1981; Cranford, 1999; Møhl, 2001; Møhl *et al.*, 2003), an increase in aperture relative to that of the dolphin of the same order of magnitude as the increase in wavelength thus seems to be realized. The radiation pattern of the piston model applied to the sperm whale is given in Fig. 10(a). It is seen that this function is generally wider than the function generated by the data points up to about 50 degrees. No mechanism has been identified that could bias the measurements towards smaller, angular estimates. Therefore, it is suggested that the sperm whale's sound-transmitting mechanism could be more sophisticated than what may readily be explained by the theory of the plane piston. This observation may support the bent-horn model (Møhl, 2001; Møhl *et al.*, 2003) that sees the spermaceti organ and junk compartment as two connected tubes, forming a bent, conical horn.

The bend is at the frontal sac (Fig. 11), directing sound generated at the monkey lips (Madsen *et al.*, 2003) from the spermaceti compartment into the junk. A conspicuous feature of the latter is the wafer-like bodies of spermaceti, often referred to as lenses, regularly spaced along the distal part of the junk. It is speculated that these wafers may have a function of adjusting the phase of the p1 pulse over the entire cross section of the horn, producing a flat wavefront at the exit surface. The combined length of this horn is twice the length of the spermaceti organ, or about 10 m in an adult male, and the assumed aperture (the frontal, flat termination of the junk) is about 0.8 m as mentioned above. It is proposed that the function of this horn, as well as the evolutionary drive behind its formation, is the generation of the on-axis, narrow-beam, monopulsed click.

It follows from the procedure for generating the composite radiation pattern presented in Fig. 10 that clicks with ASLs lower than 229 dB *re*: 1 μPa rms will not enter the computation. A consequence of this is that if the whale has control over the width of the beam and an ability to trade SL for beamwidth, such effects will not show up in the data. The same will happen if the whale turns down click amplitude. Control of click amplitude has recently been reported from an experiment with a suction cup fixated hydrophone on a diving sperm whale (Madsen *et al.*, 2002a). Fine control of the anatomical structures in the nasal complex is indicated by the observation by Oelschläger and Kemp (1998) that the trigeminal and facial nerves innervating the nasal musculature in the sperm whale, have three to eight times the count of fibers found in other odontocetes. The muscles and tendons associated with the spermaceti complex are massive and subdivided into small bundles as thick as a finger in adult specimens. Thus, from an anatomical viewpoint it seems likely that the shape of the spermaceti complex can be changed, possibly controlling the beam pattern. However, the whale also requires controlling elements to adjust the type of clicks, their repetition rate and amplitude (Madsen *et al.*, 2002a); hence, the mere presence of controlling nerves and muscles is not exclusively suggestive of a beamforming mechanism.

C. The acoustic properties of the on-axis p1 pulse

Only about one in a thousand clicks recorded was of the on-axis, monopulsed type (Fig. 7a). Thus, the likelihood of recording such clicks with conventional, single hydrophone methods (Backus and Schevill, 1966; Levenson, 1974; Weilgart and Whitehead, 1988; Gordon, 1987) or small aperture arrays (Watkins and Schevill, 1977) is lower than for large-aperture arrays. Another important factor is the often-used strategy of approaching the whales when surfaced and starting the recordings after fluking. This strategy allows for photo identification of the whale, but also increases the likelihood of off-axis recordings, as pointed out by Goold and Jones (1995). In addition to the unfavorable statistics for on-axis recordings, there is the problem of their extreme intensity, requiring a large dynamic range in the recording chain, if the much more abundant off-axis clicks and the N&H set of on-axis clicks are also to be properly recorded. And, the off-axis clicks must be recorded, both for IPI and ICI (interclick interval) determinations, and for TOAD determinations. Our earlier recordings of this species (Møhl and Amundin, 1991; Møhl *et al.*, 2000) had putative on-axis clicks distorted, thus excluding estimation of true ASLs.

All these technicalities add up to an explanation why the classical picture of sperm whale usual clicks is that of multipulsed, low-intensity, nondirectional sounds. The present data suggest that this picture is based on sound energy recorded off the main beam and, therefore, of no relevance to the echolocation capabilities of the species. Realizing that 99.6% of the energy is contained in the p1 pulse, the click is essentially single pulsed, as is the illumination of sonar targets on which the whale may train its beam.

The intensity of a sound with a source level of 235 dB *re*: 1 μPa rms may not readily be appreciated outside the community of underwater acousticians. This is the most intense sound recorded from any animal. Converted to sound pressure of a similar intensity in air, the level corresponds to 173 dB SPL (*re*: 20 μPa). The intensity is 10–14 dB above what can be measured 1 m in front of the muzzle of a powerful rifle. The peak power required for generating an omnidirectional pulse in water with a source level of 235 dB *re*: 1 μPa (rms) is 2 MW (at a conversion efficiency of 100%). Accepting a DI of 27 dB reduces this number to 4 kW, which is still a truly remarkable sound power value, possibly indicating the DI to be underestimated. The megawatt value is given here only to illustrate the consequences if the concept of low or no directionality in sperm whale usual clicks (Watkins, 1980) is applied to the SLs reported in this paper.

The duration of the p1 pulse as measured at the -10-dB limits of the envelope function is about 100 μs. This is slightly longer than the 50–80 μs given for bottlenose dolphins, but in the low end of the range of 50–400 μs for 41 odontocete species (Au, 1993, Table 7.2). Another property of the p1 on-axis pulse shared with echolocation clicks of dolphins (Au, 1980) is the broadening and flattening of the spectrum on-axis relative to off-axis recordings, as shown in Fig. 6. The spectral peaks and notches seen in the off-axis clicks of Fig. 6 are also observed in off-axis recordings of dolphins, were they may be explained by multipath transmission inside the sound production organ (Au, 1993).

D. Sonar properties of the on-axis p1 pulse

Since the monopulsed sperm whale click has properties that are particularly well suited for long-range sonar, this section will explore this putative sonar function. It is recognized that a formal proof of the use of biosonar in any free-ranging species is hard to obtain. This fact should be kept in mind, but not prevent the analysis from being made.

Sonars can be evaluated from the set of sonar equations (Urick, 1983). Using an elaborate form of the sonar equation (the transient form), predictions on dolphin sonar and actual measurements of performance agree quite closely (Au and Penner, 1981; Kastelein *et al.*, 2000). However, the basic sonar equation for the noise-limited case, using intensity terms on dB form, is instructive for illustrating the significance and interaction of the parameters. Specifying the source level (SL), transmission loss (TL, $40 \log r + r\alpha$, α being absorbtion), noise level (N), receiver directionality (DI), and target strength (TS), the detectability (DT) of a given target and range can be predicted (Urick, 1983)

$$\mathrm{DT} = \mathrm{SL} - 2\mathrm{TL} + \mathrm{TS} - \mathrm{NL} + \mathrm{DI}.$$

Using a target strength of a single squid (*Loligo*) of -40 dB (Schmidt, 1954), a transmission loss consisting of the two-way spreading loss plus an absorption of 1.5 dB/km, an SL of 235 dB *re*: 1 μPa, a spectrum level of background noise at sea state 1 of 35 dB *re*: 1 μPa applied over the cBWrms of 4.1 kHz of the p1 pulse around its centroid frequency, and finally assuming a directionality of the ear like that of dolphins (21 dB at 120 kHz; Au and Moore, 1984), a DT for this *Loligo* will be in the order of 20 dB at a range of 1 km. The main factor behind this remarkable DT at such a long range is the source level. The choice of masking bandwidth of the noise has minor effects but does include a hypothesis about the detector (Menne and Hackbart, 1986). Here, an energy detector is assumed. It is noted that the absorption term has a minimal impact due to the relatively low frequencies of the pulse spectrum. If instead a pulse with the spectrum of that of a dolphin (100 kHz) were used *ceteris paribus*, the detectability would be reduced by about 60 dB by absorption at this range.

Another conclusion from applying the noise-limited form of the sonar equation is that noise is unlikely to be the primary limitation of the putative sonar. Instead, clutter or reverberation caused by reflectors other than the target is a likely limiter. Effects of clutter are reduced with increased directionality. The narrow beam suggested by the data in Fig. 10 might be seen in this light. No information about the hearing directionality of sperm whales has been reported.

The classical scatter function of reflection versus wavelength divided by target cross section (Urick, 1983) show a diminishing return for targets with radii of less than 3 cm for p1 pulse signals. Sperm whales are remarkable for eating prey of all sizes from sardines to sharks several meters long (Berzin, 1972). Echoes from the p1 pulses could in theory be reflected from a single sardine without excessive attenuation by entering the Rayleigh scattering region.

Dolphins scan their surroundings with their sonar beam, using click rates on the order of 100 pulses per second. Since the clicking rate of sperm whales for the kind of clicks dealt

with here is in the order of 1 pulse per second, while the directionality of the clicks may be as high as that of dolphins, it follows that the acoustic images of the surroundings obtainable by these two types of biosonar must be quite different (that of the sperm whale being more patchy). However, as pointed out by Dubrovsky (2001, personal communication), a possible analogy might be to vision, where foveal vision is only covering a few degrees, with extrafoveal vision adding coarse information to the general picture. Thus, the low-level sidelobes in Fig. 10 may still convey information to the whale from nearby objects or from objects with high target strengths, although not as detailed as that from the main beam. Based on such ''extrafoveal'' information, the whale might choose to train its narrow beam towards targets of potential interest. Another possibility, which cannot be examined with the present data, is that the directional pattern may not be fixed but controlled by the whale, adapting DI to the detection task at hand. Finally, click series with high repetition rates called creaks have been reported (Weilgart and Whitehead, 1988; Gordon, 1987), and tentatively ascribed a function similar to buzzes from echolocating bats (Madsen *et al.*, 2002b).

In summary, the properties of the on-axis p1 pulse: the high-SL values, low absorption, high directionality, low time-bandwidth product, and geometric backscatter properties for target with radii down to 3 cm are seen individually and combined to be adaptations for maximizing sonar range for prey detection. The slow click rate indicates long range, but also that the angular sampling of the surroundings must be limited. The findings presented here lead to conclusions about the sonar properties of the sperm whale click that are at odds with conclusions of previous work. Since the discrepancies are rooted in the properties of the p1 pulse as described *de novo* here, comparisons to previous assessments are not meaningful.

E. The monopulse click and its relation to the Norris and Harvey theory

The interpretation of the N&H set (Fig. 1) is that the p0 pulse signals the primary event at the monkey lips, transmitted as a leakage directly to the medium. The interval between onset of this pulse and p1 is reportedly less than that of the remaining pulses of the set (Møhl and Amundin, 1991). The main pulse is p1, being shaped by traveling through the spermaceti and junk compartments once. The remaining pulses are stray energy from the p1 pulse, making the two-way travel inside the nasal structures an additional number of times. The view presented above on the intrinsic monopulsed nature of the sperm whale on-axis click and its implications for the way the sound generator works (the bent-horn model) should not be perceived as an alternative to the Norris and Harvey model, but rather as an extension of it. The original Norris and Harvey (1972) model for sound generation in the sperm whale has been successful in explaining the mechanism behind the interpulse intervals. The bent-horn model is a descendant of the Norris and Harvey model, incorporating evidence not available when the original model was formed. Such evidence is (1) the observation of p0 as an indicator of the primary sound generation event (Møhl and Amundin, 1991), (2) the observation that the pulses (except for p0) tend to be in phase with each other (Møhl and Amundin, 1991), as opposed to having even-numbered pulses being phase reversed relative to uneven-numbered pulses as in the analog model constructed by Norris and Harvey; (3) the inferred addition of the junk to the pathway of the sound (Møhl and Thiele, 1981; Cranford, 1999); (4) extreme source levels and high directionality of the clicks (Møhl *et al.*, 2000; Thode *et al.*, 2002; this paper), and (5) the monopulsed nature of the on-axis click. Observation 2 suggests that p0, being the ancestor to the rest of the N&H set, is largely contained in the system. Only a tiny fraction [see Fig. 7(b)] is leaking out directly from the source (the monkey lips; Madsen *et al.*, 2003) to the medium, the distal sac possibly acting as a sound screen. The bulk of the energy is traveling rearwards towards the frontal sac, where it is reflected and directed forward through the junk, exiting as p1 (see Fig. 1). The p2 pulse, again of insignificant amplitude, is proposed to be a fraction of the p1 pulse that does not get into the junk but is returned to the spermaceti organ, then being reflected a second time at the distal sac and subsequently a third time by the frontal sac. Each reflection introduces a phase shift of 180 deg. The higher numbered pulses are repeats of the history of p2. According to this scheme, p1 is phase reversed once, p2 three times, p3 five times. Thus, all pulses will appear as having the same phase. It should be emphasized that this relationship is not seen in all clicks.

The off-axis click properties known to Norris and Harvey were quantitatively quite different from the on-axis properties, likely to be those that matter to the sperm whale. Still, Norris and Harvey came up with a model containing all the essential mechanisms behind click generation in the sperm whale. It is remarkable that their modeling was in fact facilitated by what would now appear to be off-axis signals.

V. CONCLUSIONS

The bent-horn model adds to the collection of quite divergent ideas on the functional anatomy and evolutionary drive behind the development of the nasal complex in sperm whales. Other proposals are: a hydrostatic organ (Clarke, 1970), the single tube, multipulse sound generator (Norris and Harvey, 1972), a nitrogen sink (Schenkkan and Purves, 1973), a device signaling sexual qualities (Gordon, in Goold and Jones, 1995; Cranford, 1999), and a ramming device in male–male fighting (Carrier *et al.*, 2002). While the bent-horn model is based on the Norris and Harvey scheme, it goes further by linking the extraordinary anatomical proportions and complexity of the sperm whale head with the equally extraordinary acoustic properties of the monopulsed click.

ACKNOWLEDGMENTS

We thank the crews, skippers, and owners of MALLEMUKKEN, SIGNE RINK, IOLAIRE, and R/V NARHVALEN for providing the ''backbone'' of the array. We also thank recordists M. F. Christoffersen, O. Damsgaard, L. A. Miller, B. K. Nielsen, M. H. Rasmussen, M. Simon, A. Surlykke, J.

Tougaard, and F. Ugarte. N. U. Kristiansen developed and assembled the electronics. Andenes Whale Safari and Svein Spjelkavik, Hisnakul, Andenes, provided shore base facilities and support. The Danish Research Council, through The Center for Sound Communication, SDU, funded this work.

[1]Several types of sperm whale clicks can be distinguished (Gordon, 1987; Weilgart and Whitehead, 1988). The most abundant type, and the one, with which this paper deals, is called ''usual clicks.'' Additional types are discussed in Madsen *et al.* (2002b).

[2]Au (1993, p. 130) states the pp measure of a *Tursiops* click is 15.5 dB above the true rms measure.

[3]$E=(1/\rho c)\Sigma_i p_i^2 \Delta t$, where ρ is the density, c is the sound velocity, p_i is the ith sample of the sound pressure, and Δt is the sample interval.

Au, W. W. L. (**1980**). ''Echolocation signals of the Atlantic bottlenose dolphin (*Tursiops truncates*) in open waters,'' in *Animal Sonar Systems*, edited by R. G. Busnel and J. F. Fish (Plenum, New York), pp. 251–282.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer, New York).

Au, W. W. L., and Penner, R. H. (**1981**). ''Target detection in noise by echolocating Atlantic bottlenose dolphins,'' J. Acoust. Soc. Am. **70**, 251–282.

Au, W. W. L., and Moore, P. W. L. (**1984**). ''Receiving beam patterns and directivity indices of the Atlantic bottlenose dolphin, *Tursiops truncatus*,'' J. Acoust. Soc. Am. **75**, 255–262.

Au, W. W., Moore, P. W., and Pawloski, D. (**1986**). ''Echolocation transmitting beam of the Atlantic bottlenose dolphin,'' J. Acoust. Soc. Am. **80**, 688–691.

Backus, R. H., and Schevill, W. E. (**1966**). ''Physeter clicks,'' in *Whales, Dolphins, and Porpoises*, edited by K. S. Norris (University of California Press, Berkeley), pp. 510–528.

Berzin, A. A. (**1972**). *The Sperm Whale* (Israel Program for Scientific Translations, Jerusalem).

Carrier, D. R., Debahn, S. M., and Otterstrom, J. (**2002**). ''The face that sunk the Essex: Potential function of the spermaceti organ in aggression,'' J. Exp. Biol. **205**, 1755–1763.

Ciano, J. N., and Huerle, R. (**2001**). ''Photo-identification of sperm whales at Bleik Canyon, Norway,'' Marine Mammal Sci. **17**(1), 175–180.

Clarke, M. R. (**1970**). ''Function of the spermaceti organ of the sperm whale,'' Nature (London) **228**, 873–874.

Cranford, T. W. (**1999**). ''The sperm whale's nose: Sexual selection on a grand scale?'' Marine Mammal Sci. **15**, 1133–1157.

Cranford, T. W., Amundin, M., and Norris, K. S. (**1996**). ''Functional morphology and homology in the odontocete nasal complex: Implications for sound generation,'' J. Morphol. **228**, 223–285.

Dubrovsky, N. (**2001**). Personal communication.

Dunn, J. L. (**1969**). ''Airborne measurements of the characteristics of a sperm whale,'' J. Acoust. Soc. Am. **46**, 1052–1054.

Fristrup, K. M., and Harbison, G. R. (**2002**). ''How do sperm whales catch squids?'' Marine Mammal Sci. **18(1)**, 42–54.

Goold, J. C., and Jones, S. E. (**1995**). ''Time and frequency domain characteristics of sperm whale clicks,'' J. Acoust. Soc. Am. **98**, 1279–1291.

Gordon, J. C. (**1987**). ''The behaviour and ecology of sperm whales off Sri Lanka,'' Ph.D. thesis, University of Cambridge.

Gordon, J. C. (**1991**). ''Evaluation of a method for determining the length of sperm whales (*Physeter catodon*) from their vocalizations,'' J. Zool. **224**, 301–314.

Kastelein, R. A., Au, W. W., and de Haan, D. (**2000**). ''Detection distances of bottom-set gillnets by harbor porpoises (*Phocoena phocoena*) and bottlenose dolphins (*Tursiops truncatus*),'' Mar. Environ. Res. **49**, 359–375.

Levenson, C. (**1974**). ''Source level and bistatic target strength of the sperm whale (*Physeter catodon*) measured from an oceanographic aircraft,'' J. Acoust. Soc. Am. **55**(5), 1100–1103.

Madsen, P. T., and Møhl, B. (**2000**). ''Sperm whales (*Physeter catodon* L.) do not react to sounds from detonators,'' J. Acoust. Soc. Am. **107**, 668–671.

Madsen, P. T., Payne, R., Kristiansen, N. U., Wahlberg, M., Kerr, I., and Møhl, B. (**2002a**). ''Sperm whale sound production studied with ultrasound time/depth-recording tags,'' J. Exp. Biol. **205**, 1899–1906.

Madsen, P. T., Wahlberg, M., and Møhl, B. (**2002b**). ''Male sperm whale (*Physeter catodon*) acoustics in polar waters: Implications for echolocation and communication,'' Behav. Ecol. Sociobiol. **53**, 31–41.

Madsen, P. T., Carder, D. A., Au, W. W. L., Møhl, B., Nachtigall, P. E., and Ridgway, S. H. (**2003**). ''Sound production in neonate sperm whales,'' J. Acoust. Soc. Am. **113**(6), 2988–2991.

Medwin, H., and Clay, C. C. (**1998**). *Fundamentals of Acoustical Oceanography* (Academic, New York).

Menne, D., and Hackbarth, H. (**1986**). ''Accuracy of distance measurement in the bat *Eptesicus fuscus*: Theoretical aspects and computer simulations,'' J. Acoust. Soc. Am. **79**, 386–397.

Møhl, B. (**2001**). ''Sound transmission in the nose of the sperm whale, *Physter catodon*. A *post mortem* study,'' J. Comp. Physiol., A **187**, 335–340.

Møhl, B., and Amundin, M. (**1991**). ''Sperm whale clicks: Pulse interval in clicks from a 21-m specimen,'' in *Sound Production in Odontocetes, with Emphasis on the Harbour Porpoise, Phocoena phocoena*, edited by M. Amundin (Stockholm).

Møhl, B., and Thiele, L. (**1981**). ''Hvaler og støj i havet (Whales and underwater noise),'' Fisk og Hav **39**, 49–55.

Møhl, B., Surlykke, A., and Miller, L. A. (**1990**). ''High intensity narwhal clicks,'' in *Sensory Abilities of Cetaceans*, edited by J. Thomas and R. Kastelein (Plenum, New York), pp. 295–304.

Møhl, B., Wahlberg, M., Madsen, P. T., Miller, L. A., and Surlykke, A. (**2000**). ''Sperm whale clicks: Directionality and source level revisited,'' J. Acoust. Soc. Am. **107**(1), 638–648.

Møhl, B., Wahlberg, M., and Heerfordt, A. (**2001**). ''A GPS-linked array of independent receivers for bioacoustics,'' J. Acoust. Soc. Am. **109**(1), 434–437.

Møhl, B., Madsen, P. T., Wahlberg, M., Au, W. W. L., Nachtigal, P. N., and Ridgway, S. R. (**2003**). ''Sound transmission in the spermaceti complex of a recently expired sperm whale calf,'' ARLO **4**(1), 19–24.

Norris, K. S., and Harvey, G. W. (**1972**). ''A theory for the function of the spermaceti organ of the sperm whale,'' NASA **SP-262**, 397–417.

Oelschläger, H. A., and Kemp, B. (**1998**). ''Ontogenesis of the sperm whale brain,'' J. Comp. Neurol. **399**, 210–228.

Schenkkan, E. J., and Purves, P. E. (**1973**). ''The comparative anatomy of the nasal tract and the function of the spermaceti organ in the *Physeteridae (Mammalia, Odontoceti)*,'' Bijdragen tot de Dierkunde **43**, 93–112.

Schmidt, P. F. (**1954**). ''Further measurements of the sound scattering properties of several marine organisms,'' Deep-Sea Res. **2**, 71–79.

Spiesberger, J. L., and Wahlberg, M. (**2002**). ''Probability density functions for hyperbolic and isodiachronic location,'' J. Acoust. Soc. Am. **112**(6), 3046–3052.

Thode, A., Mellinger, D. K., Stienessen, S., Martinez, A., and Mullin, K. (**2002**). ''Depth-dependent acoustic features of diving sperm whales (*Physeter macrocephalus*) in the Gulf of Mexico,'' J. Acoust. Soc. Am. **112**(1), 308–321.

Urick, R. J. (**1983**). *Principles of Underwater Sound* (McGraw-Hill, New York).

Wahlberg, M. (**2002**). ''The acoustic behaviour of diving sperm whales observed with a hydrophone array,'' J. Exp. Mar. Biol. Ecol. **281**, 53–62.

Wahlberg, M., Lettevall, E., and Medlund, L. (**1995**). ''Estimating the length of sperm whales from interpulse intervals in the clicks,'' Eur. Res. Cetaceans **9**, 38–40.

Wahlberg, M., Møhl, B., and Madsen, P. T. (**2001**). ''Estimating source position accuracy of a large-aperture hydrophone array for bioacoustics,'' J. Acoust. Soc. Am. **109(1)**, 397–406.

Watkins W. A. (**1980**). ''Acoustics and the behavior of sperm whales,'' in *Animal Sonar Systems*, edited by R. G. Busnel and J. F. Fish (Plenum, New York), pp. 283–289.

Watkins, W. A., and Schevill, W. E. (**1977**). ''Spatial distribution of *Physeter catodon* (sperm whales) underwater,'' Deep-Sea Res. **24**, 693–699.

Weilgart, L., and Whitehead, H. (**1988**). ''Distinctive vocalizations from mature male sperm whales (*Physeter macrocephalus*),'' Can. J. Zool. **66**, 1931–1937.

Whitney, W. (**1968**). ''Observations of sperm whale sounds from great depths,'' Marine Physical Laboratory, Scripps Institute of Oceanography No. 1–9. MPL-U-11/68.

Multiecho processing by an echolocating dolphin

Richard A. Altes[a)]
Chirp Corporation, 8248 Sugarman Drive, La Jolla, California 92037

Lois A. Dankiewicz
Science Applications International Corporation, 3990 Old Town Avenue, Suite 208A, San Diego, California 92110

Patrick W. Moore
SSC San Diego, Code 23502, 53560 Hull Street, San Diego, California 92152

David A. Helweg
Pacific Island Ecosystems Research Center, 3190 Maile Way, Room 408, Honolulu, Hawaii 96822

(Received 29 April 2002; revised 18 April 2003; accepted 12 May 2003)

Bottlenose dolphins (*Tursiops truncatus*) use short, wideband pulses for echolocation. Individual waveforms have high-range resolution capability but are relatively insensitive to range rate. Signal-to-noise ratio (SNR) is not greatly improved by pulse compression because each waveform has small time–bandwidth product. The dolphin, however, often uses many pulses to interrogate a target, and could use multipulse processing to combine the resulting echoes. Multipulse processing could mitigate the small SNR improvement from pulse compression, and could greatly improve range-rate estimation, moving target indication, range tracking, and acoustic imaging. All these hypothetical capabilities depend upon the animal's ability to combine multiple echoes for detection and/or estimation. An experiment to test multiecho processing in a dolphin measured detection of a stationary target when the number N of available target echoes was increased, using synthetic echoes. The SNR required for detection decreased as the number of available echoes increased, as expected for multiecho processing. A receiver that sums binary-quantized data samples from multiple echoes closely models the N dependence of the SNR required by the dolphin. Such a receiver has distribution-tolerant (nonparametric) properties that make it robust in environments with nonstationary and/or non-Gaussian noise, such as the pulses created by snapping shrimp. [DOI: 10.1121/1.1590969]

PACS numbers: 43.80.Lb, 43.66.Gf [WA]

I. INTRODUCTION

Active echolocation allows bottlenose dolphins (*Tursiops truncatus*) to investigate their surroundings using hearing (see Au, 1993 for review). Multiple broadband, short-duration acoustic ''clicks'' are emitted by the dolphin. Interaction of the emitted signals with an object causes echoes to return to the animal. Echo characteristics are influenced by the location, orientation, and physical attributes of the object. By listening to these returning echoes, dolphins are able to locate and identify elements in their environment that might be difficult to detect visually.

Because an echo is potentially generated for every click that impinges on an object, the amount of information available to the dolphin increases as more click–echo pairs are produced. Much research has focused on the information contained in the click–echo pair and how it is used by the dolphin (Au, 1993; Au *et al.*, 1988; Busnel and Fish, 1980; Helweg *et al.*, 1996; Nachtigall and Moore, 1988; Thomas and Kastelein, 1990). The manner in which multiple echoes clarify or add information, and how the dolphin utilizes this information, is less clear (Dankiewicz *et al.*, 2002; Moore *et al.*, 1991; Roitblat *et al.*, 1991). Dependence of detection performance upon the number of available echoes has been demonstrated in the big brown bat (Surlykke, 1998).

Several theories exist regarding object detectability as a function of the number of observations available to a receiver. Dating back to the 1950's, several authors have investigated detection of multiple acoustic signals in noise. Green and Swets (1988) proposed two theories to account for the influence of multiple observations on signal detection performance. The first, termed the observation-integration model, assumes that the subject is able to retain information from successive presentations over a certain time period. Detectability is improved as long as the subject is able to successively integrate information from each stimulus presentation. The second model is based on threshold theory, and is comparable to the ''multiple looks'' model of temporal integration (Viemeister and Wakefield, 1991). In this model, each stimulus presentation can independently excite the sensory system. Given that the subject's momentary threshold varies with time, the likelihood of the stimulus exceeding the momentary threshold increases with the number of stimulus presentations.

Data obtained by Swets *et al.* (1964) and Swets and Green (1964) lend support to the integration model, and show that performance generally increases proportionally to the square root of the number of stimulus presentations. In a

[a)]Author to whom correspondence should be addressed. Electronic mail: altes@att.net

study examining the effect of multiple observations on sensory thresholds, Schafer and Shewmaker (1953) also found that thresholds decreased in proportion to the square root of the number of presentations. The integration model implies that the detectability index of a set of N presentations equals the square root of the sum-of-squares of the detectability indices for the individual presentations (Green and Swets, 1988). If the detectability indices for the individual presentations are identical, then the detectability index of a set of N presentations equals $\sqrt{N}$ times the detectability index of a single presentation. The $\sqrt{N}$ dependence follows from the definition of the detectability index, as given in the Appendix. Although many different integration models are possible (e.g., linear summation, energy detection, and binary summation), all such models have detectability indices that vary as $\sqrt{N}$.

Multiecho combining is relevant to many sonar capabilities, e.g., range-rate estimation and moving target indication (MTI) with short-duration, *Tursiops*-like waveforms, target tracking, and acoustic imaging in two or three dimensions. A logical step to investigate such capabilities in dolphins is to perform a critical experiment that ascertains whether the dolphin is capable of the simplest echo-combining task, which is to use multiple echoes from a stationary target to improve detection performance. If an accurate receiver model can be found, i.e., a model that accurately describes the dolphin's N-echo stationary-target detection performance, then this model may be applicable to more sophisticated dolphin echo-combining operations.

The current study was thus designed (i) to test the hypothesis that dolphins combine echoes to improve signal detectability; and (ii) to find the best receiver model to describe the dolphin's performance. A dolphin was trained to report detection of synthetic echoes generated by computer in response to the dolphin's clicks, placing the number of echoes available to the dolphin under experimental control. The dolphin's signal-detection performance was assessed when 1,2,4,8, and 16 echoes were made available. Although the available number of echoes (N) was preset, the dolphin's click emission rate was not controlled. The number of emitted dolphin clicks thus could be much larger than the number N of available echoes. During a test session, half of the trials contained synthetic echoes in noise and half contained noise only. Echo amplitudes were systematically decreased until detection fell to chance. At least two such thresholds were taken at each N level.

The results are summarized by plotting the signal-to-noise ratio (SNR) required for detection as a function of the number of available synthetic echoes that could be used by the dolphin. This experimental function is compared to the theoretical detection performance of three receiver models operating in additive, white, zero-mean Gaussian noise. All the receiver models initially are assumed to operate on one time sample (or range sample) from each available simulated echo, yielding N time samples altogether, where N is the number of available simulated echoes. The three models are linear summation, square-law summation (energy detection), and summation of binary-quantized sample values (binary M-out-of-N detection).

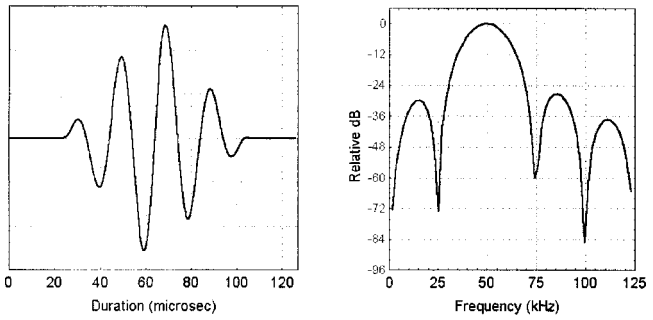


FIG. 1. Enlargement of GO stimulus pulse and corresponding spectrum. When more than one pulse was allowed per trial, pulse separation was constrained to be no smaller than 8 ms.

II. METHODS

A. Subject

The subject was a 17-year-old female Atlantic bottlenose dolphin (*Tursiops truncatus*; ''CAS''). Floating pen enclosures on San Diego Bay, Space and Naval Warfare Systems Center were utilized for housing and experimental sessions. The subject resided with a small dolphin group but was separated from them when sessions were conducted. CAS' hearing was measured previously and shown to be normal (Brill *et al.*, 2001).

B. Synthetic echoes and noise

Conditions for behavioral responding were contingent upon two types of computer-generated stimuli. The NO-GO stimulus consisted of Gaussian noise with flat power spectrum over the echolocation bandwidth of the dolphin (95 dB SPL re: 1 μPa^2/Hz between 10 and 150 kHz). This white noise was present for the 4-s trial duration. The ambient noise in San Diego Bay had a power spectrum level decreasing from approximately 80 dB re: 1 μPa^2/Hz at 10 kHz to approximately 60 dB re: 1 μPa^2/Hz at 100 kHz, measured with one-octave spectrum analysis filters. The ambient noise level has increased with time and is thus larger than the level reported in Au (1993). The directivity of the dolphin's receiver (Au, 1993) further reduced the effective ambient noise level relative to the NO-GO stimulus. At 50 kHz, the ambient noise level was approximately 70 dB re: 1 μPa^2/Hz. This ambient noise level was 25 dB below the NO-GO stimulus level, and was 38 dB below the NO-GO stimulus level when the dolphin's directivity index is considered.

The GO stimulus included 1,2,4,8,16,32, or 64 pulses embedded in the white noise. The 32- and 64-pulse conditions were utilized during training phase sessions only. The number of pulses for GO stimulus trials did not vary within a session. Each pulse was a triangle-windowed 50-kHz 80-μs sinusoid (Fig. 1) delivered in response to the dolphin's outgoing echolocation click. An 8-ms click-pulse delay was inserted to simulate a 6-m range. The total range of the artificial echoes was 7m, counting propagation time between the transducers and the dolphin. Although the dolphin's click emission rate was not experimentally controlled, it was influenced by this imposed range parameter (Penner, 1988). Pulse source level was manipulated to determine CAS' de-

tection thresholds above the noise floor. No attempt was made to associate or equate the artificial stimuli with echoes encountered under nonexperimental conditions.

C. Apparatus

Synthetic echoes were generated and delivered by an electronic synthetic echo system (SES). One electronic echo was delivered for every click emitted by the subject up to the maximum N allowed for that session per trial. The dolphin's clicks were detected by a Reson TC4013 hydrophone located 0.64 m directly in front of her melon and triggered a single electronic echo if the received level exceeded 170 dB re: 1 μPa. Clicks were bandpass filtered (3–300 kHz) and amplified by 54 dB before reaching a multifunction board (National Instruments PCI MIO-16E-1; on Pentium PC) where triggering of synthetic echoes previously stored to RAM occurred. Upon receiving a trigger, the SES converted a digital waveform to an analog signal that was then filtered (10–200 kHz) and amplified (20 dB) by a DL Electronics 4302 filter/amplifier. Analog echoes were added to the white noise using custom hardware and projected to the subject by a second TC4013 hydrophone located 0.7 m beyond the trigger hydrophone. The echo stimulus thus emanated from a transducer that was 1.34 m from the dolphin's melon, located on a horizontal line directly in front of the melon. A 7-m echo range was simulated by insertion of an 8-ms delay between the trigger event and output of an echo (12-m electronic delay plus 2-m propagation delay, divided by 2). System calibration included SPL measurements of TC4013 electronic echo projection by an ITC 6030 omnidirectional hydrophone located at the subject's test station position. Surface reflections were absorbed and dispersed by a cluster of nylon-bristle brushes placed at the water surface midway between the dolphin and the transducers.

D. Session procedure

CAS was positioned at an intertrial station in front of an experimenter before a trial. At the start of each trial, CAS was cued to submerge into a test station hoop 1.35 m below the water surface by the experimenter's hand gesture. An acoustically opaque screen (a PVC sheet covered with closed-cell foam neoprene) placed in front of the hoop was removed and the SES simultaneously activated, initiating white noise and permitting the dolphin to begin echolocating. The 4-s white-noise burst defined the trial duration for the dolphin. To report a signal-present condition (GO response), the dolphin immediately moved to a nearby paddle and touched it with her rostrum. To indicate the absence of a signal (NO-GO response), she remained stationary in the hoop for the trial duration (4 s). If CAS did not begin movement toward the paddle before the end of the 4-s window, her response was classified as NO-GO. CAS typically initiated a GO response within 1–2 s. Tone and fish rewards were given for every correct response. An equal number of GO and NO-GO trials was presented in a randomized Gellermann series (Gellermann, 1933). The likelihood of a GO following a NO-GO (or the reverse) followed a 0.5 first-order conditional probability for every ten-trial block. The dolphin's motivation to perform reliably was assessed by ten warm-up trials before every session, with an 80%-correct response rate required in order for a test session to ensue. No more than one experimental session was conducted in a day.

E. Threshold titration

Thresholds were estimated for both training and testing phases by using a signal amplitude titration method (up/down staircase) that was contingent upon the dolphin's responses to GO stimulus trials (Moore and Schusterman, 1987). During the sessions, the experimenter manipulated SPL by changing the voltage value of the synthetic echo amplitude. Initially, GO signal amplitude was held constant and easily discernible for the first ten trials of the session. After the first ten trials, 0.2-V decrements in signal amplitude were made until the dolphin responded incorrectly. Amplitude was raised in 0.1-V increments until the dolphin detected the signal again. All subsequent amplitude adjustments were in 0.1-V steps, with decrements made after every correct GO response and increases after every incorrect response. A change in direction of amplitude adjustment constituted a reversal, and a threshold estimate was taken after ten reversals were acquired by calculating the mean decibel level at those reversal points (50%-correct detection rate). As CAS became experienced with the task, and echo amplitudes were close to the white-noise floor, the titration deltas were changed to 0.05 V. Logarithmic steps (constant $\Delta v/v$) are more compatible with an animal's sensitivity to differences than constant Δv steps, but constant steps are approximately proportional to logarithmic steps when the steps are small relative to the threshold level ($\Delta v \ll v$).

F. Animal training

Training the stimulus-response contingency was accomplished by imposing minimal restrictions on the GO stimulus variables in an effort to highlight differences from the NO-GO stimulus. A generous number of synthetic echoes were provided ($N=256$) and signal amplitude was held approximately 40 dB above the noise floor (1.0 V). Stimuli were presented in three to six same-trial blocks, e.g., four NO-GO trials followed by six GO trials. Approximately four sessions were required before the appropriate responding was observed. Trial type was then randomized as described previously. Once responding was stable, CAS was introduced to a reduction in N. One to eight sessions (s) were conducted at successively lower N levels as follows: $N=64$ ($s=8$); $N=32$ ($s=3$); $N=16$ ($s=5$); $N=8$ ($s=1$); $N=4$ ($s=5$); $N=2$ ($s=2$); and $N=1$ ($s=3$). Thresholds were estimated during the final three sessions at $N=64$, 16, 4, and 1, and during the final session only at $N=32$, 8, and 2. Once CAS demonstrated stable performance at the minimum N level ($N=1$), as indicated by the threshold session results at $N=1$, no further training was undertaken.

G. Testing

Exposure to all experimental conditions was completed in the training phase so that testing-phase thresholds were free of novelty effects. Two ($N=1,4,8,16$) or three ($N=2$)

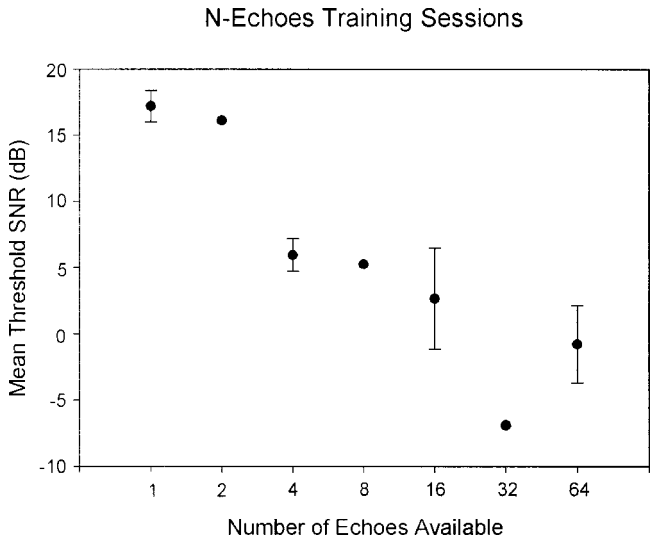


FIG. 2. Average detection threshold in decibels obtained during training sessions when the number of available echoes per trial (N) was 1, 4, 16, 64 (number of measured thresholds = 3) or 2, 8, 32 (number of measured thresholds = 1). Error bars for the trials with three measured thresholds represent SEM (standard error of the mean).

final thresholds were obtained in which signal detection performance as a function of N was assessed. N was held constant during a session while signal amplitude was titrated as described previously.

H. Calculating thresholds

Recall that the experimenter manipulated SPL by changing the voltage value of the synthetic echo amplitude. Synthesized white, Gaussian noise was held constant at 95 dB re: 1 μPa^2/Hz. Thresholds were computed as signal-to-noise ratios (SNR), the required echo amplitudes $A(N)$ for detection with N available echoes, divided by the rms noise power. The bandwidth for rms noise power was estimated using Q derived from critical band measures of the bottlenose dolphin receiver. Q was approximately 2.2 for signals with center frequency of 60 kHz (Au and Moore, 1990). The synthetic signals used in this study had center frequency of 50 kHz; thus, noise bandwidth was estimated to be approximately 22.72 kHz. The calibrated system permitted conversion of the voltages $A(N)$ to dB, thereby allowing computation of SNR in dB by subtracting rms noise power (dB) from synthetic echo amplitude $A(N)$ (dB). Importantly, note that SNR was computed *per echo*, without weighting for the number of available echoes N.

III. RESULTS

A. Animal training

Figure 2 shows results of the initial detection threshold sessions that were conducted at each N level. N is presented on the horizontal axis, with sessions represented left-to-right in the opposite order in which they were conducted. Detection performance was strong for N values of 64, 32, and 16, although sporadic threshold elevations were seen. It is likely that these variations represent CAS' growing familiarization with the manipulation of N while the thresholds were being

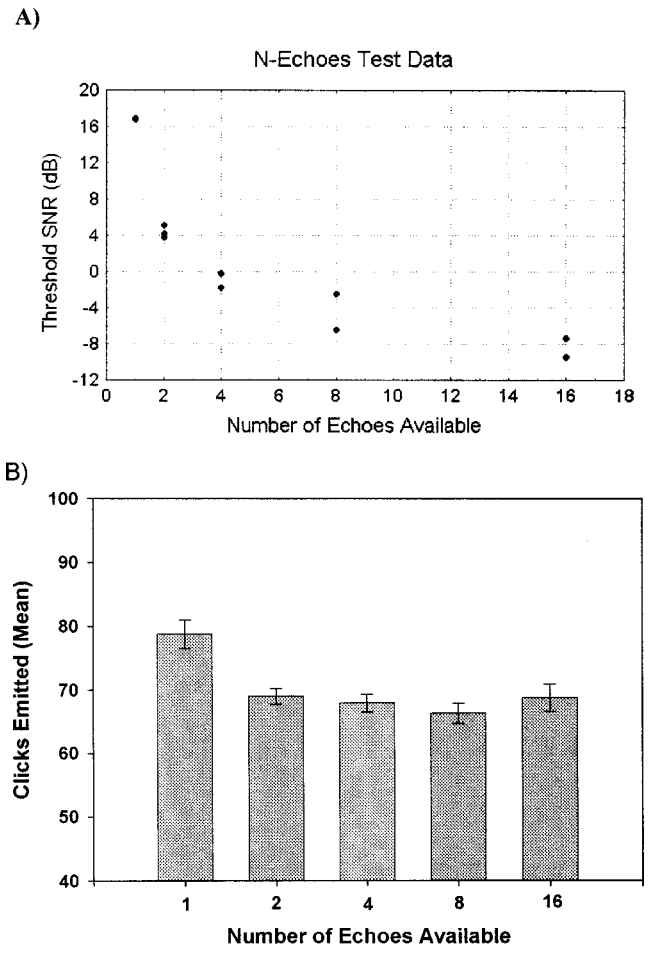


FIG. 3. Test session data: (A) Final detection thresholds in decibels when the number of available echoes per trial (N) was 1, 4, 8, 16 (number of measured thresholds = 2) or 2 (number of measured thresholds = 3). The two measured thresholds at $N = 1$ were equal and the dots at $N = 1$ therefore overlap. (B) Average number of clicks emitted per trial at each N level (pooled sessions). Error bars represent SEM. Significantly more clicks were emitted at $N = 1$ compared to all other levels (Tukey–Kramer, $\alpha = 0.05$).

titrated. At N values of 8 and 4, SNR required for detection increased, and was highest when N was held at 2 and 1. The mean false-alarm rate for the threshold sessions was 0.088 (s.d. = 0.069). Click emission was tracked for every trial and results showed that CAS always emitted enough clicks to receive the maximum number of echoes that were allowed (mean clicks per trial = 80).

B. Threshold testing

The top panel of Fig. 3 summarizes CAS' detection thresholds that were estimated during the test phase of the experiment. N is presented on the horizontal axis, with successive sessions represented in left-to-right order. The two estimated threshold SNR values for $N = 1$ were identical, and are thus represented by a single point in Fig. 3. Detection thresholds were lower overall at each N level when compared to the training session thresholds, perhaps due to increased familiarization with the task. Mean false-alarm rate was 0.034 (s.d. = 0.042). The thresholds are well behaved, with SNR required for detection falling off monotonically as the number of echoes (N) is increased. Recall that SNR is

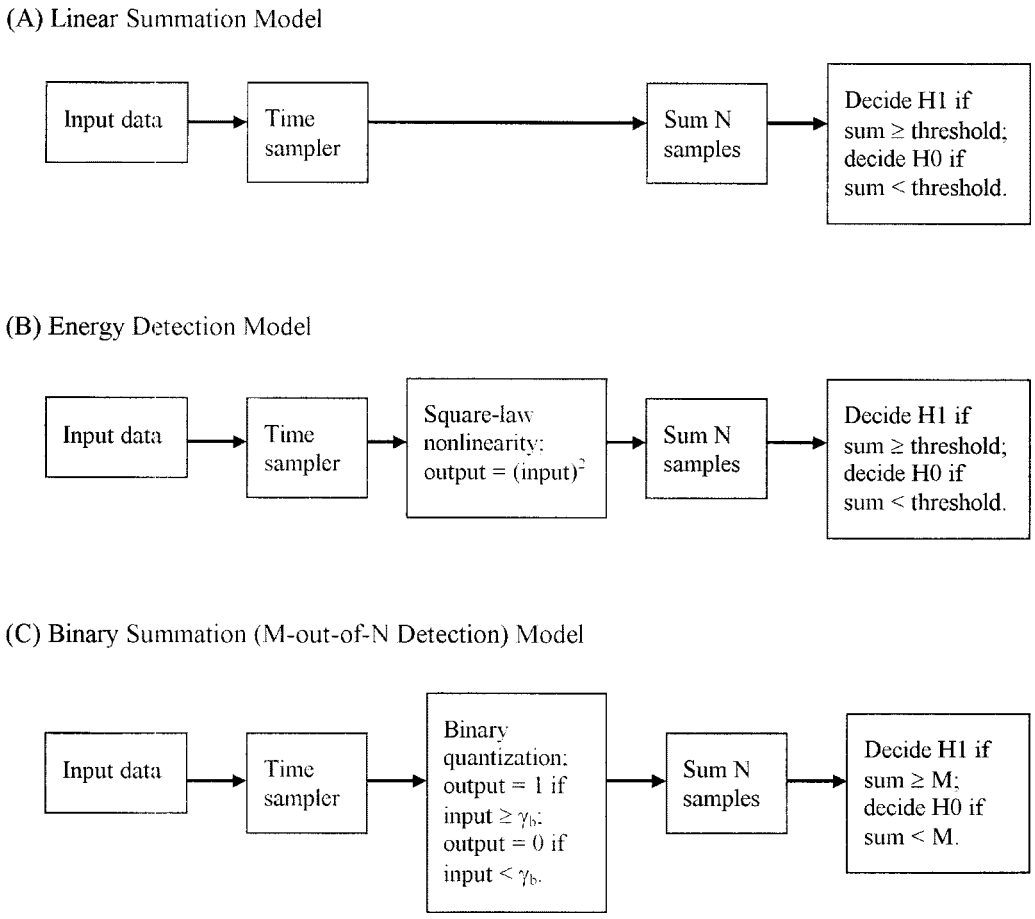


FIG. 4. Receiver models that are compared with dolphin N-echo detection data. (A) linear summation; (B) energy detection; (C) binary summation (M-out-of-N detection). H1 is the signal-plus-noise hypothesis corresponding to the GO stimulus. H0 is the noise-only hypothesis corresponding to the NO-GO stimulus.

measured per echo. The trend is consistent with those from human listeners showing that detection improved as the number of signals was increased (Green and Swets, 1988; Swets and Green, 1964). These results support the inference that CAS was able to combine multiple echoes in her biological signal-processing system.

The lower panel of Fig. 3 summarizes the distribution of clicks emitted by CAS that were above 170 dB during each session. Analysis by a one-way ANOVA showed a difference in click production as a function of N, $F(4,524)=5.4$, $p<0.0003$. Comparison among the means using the Tukey–Kramer test revealed that CAS emitted significantly more clicks for the $N=1$ condition than all the others ($\alpha=0.05$, two-tail), supporting the notion that this condition was more difficult than $N=2$, 4, 8, and 16. Further evidence for difficulty at $N=1$ is that the change in threshold for the $N=1$ condition compared to the $N=2$ condition was 12 dB, whereas the change in threshold between the other conditions ($N=2$ through $N=16$) was almost a consistent 4-dB change. The mean number of clicks emitted during all testing sessions was 70, only a slight decrease from the average number emitted during training sessions. CAS always emitted more clicks than echoes that were available, thus ensuring that she received all available synthetic echoes.

C. Receiver models

Various models of animal echolocation have been employed to understand the signal-processing operations that may be used by the animals and to guide the design of broadband sonar systems that attempt to emulate animal capabilities. Receiver models that incorporate summation or integration are relevant to this inquiry. Well-known integration models pertain to summation over intervals in range/delay/time (critical intervals; Vel'min and Dubrovskiy, 1976) and over intervals in frequency (critical bands; Johnson, 1968). The synthetic echo in Fig. 1 fits within a single critical interval and a single critical band for a critical bandwidth of 22.72 kHz at a frequency of 50 kHz (Au and Moore, 1990). This study addresses integration along a different dimension, correspnding to the number of click–echo pairs (Floyd, 1980; Surlykke, 2003). A critical N value, corresponding to the maximum number of echoes that can be integrated by the dolphin, has yet to be determined. Figure 3 implies that the critical N is greater than 16.

Three integration models are considered here in order to better understand the SNR required by a dolphin for target detection when the number of available echoes is varied. These models correspond to linear summation, energy detection, and summation of binary-quantized echo data (binary M-out-of-N detection). The operations performed by the three models are illustrated in Fig. 4.

The N-echo detection performance of the three receivers is predicted by the analysis in the Appendix. For the linear summation, energy detection, and binary M-out-of-N receivers, the required echo amplitudes $A(N)$ for detection with N

available echoes, divided by the rms noise power σ, are

$$[A(N)/\sigma]_{\text{lin}} = c_l/\sqrt{N}, \tag{1}$$

$$[A(N)/\sigma]_{\text{egy}} = (c_e/\sqrt{N})[1+\sqrt{1+(2N/c_e^2)}]^{1/2}, \tag{2}$$

$$[A(N)/\sigma]_{\text{bin}} = \text{erfc}_*^{-1}(p_0) - \text{erfc}_*^{-1}(p_1), \tag{3}$$

where c_l and c_e are constants, $\text{erfc}_*^{-1}(\bullet)$ is the inverse complementary error function (Fig. 5), p_0 is the probability that the threshold level of the binary quantizer is exceeded when only noise is present (the H0 hypothesis), and p_1 is the probability that the binary quantizer threshold is exceeded when both signal and noise are present (the H1 hypothesis). The function $\text{erfc}_*^{-1}(p)$ is related to the probit transformation (Collett, 1952),

$$\text{erfc}_*^{-1}(p) = \text{probit}(1-p). \tag{4}$$

In the binary summation model, probability p_1 depends upon p_0, N, and a constant c_b

$$p_1 = \frac{(2p_0 + c_b/N) + \sqrt{(2p_0 + c_b/N)^2 - 4p_0(1+c_b/N)[p_0(1+c_b/N) - c_b/N]}}{2(1+c_b/N)}. \tag{5}$$

For a prespecified value of p_0, all the $A(N)/\sigma$ expressions in (1)–(3) depend on a constant (c_l, c_e, or c_b) and on the number of available echoes, N.

D. Comparison of theoretical performance with dolphin data

To compare the receiver models with dolphin detection data, the parameters c_l, c_e, p_0, and c_b are adjusted to provide a minimum mean-square error (MMSE) fit between the values of $[A(N)/\sigma]_{\text{model}}$ in (1)–(4) and the average experimental value of $A(N)/\sigma$ for each N value. The correlation coefficients between the average data points and their theoretical counterparts are computed for each model. For visual comparison, the data points and theoretical curves are plotted together in Fig. 6 on a decibel scale, showing $20\log_{10}[A(N)/\sigma]$ vs N. Figure 6 illustrates that the best fit (by far) is obtained with the binary M-out-of-N receiver.

The best-fit parameters and data-model correlation coefficients r are as follows:

Linear summation model: $c_l = 4.58$, $r_l = 0.9055$, (6)

Energy detection model: $c_e = 3.12$, $r_e = 0.9095$, (7)

Binary M-out-of-N detection:

$$p_0 = 0.5,\ c_b = 0.999\,999\,3,\ r_b = 0.9997. \tag{8}$$

The more accurate specification of c_b is necessitated by receiver operation on a steep part of the curve in Fig. 5 when $N=1$, as discussed in the Appendix.

The binary M-out-of-N detector seems to have an unfair advantage because two parameters can be varied instead of one, providing an extra degree of freedom for data fitting. The extra degree of freedom is eliminated by choosing a prior value for p_0. Choosing a fixed p_0 value is equivalent to choosing a threshold for binary quantization. The most appropriate prior choice for the binary quantization threshold is zero, which implies that $p_0 = 0.5$ for all symmetric, zero-mean noise distributions, independent of the noise power σ^2.

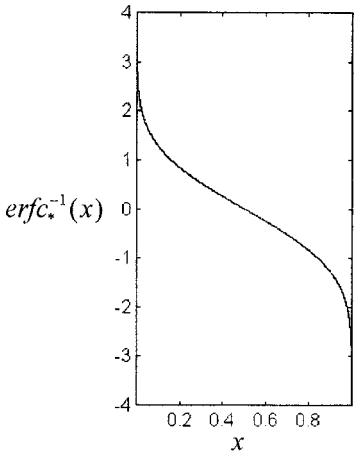


FIG. 5. The inverse complementary error function $\text{erfc}_*^{-1}(x)$.

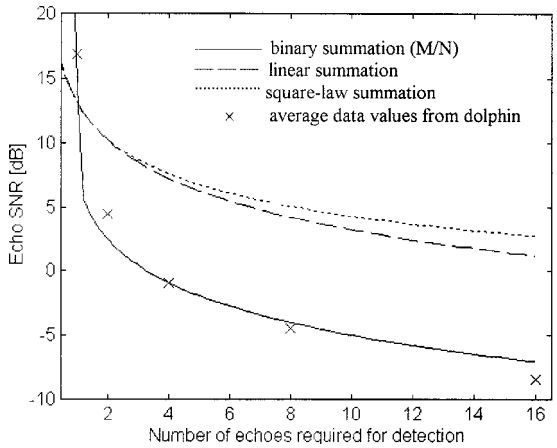


FIG. 6. MMSE (minimum mean-square error) fits between three receiver models and dolphin N-echo detection data. The MMSE algorithm compares models to noise-normalized data amplitudes $A(N)/\sigma$, but the curves are shown on a dB scale corresponding to $\text{SNR[dB]} = 20\log_{10}[A(N)/\sigma]$.

Variation of p_0 can be used to check the results, since the best p_0 value for nonparametric operation is known to be 0.5.

IV. DISCUSSION

A. Estimation of detectability index

The analysis in the Appendix indicates that the constants c_l, c_e, and c_b are related to the corresponding detectability indices by the equations

$$c_l = d_l; \quad c_e = d_e; \quad c_b = d_b^2/2. \tag{9}$$

The detectability indices corresponding to the MMSE estimates of c_l, c_e, and c_b are

$$d_l = 4.58; \quad d_e = 3.12; \quad d_b = 1.41, \tag{10}$$

for the linear summation, energy detection, and binary M-out-of-N receivers, respectively. These performance measures are based on a restricted set of N values equal to 1, 2, 4, 8, and 16. The receiver model with the best fit to the data has the worst performance in zero-mean Gaussian noise.

B. Distribution tolerance of the binary summation model

The linear receiver is optimum for Gaussian noise with known variance, but the binary M-out-of-N processor has distribution-tolerant (nonparametric) properties. The false-alarm rate of the binary M-out-of-N receiver is insensitive to time-varying noise power and to the shape of any symmetric, zero-mean noise distribution. If the binary M-out-of-N receiver is a viable model for dolphin multiecho processing, then the dolphin has traded optimality in Gaussian noise with specified noise power for robustness with respect to the distribution and power of the noise. The performance disparity between binary and linear summation is not large if many echoes are used. Figure 6 implies that large N is associated with small SNR for all of the models. For large N and small SNR, it is shown in the Appendix that $d_l \approx 1.25 d_b$.

C. Preprocessing with an auditory transduction model

Figure 6 illustrates the performance of an M-out-of-N receiver with zero binary quantization threshold. This performance is unaffected by preprocessing input data with a sign-preserving zero-memory nonlinear transformation. Two examples of such a transformation are (1) a half-wave rectifier and (2) membrane potential as a function of the displacement of either inner or outer hair cells (Russell *et al.*, 1986; Mountain and Hubbard, 1996). The binary summation model is insensitive to the nonlinear signal transformation that occurs during cochlear transduction from acoustic waveforms to neuronal excitations.

An envelope preprocessor is approximated by a weighted average of neighboring half-wave rectified data samples. A receiver model that uses envelope detection prior to binary quantization and M-out-of-N detection cannot be ruled out with current data.

D. Phase sensitivity

Phase sensitivity of the dolphin N-echo receiver model is still an open question. If the data are linearly processed, half-wave rectified, or passed through a zero-memory hair cell model before binary quantization, then phase sensitivity depends upon maintaining an accurate sampling time relative to the time of signal transmission. Multiple, parallel M-out-of-N detectors can be used to test hypothesized sampling times, to compensate for sampling time errors, and to generate range-tracking information. If an envelope detector that forms a weighted sum of neighboring half-wave rectifier outputs is used as a preprocessor, then receiver tolerance to sampling time errors is increased but phase sensitivity is reduced.

To test phase sensitivity, the waveform in Fig. 1 can be replaced with a signal that has a short-duration, high-amplitude positive peak followed by (or surrounded by) long-duration, small-amplitude negative components with the same total area as the positive peak. Phase reversal of this waveform (multiplication by -1) should affect the detection performance of a phase-sensitive receiver.

E. Postdetection integration

The binary M-out-of-N receiver model is equivalent to a postprocessor for a detector that makes an H1 versus H0 decision at each range sample, for each click–echo pair. The binary decision variable at a given range is integrated or counted over successive click-echo pairs. A different strategy is followed by the linear and quadratic summation models, which do not implement a decision until all relevant data are summed. The latter strategy is generally regarded as superior because the level of each detector output is preserved and no information is lost via premature decision making. A large transient interference pulse (e.g., from a snapping shrimp), however, will have a much larger effect on linear or quadratic summation than upon summation of binary decision variables (Bullock, 1986). In San Diego Bay and many other locations, snapping shrimp are an important source of interference for dolphin echolocation (Au and Banks, 1998). Aside from interference considerations, the dynamic range tolerance provided by binary quantization may be important for detection, tracking, and acoustic imaging of prey with large aspect-dependent variation in target strength.

F. Biological neural networks

The binary M-out-of-N detector model is a basic building block for neural networks that process action potentials, which are binary, all-or-none signals. A neuronal version of a binary M-out-of-N processor could use binary sampling intervals corresponding to the width of an action potential spike. If the intensity of a stimulus is encoded by the density of action potential spikes and/or the duration of a spike sequence, the binary M-out-of-N processor can function as a stimulus intensity or amplitude decoder, despite the amplitude insensitivity associated with binary quantization.

G. Polarity coincidence correlation and binaural localization models

The binary M-out-of-N detector is a polarity coincidence correlator with a constant, unit reference function. The polarity coincidence correlator is well known for its relative insensitivity to the probability distribution of input data (Wolff *et al.*, 1962). A neurophysiological model for binaural localization (interaural time delay estimation) uses coincidence of the excitations in two neural delay lines, one from each ear (Jeffress, 1948; Konishi, 1993; Colburn, 1996). This binaural model is similar to polarity coincidence correlation and thus to the binary M-out-of-N receiver model. In biological sonar systems, interaural polarity coincidence correlation can be used for azimuth estimation (and for elevation estimation if the animal rolls by 90 degrees). Range tracking can be implemented via cross correlation of successive echoes, using a polarity coincidence correlator.

H. Binary quantization and zero crossings

A binary waveform representation preserves information about real zero crossings, which are important signal attributes (Kedem, 1994; Marr, 1982; Requicha, 1980; Voelcker, 1966a, 1966b). A polarity coincidence correlator can use these attributes for detection, estimation, classification, and decomposition via Haar functions (Hagen and Farley, 1973; Vetterli and Kovacevic, 1995).

I. Capabilities of a multipulse sonar receiver that uses binary summation

The ability of a dolphin to combine information from multiple pulse–echo pairs is necessary for advanced signal-processing capabilities. One of these capabilities is acoustic imaging via a simplified version of synthetic aperture sonar (SAS) processing (Altes, 1995; Altes *et al.*, 1998; Altes, 2003). SAS images can be formed by adding the echo sample at each range to an appropriate pixel in a two- or three-dimensional image. The image is sequentially constructed as multiple echoes are obtained from different aspects. High-resolution SAS images have been created from binary-quantized sonar echo envelopes (with a nonzero quantization threshold), using dolphin-like transmitted waveforms (Altes, personal observation). After N echoes are processed, each pixel level in such an image represents the response of a binary M-out-of-N receiver.

The dolphin may use multipulse processing to estimate range-rate, implement a moving target indicator (MTI), track targets, and perform acoustic imaging in two or three dimensions. Figure 6 implies that the dolphin can perform robust integration along a constant-range line in the range, echo-number (R,N) plane. Range-rate estimation, tracking, and acoustic imaging involve integration along other lines or curves in the R,N plane. The simplest moving target indicator computes the difference between successive detector outputs along a constant-range line in the R,N plane. A more sophisticated MTI uses a weighted sum of such outputs, with positive and negative weights.

J. Adaptability of the dolphin receiver

The power spectrum of artificially added Gaussian noise (white over the dolphin's echolocation bandwidth) was 95 dB re: 1 μPa^2/Hz. When an extra 13 dB is added to account for the dolphin's directivity index, the artificial Gaussian noise was 108 dB above spatially uniform noise with a power spectral density of 1 μPa^2/Hz, and 38 dB above the average ambient noise level of 70 dB re: 1 μPa^2/Hz. The ambient level included average snapping shrimp interference and the sounds of other dolphins in the area. Despite the 38-dB difference, the analysis indicates that the dolphin's receiver did not adapt to operation in Gaussian noise as opposed to strong transient interference. This apparent lack of adaptability can be explained by the large disparity between peak and average levels of transient sounds. At a distance of 1 meter, the power spectral density of a single snapping shrimp pulse is between 105 and 111 dB re: 1 μPa^2/Hz (Au and Banks, 1998; Versluis *et al.*, 2000). Spherical spreading decreases this level by 20 log r if the shrimp is r meters from the receiver. A spatially averaged interference power spectral density level of 108 dB re: 1 μPa^2/Hz is required to make the interference power spectrum equal to the power spectrum of the artificially added Gaussian noise at the input to the dolphin's receiver. The possible presence of nearby snapping shrimp thus could have constrained the dolphin's receiver design. The experiment was performed in an area with floating walkways to support trainers and equipment. Snapping shrimp appear to congregate in such areas (Ferguson and Cleary, 2001).

V. CONCLUSION

Synthetic echoes in additive noise were used to estimate the SNR required by a dolphin for detection when the number of available echoes was varied. A close fit to the dolphin's performance data was obtained with a receiver model that sums binary-quantized time samples from N available echoes. This detector does not perform as well as linear summation for Gaussian noise with known average power, but its false-alarm rate is distribution tolerant and nonparametric with respect to variable noise power. The dolphin's acoustic environment is, in fact, notoriously non-Gaussian and nonstationary (Urick, 1975). The close fit of the binary M-out-of-N model to the dolphin detection data, together with the relatively poor fit of the linear and energy summation models, implies that the dolphin trades optimality (in Gaussian noise with known power) for robustness.

Further experiments are needed to determine whether the dolphin's time sampling is sufficiently precise to allow phase sensitivity in the context of the binary summation model. Even without phase sensitivity, binary summation can be used for acoustic imaging. Binary M-out-of-N detection is a special case of polarity coincidence correlation (a model for binaural localization), and it is similar to operations performed by biological neural networks. Another question is whether the dolphin's receiver can adapt to optimum operation in Gaussian noise when snapping shrimp are not present.

ACKNOWLEDGMENTS

We thank Justine M. Zafran and Michael D. Massimi for their assistance during the behavioral data collection phase of this study. This research was made possible in part thanks to support from Dr. John Tague and Dr. Robert Gisiner (Office of Naval Research 321US and 342). Helpful critiques were provided by Dr. Magnus Wahlberg (Department of Zoophysiology, Aarhus University, Denmark) and Dr. David Kastak (University of California at Santa Cruz).

APPENDIX: DETECTABILITY INDEX DERIVATIONS

The output mean values and variances from each receiver model in Fig. 4 can be used to predict the detectability index at the receiver output. The detectability index is closely related to the Fisher ratio and to the t-statistic. Receiver performance (detection, false alarm, and error probabilities) can be computed from the detectability index if the output distributions are Gaussian with equal variance under noise-only and signal-plus-noise conditions (Van Trees, 1968). This condition seldom applies to nonlinear receivers. Detection performance, however, almost always varies monotonically with the detectability index, which is defined as the difference between the output mean value when hypothesis H1 (signal plus noise) is true and the output mean value when H0 (noise alone) is true, divided by the square root of the average output variance for the two hypotheses.

Addition of N independent receiver outputs with fixed SNR causes the mean output and variance to be multiplied by N, for both H1 and H0. The detectability index for the sum is then the detectability index for a single observation multiplied by the square root of N. Equivalently, the detectability index of the sum is the square root of the sum-of-squares of the detectability indices for the individual observations as in Green and Swets (1988), regardless of the receiver model, for constant SNR.

In the following analysis, the detectability index of each receiver model is assumed to vary monotonically with detection performance and to be constant for all values of N, the number of echoes available for detection. A constant detectability index for all N values implies that SNR decreases with N. These assumptions imply that each receiver model (as well as the dolphin) uses a consistent performance criterion (detection, false alarm, and error probability) for decision making at all N values.

Since the detectability index d depends upon SNR and N, it should be possible to obtain an expression for SNR $=20\log[A(N)/\sigma]$ as a function of d and N. The noise-normalized echo amplitude required for detection, $A(N)/\sigma$, depends upon the detectability index d and the number of available echoes N. For the binary summation model, $A(N)/\sigma$ also depends upon the threshold γ_b for binary quantization, or equivalently, on the probability p_0 of a threshold crossing when H0 is true. In general

$$[A(N)/\sigma]_{\text{theory}} = f_{\text{recvr}}(d, N, \gamma_b), \tag{A1}$$

where the function $f_{\text{recvr}}(d, N, \gamma_b)$ depends upon the receiver model.

The experimental results yield the average signal-to-noise ratio (SNR) in decibels at specific N values, and

$$[A(N)/\sigma]_{\text{expt}} = 10^{\text{SNR}/20}. \tag{A2}$$

For each receiver model, a gradient descent algorithm is used to find the value of d (and p_0 or γ_b if one of these quantities is not prespecified) that minimizes the mean-square difference between $[A(N)/\sigma]_{\text{theory}}$ and $[A(N)/\sigma]_{\text{expt}}$ at the N values that were used in the experiment. The resemblance between a model and the experimental data is quantitatively represented by a correlation coefficient (Hays, 1994) computed from the $A(N)/\sigma$ values of the best-fit model and the data at the experimental N values.

The linear summation model computes the function

$$h(\underline{x}) = (1/N)\sum_{i=1}^{N} (x_i + A), \tag{A3}$$

where A is the sampled signal value and $\{x_i; i=1,...,N\}$ are independent identically distributed noise samples (one from each echo) with zero mean and variance σ^2; $E(x_i)=0$ and $E(x_i x_j)=\sigma^2$ if $x_i = x_j$, and zero otherwise. $E(x)$ is the ensemble expected value of x, and $\underline{x}$ is the set of noise samples $x_1, x_2, \ldots, x_N$. It follows that

$$E[h(\underline{x})] = (1/N)\sum_{i=1}^{N} E(x_i + A) = A, \tag{A4}$$

and

$$\begin{aligned} E[h^2(\underline{x})] &= (1/N^2)\sum_{i=1}^{N}\sum_{j=1}^{N} E[(x_i + A)(x_j + A)] \\ &= (1/N^2)\left[\sum_{i=1}^{N} E[(x_i + A)^2] \right. \\ &\quad \left. + \sum_{i=1}^{N}\sum_{j\neq i} E[(x_i + A)(x_j + A)]\right] \\ &= (\sigma^2/N) + A^2. \end{aligned} \tag{A5}$$

The variance of the averaged, linearly transformed data is then

$$\text{Var}[h(\underline{x})] = E[h^2(\underline{x})] - E^2[h(\underline{x})] = \sigma^2/N. \tag{A6}$$

When hypothesis H1 is true, the data consist of signal plus noise with $A \neq 0$. If H0 is true, the data consist of noise alone $(A=0)$. The corresponding detectability index is

$$\begin{aligned} d_l &= \frac{|E[h(\underline{x})|H1] - E[h(\underline{x})|H0]|}{[(1/2)\{\text{Var}[h(\underline{x})|H0] + \text{Var}[h(\underline{x})|H1]\}]^{1/2}} \\ &= \sqrt{N}A/\sigma. \end{aligned} \tag{A7}$$

In the psychophysical literature, d' is used instead of d. The prime is omitted here in order to simplify notation in the equations.

The linear summation model is evaluated by adjusting the constant $c_l = d_l$ in the equation

$$[A(N)/\sigma]_{\text{lin}} = c_l/\sqrt{N}, \tag{A8}$$

to obtain a minimum mean-square error (MMSE) fit of $[A(N)/\sigma]_{\text{lin}}$ to $[A(N)/\sigma]_{\text{expt}}$ for the experimental values of N. The resulting best fit is then evaluated via the correlation coefficient between $[A(N)/\sigma]_{\text{lin}}$ and $[A(N)/\sigma]_{\text{expt}}$.

For an average of squared data samples (energy detection)

$$h(\underline{x})=(1/N)\sum_{i=1}^{N}(x_i+A)^2, \tag{A9}$$

$$E[h(\underline{x})]=(1/N)\sum_{i=1}^{N}E(x_i+A)^2=\sigma^2+A^2, \tag{A10}$$

and

$$\begin{aligned}E[h^2(\underline{x})]&=(1/N^2)\sum_{i=1}^{N}\sum_{j=1}^{N}E[(x_i+A)^2(x_j+A)^2]\\&=(1/N^2)\Bigg[\sum_{i=1}^{N}E[(x_i+A)^4]\\&\quad+\sum_{i=1}^{N}\sum_{j\neq i}E[(x_i+A)^2(x_j+A)^2]\Bigg],\end{aligned} \tag{A11}$$

where

$$\begin{aligned}E[(x_i+A)^4]&=E(x_i^4)+4E(x_i^3)A+6E(x_i^2)A^2\\&\quad+4E(x_i)A^3+A^4.\end{aligned} \tag{A12}$$

For a zero-mean Gaussian random variable, $E(x_i)=0$, $E(x_i^2)=\sigma^2$, $E(x_i^3)=0$, and $E(x_i^4)=3\sigma^4$. It follows that

$$\begin{aligned}E[h^2(\underline{x})]&=(1/N)(3\sigma^4+6\sigma^2A^2+A^4)\\&\quad+[(N-1)/N](\sigma^2+A^2)^2,\end{aligned} \tag{A13}$$

and

$$\begin{aligned}\text{Var}[h(\underline{x})]&=E[h^2(\underline{x})]-E^2[h(\underline{x})]\\&=(2\sigma^2/N)(\sigma^2+2A^2).\end{aligned} \tag{A14}$$

The detectability index is then

$$\begin{aligned}d_e&=\frac{|E[h(\underline{x})|H1]-E[h(\underline{x})|H0]|}{[(1/2)\{\text{Var}[h(\underline{x})|H0]+\text{Var}[h(\underline{x})|H1]\}]^{1/2}}\\&=\sqrt{N/2}\frac{(A/\sigma)^2}{\sqrt{1+(A/\sigma)^2}}\\&\cong\sqrt{N/2}|A/\sigma|\ \text{ at high SNR } (|A/\sigma|\gg 1)\\&\cong\sqrt{N/2}(A/\sigma)^2\ \text{ at low SNR } (|A/\sigma\ll 1).\end{aligned} \tag{A15}$$

For a general square-law model with no assumptions about SNR, the first equation in (A15) can be written

$$(N/2)x^2-d_e^2x-d_e^2=0, \tag{A16}$$

where $x=(A/\sigma)^2$. Solving for x yields

$$A(N)/\sigma=[(d_e^2/N)\pm(d_e/N)\sqrt{d_e^2+2N}]^{1/2}. \tag{A17}$$

To obtain a real-valued $A(N)/\sigma$, the plus-or-minus operation in (A17) must always be plus, and

$$[A(N)/\sigma]_{\text{egy}}=(c_e/\sqrt{N})[1+\sqrt{1+2(c_e/\sqrt{N})^{-2}}]^{1/2}, \tag{A18}$$

where $c_e=d_e$. To evaluate the general square-law model, the constant c_e in (A18) is adjusted to obtain an MMSE fit of $[A(N)/\sigma]_{\text{egy}}$ to $[A(N)/\sigma]_{\text{expt}}$ for the experimental values of N.

For the M-out-of-N receiver model, the binomial distribution (Papoulis, 1965) describes the probabilities of various numbers of ones and zeros at the output of the binary quantizer for N echoes. Let p_1 equal the probability that the binary random variable equals 1 (the sampled data value is greater than the binary quantization threshold, γ_b) when an echo is present (the signal plus noise condition, H1). Let p_0 equal the probability that the binary random variable is 1 when the echo is absent (the noise alone condition, H0). For one binary sample from each of N echoes

The expected number of ''ones'' with echo present (H1) equals Np_1.
The expected number of ''ones'' with echo absent (H0) equals Np_0.
The variance of the distribution of the number of ''ones'' given H1 equals $Np_1\ (1-p_1)$.
The variance of the distribution of the number of ''ones'' given H0 equals $Np_0\ (1-p_0)$.

The detectability index is the difference in means divided by the square root of the average variance.

$$d_b=\frac{Np_1-Np_0}{\sqrt{(1/2)[Np_1(1-p_1)+Np_0(1-p_0)]}}. \tag{A19}$$

As in the previous models described by (A8) and (A15), the detectability index is proportional to the square root of N if p_1 is constant (constant SNR).

If a threshold value γ_b is used to convert echo samples into binary data, the probability that the binary random variable equals 1 when the signal is absent (noise alone) is

$$p_0=\text{erfc}_*(\gamma/\sigma), \tag{A20}$$

and the probability that the binary random variable equals 1 when the signal is present is

$$p_1=\text{erfc}_*\{[\gamma-A(N)]/\sigma\}, \tag{A21}$$

where σ is the rms noise power and the complementary error function $\text{erfc}_*(x)$ is the integral of a zero-mean, unit variance normal distribution between x and infinity.

From (A20), the threshold level is

$$\gamma=\sigma\,\text{erfc}_*^{-1}(p_0), \tag{A22}$$

where $\text{erfc}_*^{-1}(p_0)$ is the inverse complementary error function of p_0 as in Fig. 5. Similarly, (A21) can be solved for $A(N)$:

$$A(N)=\gamma-\sigma\,\mathrm{erfc}_*^{-1}(p_1). \tag{A23}$$

Substituting (A22) into (A23),

$$[A(N)/\sigma]_{\mathrm{bin}}=\mathrm{erfc}_*^{-1}(p_0)-\mathrm{erfc}_*^{-1}(p_1). \tag{A24}$$

Letting $c_b=(1/2)d_b^2$ and solving (A19) for p_1 yields

$$p_1=\frac{(2p_0+c_b/N)\pm\sqrt{(2p_0+c_b/N)^2-4p_0(1+c_b/N)[p_0(1+c_b/N)-c_b/N]}}{2(1+c_b/N)}. \tag{A25}$$

The plus-or-minus operation in (A25) must always be plus in order for p_1 to be non-negative. For a given level of detection performance (e.g., a given percentage of correct decisions), the detectability index is constant and the parameters c_b and p_0 in (A24)–(A25) can be adjusted to obtain an MMSE fit of $[A(N)/\sigma]_{\mathrm{bin}}$ to $[A(N)/\sigma]_{\mathrm{expt}}$. The resulting values of $[A(N)/\sigma]_{\mathrm{bin}}$ yield a much better fit to the dolphin data than can be obtained via linear summation or energy detection models.

The binary M-out-of-N receiver can be easily compared with linear summation for small SNR, which corresponds to a large number N of available echoes. For zero binary quantization threshold ($\gamma_b=0$) and for small SNR

$$p_0=1/2, \tag{A26}$$

and

$$\begin{aligned}p_1&=\mathrm{erfc}_*[(\gamma_b-A)/\sigma]\\&=\int_{-A/\sigma}^{\infty}(2\pi)^{-1/2}\exp(-y^2/2)dy\\&=(1/2)+\int_0^{A/\sigma}(2\pi)^{-1/2}\exp(-y^2/2)dy\\&\approx p_0+(A/\sigma)\left[(d/dx)\int_0^x(2\pi)^{-1/2}\exp(-y^2/2)dy\right]_{x=0}\\&=p_0+[A/(\sqrt{2\pi}\sigma)].\end{aligned} \tag{A27}$$

Substituting (A26) and (A27) into (A19) yields

$$d_b\approx\frac{\sqrt{N}A/\sigma}{\sqrt{\pi[(1/2)-(A/\sqrt{2\pi}\sigma)^2]}}\approx(2/\pi)^{1/2}d_l. \tag{A28}$$

At low SNR (large N), $d_b\approx 0.8d_l$. The performance of the binary M-out-of-N receiver is slightly worse than that of the linear summation receiver for a large number of echoes, and a slightly larger SNR should be required for detection. This comparison pertains to additive zero-mean Gaussian noise with known variance (known expected noise power). For non-Gaussian and/or nonstationary noise, the binary M-out-of-N receiver may be superior to linear summation.

For a small number of echoes, performance prediction of the binary M-out-of-N model involves a nonlinear transformation that greatly increases the required SNR. The binary summation data fit illustrated in Fig. 6 corresponds to $p_0=1/2$ and $c_b=0.999\,999\,3$ in (A25). The parameter c_b is written with high accuracy because a small change in c_b is associated with a large change in $\mathrm{erfc}_*^{-1}(p_1)$ when p_1 is close to unity. Substituting $p_0=1/2$, and $c_b=1.0$ into (A25), results in the equation

$$p_1=(1/2)+[2(1+N)]^{-1/2}. \tag{A29}$$

Substituting this expression for p_1 into (A24) with $p_0=1/2$ yields

$$[A(N)/\sigma]_{\mathrm{bin}}=-\mathrm{erfc}_*^{-1}\{(1/2)+[2(1+N)]^{-1/2}\}. \tag{A30}$$

The argument of the inverse complementary error function in (A30) is unity when N equals one and approaches $\frac{1}{2}$ as N becomes very large. As indicated in Fig. 5, the inverse complementary error function is unbounded when its argument equals zero, decreases monotonically as its argument increases, passes through zero when the argument is $\frac{1}{2}$, and goes to $-\infty$ when the argument equals 1,

$$\mathrm{erfc}_*^{-1}(0)=\infty;\quad \mathrm{erfc}_*^{-1}(1/2)=0;\quad \mathrm{erfc}_*^{-1}(1)=-\infty. \tag{A31}$$

Since c_b is slightly less than 1 for an MMSE fit to the data, the negative-inverse complementary error function is not unbounded for $N=1$, but is very large. This extremely nonlinear behavior allows the binary M-out-of-N model to closely approximate the large SNR required by the dolphin when only one echo is available ($N=1$).

Receiver comparisons can be further investigated by ROC (receiver operating characteristic) computation. The ROC is a plot of detection versus false-alarm probabilities for various threshold settings. For the binary M-out-of-N detector, threshold settings are limited to integer values of M between 1 and N. The probabilities of detection and false alarm for a given M value are

$$\begin{aligned}&P_{D,M/N}=\sum_{k=M}^{N}\binom{N}{k}p_1^k(1-p_1)^{N-k},\\&p_1=\mathrm{erfc}_*[(\gamma_b-A)/\sigma],\end{aligned} \tag{A32}$$

$$P_{F,M/N}=\sum_{k=M}^{N}\binom{N}{k}p_0^k(1-p_0)^{N-k},\quad p_0=\mathrm{erfc}_*(\gamma_b/\sigma). \tag{A33}$$

The false-alarm probability is independent of noise power σ^2 if the threshold for binary quantization γ_b equals zero. For zero γ_b, the false-alarm rate of the binary M-out-of-N receiver can be changed by adjusting the number of binary threshold crossings M that are required for detection. As M is increased, the false-alarm rate decreases, and the detection probability also decreases.

For linear summation

$$P_{D,\mathrm{lin}} = \mathrm{erfc}_*[\sqrt{N}(\gamma - A)/\sigma], \tag{A34}$$

$$P_{F,\mathrm{lin}} = \mathrm{erfc}_*(\sqrt{N}\gamma/\sigma). \tag{A35}$$

Dependence of the false-alarm rate on noise power can again be eliminated by setting the threshold of the linear receiver equal to zero. In the linear case, however, this threshold setting yields a large false-alarm rate that cannot be changed without introducing noise dependence. For linear summation with a constant false-alarm rate that is unequal to 0.5, noise power must be estimated and the threshold value must be adjusted accordingly.

Altes, R. A. (**1995**). ''Signal processing for target recognition in biosonar,'' Neural Networks **8**, 1275–1295.

Altes, R. A. (**2003**). ''Synthetic aperture and image sharpening models for animal sonar,'' *Echolocation in Bats and Dolphins*, edited by J. Thomas, C. Moss, and M. Vater (University of Chicago Press, Chicago), pp. 492–500.

Altes, R. A., Moore, P. W., and Helweg, D. A. (**1998**). ''Tomographic image reconstruction of MCM targets using synthetic dolphin signals,'' Tech. Doc. 2993, Space and Naval Warfare Systems Center, San Diego, CA 921152-5001.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer, New York).

Au, W. W. L., and Banks, K. (**1998**). ''The acoustics of the snapping shrimp *Synalpheus parneomeris* in Kaneohe Bay,'' J. Acoust. Soc. Am. **103**, 41–47.

Au, W. W. L., and Moore, P. W. B. (**1990**). ''Critical ratio and critical bandwidth for the Atlantic bottlenose dolphin,'' J. Acoust. Soc. Am. **88**, 1635–1638.

Au, W. W. L., Moore, P. W. B., and Pawloski, D. A. (**1988**). ''Detection of complex echoes in noise by an echolocating dolphin,'' J. Acoust. Soc. Am. **83**(2), 662–668.

Brill, R. L., Moore, P. W. B., and Dankiewicz, L. A. (**2001**). ''Assessment of dolphin (*Tursiops truncatus*) auditory sensitivity and hearing loss using jawphones,'' J. Acoust. Soc. Am. **109**(4), 1717–1722.

Bullock, T. H. (**1986**). ''Significance of findings on electroreception for general neurobiology,'' in *Electroreception*, edited by T. H. Bullock and W. Heiligenberg (Wiley, New York), pp. 661–674.

Busnel, R.-G., and Fish, J. F., editors (**1980**). *Animal Sonar Systems* (Plenum, New York).

Colburn, H. S. (**1996**). ''Computational models of binaural processing,'' in *Auditory Computation*, edited by H. L. Hawkins, T. A. McMullen, A. N. Popper, and R. R. Fay (Springer, New York), pp. 332–400.

Collett, D. (**1952**). *Modeling Binary Data*, 2nd ed. (Chapman and Hall/CRC, Boca Raton, 2003), pp. 56–57.

Dankiewicz, L. A., Helweg, D. A., Moore, P. W., and Zafran, J. M. (**2002**). ''Discrimination of amplitude-modulated synthetic echo trains by an echolocating bottlenose dolphin,'' J. Acoust. Soc. Am. **112**, 1702–1708.

Ferguson, B. G., and Cleary, J. L. (**2001**). ''*In situ* source level and source position estimates of biological transient signals produced by snapping shrimp in an underwater environment,'' J. Acoust. Soc. Am. **109**, 3031–3027.

Floyd, R. W. (**1980**). ''Models of cetacean signal processing,'' in *Animal Sonar Systems*, edited by R. G. Busnel and J. F. Fish (Plenum, New York), pp. 615–623.

Gellermann, L. W. (**1933**). ''Chance orders of alternating stimuli in visual discrimination experiments,'' J. Gen. Psychol. **42**, 206–208.

Green, D. M., and Swets, J. A. (**1988**). *Signal Detection Theory and Psychophysics* (Peninsula, Los Altos, CA).

Hagen, J. B., and Farley, D. T. (**1973**). ''Digital-correlation techniques in radio science,'' Radio Sci. **8**, 775–784.

Hays, W. L. (**1994**). *Statistics*, 5th ed. (Harcourt-Brace, Fort Worth), pp. 597–660.

Helweg, D. A., Roitblat, H. L., Nachtigall, P. E., and Hautus, M. J. (**1996**). ''Recognition of three-dimensional aspect-dependent objects by an echolocating bottlenosed dolphin,'' J. Exp. Psychol. Anim. Behav. Process **22**, 19–31.

Jeffress, L. A. (**1948**). ''A place theory of sound localization,'' J. Comp. Physiol. Psychol. **41**, 35–39.

Johnson, C. S. (**1968**). ''Masked tonal thresholds in the bottlenosed porpoise,'' J. Acoust. Soc. Am. **44**, 965–967.

Kedem, B. (**1994**). *Time Series Analysis by Higher Order Crossings* (IEEE, New York).

Konishi, M. (**1993**). ''Listening with two ears,'' Sci. Am. **268**(4), 66–73.

Marr, D. (**1982**). *Vision* (Freeman, San Francisco).

Moore, P. W. B., Roitblat, H. L., Penner, R. H., and Nachtigall, P. E. (**1991**). ''Recognizing successive dolphin echoes with an Integrator Gateway Network,'' Neural Networks **4**, 701–709.

Moore, P. W. B., and Schusterman, R. J. (**1987**). ''Audiometric assessment of Northern fur seals, *Callorhinus ursinus*,'' Marine Mammal Sci. **3**, 31–53.

Mountain, D. C., and Hubbard, A. E. (**1996**). ''Computational analysis of hair cell and auditory nerve processes,'' in *Auditory Computation*, edited by H. L. Hawkins, T. A. McMullen, A. N. Popper, and R. R. Fay (Springer, New York), pp. 121–156.

Nachtigall, P. E., and Moore, P. W. B., editors (**1988**). *Animal Sonar: Processes and Performance* (Plenum, New York).

Papoulis, A. (**1965**). *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, New York), pp. 102, 158.

Penner, R. H. (**1988**). ''Attention and detection in dolphin echolocation,'' in *Animal Sonar: Processes and Performance*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum, New York), pp. 707–713.

Requicha, A. (**1980**). ''Zeros of entire functions,'' Proc. IEEE **68**, 308–328.

Roitblat, H. L., Moore, P. W. B., Nachtigall, P. E., and Penner, R. H. (**1991**). ''Natural dolphin echo recognition using an Integrator Gateway Network,'' Adv. Neural Inf. Process. Systems **3**, 273–281.

Russell, I. J., Cody, A. R., and Richardson, G. P. (**1986**). ''The responses of inner and outer hair cells in the turn of the guinea-pig cochlea and in the mouse cochlea grown *in vitro*,'' Hear. Res. **22**, 199–216.

Schafer, T. H., and Shewmaker, C. A. A. (**1953**). ''A comparative study of the audio, visual and audio–visual recognition differentials for pulses masked by random noise,'' Naval Electronics Laboratory, Report 372.

Surlykke, A. (**2003**). ''Detection thresholds depend on the number of echoes in the Big Brown Bat,'' *Echolocation in Bats and Dolphins*, edited by J. Thomas, C. Moss, and M. Vater (University of Chicago Press, Chicago), pp. 368–273.

Swets, J. A., and Green, D. M. (**1964**). ''Sequential observations by human observers of signals in noise,'' in *Signal Detection and Recognition by Human Observers*, edited by J. A. Swets (Wiley, New York), pp. 221–242.

Swets, J. A., Shipley, E. F., McKey, M. J., and Green, D. M. (**1964**). ''Multiple observations of signals in noise,'' in *Signal Detection and Recognition by Human Observers*, edited by J. A. Swets (Wiley, New York), pp. 201–220.

Thomas, J. A., and Kastelein, R. A., editors (**1990**). *Sensory Abilities of Cetaceans: Laboratory and Field Evidence* (Plenum, New York).

Urick, R. J. (**1975**). *Principles of Underwater Sound*, 2nd ed. (McGraw-Hill, New York), pp. 189–193.

Van Trees, H. L. (**1968**). *Detection, Estimation, and Modulation Theory* (Wiley, New York), pp. 36–38.

Vel'min, V. A., and Dubrovskiy, N. A. (**1976**). ''The critical interval of active hearing in dolphins,'' Sov. Phys. Acoust. **22**, 351–352.

Versluis, M., Schmitz, B., von der Heydt, A., and Lohse, D. (**2000**). ''How snapping shrimp snap: Through cavitating bubbles,'' Science **289**, 2114–2117.

Vetterli, M., and Kovacevic, J. (**1995**). *Wavelets and Subband Coding* (Prentice-Hall, Englewood Cliffs), pp. 208–213.

Viemeister, N. F., and Wakefield, G. H. (**1991**). ''Temporal integration and multiple looks,'' J. Acoust. Soc. Am. **90**, 858–865.

Voelcker, H. (**1966a**). ''Toward a unified theory of modulation, I,'' Proc. IEEE **54**, 340–353.

Voelcker, H. (**1966b**). ''Toward a unified theory of modulation. II,'' Proc. IEEE **54**, 735–755.

Wolff, S. S., Thomas, J. B., and Williams, T. (**1962**). ''The polarity coincidence correlator: A nonparametric detection device,'' IRE Trans. Inf. Theory **IT-8**, 1–19.